The establishment size-wage premium : evidence from France

Syeda Batool

Université Paris 1 - Panthéon Sorbonne
U.F.R de Sciences Economiques

2013

Numéro attribué par la bibliothèque
|||||||||||||

Thèse

Pour le doctorat de Sciences Economiques

Présentée et soutenue publiquement par

Syeda Batool

Le 25 Octobre 2013

The Establishment Size-Wage Premium:

Evidence from France

Directeur de Thèse :

Michel Sollogoub: Professor, Université Paris-1 Panthéon-Sorbonne

Jury :

Philip Merrigan : Professor, Université du Québec à Montréal

Claude Mathieu: Professor, Université de Paris-Est Créteil

Catherine Sofer: Professor, Université Paris-1 Panthéon-Sorbonne

François Gardes: Professor, Université Paris -1 Panthéon-Sorbonne

Acknowledgements

The most rewarding achievement in my life is the completion of my doctoral dissertation. When I naively stepped onto this path of self-discovery, I hardly understood the academic, emotional and physical difficulties in doing a qualitative research and writing an academic paper. I have received much useful advice throughout the writing of my thesis, but all the faults that remain are obstinately my own.

This dissertation would never have come to fruition without the support of many individuals, and it is with pleasure and gratitude that I acknowledge their efforts.

I would like to take this opportunity to express my immense gratitude to my research supervisor, Michel Sollogoub, for his sincere and selfless support, prompt and useful advice during my research. He has given me a lifetime unforgettable memory of his benevolence, patience and intelligence. It is like we went through all the difficulties together in these so many years and it was not an easy task.

I greatly appreciate the comments on the earlier version of this dissertation provided by the members of my PhD committee. I would like to express my gratitude to Philip Merrigan, Claude Mathieu, Catherine Sofer and François Gardes. Their bright academic minds amazed me and they helped me develop my thesis with constructive comments, encouragement, and support. I wish to especially acknowledge Catherine Sofer and François Gardes for their guidance and encouragement along the way uptill now.

Equally, I would also like to thank University Paris-1, Paris school of economics, Centre d'Economie de la Sorbonne for providing me this great opportunity to work on this dissertation as well as the department of Applied Microeconomics and Econometrics for support. Many people whom I work with at MSE need to be acknowledged especially Antoin Terracol and Helene Huber. I would also like to acknowledge efforts of Elda André, Loïc Sorel and Viviane Makougni for their nice nature and administrative assistance.

I would like to thank my friends Omar, Claire, Catarina, Mariska, all friends of FBL fifth floor (2011-2012), Nazia, Natalia, Arslan, Muhammad Khan, Naeem for their continuous support. I want to thank my collegues at 4th floor of MSE (office 402, 403 and 413) for their support especially Noemi and, equally, I am greatful to Muniza, Salman, Clemence, Irene, Hela, and Gaelle for their guidance in the first year of my PhD. I will forever be thankful to Ata for his unconditional care and support throughout my thesis.

Last, but definitely not the least, I am greatly indebted to my family. It was my parents' unconditional love, care, and tolerance which made the hardship of writing the thesis worthwhile. I would like to acknowledge the endless support of my five sisters, Shazia, Fouzia, Nazia, Tasmia and Asma and three brothers, Kifayat, Nazakat and Aqib. Without their support, I do not think that I could overcome the difficulties during these years. I truly thank my family for sticking by my side; even when I was irritable and depressed. I dedicate this thesis to my mother, Amina, and my father, Zahoor.

Avertissement

Mis a part l'introduction, la revue de litterature et la conclusion de cette these, les differents chapitres sont issus d'articles de recherche rediges en anglais et dont la structure est autonome. De ce fait certaines explications et les références correspondants sont repétés à plusieurs reprises dans le corps de la these.

Notice

Except the general introduction, the survey of litterature and the conclusion, all chapters of this thesis are self-containing research articles. Moreover, some explanations, like corresponding literature, are repeated at different times in the thesis.

General Contents

Résumé

Selon Moore (1911), « plus la taille de l'établissement augmente, plus s'améliore la condition de l'ouvrier dans toutes les dimensions». De même Oi & Idson (1999) ont estimé qu' « un travailleur, qui est titulaire d'un emploi dans une grande entreprise, est payé un salaire plus élevé, reçoit des avantages sociaux plus généreux, obtient plus de formation et évolue dans un environnement de travail plus propre, plus sûr et généralement plus agréable ».

Observer que des personnes ayant des caractéristiques économiques équivalentes reçoivent des rémunérations différentes pour des prestations équivalentes toujours été un casse-tête pour les économistes du travail. Les économistes considèrent que si les travailleurs et les postes qu'ils occupent sont similaires et qu'il existe de la mobilité des travailleurs entre les entreprises et les postes, les salaires devraient être les mêmes. La diversité des emplois et des entreprises augmente la complexité de la relation employeur-employé. Selon Oi & Idson (1999), le marché du travail n'est pas un lieu d'échange unique. Il y a une multiplicité de « marchés » dans lequel un emploi est défini par une relation travailleur-entreprise. Les emplois diffèrent non seulement parles tâches que l'employé doit effectuer, mais aussi par les obligations d'un employeur doit à ses employés. Ainsi, la taille de l'entreprise-employeur n'est pas le seul facteur déterminant des salaires. Ces derniers devraient, au contraire, être déterminés par la diversité des emplois et des salariés à la recherche de ces emplois. Cependant, on observe des salaires différents pour des employés semblables dans des entreprises de tailles différentes après prise en compte des deux caractéristiques de l'employeur-employé. Ceci signifie que quelque chose manque pour expliquer cette relation employeur-employé. Cette question reste, à ce jour, non résolue. Pour comprendre le rôle des agents du marché du travail, il est important de connaître la relation causale entre la taille de l'employeur et les salaires des travailleurs de cet établissement. Cette thèse a pour objectif d’investiguer cette relation causale qui lie la taille de l'entreprise employeur et le salaire.

La « différence de salaire par taille » représente la prime salariale versée à des employés travaillant dans des établissements de grande taille après avoir contrôlé pour l'individu et l'employeur des caractéristiques communes à tous les employés de cette entreprise ou établissement. Les caractéristiques des employeurs et des employés ainsi que leurs interactions sont importantes pour identifier la relation positive entre la taille et les salaires. Ainsi, il est très important pour les économistes du travail d'identifier comment les caractéristiques des employeurs et des employés interagissent pour créer la prime de salaire par taille de l’employeur (désormais PSTE).

Le cas du marché du travail concurrentiel est hypothétique. C’est, par conséquent, un défi pour les économistes du travail de formuler des politiques économiques pour aborder les cas d'un marché du travail non concurrentiel. L'hétérogénéité des travailleurs et des employeurs influencent les salaires à un niveau qui dépasse les conclusions des théories traditionnelles de la fixation des salaires. Il existe une vaste littérature sur les écarts de salaires entre les employés aux caractéristiques semblables. Ces différences sont surtout étudiées dans le contexte de différences de sexe, de race et de l'ethnicité. Une autre partie de la littérature se concentre sur l'inégalité des salaires selon la taille de l'employeur. Ce domaine est particulièrement important pour les économistes du travail afin de formuler des politiques liées à la rémunération, aux emplois, à l’industrialisation et aux parcours professionnels.

Alors que l'écart salarial par taille est considéré comme le résultat de la corrélation entre les caractéristiques des employeurs et celles des employés avec la taille des entreprises ou des établissements, les raisons de l'impact de la taille sur les salaires ne font pas l’objet d’un consensus parmi les économistes. Différentes hypothèses ont été formulées et testées afin de déterminer l'ampleur et les causes de la PSTE. Celle-ci est expliquée soit en termes de disposition des grandes entreprises à embaucher des travailleurs de haute qualité, soit par une nécessité de compenser des conditions de travail plus pénibles dans les grandes entreprises, ou encore comme un salaire d'efficience permettant d'accroître la productivité des travailleurs et d'inciter à l'effort au travail, ou d'éviter les coûts de surveillance et de contrôle. En outre, l'existence d'un écart salarial-taille peut être attribué à des tentatives visant à éviter la syndicalisation ou, plus simplement, octroyé par des employeurs plus rentables et monopolisés. Toutes les études ont révélé qu'aucune des variables explicatives (à la droite de l'équation), qu'elles soient liées à l'employeur, ou aux caractéristiques des employés, n'explique convenablement l'écart salarial par taille. Par conséquent, il est généralement considéré comme un facteur non mesuré dans le terme d'erreur. Tout ceci traduit que la PSTE est un puzzle non résolu. Les explications théoriques sont décrites en détail dans le chapitre-1.

Toutes les hypothèses théoriques sur la PSTE reposent soit sur les différentiels compensateurs associés aux caractéristiques de l'employeur et de l'emploi soit sur la qualité mesurée ou non mesurée de la main-d'œuvre (2000 Criscuolo). Les explications néoclassiques se concentrent sur la qualité du travail et des conditions de travail, tandis que les explications institutionnelles se tournent vers des facteurs, tels que, le pouvoir de marché et la volonté d'empêcher la syndicalisation. Brown et Medoff (1989) ont conclu que « la différence de salaire liée à la taille de l'entreprise semble être importante et omniprésente ; mais notre analyse nous laisse mal à l'aise car nous sommes incapables de l'expliquer, ou du moins la partie de la différence qui n'est pas expliquée par des indicateurs observables de la qualité du travail. » Ni la compensation différentielle pour les conditions de travail, ni le pouvoir sur le marché des produits ni la menace de syndicalisation n'ont pu expliquer la prime de taille.

Cette thèse étudie la relation employeur-employé selon la taille de l'employeur pour le marché du travail français en utilisant des données de coupe transversale de l'enquête sur le cout du travail et des salaires (appelé ECMOSS) de 1992 à 2006. Le marché du travail français est dual par nature. Le système de fixation des salaires est particulièrement complexe car il dépend simultanément de la politique salariale nationale et des négociations collectives à d'autres niveaux. De plus, la question de la sélection n'a pas été étudiée en détail en utilisant des données transversales faisant correspondre les employeurs et les employés

Avant d'aller plus loin, il est important de se pencher sur deux difficultés. Premièrement, comment mesurer la taille : par le nombre d'emplois, la quantité d'actifs ou le volume des ventes ? La taille mesurée par le nombre d'emplois a été utilisée dans la littérature en raison de sa disponibilité dans les données bien que la productivité du travail puisse varier à travers le temps et les entreprises ce qui a tendance à rendre cette mesure inappropriée. Celle-ci reste cependant meilleure qu'un ratio capital/travail ou que le volume des ventes comme Oi & Idson (1999) l'ont mis en évidence : le ratio capital/travail varie largement entre les industries en raison de différences au niveau des technologies de production et les comparaisons inter temporelles entre industries sont compliquées lorsqu'on utilise les chiffres de vente comme mesure de la taille. Par conséquent, tout en ayanten tête le problème de la disponibilité des données, nous avons adopté l'emploi comme mesure de la taille. Deuxièmement, quel est le niveau de l'unité approprié pour l'analyse? Nous avons utilisé des données au niveau de l'établissement plutôt qu'à celui de la firme en suivant un raisonnement en ligne avec Weiss (1966). Premièrement, la taille de l'employeur semble principalement importer de par son impact au niveau du marché local qui dépend de la taille de l'établissement plutôt que l'entreprise entière. Deuxièmement, les établissements sont assignés à une industrie de façon plus précise. Enfin, les données au niveau de l'établissement sont disponibles pour un nombre plus important d'industries qu' au niveau de la firme.

Parmi les autres facteurs, le choix de la technique de production, les coûts fixes et/ou les différences en termes de localisation peuvent créer de l'hétérogénéité technologique parmi les industries et parmi les usines dans une même industrie. Diverses sources d'hétérogénéité technologique peuvent induire une sélection par les capacités des travailleurs, ce qui mène à une différence en termes de salaire au travers des établissements d'une même industrie. Les mécanismes de sélection qui affectent l'éventail des compétences de façon différenciée par les employeurs génèrent des différences en termes de dispersion par classe de taille. Davis and Haltiwanger (1996) and Oi (1999) avancent que de grandes entreprises engagent des travailleurs avec des compétences plus pointues et d'un niveau plus homogène car les employeurs de grande taille dépendent plus intensivement de technologies de production standardisées. Garen (1985) va dans le même sens en avançant que la dispersion salariale intra-site baisse avec la taille du site en raison de coûts de détection et de surveillance plus élevés impliquant une plus faible sensibilité des salaires au niveau de compétence pour les plus grands sites d'exploitation. Par conséquent, plusieurs facteurs associés avec la taille peuvent affecter la structure salariale à l'intérieur d'un même établissement comme entre établissements. Ici, nous nous concentrons sur les différentiels de salaire inter établissement et laissons de côté les dispersions intra établissement.

- **Auto Sélection et Tri non aléatoire**

La principale difficulté économétrique dans l'explication de la relation qui lie la taille de l'employeur et le salaire provient de la sélection non aléatoire des travailleurs dans des établissements de tailles différentes. Pendant longtemps, ce problème a été ignoré dans la littérature : de nombreuses études étaient simplement basées sur des régressions OLS afin d'estimer la prime de salaire en ignorant le phénomène de sélection non aléatoire des travailleurs. Un problème commun dans l'estimation empirique de données transversales

(cross-section data) se trouve dans l'existence d'une endogénéité potentielle des variables explicatives du côté droit de l'équation. L'estimateur des moindres carrés ordinaires mène à des estimations biaisées et inconsistantes en présence d'un biais d'endogénéité. Les sources de biais potentielles incluent les variables omises, la simultanéité et les erreurs de mesures (Wooldridge, 2002).

La source la plus importante d'endogénéité dans la relation taille-salaire est le biais de sélection généré par un biais de variable omise. Ceci se produit lorsque des observations potentielles restent non observées. Par exemple, lorsque des individus travaillant dans un établissement de grande taille ne sont pas représentés par un échantillon aléatoire de toute la population. Cette exclusion d'observation potentielle peut causer le biais de l'estimation OLS des paramètres du modèle. Si les différences non-observables en productivité affectent l'allocation, négliger cet élément pourrait biaiser les estimations des effets de la tailles de l'établissement sur les salaires.

Le problème du biais de sélection survient lorsqu'on utilise un échantillon non aléatoire de la population générale et que l'on en infère ce qu'un individu moyen aurait expérimenté s'il avait été sélectionné aléatoirement. Lorsqu'on a affaire à l'effet causal de travailler dans un établissement de grande taille, il convient de faire avec soin une distinction entre la modélisation des effets potentiels et les méthodes empiriques nécessaires à l'identification de ces effets. La relation causale entre la taille de l'employeur et les salaires demande une analyse appropriée des facteurs observables et non observables dans les équations de résultat et de sélection afin d'identifier le modèle causal à partir des données et de clarifier la nature des hypothèses d'identification car ce sont ces facteurs non observables qui donnent naissance au biais de sélection dans l'analyse causale. Afin de résoudre ce biais de sélection, il convient de prendre en compte à la fois le traitement du résultat et celui des mécanismes de choix. Le statut de traitement consiste à « être dans un établissement de grande taille ». Par conséquent, afin d'estimer les paramètres avec ces mécanismes, les facteurs non observables dans le résultat et les équations de choix de traitement doivent être pris en compte.

Nous ne prétendons pas que cette thèse de doctorat va combler les lacunes dans la littérature pour le cas de la France. Cependant, certaines découvertes intéressantes peuvent ouvrir la voie vers de nouvelles avancées. Cette thèse contribue à l'étude de la relation taille-salaire pour le marché du travail français par l'utilisation de données transversales (cross-section) utilisée dans l'optique d'observer l'importance et les sources de la prime salariale octroyée dans les établissements de grande taille en France. La principale difficulté d'une telle démarche consiste à gérer le problème de biais de sélection associé à la relation taille-salaire. Des méthodes différentes sont employées afin d'étudier la sélection non aléatoire des travailleurs dans des établissements de tailles différentes. Nous testons des techniques d'appariement (Matching) à travers l'utilisation du score de propension (Propensity score) ainsi que des effets fixes de cohorte (Cohort fixed effects) et des méthodes de différence première sont estimées afin de générer des données en pseudo-panel à partir de la répétition de données transversales.

- **Données**

Les données sont tirées de deux séries d'enquêtes ; l'une sur le coût du travail (ECMO) et l'autre sur la structure des salaires (ESS), conjointement appelées ECMOSS (enquête sur le coût du travail et la structure des salaires) pour le secteur non agricole privé collectées par l'INSEE. Par conséquent, les résultats des estimations reflètent la structure des salaires et l'impact de la taille des employeurs sur les salaires dans le secteur privé. Les implications pour le secteur public peuvent être différentes. L'enquête vise à examiner la structure de rémunération et de temps de travail des employés. Elle fournit des informations détaillées pour analyser les disparités salariales. Elle se compose de deux parties: une est l'enquête sur la structure des salaires (appelée ESS), et l'autre présente la mesure et la décomposition des coûts de main-d'œuvre (appelée ECMO).

L'enquête sur le coût de la main-d'œuvre (ECMO) est conçue pour examiner d'un côté, un coût annuel moyen de la main-d'œuvre par salarié, et de l'autre, un coût horaire moyen de la main-d'œuvre par heure effectivement travaillée. Elle fournit la structure détaillée de ces coûts par secteur, par activité économique, par région d'implantation désétablissements par taille de l'établissement. L'enquête sur la structure des salaires (ESS) vise à fournir, pour un échantillon d'employés, des données individuelles sur : les salaires, ses composantes, les déterminants des salaires et les caractéristiques du milieu du travail. Pour cette thèse, nous avons utilisé des enquêtes pour les années 1992, 1994, 2002, 2005 et 2006. Il n'y a pas de données compilées disponibles entre 2006 et maintenant. Les descriptions détaillées des données ainsi que la définition et la construction des variables sont présentées dans le chapitre -2 de statistiques descriptives.

L'échantillon est constitué de 14.000 établissements et environ 140.000 employés du secteur privé non agricole. Les employeurs répondent à un questionnaire décrivant de nombreuses caractéristiques du milieu du travail et donnent des informations sur un échantillon aléatoire de leurs employés. Cet ensemble de données comporte de nombreuses observations et une grande variété de caractéristiques de l'employeur et de l'employé qui peuvent être utilisés comme des instruments intéressants pour les recherches sur le coût du travail et la structure des salaires.

Il s'agit d'une base de données très riche comprenant des caractéristiques socio-économiques des travailleurs ainsi que les caractéristiques des établissements. Il n'existe aucun autre ensemble de données qui fournit en même temps les informations sur la taille de l'établissement, sa principale activité, sa situation géographique, sa structure salariale et la composition de ses salaires. En outre, on peut trouver des informations détaillées sur l'éducation, la profession, la répartition par secteur, l'âge, la nationalité, la situation familiale et le nombre d'enfants à charge des travailleurs.

Les principales observations des données sont les suivantes :

- Le nombre d'observations pour chaque année, après avoir enlevé les valeurs aberrantes dans diverses enquêtes, est de 74 696 (1992), 30 216 (2002), 51 272 (2005) et 53 508 (2006).
- Le salaire horaire moyen à diverses enquêtes est 71francs (1992), 17euros (2002), 20euros (2005) et 21euros (2006).
- Le salaire horaire moyen augmente avec la taille de l'employeur. La différence de salaire horaire moyen selon la taille de l'employeur devient plus importante lorsque la limite de taille passe à 50 ou à 200 employés.
- La population des hommes par rapport aux femmes, dans l'échantillon, est d'environ 60 % dans tous les sondages. La différence est plus grande dans l'ensemble par rapport aux établissements de petite et moyenne taille. Les salaires en moyenne pour les travailleurs de sexe masculin sont plus élevés que ceux des travailleurs de sexe féminin et l'écart est d'autant plus important chez les employeurs de grande taille par rapport aux établissements de petite et moyenne taille. Une plus forte proportion de femmes travaille dans les établissements de petite et de moyenne taille .Pour les travailleurs de sexe masculin, la proportion la plus élevée est associée aux employeurs de moyenne et de grande taille.

- Le salaire horaire moyen des travailleurs mariés est plus élevé que ceux des célibataires ou de ceux qui ont un autre statut familial. La proportion de travailleurs mariés est plus importante chez les gros employeurs par rapport aux établissements de taille petite ou moyenne. Le salaire horaire moyen pour un travailleur féminin célibataire est supérieur à celui d'un travailleur féminin marié. L'inverse est vrai pour les travailleurs de sexe masculin.
- Le secteur le plus important est le secteur des services. A l'inverse, le secteur le plus petit est le commerce. Dans les établissements de grande taille, le secteur le plus important est le secteur manufacturier (sauf dans le sondage de 2005 où le secteur le plus important est, par établissements, le secteur des services), alors que, dans les établissements de taille petite et moyenne, le secteur le plus important est le secteur des services.
- Entre secteurs, le salaire horaire moyen est plus élevé dans le secteur manufacturier par rapport aux secteurs des services et du commerce, sauf en 2002 où le salaire horaire moyen dans le secteur des services est plus élevé comparé aux secteurs de l'industrie manufacturière et du commerce. Dans les grands établissements, le salaire horaire moyen le plus élevé est dans le secteur manufacturier pour les années1992 et 2006. Pour l'année 2005, les salaires horaires moyesn dans l'industrie manufacturière et les services sont identiques.
- Le niveau d'éducation s'est amélioré en moyenne sur les années d'enquête. Entre 2002 et 2006, la majorité de la population dans l'échantillon a acquis le niveau d'éducation supérieure à la différence de 1992, où la majorité de la population détiennent l'enseignement technique court et primaire.
- Les augmentations moyennes de salaire horaire avec le niveau d'éducation restent plus élevées pour les établissements de grande taille. Les petits établissements ont

une plus grande capacité à absorber les personnes peu instruites que les grands employeurs.

- Dans les établissements de petite taille, les professions les plus nombreuses sont les cols bleus et cols blancs peu qualifiés, tandis que, dans les établissements de grande taille les principales professions sont les cols bleus et cols blancs hautement qualifiés. Dans l'enquête de 2005 et 2006, la majorité de la population dans l'échantillon fait partie des professions liés à la gestion et des professions intellectuelles supérieures suivies par des emplois cols blancs hautement qualifiés entre établissements.
- Le salaire horaire moyen dépend de la profession. Pour les professions de niveau supérieur, le salaire horaire est plus élevé. Pour toutes les professions, le salaire horaire dans les établissements de grande taille est supérieur.
- L'Ile de France est la région où est installée la plus grande proportion d'établissements. Elle est suivie par la région Rhône Alpes et la Méditerranée.
- Parmi toutes les régions, le salaire horaire dans les établissements de grande taille est plus élevé. La différence de salaire horaire entre les petites et moyennes entreprises est faible dans les régions intermédiaires.
- L'âge moyen de la population dans l'échantillon varie entre39 et 42 ans selon les enquêtes. La majorité de la population se situe entre 31 et 40 ans d'une part, et d'autre part entre 41 et 50 ans d'âge. Dans chaque groupe d'âge, les augmentations de salaire horaire varient avec la taille de l'employeur. Le salaire horaire moyen augmente proportionnellement à l'âge.
- L'ancienneté moyenne dans l'établissement est de 11-12 ans selon les enquêtes. Une majorité des personnes de l'échantillon a passé entre 0-5 ans dans l'emploi en

cours, alors que, dans les établissements de grande taille le plus grand échantillon de travailleurs ont passé de 11 à 20 ans dans l'emploi actuel.

- La majorité des contrats dans toutes les enquêtes et dans tous les établissements sont fixés dans les contrats de travail à durée déterminée. Le salaire horaire moyen du même type de contrat dans son ensemble est plus élevé par rapport aux moyennes et petites entreprises.
- L'expérience du marché du travail total moyen est de 20 ans dans les données.

Objectifs et principaux résultats

Les objectifs de cette thèse sont multiples. La thèse vise à répondre aux questions suivantes. (Dans chaque chapitre les objectifs connexes sont décrits en détail):

- Quelle est la relation causale entre la taille de l'employeur et le salaire de l'employé ?
- Quelle est l'ampleur de l'écart de taille-salaire employeur en France avec des données au niveau des établissements ?
- Quelle est l'ampleur de la PSTE en présence de sélection non aléatoire ? Comment les caractéristiques observées et non observées des travailleurs sont récompensés par la taille de l'employeur ?
- Quel modèle de sélection est le plus approprié ?
- Est-ce que l'écart salarial-taille est, en fait, un écart de rémunération entre hommes et femmes? Quels sont les facteurs constituants cet écart salarial et quelle est la part des composants expliqués et inexpliquées dans l'écart de rémunération entre les sexes ?

- Quelles sont les implications pour les pseudos méthodes d'estimation de panel dans la littérature reliant la taille et le salaire ?

La thèse se compose de six chapitres : chaque chapitre s'inscrit dans le prolongement du précédent. Le premier chapitre passe en revue la littérature sur la relation entre l'employeur, la taille et les salaires. Le deuxième chapitre explore les données et présente des statistiques descriptives. Dans le troisième chapitre, le différentiel de salaire global est identifié après avoir tenu compte du biais dû aux caractéristiques observables des travailleurs qui sont mis en correspondance dans les établissements de grande et de petite taille et la différence de salaire est analysée. Ce chapitre traite également de l'importance relative des bonus, du paiement d'heures supplémentaires et des indemnités par comparaison entre le salaire horaire brut et le salaire de base. Le troisième chapitre nous laisse avec le problème de biais de sélection non résolu ou tri non aléatoire des travailleurs. Donc, dans le quatrième chapitre, un modèle de maximum de vraisemblance à avec information complète est présenté :une sélection non aléatoire des travailleurs est étudiées l'ampleur de la PSTE est analysée. En outre, sont également étudiés, par groupe de taille d'établissement, la rémunération pour compétences mesurées et non mesurés. Dans le cinquième chapitre l'écart salarial est décomposé en composante expliquée et inexpliquée dans les grands et les petits établissements. Enfin, une analyse de cohorte est présentée dans le chapitre six, en utilisant les ensembles de données employeur-employé appariée pour estimer les effets fixes. L'objectif principal de la thèse est l'étude de données transversales et des méthodes, mais dans la dernière partie est analysée la possibilité de construire des Pseudo-données de panel à l'aide de la méthode de Deaton.

Ci-dessous une brève description des quatre articles (chapitre 3-6) est présentée.

• La PSTE : une analyse à l'aide de la méthode de l'appariement par le score (Propensity Score Matching PSM)

Dans ce chapitre, on appliquera la méthode des moindres carrés ordinaires(MCO) et de l'appariement par le score (PSM). L'objectif est d'estimer l'ampleur de la PSTE en contrôlant les caractéristiques observables de l'individu et de l'employeur et d'étudier la différence de salaire des personnes travaillant dans les établissements de grande taille avec les travailleurs appariés travaillant dans les établissements de petite taille. Les MCO avec salaire horaire brut donnent environ 7% de l'écart salarial par rapport à la variable dummy de grand établissement de taille, tandis que la prime disparaît presque à 0,8% pour le salaire horaire net L'appariement conditionné sur le score de propension des covariables observées réduit la différence de salaires entre les personnes qui travaillent dans des établissements de grande et de petite taille. Une différence de salaire avant appariement de 17 % entre grands et petits et une différence de salaire après appariement de 9 % pour le salaire horaire brut sont trouvés. Pour le salaire horaire de base, le PSM montre une différence de salaire moyen de 8 % et de 2,5 % après appariement. La différence de salaire entre les établissements de grande et de petite taille est principalement due à la forte proportion de travailleurs du sexe masculin, qui travaillent dans le secteur manufacturier et dans les emplois de cols bleus. Les résultats sont robustes à différents algorithmes d'appariement. La prime de salaire employeur taille est plus le résultat de la politique de rémunération de l'employeur, que la différence de salaire elle-même. Les résultats montrent que les préférences des "gros" employeurs pour les travailleurs ayant une capacité de gain élevée et des caractéristiques distinctes pour des établissements de grande taille déterminent la prime salariale qui tient compte de l'hétérogénéité de l'employeur pour récompenser les travailleurs similaires différemment.

L’objectif de ce chapitre est d'estimer l'ampleur et les sources de l'écart salarial des employés travaillant dans la zone de grande taille comparées à d’autres employés semblables travaillant dans les établissements de petite taille. OLS et PSM sont utilisés pour identifier des écarts salariaux attribuables aux caractéristiques des travailleurs et des employeurs.

Les principales conclusions sont les suivantes :

- Les MCO montrent un effet de taille de 7 % de salaire horaire brut en contrôlant pour les caractéristiques de l'individu et de l'employeur. D'un autre côté, PSM montre une différence de salaire brut moyen de 17 % avant appariement à comparer avec une différence de 9 % après appariement entre les grands et les petits établissements. Les MCO sous prédisent la relation entre la taille de l'employeur et le salaire. Ainsi, l'analyse de régression avec l'hypothèse de linéarité mais au-delà du support commun sous prédit la prime salariale versée par les grands établissements.
- L'appariement conditionné sur le score de propension des covariables observées réduit la différence de salaire moyen entre ceux qui travaillent dans des établissements de grande taille et ceux qui travaillent dans les établissements de petite taille. Ceci reflète que les gros employeurs apprécient plus les caractéristiques observables. 50 % de la différence de salaire moyen s'explique par des différences dans les caractéristiques observables des travailleurs du Grand établissement, alors que, la différence de salaires nets est entièrement la différence entre des caractéristiques observées entre les grands et les petits établissements.
- De ces deux mesures de salaires, deux aspects évidents émergent: premièrement, les gros employeurs préfèrent des travailleurs avec une grande capacité de gain.

Deuxièmement, les établissements de grande taille déterminent la prime salariale sous forme de compensations et payent des paquets (primes, indemnités, sur les paiements de temps etc.).Nous pouvons dire que la prime de salaire employeur par taille est surtout le résultat de la politique de rémunération de l'employeur plus que la différence de salaire elle-même.

- L'effet de prime de salaire de taille de l'employeur en fonction des caractéristiques observables est plus évident pour les hommes, dans des emplois de cols bleus et dans les grandes industries.
- La différence de salaires par taille de l'employeur est plus présente dans les professions de niveau inférieur par rapport aux professions de niveau supérieures.
- Les informations disponibles reliées aux syndicats et aux conventions collectives de travail ne sont pas suffisantes pour déterminer le rôle des conventions collectives dans l'écart de taille-salaire employeur.
- La différence de salaire brut et net peut expliquer la différence restante, mais la mesure de la politique de rémunération davantage liée à la taille de l'employeur n'est pas claire.

Le PSM a été largement utilisé dans plusieurs types de recherches d'évaluation de programme. Néanmoins, cette technique fonctionne avec certaines restrictions et limitations. Comme Heckman et al.1997 ont montré PSM peut récupérer uniquement les effets moyens et ne peut pas répondre aux questions relatives aux effets redistribuais du programme, tels que le pourcentage des participants au programme qui en bénéficient.

En outre, PSM ne peut pas estimer l'effet du traitement moyen local (‘local average treatment effect’LATE), qui est l'incidence moyenne du programme sur ceux dont l'état de participation change en raison d'un changement de politique. Les méthodes de variables

instrumentales peuvent supprimer des biais de sélection en raison de la confusion observées et non observées mais des instruments plausibles ne sont pas disponibles dans de nombreuses circonstances. Les méthodes du Score de propension (PS) et de régression peuvent traiter les biais de sélection sous « unconfoundness », mais ils supposent également que la forme fonctionnelle de la régression ou le taux de propension sont correctement spécifiés.

La poursuite des travaux peut être faite pour traiter les caractéristiques non prises en charge des travailleurs qui entraînent la sélection non aléatoire des travailleurs dans l'ensemble des employeurs de tailles différentes. MCO et PSM ne résolvent pas le problème d'endogénéité ni le lien de causalité de la taille de l'employeur et le salaire. Il est nécessaire d'estimer le LATE grâce à des modèles de sélection qui traitent de sélection non aléatoire entre des groupes de taille.

• Sélection non aléatoire et prime taille-rémunération de l'employeur dans les établissements français

Ce chapitre prend en compte la sélection non-aléatoire des travailleurs hétérogènes dans les employeurs de différentes tailles. Une estimation conjointe de la fonction de maximum de vraisemblance et des modèles de switching-régression sont utilisées pour cette étude pour voir l'ampleur et les sources de l'écart taille- rémunération sur le marché du travail français. Les résultats montrent une sélection négative dans l'établissement de grande taille et positive dans l'établissement de petite taille ce qui semble indiquer que la rémunération conditionnelle est inférieure au salaire inconditionnel pour le travailleur d'établissement de grande taille et vice versa. On en conclut que la non prise en charge des facteurs non-observables ont de fortes incidences dans les établissements de petite taille alors que les facteurs observés ont de fortes incidences dans les établissements de grande taille. La prime

salariale et l'effet de sélection existent pour les salaires horaires bruts et pour les travailleurs de sexe masculin.

Ce chapitre tente d'expliquer la sélection non aléatoire des travailleurs employeurs de différentes tailles afin d'examiner l'écart de taille-rémunération dans les établissements français. Des méthodes d'estimation différentes telles que la méthode de maximum de vraisemblance à information complète (FIML) et des procédures d'estimation en deux étapes d'Heckman sont utilisées pour analyser et comparer les résultats de salaire selon la taille de l'employeur. Bien que nous n'étions pas en mesure d'obtenir des données de panel ou d'instruments parfaits, l'impact des biais de sélection n'est ni laissé au hasard ni ignoré mais, au contraire, il est explicitement utilisé et modélisé dans l'équation d'estimation du salaire horaire.

Les principales conclusions sont les suivantes :

- Les modalités de sélection dans le modèle Heckman ne sont généralement pas significatives tandis que le FIML montre la forte corrélation négative pour toute la population et pour l'échantillon des hommes. Deux choses ressortent de cette analyse ; le choix du modèle de sélection contient les emplois maximaux et les caractéristiques observables qui peuvent rendre l'effet de sélection non significatif dans certains cas ou le choix des instruments n'est pas bon pour contrôler l'hétérogénéité non observée.
- L'effet de prime et de sélection de taille est élevé chez les travailleurs de sexe masculin dans la taille différente des établissements. Chez les travailleuses, la composante de sélection n'est pas significative pour le plus grand groupe.
- Une sélection négative dans un établissement de grande taille et positive dans un établissement de petite taille est trouvée, ce qui semble indiquer que le salaire inconditionnel est inférieur à la rémunération conditionnelle pour le travailleur

d'établissement de petite taille. En revanche, le salaire conditionnel est inférieur au salaire inconditionnel pour les employés travaillant dans l'établissement de grande taille. La non prise en charge des facteurs inobservables aune grande incidence dans les établissements de petite taille alors que les facteurs observés ont de grandes incidences dans les établissements de grande taille.

- Il y a un fort besoin de tester le rôle des syndicats pour étudier les comportements de gros employeurs. Ce chapitre ne parvient pas à analyser le rôle du système de négociation collective de la France avec l'information fournie. Les données ne nous permettent pas de tester les théories de l'écart de salaire employeur taille, mais les résultats sont dans la même ligne que d'autres études dans la littérature de taille-salaire (Idson et Feaster 1990 et Lluis 2008).
- Généralement le FIML est considéré comme plus efficace par rapport à Heckman, mais les modèles de switching-régression ont été très populaires dans la littérature de PSTE traitant de la sélection ce qui nous permet d'observer les salaires sous différents régimes avec effet de sélection. Par conséquent, nous pouvons conclure que les deux modèles sont équivalents et peuvent être utilisés côte à côte, en particulier dans les travaux de recherche appliquée.

Compte tenu de l'importance de la correction pour les biais de sélection dans la littérature de salaire-taille, l'absence de conclusions définitives est préoccupante. Plusieurs recherches devront mettre l'accent sur la relation de causalité de l'écart de salaire employeur taille. La question de déterminer les forces et les faiblesses de chaque méthode et les conditions optimales dans lesquelles chaque méthode doit être utilisée reste sans réponse.

Le Prochain chapitre décompose les différences de salaire entre les sexes dans toutes les catégories de taille afin de savoir si l'écart salarial de taille est en fait un écart salarial par genre

• Décomposition de salaire entre les sexes différenciés selon la taille de l'employeur

La décomposition des écarts salariaux a été étudiée par de nombreux auteurs dans le contexte de différence de sexe, de race, d'origine ethnique, etc. Mais les écarts salariaux en décomposition selon la taille de l'employeur n'ont pas été étudiés en détail. La procédure d'estimation de Heckman en deux étapes est utilisée pour identifier les paramètres et ensuite la décomposition de salaire selon les procédures d'Oaxaca (1973) Blinder (1973) est appliquée pour les équations de régression. L'objectif consiste à décomposer l'écart de salaire entre les sexes selon la taille de l'employeur afin de comparer les profils de l'écart salarial entre les sexes dans les différentes tailles des employeurs. La ségrégation du lieu de travail est considérée et l'effet des différences de caractéristiques personnelles sur l'écart salarial est décortiqué avec l'effet de sélection dans les différents établissements des femmes et des hommes.

L'écart salarial est plus grand dans les grands établissements par rapport aux établissements de petite taille, mais dans tous les cas, la plus grande partie de l'écart salarial demeure inexpliquée. Les régressions ajustées pour sélection augmentent écart salarial en grande taille mais la preuve de la sélection aléatoire se trouve seulement chez les travailleurs masculins et aucune modalité de sélection n'est importante pour les femmes. L'écart salarial existe dans les deux mesures du salaire, brut et basique, l'écart salarial augmente la différence des augmentations de salaire brut et su salaire de base.

Une conclusion importante de cette analyse est la ségrégation des femmes dans les lieux de travail à basses rémunérations. Il y a des stéréotypes dans l'allocation des femmes dans des emplois particuliers qui résulte en des salaires peu élevés par rapport aux hommes. Le comportement des employeurs est discriminatoire envers les femmes et ils offrent aux hommes un avantage indu dans le même travail. Deux facteurs sont importants pour expliquer l'écart de rémunération entre les sexes selon la taille de l'employeur. Tout d'abord la ségrégation des femmes dans les lieux de travail à basses rémunérations; la sélection des employés basée sur des stéréotypes entrave l'évolution des carrières des femmes. Deuxièmement, le comportement de l'employeur est discriminatoire envers les femmes. Les femmes sont surreprésentées dans les emplois faiblement rémunérés: il y a prévalence de la ségrégation horizontale qui conduit à des salaires bas et augmente l'écart salarial entre les sexes. Dans une certaine mesure il y a aussi la ségrégation verticale comme seules les femmes sont sous-représentées dans les professions à salaire élevés.

L'étape suivante est la décomposition par cohortes d'âge et par durée de travail des deux groupes . Des travaux supplémentaires devraient examiner des méthodes de décomposition différentes, la décomposition de quintile et avec différents réglages de correction de la sélection, comme l'option « Oaxaca » (sous Stata) ne permet pas d'utiliser de nombreux types de décompositions. Une grande partie de l'écart salarial demeure inexpliquée. La poursuite des travaux peuvent également être faite pour tenir compte en même temps de la ségrégation professionnelle, de la ségrégation sur le lieu de travail- et décomposer les écarts de salaire entre les sexes

- **Une estimation de pseudo-panel de l'écart taille- rémunération en présence de biais de sélection**

Une analyse de cohorte, générant des Pseudo-données de panel pour faire face à l'hétérogénéité individuelle non observée, est utilisée dans de nombreuses disciplines. La même méthode est utilisée dans ce chapitre pour examiner la relation entre la taille de l'employeur et les salaires des employés sur le marché du travail français. Des coupes successives de groupe de données appariées employeurs-employés tirées de l'ECMOSS 2005-06 pour la France sont utilisés pour la construction de cohorte et l'ECMOSS 2002 est ajouté ensuite afin de comparer les résultats. Le chapitre présente les effets fixes, les effets aléatoires et les premières estimations de la différence d'impact de taille sur le salaire basé sur l'analyse de la cohorte des travailleurs qui sont nés entre 1928 et 1986. Quatre cohortes différentes de l'ensemble de données sont générées en prenant les différents groupes de caractéristiques individuelles invariantes dans le temps. Les résultats soutiennent systématiquement l'hypothèse qu'au sein d'une cohorte, il y a un effet positif et significatif de la taille de l'employeur sur le salaire. L'impact est fort dans les établissements de moyenne taille. La méthode d'estimation préférée est celle des effets fixes comme l'hétérogénéité non observée est la principale source de biais de sélection. Les Résultats peuvent être améliorés en augmentant le nombre d'années.

Ce chapitre montre comment des Pseudo-données de panel peuvent être construites pour remédier au manque de vraies données de panel. L'analyse de cohorte est présentée afin d'étudier la relation causale entre la taille et les salaires. Générer un ensemble de données agrégées de panel basées sur un individu constant ou sur des caractéristiques des employeurs montrent des constatations intéressantes. Quatre différents types d'ensembles de données agrégées sont générés selon les caractéristiques individuelles qui ne changent pas au fil du temps. Les résultats dans chacun des types de données sont différents de l'autre comme prévu.

Les résultats soutiennent systématiquement l'hypothèse qu'au sein d'une cohorte, il y a un effet positif et significatif de la taille de l'employeur sur le salaire. L'impact est fort dans les établissements de taille moyenne. Comme la taille des employeurs augmentent, les salaires augmentent et cet effet reste important même après le contrôle des caractéristiques de l'employeur et l'employé.

Une comparaison entre le modèle à effets aléatoires et modèle à effets fixes a révélé que les effets fixes étaient statistiquement significatifs. Par conséquent, le modèle à effets fixes forme le noyau de notre analyse. Des éclairages supplémentaires ont été fournis par les résultats produits par les modèles transformation et des différences premières, qui ont utilisé des ensembles de données transformées. Les effets aléatoires, les effets fixes, entre les modèles de transformation et FD tous ont montré un effet significatif et positif de taille sur les salaires.

En utilisant toutes les quatre méthodes d'estimation, on ne saurait dire que cette méthode élimine les biais d'hétérogénéité inobservables des cohortes qui changent au fil du temps. Des cohortes de différentes données ont été produites et les résultats ont été comparés. Les Pseudo-données de panel préféré sont la combinaison du sexe, de région et de génération. L'importance, la direction et le signe des variables sont corrects. La méthode d'estimation préférée est celle des effets fixes. Cela correspond au fait que les effets fixes captent l'hétérogénéité en soustrayant la moyenne à ces données. Comme l'hétérogénéité non observée est la principale source de biais de sélection. Les résultats des effets aléatoires et des effets inter ne sont pas fiables. Ils exagèrent l'effet véritable et sont biaisés. Les effets inter sont biaisés en raison de variables latentes permanentes.

Afin de réaliser une meilleure estimation, on aurait besoin d'augmenter le nombre de cellules en ayant plus de données traversa les à des délais plus courts entre elles. En outre, il

serait également préférable de construire des pseudos cellules par différentes méthodes, par exemple, en utilisant la méthode du réseau neuronal ce qui permet de définir les cellules plus homogènes.

Conclusion

Cette thèse a pour objectif d'examiner pourquoi les gros employeurs paient des salaires plus élevés à des travailleurs similaires employés dans des établissements de taille plus petite quand l'assignation des travailleurs se fait de façon aléatoire. L'objectif principal est de déterminer l'ampleur de l'impact de la taille sur le salaire en présence de sélection et de tester différentes méthodes pour expliquer la relation entre la taille de l'employeur et le salaire. Les objectifs de cette thèse de doctorat sont multiples. Chaque chapitre amène de nouvelles questions auxquelles le chapitre suivant tente de répondre. De manière générale, tous les chapitres explorent la relation entre la taille de l'employeur et le salaire en présence d'un biais de sélection. Cette question de recherche n'est pas nouvelle en économie du travail, ni les méthodes économétriques utilisées. Cependant, il s'agit d'enrichir cette littérature sur le marché du travail français. En outre, à notre connaissance, les méthodes appliquées dans cette thèse n'ont jamais été utilisées auparavant pour étudier cette question, comme par exemple les méthodes PSM, FIML, pseudo panel etc.

Les éléments institutionnels du marché du travail français montrent que la taille de l'employeur est certainement un élément important des politiques d'emploi, de structure des salaires et d'emploi. Les conclusions sont pertinentes au sein de cette littérature sur la relation taille-écart salarial en France, car tout d'abord, peu d'études sur la question ont été menées et en second lieu, les biais de sélection d'échantillon ont été négligés dans les études utilisant des données transversales. C'est pourquoi, en France, la question fondamentale est de trouver

un bon équilibre entre le respect de l'autonomie du travail, les représentants de la direction et l'intervention de l'État pour compenser les défauts de cette autonomie. Cette thèse ne cherche pas à formuler ou proposer des changements de politique liés à la relation "employé-employeur" en France, mais plutôt à étudier la dynamique de cette relation qui peut aider à réorienter la politique.

Le texte est composé de six chapitres principaux, parmi lesquels quatre présentent les résultats des travaux de recherche. Le premier chapitre consiste en une revue de la littérature sur la relation entre la taille de l'employeur et les salaires. Le deuxième chapitre explore les données en présentant les statistiques descriptives des quatre coupes de l'enquête … (1992, 2002, 2005 et 2006) ainsi que des statistiques descriptives sur l'égalité des sexes. Dans le troisième chapitre, le différentiel de salaire global est identifié par la méthode des MCO puis la méthode « propensity score matching » où les employés des entreprises de grande taille sont jumelés à ceux qui travaillent dans les établissements de petite taille, en fonction des caractéristiques observables. Cela nous permet donc d'analyser l''écart salarial. Ce chapitre traite également de l'importance relative des bonus, du paiement d'heures supplémentaires et des indemnités par comparaison du salaire horaire brut et de base. Ce troisième chapitre n'a cependant pas permis de résoudre le problème du biais de sélection ou du tri non-aléatoire des travailleurs. Par conséquent, le quatrième chapitre va prendre en compte ce problème grâce à l'utilisation du modèle de maximum de vraisemblance avec information complète. Dans ce modèle, la sélection non-aléatoire des travailleurs est étudiée et l'ampleur de la prime salariale-taille est analysée. Des modèles à deux équations, modèle de sélection et modèle de salaires, sont analysés conjointement pour contrôler pour le maximum d'information disponible.

Les instruments pour les restrictions d'exclusion sont utilisés pour prédire la taille de l'établissement. Deux types d'instruments sont utilisés: le nombre d'enfants à charge en-dessous de dix-huit ans et l'interaction entre la taille de l'industrie et la région du travailleur où l'établissement est basé. En outre, les récompenses pour les compétences mesurées et non mesurées par groupe sont étudiés en utilisant un « Switching model » endogène où l'inverse du ratio de Mills est calculé par la procédure d'Heckman en deux étapes. Ensuite, nous nous demandons dans le cinquième chapitre si la différence de salaires n'est en fait pas un écart salarial de genre, et les salaires masculins et féminins sont comparés par catégories de taille différente. Enfin, l'analyse de cohorte est présenté dans le chapitre 6 en utilisant l'ensemble des données les plus récentes en coupe transversale afin d'estimer les effets fixes. Les effets fixes sont comparés avec les différences premières, les effets aléatoires et les modèles de transformation de données.

On observe que la PSTE prévaut encore avec différentes sources de données, différents lieux géographiques et différentes structures de composition de la main-d'œuvre. Différentes explications théoriques sont valables pour les différents pays selon l'état du marché du travail local. Parmi toutes ces études, en général, aucune attention n'a été accordée au problème d'endogénéité lié au fait qu'il existe une corrélation entre le terme d'erreur et la variable explicative. Il existe peu d'études ayant utilisé des données longitudinales afin de prendre en compte l'hétérogénéité des employeurs et des travailleurs, mais en revanche, la majorité des études en coupe ignorent le problème de biais de sélection. Les principaux facteurs non observés affectant la relation taille-salaire sont des facteurs de sélection qui entraînent une répartition non aléatoire des travailleurs selon la taille de l'employeur. Ceci rend à son tour la variable taille endogène dans l'équation des salaires et l'omission de ce problème de sélection aléatoire biaise les paramètres.

Il convient de mener deux procédures pour faire face à l'endogénéité : tout d'abord, connaître l'ampleur du problème et ensuite de le supprimer. Nous devons clarifier quelle est la causalité la plus importante dans cette relation, le lien de causalité sur le temps ou la causalité sur la section transversale. Ce travail porte sur la causalité sur coupe transversale puisque les données ne sont pas suffisantes pour aborder la causalité sur séries temporelles. Différentes méthodes économétriques servent à obtenir plus d'informations sur la causalité. La méthode correspondante identifie l'étendue du problème. À l'aide de modèles de sélection et de décompositions de salaire, nous sommes en mesure d'étudier les différents comportements des employeurs de taille différente. Nous savons donc comment de petits, moyens et grands employeurs rémunèrent les caractéristiques observées et non observées des travailleurs. En outre, l'ampleur de l'effet de la taille en présence de biais de sélection et les diverses composantes de la différence de salaire sont identifiés. L'inférence causale dépend du choix du modèle de la sélection et du choix des instruments. Les résultats peuvent être améliorés en utilisant des instruments plus pertinents et valides. Bien que l'endogénéité ne soit pas totalement supprimée, elle est clairement identifiée et les méthodes pour y faire face sont mises en évidence.

Les principaux résultats sont les suivants : la taille des entreprises est un élément clé pour comprendre le marché du travail français. Une relation positive et significative entre la taille de l'employeur et les salaires des employés se dégage, au travers de chacune des méthodes utilisées. La comparaison des deux mesures des salaires montre une nette différence dans la prime salariale. Le calcul des résultats en distinguant les deux mesures du salaire nous montre comment les composantes salariales sont liées à la taille de l'employeur. La prime salariale est fortement liée à la structure de rémunération des gros employeurs. L'ampleur de l'impact du salaire taille devient négligeable lorsque nous utilisons le salaire horaire de base (net). Les salaires bruts sont liés aux bénéfices des entreprises, reflétant que

les gros employeurs gagnent plus de bénéfices par rapport aux petits employeurs et les bénéfices sont partagés avec leurs travailleurs. Il est observé que la différence taille de l'employeur salaire est davantage une différence de politique de rémunération des employeurs.

La discussion sur l'effet de la taille sur les salaires est incomplète tant qu'il n'est pas tenu compte de la répartition non aléatoire des travailleurs. Les méthodes de sélection ont donc été appliquées pour comparer les rendements des caractéristiques observables et non observables des travailleurs selon les différentes tailles d'établissements. Le coefficient de corrélation et l'inverse du ratio de Mills apparaissent négatifs en utilisant les procédures FIML et d'Heckman, montrant que les variables non observables dans l'équation de sélection et l'équation de salaire sont corrélées négativement. Ainsi, les facteurs non observés sont largement valorisés dans les petits établissements tandis que les facteurs observés ont des gains élevés dans les établissements de grande taille. Les besoins moyens de compétences dans les grands établissements sont faibles par rapport aux petits établissements. Les résultats confirment que les grandes entreprises choisissent de meilleurs travailleurs et les salaires plus élevés sont basés sur les caractéristiques observables. Même après avoir contrôlé pour les caractéristiques observées et non observées, la prime salariale existe toujours et est plus forte pour les travailleurs masculins. Cela peut être le résultat de l'hétérogénéité de l'employeur.

Comme suggéré par les modèles du salaire d'efficience, si les entreprises de différentes tailles diffèrent dans leur facilité à surveiller et à contrôler les agents, dans leur coût de formation, ou dans le travail d'équipe, elles peuvent trouver rentable de payer des salaires différents pour des travailleurs identiques (Morissette (1993). L'écart salarial entre les sexes est plus grand dans les établissements de petite taille, mais dans tous les cas, la majorité de l'écart salarial entre hommes et femmes reste inexpliqué .Dans les grandes entreprises, les salaires sont généralement négociés entre l'entreprise et le syndicat, le syndicat tentant de faire pression sur l'entreprise afin d'obtenir des salaires plus élevés, mais la conclusion ne soutient pas l'hypothèse de la menace d'union puisque les informations disponibles sur la syndicalisation ne sont pas suffisantes pour observer l'impact sur la structure des salaires de l'employeur .

Les méthodes PSM et MCO sont équivalentes, mais traitent uniquement des observables et ne résolvent pas le problème du biais de sélection. Heckman, Smith et Clements (1997) font observer que PSM ne peut répondre aux questions relatives aux effets redistributifs du programme. PSM traite uniquement des effets moyens. Parmi toutes les méthodes, les modèles de sélection sont meilleurs pour prédire la relation causale entre la taille de l'employeur et le salaire si de bons instruments sont disponibles. Pour les effets du traitement hétérogène, les modèles de sélection sont meilleurs par rapport aux 2SLS et GMM. Généralement FIML est considéré comme plus efficace comparativement à Heckman mais dans la littérature, les modèles de Switching sont réputés comme meilleurs pour traiter des différences de salaire selon la taille de l’employeur en corrigeant la sélection nous permettant d'observer les salaires sous différents régimes avec effet de sélection. Par conséquent, les deux modèles sont équivalents et peuvent être utilisés côte à côte, en particulier dans les travaux de recherche appliquée. La méthode du pseudo-panel a résolu le problème d'indisponibilité de données de panel, mais l’utilisation de quelques années de données en coupe ne donne pas de résultats pertinents. Il existe d'autres méthodes de construction de données de pseudo-panel. Les résultats peuvent être plus intuitifs si l'on compare les différentes méthodes de construction des cellules de cohorte.

Comme mentionné ci-dessus, il est difficile de dire que tous les objectifs ont été atteints, beaucoup de nouvelles questions apparaissent et plusieurs extensions possibles pourraient améliorer la qualité du travail et permettre de faire progresser la recherche dans ce

domaine. Les modèles de sélection sont les plus pertinents pour ce type de question lorsque l'on travaille avec des données en cross-section, mais il est très important de disposer de bons instruments pour prédire le modèle de sélection. Nous devons alors rechercher davantage d'instruments. Des modèles de régression avec sélection pourraient être étudiés en détail (Bourguignon et al. 2007) afin d'améliorer les résultats. Il serait également intéressant d'exploiter les informations détaillées sur les éléments de structure de rémunération chez les employeurs de grande taille et d'identifier clairement quels avantages en nature sont les plus sensibles à la taille et faire la différence entre la nature volontaire et involontaire des heures de travail supplémentaires. Pour le choix des méthodes, des travaux supplémentaires pourraient être menés à l'aide de traitements multiples dans le score de propension correspondant. La méthode de « neural network » pourrait être appliquée pour construire des cellules homogènes et les résultats pourraient être comparés à ceux obtenus par la méthode de Deaton des Pseudo panel.

GENERAL INTRODUCTION

1. Introduction

According to Moore (1911),"as the size of establishment increases, the condition of the laborer improves in all directions". Similarly Oi & Idson (1999) were of the view that “a worker, who holds a job in a large firm, is paid a higher wage, receives more generous fringe benefits, gets more training, and is provided with a cleaner, safer, and generally more pleasant work environment”.

It has always been a puzzle for labor economists to see that people with equivalent economic characteristics receive different payments for the labor services they sell. The economists consider that if the workers are equivalent and the posts also and if there is mobility of workers among firms and posts, the wages should be the same. The diversity of jobs and firms brings complexity to employer-employee relationship. According to Oi & Idson (1999), the labor market is not a single place of exchange. It is a multiplicity of "markets" in which a job is defined by a worker-firm attachment. Jobs differ not only in the tasks that an employee must perform, but also in the obligations of an employer to her employees. Thus, size of the employer is not the sole determinant of wages. Rather; it should come through the diverse nature of jobs and employees looking for those jobs. However, we observe different wages for similar employees working in different sizes of employers after controlling for both employer-employee characteristics: this implies that something is missing to explain this employer-employee relationship. This makes it an unsolved puzzle. For understanding the role of labor market agents, it is important to know the causal relationship of employer size and wages of workers linked with that establishment. This dissertation is centered on the causal relationship of employer size and wage.

The "size wage differential" represents the wage premium paid to employees working in big size of establishments after controlling for individual and employer characteristics common to all employees in that firm or establishment. Both employers' and employees' characteristics and their interactions are important for generating the positive relation between size and wage. Thus, it is very important for labor economists to identify how the characteristics of employers' and employees' are interacted to create the size wage premium (henceforth ESWP).

The case of competitive labor market is hypothetical and this challenges labor economists to formulate economic policies to tackle with non-competitive labor markets. The heterogeneity of workers and employers influences wage to a level which is above the level predicted by the traditional theories of wage determination. There is a vast literature on the wage differentials between employees of similar characteristics. Mostly it is studied in the context of gender, race, and/or ethnicity; while there is another set of literature on wage inequality depending on the size of employer[1]. This area is particularly important for labor economists in order to formulate policies related to compensation, employment, industrialization and occupations.

Although the size-wage gap is seen as the correlation of employers' or employees' characteristics with size of firms or establishments, there is less agreement on the reasons of the impact of size on wages. Various hypotheses have been formulated and tested to determine the magnitude and causes of ESWP. It is explained either in terms of large employers hiring high quality workers[2] , or as the compensation to worst working conditions in the large firms[3], or as an efficiency wage to increase workers' productivity or to invoke effort at work[4] or to avoid monitoring costs[5]. Furthermore, the size-wage gap is attributed to

---

[1]*Such studies include (Moore, 1911), (Lester, 1967), (Brown & Medoff, 1989), Brown et al(1990), (Idson & Feaster, 1990), (Oi & Idson, 1990), (Groshen, 1991)(Main & Reilly, 1993), (Stephen & Melissa, 1997), (Mizala & Romaguera, 1998 ,(Troske, 1999),(Criscuolo, 2000),(Paez, 2003), (Lluis & Ferre, 2004), (Lallemand & Plasman, 2005), (Fathi & FitzRoy, 2006), (Lallemand & Plasman, 2005)***,***(Lane, Salmon, & Spletzer, 2007), (Pedace, 2008), (Feng, 2009) and many others.*

[2]*A non-exhaustive list of such studies includes Shinohara (1962), Griliches (1969), Hamermesh (1980), Foss (1981), Oi (1983), Brown and Medoff, (1989), Bayard & Troske (1999), Troske (1999), Lluis & Ferrer (2004), Silva (2004), Lluis (2008) etc.*

[3]*by Lester (1967), Master (1969), Scherer (1976), Stafford (1980), Mellow (1982) and Lane & Spletzer (2007).*

[4]*Doeringer and Piore (1971), Oi & Idson (1999), Lazear (1995), Criscuolo (2000)*

the effort to avoid unionization[6] or is shared as rent by more profitable and monopolized employers.[7] All the studies found that none of the variable on the right hand side of the equation, whether related to employer or to the employee characteristics, explains as such the size-wage gap. Therefore, it is considered as unmeasured factor in the error term that makes it an unsolved puzzle. The theoretical explanations are described in detail in chapter-I.

All the theoretical hypotheses about size-wage premium are based on either compensating differentials related to employer and job characteristics or measured or unmeasured quality of labour (Criscuolo 2000). Neoclassical explanations focus on labour quality and working conditions, while institutional explanations turn to factors, such as, market power and union avoidance. Brown and Medoff (1989) concluded that "the size-wage differential appears to be both sizeable and omnipresent; our analysis leaves us uncomfortably unable to explain it, or at least the part of it that is not explained by observable indicators of labour quality.' Neither compensating differentials for working conditions, nor product market power or the threat of unionization were able to explain the size premium.

The present work studies the employer-employees relationship depending on the size of the employer for French labor market using cross section data from the labor cost and wage structure survey called ECMOSS from 1992 to 2006. French labor market is dual in nature. The French system of wage setting is particularly complex because it depends simultaneously on state level wage policies and collective bargaining at other levels. Excess labor turnover, greater wage inequality, lower human capital in terms of training, higher unemployment risk and growing number of unstable jobs along with apparently very protective legal environment makes it a complex labor market structure. Further, the question of selection has not been studied in detail using cross section matched employer-employee data.

Before going further, two issues are important to be addressed. First, how do we measure size: by employment, by assets or by sales? Size measured by employment has been used in the literature due to availability of data although labour productivity can vary over time and across firms which makes it an inappropriate measure. But still it is better than capital/ labour ratio or sales as Oi and Idson (1999) have pointed out: the capital/labour ratio varies widely across industries because of differences in the technology of production and similarly

[5] *Kruse (1992) , Piekkola (2000) and Fujiwara-Greve and Greve (2004).*
[6] *Brown & Medoff (1989).*
[7] *Weiss (1966), Mellow (1982), Kruger and Summers (1989), Fakhfakha & FitzRoyb (2002).*

comparisons across industries and over time is complicated when one uses sales as measure of size. Therefore, keeping in view data availability, we have adopted employment as a measure of size. Second, what is the appropriate unit of analysis, the firm or the establishment?[8] There are studies on both types of units. We have used data on establishment instead of firm and our reasoning is in line with Weiss (1966), firstly, employer size seems to mainly matter of its impact on local labour markets which depends on the size of plants rather than the whole firm; secondly, establishments are more accurately assigned to industries than firms are; and lastly because plant-size data are available for a bigger number of industries than are firm-size data.

It is equally important to determine the consistent determinant of wage between firm and establishment. There are many factors that affect the wage determination process, for example, work standards, legislation, performance, sunk cost etc. A firm normally sets wage for its employees based on standards. Large plants may also pay higher wages and set wage standards keeping in view the division of labor as absenteeism or bad performance can be costly at large plants. In large manufacturing firms, only a minority of production workers' wages are set unilaterally by the firms because wages are usually negotiated between the firms and a union. Industry level agreements and standards apply to all plants but the legislation concerning the threshold may affect the employment and productivity decisions for firms. The threshold legislation may increase firing costs. Firms may choose to remain below the threshold limit and produce less than the optimal level. Similarly, the wage gap between large and small firms is also explained by the performance of the firms. The productivity performance of large firms is good than small firms. Sunk costs can also affect the firm's exit decision and can act as a barrier to exit. The strength of barriers depends on the level of sunk costs (Blanchard et al. (2010). The exit process of firms may depend on the firms' performance and barriers to exit which further depends on the sunk costs. A small firm can choose low wage because of the possibility of exit and thus stays at lower wage in the market.

[8] *A firm or enterprise is an actual registered company, association or trust; whereas an establishment is each physical location, where business is conducted or where services or industrial operations are performed. (For example: a branch, a factory, a plant, operating office, mill store, hotel, movie theatre, mine, farm, and administrative office.*

Among other factors, the choice of production technique, sunk costs and/or locational differences can create technological heterogeneity among industries and among plants within industries. Various sources of technological heterogeneity can induce sorting by worker ability among plants, which in turn leads to differences in the wage structure across plants. A general point is that sorting mechanisms that differentially affect the skill mix by employer size lead to differences in dispersion by size class. Davis and Haltiwanger (1996) and Oi (1999) argue that larger employers hire workers with greater mean skill levels and more homogeneous skill levels because larger employers rely more intensively on standardized production technologies that call for homogeneous labor. Garen (1985) argues the same point that within-plant wage dispersion falls with plant size because higher screening and monitoring costs imply less sensitivity of wages to ability at larger plants. Hence, multiple factors are associated with the size that can affect the wage structure of employees within establishments and across establishments. Here, we will focus on wage differentials across establishments and will keep aside within plant wage dispersions.

2. Self-selection and nonrandom sorting

The key econometric difficulty in explaining the employer size and wage relationship comes from nonrandom selection of workers into establishments of different size. For a long time, this problem has been ignored in this literature and many studies were based simply on OLS to estimate the wage premium while ignoring the nonrandom assignment of workers. The common problem in the cross-section empirical work, when dealing with individual data, is the potential endogeneity of explanatory variables on the right hand side of the equation. Ordinary least squares estimates will yield biased and inconsistent estimates in the presence of endogeneity bias. The sources of potential endogeneity bias may include omitted variables, simultaneity, and measurement error (Wooldridge, 2002).

The most important source of endogeneity in size-wage relationship is the sample selection bias generated through omitted variable bias. This results whenever some potentials observations remain unobserved.[9] For example, people working in a big size of the establishment may not be a random sample of all population. Employees may have self-selected themselves into establishments of big sizes. Such exclusion of potential observations

[9] *The nonrandom selection into different size of employers has been overlooked in the literature on the size-wage relationship. Studies focused on to reduce the magnitude of the size-wage impact by adding employees' and employers' control but the effect of unobserved characteristics is overlooked.*

may cause OLS parameters of the model to be biased. If unobservable differences in productivity affect the allocation, neglect of these could impart a bias in the estimate of the effect of employer size on wages (Oi, 1999).

The topic of 'self-selection' in labor economics was first discussed by Roy (1951) by giving an example of workers selecting between fishing and hunting.[10] Later, Borjas (1987) presented the self-selection model for the immigration decision to the United States. Selection models defined for potential outcomes with explicit treatment assignment mechanisms were developed by Gronau (1974) and Heckman (1976, 1978, and 1979) to study the selection bias in the causal inference.

Heckman (2005) [11] discusses the historical developments of causal inferences, classical problems of estimating causal effects, and focuses on the importance of econometric causality[12]. Causal parameters and causal inferences in economics are motivated by policy questions.[13] Nevertheless, we cannot leave the problem of causality aside in explaining the employer size and wage relationship. Employees working in big size firms are similar to the ones who do not, except for being in big size (treatment status) and the wage (the outcome associated with treatment status). The problem of causal inference is to assess whether changing the size (manipulation of the treatment), holding all other factors constant, affects wage (outcomes).[14]

---

[10] *Roy concluded that the key determinants of self-selection are the distribution of skills, correlation among the skills in the population, the technologies for applying these skills and consumer tastes for different types of outputs.*

[11] *Heckman, J. J. 2005. "The scientific model of causality, Sociological Methodology", 35, 1(97).*

[12] *Economists focus on causality from the perspective of policy evaluation. Causal parameters and causal inferences in economics are motivated by policy questions. The econometric approach develops explicit models of outcomes where the causes of effects are investigated and the mechanisms governing the choice of treatment are analyzed. (Heckman, J.J (2008). 'Econometric Causality' International Statistical Review, 2008, (76) 1, 1-27).*

[13] *A standard method for evaluating social programs uses the outcomes of nonparticipants to estimate what participants would have experienced had they not participated. The difference between participant and nonparticipant outcomes is the estimated gross impact of a program reported in many evaluations. The outcomes of nonparticipants may differ systematically from what the outcomes of participants would have been without the program, producing selection bias in estimated impacts. (James Heckman and Hidehiko Ichimura, Jeffrey Smith and Petra Todd, 1998. "Characterizing Selection Bias Using Experimental Data, Econometrica, Vol. 66, No. 5 (September, 1998) 1017-1098).*

[14] *The econometric approach to causal inference distinguishes among (1) defining counterfactuals, (2) identifying causal models without sampling variations, from data of population, and (3) identifying causal models with sampling variations, from actual data.*

The problem of selection bias arises when using a nonrandom sample of the general population and inferring what the average person would experience if selected at random. While dealing with the causal effect of working in an establishment of big size, a careful distinction is required between modeling potential outcomes and empirical methods for identifying the effect of working in big size establishment. The causal relationship between employer size and wage demands a careful analysis of both observable and unobservable factors determining this relationship. It is important to model the relationship between the unobservable in outcome equations and selection equations to identify causal models from data and to clarify the nature of identifying assumptions, because these are the unobservable factors that give rise to selection bias in causal analysis. In order to resolve selection bias, one need to take into accounts both treatment outcomes and treatment choice mechanisms. The treatment status refers to ‘being in big size of establishment’. Thus, in order to estimate parameters with these mechanisms the unobservable factors in the outcome and treatment choice equations should be taken care of.

Idson and Feaster (1990) did sample selectivity correction and their results support the hypothesis that more dynamic workers are attracted to small firms. Applying Heckman two step methodology to 1979 CPS data for men indicated a significant positive selection bias in small firms and negative selection bias in large firms. Unobserved traits that would raise men's wages also made it more likely that they wouldbe employed in firms in the smaller size groups. The mean wages of workers in a small size category are thus higher than the mean wages that would have prevailed if workers had been randomly allocated to size categories. For Stigler (1962), workers with more ambition and energy would do better in a small firm where their performance will be noticed and rewarded. The procedure used to adjust the size-wage gap for self-selection into different size categories tacitly assumes that the"value" of the unobserved drive and motivation is the same across firm sizes. Attributes, such as individual initiative, that are productive in small firms may actually be a hindrance in large firms that organize production around structured teams. A random reallocation of workers across different size firms may thereby produce very different results than those predicted from this model.

In the studies of employer-employee relationship for the labor market in France, the question of employers’size-wage gap has been highlighted by many authors. However there is no uniform consensus on whether the size wage premium is the result of unobserved omitted characteristics of workers, is paid to compensate worst working condition or is shared as rent. The rent sharing hypothesis is explained using longitudinal employer-employee dataset for France [15]. The compensating wage differential hypothesis is also considered to be tested[16]. The behavior of labor unions in French labor market context is also studied by many authors.[17]. However, some areas still need attention to be explored in this literature particularly the self-selection mechanism.

It is not being claimed that this dissertation is going to fill gaps in this literature for France; however, some interesting findings may open doors to further advancements. This dissertation contributes to the study of the size-wage relationship for French labor market by using cross section datasets to see the magnitude and sources of the size-wage premium for France. The main concern is to deal with the problem of selection bias associated to the size wage relationship. Different methods are employed to study the nonrandom selection of workers into employers of different size including, full information maximum likelihood, switching regression models and wage decomposition method. Matching technique through propensity score is tested to observe the wage differentials between large and small employers. Cohort fixed effects and first difference is estimated by generating a pseudo panel from repeated cross sections.

In order to be informed about the specific characteristics of French labor market, the main features of French labor market are presented below. This will be followed by data description, objectives and main findings subsequently.

3. French Labor market overview

It is important to stress that the labor market is strongly regulated in France. France is a country with population of 65,350,333 on 1st January 2012, economic growth of 0.2 percent[18], inflation of 0.2 percent and with the rate of unemployment at 10.3 percent in the 3rd quarter of

[15] *Abowd, J. M., F. Kramarz, and D. N. Margolis. 1999, Margolis, David N. & Salvanes, Kjell G., 2001,John Abowd, Francis Kramarz and David Margolis 2001, Abowd, J. M., F. Kramarz, P. Lengermann, and S. Roux. 2005, Abowd, J. M., F. Kramarz, and S. Roux. 2006, Kramarz (2008).*

[16] *Lanfranchi, Joseph, Henry Ohlsson and Ali Skalli, (2002), Christophe Daniel and Catherine Sofer 1998.*

[17] *Breda Thomas, 2010, Bryson A, Forth J, Laroche P (2009), Jamet (2006). Laroche and Wechtler (2011).*

[18] *In 2012 Q3, French gross domestic product (GDP) in volume terms* increased by 0.2% after a –0.1% decrease over the previous quarter.*

2012[19].The French labor market is characterized by a weak participation rate of youth and the oldest compared to other European countries. This feature is often linked to the fact that in the 1980's and the 1990's, the government and social partners answered to a growing mass unemployment by promoting early retirements and longer studies. In 2006, around 10% of the 15-64 participants were unemployed; nearly half of them having been unemployed for more than one year. Higher unemployment risk is correlated with low level of education, youth population, female gender, and blue-collar occupation. 13.5% of the employed occupied an unstable job that is training, apprenticeship, fixed-duration or temporary contract jobs[20].

France is characterized by quite strict job protection measures, extensive coverage of collective agreements and a rather centralized system of pay determination. In France, collective bargaining in the private sector establishes industry minima for wages and employment conditions, whilst, in the public sector, unions also take part in national wage negotiations but the outcome is not legally binding for the government.[21]

3.1 Institutional Settings in France

This section presents the institutional settings and main characteristics of collective bargaining system in France. Mainly, for four times laws or modification of laws were introduced concerning unions, in 1950, 1971, 2004 and 2008.

The employment related organizations in France are tightly controlled and regulated by government. According to the Statistics Department of the French Ministry of Labor (DARES), 97.7% of the workforce was covered by a collective agreement in 2004. With a union density around 8%, France is the OECD country with both the highest coverage rate and the lowest union density (OECD Employment Outlook, 2004).

Wage determination process in general

[19] *In Q3 2012, the average ILO unemployment rate in metropolitan France and overseas departments stood at 10.3% of the active population. In metropolitan France only, with 2.8 million unemployed people, 9.9% of the active population was unemployed. The unemployment rate increased by 0.1 point q-o-q after an increase of 0.2 in Q2 2012. Source INSEE.*

[20] *Magali Beffy, Elise Coudin, Roland Rathelot, 2008*

[21] *Claudio Lucifora & Dominique Meurs, 2006.*

First, we have to see how the wages are determined in France. Traditionally, there are three different levels at which wages are determined:

- First by state regulation through the setting of an inter-professional minimum wage,
- Second, through joint autonomous regulation by collective agreements concluded at industry level (the minimum wages resulting from these negotiations are often under the level of the public minimum wage due to lack of frequent negotiations[22]) and/or collective agreements at workplace (or company) level;
- Finally, unilateral regulation by the employer.

Since 1951, French industry has been subject to a national minimum wage (called the SMIC since the revisions to the relevant law in 1971) that is indexed to the rate of change in consumer prices and to the average blue collar wage rate.[23]

In the late 1990s, the French government passed a law (called the Aubry Act) to reduce the statutory working week from 39 hours to 35 hours in order to raise the wages and employment. There was a chance that incomes of low income people may fall due to the reduction of working week hours. Hence, the government raised the SMIC rate and made it a single rate from 2002 onwards.[24]

3.2 Main characteristics of collective bargaining

- Collective bargaining covers a wide range of topics related to wages and employment conditions. Collective bargaining takes place at various levels in France: national, inter-professional, sector, section of industry and firm level. Sector-level collective agreements are very often extended to all employees of the sector. At industry level, any industrial federation affiliated to the five National Centers is entitled to represent automatically workers and to sign sector-wide collective agreements. The agreements on wages are aimed to raise the sector-level minimum wage up to the level of the SMIC (Salaire Minimum Interprofessionnel de Croissance). But they do not succeed

[22] *The public minimum wage is revisedannually, but the collective agreement is revised less frequently so it turns out that often the industry minimum wage is less than the public one. In this case the latter applies.*

[23] *Wage Structure in France 1977-96 Francis Kramarz, Sébastien Perez-Duarte Chapter in NBER book The Structure of Wages: An International Comparison (2008).*

[24] *Usually, employees work 35 hours a week. In addition, employees must not work more than: 1) An average of 44 hours a week during any 12 consecutive weeks; 2) 48 hours during any given week; 3) 10 hours a day; 4) 220 hours of overtime a year.*

in doing so and very often the SMIC is higher than the sectoral level minimum wage. This is a very serious problem for the French wage system as it hinders the bargaining between unions and employers.

- The coverage of collective agreements is vast, it applies to all those who have signed it, may it be an organization or a person and to those who belong to an union. It also applies to plants or units that are members of the signatory employers' organizations. They are extended to all employers in an industry. Therefore, the coverage rate is 95-98%.
- Bargaining also occurs at the level of the firm. Since 1982 (loi Auroux), enterprises with a union delegate must negotiate at least once a year on effective salaries and working time, although they do not have to sign an agreement.

3.3 French Legislations concerning thresholds of Size

In France, the thresholds of firm/establishment's size are very important for implementation of certain rules determining the employer-employee relationship. The state has made the threshold limit of size important for organizing in a firm or industry and it has made employers accountable for some legislation concerning employees' organizations. The most important size threshold is when a firm reaches the limit of fifty employees. This is the critical threshold when costs rise significantly as a number of labor market regulations become applicable regarding the firm's ability to adjust its labor.

The legislations related to thresholds of size can affect to limit the size, on the one hand, and promote employment stability of seniors, on the other hand. The size is an employer's decision that can affect his other decisions. Therefore, it will be interesting to see how in the presence of legislative requirements and state protection, employer's size can form the compensation structure of an establishment. The threshold of the firm's size is directly related to the question of wage and size. To avoid the direct and indirect costs related to employee representatives or work councils, employers may limit the size. For instance, many small French companies limit themselves to either 10 employees or 49 employees, because, there is an obligation for the employer to organize elections for the appointment of staff delegate at the threshold of 11 workers and for the work council elections at the threshold of 50 workers.

The size of the firm is important when dealing with the following issues:

1. The **election** for Union Delegates (Délégués du personnel) becomes compulsory once a company has 11 workers. Further, when it grows to 50 employees, it has to have a Work Council (*Délégué syndical).*
2. The probability of a firm to form and **organize unions** depends considerably on its size. One representative of a particular union is a necessary condition for its recognition in a firm but the strength can go to five representatives depending on the size of the firm. The categorization is as follows:
    a. The maximal number of representatives that can be legally recognized is 1 for workplaces with less than 1000 employees;
    b. 2 for workplaces having between 1000 and 1999 employees;
    c. 3 for workplaces having between 2000 and 3999 employees;
    d. 4 for workplaces having between 4000 and 9999 employees; and
    e. 5 for workplaces having 10000 employees or more:
3. Before **dismissing** more than nine workers, the employers, with less than 1000 employees, should arrange another employment in the group to which the firms belongs to. It is also the responsibility of the employer to train the worker to cope with the new job through assisting professional trainings and help him to adapt.
4. In case of the firm size of more than 1000 employees, then, before collective dismissal, firms should offer a “re-assignment leave” during which the worker can be re-assigned without any interruption in his contract. The employee receives an allowance, which is at least 65% of his yearly gross wage (and cannot be less than 58% of the SMIC). It is paid half by the firm and half by the State.
5. In addition to the above law concerning dismissals for dismissals up to 10 workers the employers must inform employee representatives and /or for recruitment up to 50 workers must inform the works council.
6. Firms with more than 300 employees have to **negotiate** every three years on the evolution of employment and skills.

7. Employers with more than 50 workers must enter into a company agreement regarding the employment of older employees, including measures to encourage their employment.[25]

The threshold legislation may increase firing cost as firms with 50 or more employees must formulate a "social plan," which is designed to facilitate reemployment, through training, etc. This increases the costs of employing workers at that threshold. The threshold can also affect the productivity and employment decisions of a firm. Intuitively, firms will optimally choose to remain small toavoid the regulation, so the size distribution becomes distorted with "too many" firms just below the size threshold and "too few" firms just above it. Furthermore, the distribution of productivity is also distorted:some of those firms just below the cut-off are "too productive" as they have been prevented from growingto their optimal size by the regulation. The welfare effects of these regulations becomes important when more productive firms choose to remain just below the regulatory threshold, by allocating too little employment and through reducing equilibrium wages. This can also encourage many individuals to become small entrepreneurs rather than working as employees for more productive entrepreneurs. (seeGaricano et al 2012 for detail)

The above institutional elements show that size is definitely an important element in the French labor market for employers' policies related to employment, wage structure and work councils. This institutional arrangement should be kept in mind when looking at the causes of wage inequalities in France. This information will be further utilized in the dissertation to determine the causes of wage- gap depending on size.

4. Data

The data is drawn from two sets of surveys called The Cost of Labor (ECMO) and Wage Structure Survey (ESS), jointly called ECMOSS (The Labor Cost and Wage Structure Survey) for the private nonagricultural sector collected by INSEE. Therefore, the estimation results will reflect the wage structure system and employers' size impact on wage in the private sector. The implications for public sector can be different. The survey aims to examine the pay structure and working time of employees. It provides detailed information to analyze wage disparities. It consists of two parts: wage structure survey (called ESS), and measure

[25] *JérémieGicquel. Paul Hastings LLP*

and breakdown of cost of labor (called ECMO). The survey on the cost of labor (ECMO) is designed to examine an average annual cost of labor per employee and an average hourly cost of labor per hour actually worked. It provides the detailed structure of these costs by sector, by economic activity, region where the establishment is based and by the size of the establishment. The survey on the structure of wages (ESS) aims to provide the individual data for a sample of employees on wages, its components, the determinants of wages and the workplace characteristics. The sample consists of 14,000 establishments and approximately 140,000 employees from the nonagricultural private sector. Employers respond to a questionnaire describing many workplace characteristics and give information about a random sample of their employees. This data set has many observations and a large variety of employer and employee characteristics. For this dissertation, we have used surveys for the years 1992, 1994, 2002, 2005 and 2006. No compiled data is available after 2006 until now[26]. Detail description of the data and definition and construction of variables is presented in chapter-2 of descriptive statistics.

5. Objectives and Main Findings

The objectives of this dissertation are multifold. The dissertation aims to answer the following questions. In each chapter the objectives related to it are described in detail:

- What is the causal relationship between employer size and employee's wage? Whether the size wage premium is due to working in big size establishments itself or is it due to systematic differences in personal, job and workplace characteristics across employees working in different size of establishments.
- What is the magnitude of employer size-wage gap in France with establishment level data?
- What is the magnitude of size-wage gap in the presence of nonrandom selection?How the observed and unobserved characteristics of workers are rewarded by employer size? Which selection model is most appropriate?
- Is the size-wage gap actually a gender wage gap? What factors constitute this gender wage gap and what is the share of explained and unexplained components in the gender wage gap?

[26] *These surveys are provided by Reseau Quetelet, Maurice Halbwachs Centre upon official request.*

- What are the implications for pseudo panel estimation methods in the size-wage literature?

The thesis consists of six chapters: every chapter is in continuation of the previous one. First chapter reviews the literature on the employer size and wage relationship. Second chapter explores the data and presents descriptive statistics. In the third chapter, the overall wage differential is identified after controlling for the bias on observable characteristics of workers who are matched in big and small size establishments and the wage differential is analyzed. This chapter also deals with the relative importance of bonuses, overtime payments, and allowances through comparing gross and basic hourly wage. Third chapter leaves us with the problem of unsolved selection bias or nonrandom sorting of workers. Therefore, in the fourth chapter a full information maximum likelihood model is presented where nonrandom selection of workers is studied and the magnitude of size-wage premium is analyzed. Further, rewards for measured and unmeasured skills by size group are studied. Further, in the fifth chapter the gender wage gap is decomposed into explained and unexplained component in large and small establishments. Finally, a cohort analysis is presented in chapter six using the latest successive cross section matched employer-employee datasets to estimate fixed effects. Main focus of thesis is on cross sectional data and methods but in the last part the possibility of building pseudo panel data using Deaton's method is analyzed.

Below brief findings of the main chapters, chapter 3-6, are presented below:

Chapter-3: This chapter, using 1992 survey, estimates the wage premium paid by large employers by comparing wage outcomes for employees working in the big establishments with 'matched' workers working in small size establishments through propensity score matching method. The chapter contributes to explain the relationship between establishment size and individual hourly wage with ordinary least squares and propensity score matching method. Main findings are summarized below:

- Matching conditional on propensity score of the observed covariates reduces difference in mean wage between people working in large size and those working in small size. This reflects that large employers value more observable characteristics. 50 % of the mean wage difference is explained by difference in observable characteristics of large establishment workers, whereas, net wage difference is entirely the difference of observed characteristics between large and small establishments.

- There is a preference of large employers for workers with high earning capacity. Secondly, the distinct features of large size establishments to determine the wage premium in the form of compensations and pay packages (bonuses, allowances, over time payments etc).We may say that the employer size wage premium is more the result of the employer's compensation policy than the result of basic wage differences.
- The employer size-wage effect based on observable characteristics is greater for male workers, in the blue collar jobs and in the large manufacturing plants.
- The available information in the data related to union and collective bargaining agreements is not sufficient to determine role of bargaining agreements in the employer size-wage gap.

Chapter-4: This chapter attempts to explain the nonrandom selection of workers across employers of different sizes to examine the size-wage gap in French establishments. Different estimation methods including full information maximum likelihood and Heckman two-step model are used to analyze and compare the outcomes with and without selectivity considerations. Main findings are summarized below

- The selection terms in Heckman model are mostly not significant while FIML shows strong negative correlation for all population and for male sample. Two things are evident from this analysis; either the choice of selection model contains the maximum jobs and observable characteristic that can make the selection effect not significant in some cases or the choice of instruments is not good to control for unobserved heterogeneity.
- Negative selection into large size establishment and positive selection into small size establishment is found suggesting that the unconditional wage is lower than conditional wage for the small size establishment worker. On the other hand, the conditional wage is lower than the unconditional wage for the employees working in large size establishment. The unobserved factors have high reward in small size establishments while observed factors have high rewards in large size establishments.
- Generally FIML is considered more efficient compared to Heckman but the switching regression models have been popular in the employer size wage gap

literature dealing with selection that allow us to observe wages in different regimes with selection effect. Therefore, we may conclude that both models are equivalent and may be used side by side especially in applied work.

Chapter-5: Chapter five extends the results of chapter 4 where gender wage decomposition across size and across professions is performed. The work place segregation is considered and the effect of differences in personal characteristics on the gender wage gap is disentangled from the effect of selection into different establishments of women and men. Main conclusions from this chapter follow:

- The gender wage gap is greater in large compared to small size establishments but among all the cases the larger part of the gender wage gap remains unexplained.
- The adjusted regressions for selection increases gender wage gap in large size but the evidence of nonrandom selection is only found among male workers and no selection term is significant for women.
- The wage gap exists in both measures of wage, gross and basic, the wage gap increases as the difference of gross and basic wage increases.
- There is a strong tendency for women to be associated in the low paying workplace (small size) and low paying sector (trade or services). On the one hand women segregation into low paying workplaces explains the gender wage gap and on the other hand employers' evaluation of women against men for similar characteristics is discriminatory for most of the observed characteristics.
- Women are disproportionately represented in low paid occupations. There is prevalence of horizontal segregation that results into low wages and increases gender wage gap. To some extent there is also vertical segregation as only women are underrepresented in high paying occupations.

The instruments that are used for the 3rd and 4th chapters are not available in the recent datasets of the same surveys. In order to benefit from the latest information and study the recent trend of the size-wage impact, recent surveys are utilized. Therefore, pseudo panel data estimation methods are applied to experiment whether it is possible to estimate fixed effects and capture endogeneity bias. Although endogeneity bias is not removed, it is captured by using the latest datasets. This is done in Chapter six.

Chapter-6: In previous chapters different methods were applied on the cross sectional data to study the employer-employee relationship depending on the size. Chapter-6 presents estimates of fixed effects and first differencing estimation methods to observe the wage premium. But as panel data is not available, pseudo panel data is generated and cohort fixed effects are observed on the pseudo panel data of the repeated cross sections.

- The findings consistently support the hypothesis that within a cohort, there is positive and significant effect of employer size on wage. The impact is strong in the medium scale establishments. As the size of employer increases from small scale, wage increases and this effect remains significant even after controlling employer and employee characteristics.
- A comparison between the random effects model and the fixed effects model revealed that the fixed effects were statistically significant. Consequently, the FE Model formed the core of analysis.
- Additional insights were provided by the results generated through between transformation, and first-differenced models, which used transformed data sets.
- All estimation methods (random effects, fixed effects, between transformation models and first difference) showed a significant and positive effect of size on wage.
- By using any of the four estimation methods, it cannot be said that any method eliminates bias from unobserved cohort heterogeneity that changes over time. This is a first step. One can expect different results by increasing the number of years.

6. Conclusion

In the presence of centralized system of wages and wage bargaining, the heterogeneity of employers and employees poses a challenge to solve the size-wage puzzle for French labor market. The nonrandom sorting of workers into employers of different sizes creates selection bias and the causal relationship between size and wage is difficult to explained. This dissertation aims to explain why large employers pay higher wages as compared to similar workers in small size establishments. The dissertation studies the question of selection in detail. The main focus is to determine the magnitude of size wage impact in the presence of selection, to study the contribution of each components of wage differential, to compare wage differential outcome between different groups (gender) and finally, to study the implication of pseudo panel data in this literature. Thus, the tagline is to test various methods to explain the size and wage relationship. The data is not sufficient to test the role of unions or how the unionized and non-unionized sectors differ by employer size and the influence of unions on wage.

Although the selection bias problem is very complicated, no concise solution is possible to completely remove it when working with cross section data set. However, its treatment is possible and this dissertation shows different results after taking into account the nonrandom sorting on the magnitude of size-wage premium. It is important to stress that the estimation results of this dissertation depend on the selection of explanatory variables in the selection and wage equation, the question of whether to use self-selection model and the choice of model etc. The causal inference depends on the choice of selection model and choice of instruments. The exogeneity of instruments is challenging and difficult to achieve when dealing with individual cross section data. In the end everything is endogenous as one cannot predict behaviors or personal choice exclusively. The instruments used in this study are taken and/or modified based on other studies in the same literature. Further, results could be improved by using more relevant and valid instruments. Although endogeneity is not removed, it is clearly identified and methods to deal with it are highlighted. The major limitation of this dissertation is the lack of data to truly estimate the causal effect of size on wage.

Primarily, different econometric methods are estimated to study the behaviors of firms given the information in the independent cross sections. But as mentioned above, the causal inference depends on the choice of selection model and choice of instruments. The results can be improved by using more relevant and valid instruments. Although endogeneity is not removed, it is clearly identified and methods to deal with it are highlighted. Among different methods, Heckman two-step and full information maximum likelihood provide a better framework to estimate treatment effects and should be used side by side. It is hard to choose one as both are relevant in applied studies. The Heckman two-step is famous in size-wage literature to estimate employer size-wage effect in different regimes while FIML provide more efficient way to estimate parameters.

The dissertation is organized as follows. The chapters mentioned above are presented in sequence. In the end, a general conclusion is presented which summarizes the conducted work, main results, and discusses some directions for further research.

After the bibliography, appendices are presented.

Chapter-I

THE LITERATURE REVIEW

1.1 Introduction

This chapter provides a critical survey of the economic literature on the employer size wage premium (henceforth ESWP). It examines two questions; firstly, what makes large employers pay more than small employers? In order to answer this, different theoretical hypothesis are highlighted. Secondly, what is the evidence across countries on the explanation and magnitude of the ESWP? In order to answer this, an empirical evidence of studies on testing theoretical hypothesis is provided. Particular attention is devoted on the selection bias problem.

1.1.1 Stylized Facts

Many empirical studies have shown a strong and positive relationship between employer size and wages of eployees. Moore (1911) for the first time found that wage in the large plants with 500 or more employees were 38.5% higher than employees working in small plants. Lester (1967) found that within the same industry average hourly earnings in large establishments are 20-25% above the average hourly earnings in small establishments. Brown et al. (1990) reported that hourly wages in firms with 500 or more workers are 35% above than wages in firms with less than 25 employees. Groshen (1991) found, after controlling for occupations, establishment wage differential variation from 12% in the cotton and man-made textiles industry to 58% in the industrial chemicals industry. Lane et al. (2007) found that controlling for detailed occupation, 21% of wage variation can be explained merely by knowing the individual's particular establishment. Similarly, Mizala and Romaguera (1998) reported 7 to 9% of individual wage variation due to establishment wage differentials. Troske (1999) examined industries separately; he estimated that the establishment size-wage premium is 14% for workers in manufacturing, 10% in retail trade, and 11% in services.

Results of many recent studies are also consistent with the previous studies. For instance, Lluis and Ferrer (2004) estimated 15% size-wage premium in the United States and 10% in Canada. Paez (2003) found that large firms offer on average 3 percent higher wages than small to medium size firms. Further, Lallemand and Spletzer (2005) examined the magnitude and sources of the establishment-size wage premium in five European countries and found that doubling of the establishment size increases earnings by 0.6% in Denmark, 3.0% in Belgium, 3.3% in Italy, 3.9% in Ireland, and 4.5% in Spain. Pedac (2008) reported that on average, workers in large establishments receive a 9% earnings premium and after controlling for non-wage benefits and measures of training, workers in the largest firms receive about a 4% earning premium.

All of the studies mentioned above confirm the positive relationship of employer size and wage. Various explanations from theoretical and empirical perspective to answer why large employers pay more have been presented by many authors.The factors underlying the ESWP are not very well understood. A famous study by Brown and Medoff (1989) explored several hypotheses including (i) larger employershire higher quality workers; (ii) larger employers offer inferior working conditions; (iii) larger employers are more likely to be unionized or pay higher wages to avoid unionization; (iv) larger employers have greater rents to share with workers; (v) larger employers pay higher efficiency wages in response to monitoring problems; and (vi) larger employers have relatively low applicant-to-vacancy ratios. The only hypothesis to receive empirical support in their study is worker quality differences among size classes, which account for roughly one-half of observed size-wage differentials. Thus, Brown and Medoff's findings indicate that nearly half of observed mean wage differentials by employer size remain unexplained.

Section 1.2 below presents theoretical explanations formulated to explain the ESWP. It is followed by the empirical evidence on the ESWP. The conclusion is drawn in the last section of this chapter.

1.2 Theoretical Explanations of ESWP

The theoretical hypothesis have been explained and supported by studies on wage dispersion, industry wage differentials and labor market structures. But, here, attention is devoted on the studies where main focus was on ESWP.

Sorting by ability/ Capital-skill Complementarity

The hypothesis of labor quality is based on the proposition that large employers hire higher quality workers. It is said that in large plants there is capital skill complementarity. Large manufacturing establishments utilize capital more intensively; therefore, high quality labor is required to complement skills (Foss (1981), Hamermesh (1980) and Griliches (1969). [27] in the size-wage literature observed and unobserved abilities are separately associated with different size groups.

Author and Year	Hypothesis / proposition	Data	Methodology	Relevant Conclusion
Stigler (1962)	• A worker will search for wage offers (and an employer will search for wage demands) until the expected marginal return equals the marginal cost of search. • Workers with more ambition and energy would do better in a small firm, where their performance is noticed and rewarded. Attributes, such as, individual initiative which are productive in small firms may actually be a hindrance in large firms that organize production around structured teams. • The larger the cost of search the less search will be undertaken by a worker at a given		A search model of information in the labor market(theoretical paper)	Same as hypothesis

[27]Many studies tested this hypothesis and provide counter or supporting evidence. *Zabojnik and Bernhardt (2001), Brown and Medoff (1989)* , *Bayard & Troske (1999)Troske (1999)* Troske Kenneth R., 1999, *Idson, T.L. (1996), Feng (2009), Millimet (2005)* , *Hu (2003)*

	level of dispersion of employers' wage offers. These costs will vary with various characteristics of occupations.			
Griliches (1969)	"Skill" or "schooling" is more complementary with capital than unskilled or un-schooled labor.	1960 Census of Population data for USA	OLS	• Results prove the indication of greater capital-schooling (skill) complementarity.
Garen (1985)	• Monitoring/evaluation costs rise with firm size. As a result, large firms acquire less accurate information about the abilities of their workers, and thus will rely less heavily on their own evaluation of workers than do small firms and more on other indicators of ability such as schooling. • The structure of wage compensation will vary by firm size. In particular, large firms will reward schooling more and measured ability less, and their wage schedule will have a larger intercept term thus leading to an observed positive correlation	1969 National Longitudinal Surveys.	OLS, reduced form equations	• The cost of acquiring information about personnel rises with firm size; thus large firms face numerous information problems that small firms do not. • Large firms should hire workers with easily observed productivity attributes such as education. The reward to less readily observable traits, such as, IQ, therefore, is greater in a small firm. • The returns to education are higher in big firms, while small firms pay more for higher IQ scores.

Compensating Differentials

Differences in working conditions are used as an alternate hypothesis to explain the size-wage differential when labor quality hypothesis cannot explain it. It is hypothesized that workers are compensated for the worse conditions at work which can cause workers to get higher wages in large employers. Rosen (1986) explained it as “Activities that offer favorable working conditions attract labor at lower than average wages, whereas, jobs offering unfavorable working conditions must pay premiums as offsetting compensation in order to attract workers. It is the additional compensation necessary to make the worker indifferent between the two types of jobs at a given utility index. This is also termed as “shadow" price”. Researchers have shown that large and small employers differ with regard to non-pecuniary aspects of their work environment (see Scherer, 1976; Brown and Medoff, 1989, Idson 1996). These studies generally find no support for the compensating differentials explanation of employer size wage premiums, and, in fact, find that many nonpecuniary aspects of work in larger firms are better than in

small firms.				
Brown (1980)	• The central focus of the theory of equalizing differences is the choice made by individuals with given personal characteristics among jobs with different wages and differing nonwage attributes. • In order to attract labor of a given quality, an employer offering jobs that are hazardous or otherwise undesirable must pay higher wages than employers offering jobs with more desired nonwage characteristics.	NLS Young Men's sample 1966-1971 and 1973 for USA	Panel Data methods	• The study found inconsistent support for the theory of equalizing differences. • This may be due to the omission of important dimensions of worker quality or labor markets are simply not as competitive as the theory of equalizing differences assumes. • As long as workers prefer better working conditions and higher wages, and employers hire the applicants they perceive to be most qualified, the relationship between wages and unpleasant job characteristics holding worker quality constant should still be positive.
Rosen (1983)].	• The equalizing difference model is built upon the simple and intuitively compelling idea that it is the combination of wages and job attributes that constitute the relevant "price" of labor for market analysis of jobs. • The production worker effect possibly could be equalizing on more rigid work routines and the impersonality of the work environment in large establishments, but the effects for nonproduction workers and top management suggests that these estimates		Analytical paper	

	may be capturing systematic differences in unobserved worker quality among establishments of different sizes.			
Garen (1988)	Individuals with greater human capital and higher earnings potential will experience an income effect and select jobs with less risk.	1981-1982 Panel Study of Income Dynamics and Bureau of Labor Statistics job related fatality and injury data for USA	Simultaneous equation techniques	Those with greater earnings capacity are likely to choose safer jobs, and those who experience greater returns to job may choose riskier jobs.
Unionism and Monopoly Rents Another hypothesis largely adopted and tested for the size-wage differential is the rent sharing hypothesis. It is said that the inter-industry wage differences are not compensating differences but are the results of sharing rents. The size-wage difference is considered as high paying ability of large employers. Large employers are generally monopolized employers and earn abnormal profits and this is shared as rent with their workers. Rent-sharing may arise for several reasons including collective wage bargaining or the employer's willingness to avoid unionization according to Brown et al. (1990) large employers are more likely to be the target of union drives or to replicate union behavior. Dickens (1965) argues that varying costs of union avoidance across sectors will lead some employers to offer pay premiums to avoid unionization. Employers that find it costly to defeat a union will offer competitive wages to prevent unionization. According to this theory, the industry's ease of defeating a union drive has a negative relationship with wage differentials.				
Segal (1964)	• Unions can be expected to make larger wage gains in non-competitive industries (oligopolies or monopolies) than in competitive ones. • Other things being equal, the propensity of workers to unionize is greater in large firms than small ones because of lack of closeness		Analytical paper on the union wage impact and market structure	

	to managements, and other working conditions characterizing large establishments. • Because of their size and degree of market power possessed, firms in noncompetitive industries are generally under considerable public scrutiny. This has its effect also on union ability to organize them.			
Weiss (1966)	Concentrated industries pay high annual rates for labor of particular "occupations"; and these high earnings are more than those accounted for by the personal characteristics of the labor employed	1/1000 sample of the 1960 Census of Population for USA	OLS	• Concentrated industries pay high incomes for given occupations • High wages in concentrated industries in the presence of unionism or the threat of unionism.
Masters (1969)	Wages are related to unionization and product market competition. • The wage rate is related to the labor and capital ratio in an industry where expensive capital equipment often requires careful workers since simple errors may lead to costly damage. • A union's bargaining power will often be stronger when a strike can tie up large amounts of capital. • If wages are also a small percentage of unit costs in the industries with high capital-labor ratios, then the unions will be in an especially strong position.	1963 Census of Manufacturers USA	OLS	• The results provide good empirical support for the hypothesis that the average wage rate in a manufacturing industry is positively related to the proportion of large plants in that industry. • Plant size should be given greater attention when economists seek to explain inter-industry wage differentials. • Greater probability of unionization at a large plant and more aggressive membership support for the union when the plant is large and more organized. Other things being equal,

				this greater support for the union should lead to greater bargaining power and higher wages for the workers at the larger plants.
Miller and Mulvey (1996)	The larger is a firm, the more likely it is to be highly unionized. The larger is a firm, the higher the average wage of its employees.	1993 Survey of Trainingand Education Australia	OLS, gender decomposition	• Large firms pay relatively higher wages than small ones, independent of the influence of unionism. • In estimating earnings functions which do not include a variable to control for firm size, estimated union relative wage effects will be biased upwards.
Dynamic Monopsony hypothesis				
Green and Manning (1996)	• Firms face a supply of labor which is not perfectly elastic. In this model a firm can only maintain a relatively high steady-state size if it pays a relatively high wage. • The dynamic monopsony model assumes that, though firms which pay higher wages have lower quit rates and find recruitment easier, these processes take time to affect labor supply so that there are important frictions in the labor market. • In a monopsony model workers differ in quality, the high wage firms will be more selective in their recruitment and average worker quality will be higher in large employers.	British Household Panel Survey (BHPS) 1991 and the General Household Survey (GHS) of 1983, and the establishmentlevelWorkplace Industrial Relations Surveys (WIRS) of 1984 and 1990	OLS and MLE Empirical predictions of the Burdett-Mortensen (1989) model	• A dynamic monopsony model can explain important variations in the size-wage effects by sex and union status. The effect will be larger in the non-union sector than in the union sector, and larger for women than for men.

Efficiency wage hypothesis According to efficiency wage hypothesis, labor productivity depends on the real wage paid by the firm. If wage cuts harm productivity, then cutting wages may end up raising labor costs. Any reduction in the wage paid would lower the productivity of all employees already on the job. Thus the efficiency-wage hypothesis explains involuntary unemployment. Efficiency wage models provide a complementary explanation of the ESWP. Efficiency wage is either paid to minimize turnover costs to raises workers' effort level or to enhance feelings of loyalty to their employer or this could be a strategy to attract a higher quality pool of applicants. Efficiency wage hypothesis is linked with monitoring cost explanation in a way that due to large discrepancies in technology and product quality, large employers pay more in order to avoid turnover as monitoring is more difficult in larger firms than in smaller firms. Yallen (1984).				
Kruger & Summers 1989	• A first model of efficiency wages postulates that they are paid in order to minimize turnover costs. If firms must bear part of the costs of turnover, and if turnover is a decreasing function of the wages firms pay, there may be an incentive to raise wages in order to minimize turnover costs. • A second possibility is that increasing wages raises workers effort level. Workers who are paid only their opportunity costs have little incentive to perform well since losing their job would not be costly. By raising wages, firms may make the cost of job loss larger and thereby encourage good performance. • A third model postulates that workers' feelings of loyalty to their firm increase with the extent to which the firm shares its profits with them. These feelings of loyalty may have a direct effect on productivity.	May 1979CPS and Quality of Employment Survey (QES).	OLS , wage equations	The evidence suggests that industry wage differentials are successful in eliciting better performance through reduced turnover and increased effort. Competitive considerations play an important part in wage setting controlling for labor quality and compensating differentials.

1.3 Empirical Evidence on the ESWP

The main goal of most of the empirical studies is to assess several hypotheses about the determinants of the size-wage premium. List is provided chronologically.

Author and Year	Main Question	Data	Methodology	Relevant Conclusion
Mellow (1982)	This study tests the hypothesis of unionism	Current Population Surveys 1979 for USA	OLS	• Union/ nonunion relative wage differential is much greater in small firms and/or plants. • Firm size and union membership have proportionally greater impacts on the expanded measure of compensation than on wages. • Positive association between firm size and wages is found across broad industry groupings and among firms operating in competitive product markets as well as those operating in more concentrated settings.
Brown and Medoff (1989)	The study aims to test the neoclassical and institutional hypothesis of ESWP	Current Population Survey (CPS), Quality of Employment Survey (QES), Survey of Employer Expenditures for Employee Compensation (EEEC), The Wage Distribution Survey (WDS), and The Minimum Wage Employer Survey (MWES) for USA	Ordinary Least Square (OLS)	• Those who work for larger employers receive higher wages. • Employer size premia are smallest in the highest pay grades. • Large employers hire higher quality workers. • Differences in working conditions, threat of unionization and product market power seem not to explain much of the size-wage differential.
Troske (1999)	This study examines seven	Worker-Establishment	OLS	• The results support the labor quality and

	explanations for the employer size-wage premium: Capital-Skill Complementarity, worker skill, plant age, managerial skill, monitoring, rent sharing and the skill of capital	Characteristics Database (WECD) for USA		capital-skill complementary hypothesis. • The matching of more-skilled workers together in larger plants accounts for approximately 20% of both the establishment and firm size-wage premium, while the capital skill complementarity hypothesis accounts for approximately 45% of the firm size-wage premium.
Bayard & Troske (1999)	The study aims to test: productivity differences between workers in large and small establishments and rmatchingof skilled workers in large establishments.	New Worker Establishment Characteristics Database (NWECD) 1990 for USA	OLS	• Neither productivity nor segregation by skill accounts for the premium across individual establishments. • The greater productivity of workers in larger establishments accounts for over half of the firm-size wage premium in both manufacturing and services.
Morissette (1993)	The study tests various hypothesis of ESWP	1986 Labor Market Activity Survey (LMAS) for Canada	Weighted Least Square (WLS)	• Results support the efficiency wage hypothesis but do not support the working conditions and the union avoidance hypothesis. • Large firms employ workers with more unobserved abilities. • If firms of different sizes differ in ease of monitoring workers, in training costs, or in their reliance on teamwork, they find it profitable to pay differing wages to identical workers as suggested by efficiency wage models.
Davis and	The study aims to assess	Census of Manufactures data in the Longitudinal Research	full distribution accounting	• An evidence of unionized production workers is found in large establishments

Haltiwanger (1996)	several hypotheses about the determinants of the size-wage structure	Database (LRD) and 1983 Current Population Survey (CPS) for USA	methodology of John, Murphy and Pierce [1993] to decompose total wage dispersion	• worker heterogeneity tends to rise with establishment size • Incentive-based pay mechanisms in larger establishments lead to greater wage dispersion conditional on observable characteristics.
Idson (1996)	This study investigates the long-term employment relationships inlarge plants and firms	(i)1979 May Current Population Survey (CPS), (ii) the 1973/1977 Quality of Employment Survey(QES), (iii) the second wave (1982) of the employer survey of the Employment Opportunities PilotProjects (EOPP), and (iv) the National Longitudinal Survey of Young Men (NLS).	MLE	• Greater levels of intra-firm job changes and lower failure probabilities constitute the underlying basis of size differentials in both wages and mobility • Large employers have an inherently greater capacity to establish long-term relationships with their employees due to the larger job market within the firm and their higher survival probabilities. • Results do not support the monitoring and union threat hypothesis of firm size effect.
Daniel & Sofer (1998)	This study tests theory of compensating differentials and theory of segmentation.	INSEE survey on living conditions of households(1986–87)	Wage Bargaining model, OLS, IV	• Coexistence of a negative relationship between wages and good working conditions for the whole sample (market effect) and a positive relationship in highly unionized sectors (union power effect).
Abowd et al (1999)	This study examines the role of individual and firm heterogeneity in the determination of wage rates;	DADS, large-scale administrative database of matched employer-employee information collected by INSEE	Panel data methods	• Personal heterogeneity was more important in determining the compensations as compared to firms’ heterogeneity. • The firm size effect would increase at a decreasing rate with returns to seniority would be negatively correlated with firm-specific intercept.

				• High wage firms are both worker productive and profitable while high wage workers reflect high productive workforce but no higher profitability.
Piekkola (2000)	This study examines the propositions that large firms pay higher wages that lower job search because of good reputationand long work history and reward risky workers with good performance ; and finally to test that large firms substitute high wages for highmonitoring costs.	Data on Finnish labor market during 1989-1996 on individual employees from the Employment Statistics	Panel data methods	• Firms recruiting personnel with unobserved human capital perform better. • It especially pays for large firms to recruit high wage earners, whether in terms of paying reputation wages or employing risky workers. • Large firms are more willing to retain their monopoly power and maintain large worker reallocation or fire employees that turn out to perform badly.
Criscuolo 2000	The study tests several explanations of ESWP	IAB (German institute) Employment sample for the period 1975-1995.	Panel data methods	• A positive and significant plant size-wage effect after controlling for a broad set of individual and job-related characteristics • Unmeasured individual heterogeneity contributes only in part to the existence of a positive and significant plant size-wage gap • the size-wage effect remains partly unexplained even after accounting for observed and unobserved firm and worker characteristics.
Hu (2001)	The study aims to examine the firms' different hiring behaviors and the	Benefits Supplement to the CPS, between 1979 and	• OLS and linear probability regressions. This	• large firms act strategically in their hiring practices and compensation structures to attract young workers by rewarding them more

	accompanying compensation structures.	1993.	model is of the firm's cost minimization problem within an employee search framework.	relative to small firms. • firm size-wage premium between those two size groups disappears for newly hired white collar workers aged 35 or older
Belfield and wei (2004)	This study test competing explanations for the ESWE, and compare their relative and cumulative effects.	1998 Workplace Employment Relation Survey (WERS) for UK	WLS	larger workplaces • have internal labor markets that reward effort and firm-specific capital; • allow less discretion on tasks, organize better channels of communication, conditions are less congenial; technologies are more complex • Unions allow workers to obtain more of the surplus.
Margolis & Salvanes (2001)	This paper tested the rent sharing hypothesis.	matched firm-worker panel data from France and Norway, DADS (1987 – 1995) , linked employer_employee data from statistics norway (1988 – 1995);	Bargaining model, panel data methods	• The study found a positive and significant relation between profit per worker and log annual earnings. But there are other explanations, beside rent sharing, that explain the significant coefficient. • They showed that Statistical-economic explanations of (endogeneity of profits, omitted variable biases in terms of individual productive characteristics) are slightly more successful, as instrumentation reduces the significance level in France to 89% (via an increase in the standard error of the estimate).
Abowd et al (2001)	This study examines employer-level measures	"Enquête Emploi" (Labor Force Survey) for the years	Panel data methods, difference-in-	• For France individual characteristics and establishment effects explain more of the

	of compensation and the productivity and profitability of the business in France and United States	1990 to 1998 for France Current Population Survey for US for the years 1981 to 1991.	difference estimator	variability in compensation outcomes than in the United States. • The observable and unobservable components of compensation are identically correlated in the two countries. • Higher paid workers, either because of individual characteristics or establishment effects, are employed in firms that are more productive. • Higher pay due to enterprise heterogeneity is associated with higher profitability in France but lower profitability in the United States.
Lanfranchi, Ohlsson & Skalli, (2002)	This study examines compensation structure of shift workers.	ECMOSS 1992	Switchingregression model with endogenous switching	• Workers may choose shift because of compensating wage differentials but it is also possible that they have preferences for shift work. • A wage rate for shift workers is 16 percent higher than for day workers and shift premium is significant for shift work choice; • Shift premium is significant for the choice to work shift. A 1 percentage point increase in the shift premium increases the probability of shift work by 0.87 percentage points.
Fakhfakha and FitzRoy (2002)	This study examines whether wages are systematically related to firms' ability topay, or to measures of industry	Survey on Employment structure and cost of labor, EAE 1992 for France	WLS	• Strong confirmation of pecuniary rent-sharing for basic wages • Firms appear to share rents with manual workers independently of union influence.

	profitability.			
Lallemand and Plasman 2005	This study analyses the magnitude and sources of the firm-size wage premium in the Belgian private sector	1995 Structure of Business Survey (SBS) and the 1995 Structure of Earnings Survey (SES) for Belgium	OLS	• Existence of a significant and positive firm-size wage premium, even when controlling for individual characteristics and working conditions. • Findings do not support the hypothesis that large firms match high skilled workers. • The elasticity between wages and firm size is significantly larger for white-collar workers and comparable in the manufacturing and the service sectors.
Paez 2003	This study compares wages offered by employers to fill openpositions in each of the four Colorado Front Range JobVacancy Survey regions.	Colorado's Job Vacancy Surveys 2001	OLS, ANOVA regression	• The effect of employer size on entry-level wages offered by Colorado Front Range employers is smaller than any other category of vacancy characteristic. • Firm size effect still exists and it is not explained by human capital or institutional vacancy characteristics.
Fakhfakha& FitzRoy 2005	This study tests the hypothesis of dynamic monopsony or upward sloping labor supply curves.	Large sample of French firms observed from 1986 to 1996.EAE (Enquête annuelled'Entreprises)ESE (EnquêteStructure des Emplois)	OLS, fixedeffects, random effects and Generalised Method of Moments or GMM	• Strong evidence of a much larger long run ESWE controlling for worker quality and compensating differentials with lagged wages, and for profitability (rent sharing). • Results suggest upward sloping labor supply since temporary shocks or unobserved worker quality cannot plausibly explain the ESWE. • Control for profit per employee has a highly significant positive effect on wages (-growth), which is consistent with rent sharing

				hypothesis.
Lallemand & Rycx 2005	This study examines wage dispersion in European union countries	1995 European Structure of Earnings Survey. This harmonized survey, covering four European countries (i.e. Belgium, Ireland, Italy, and Spain),	Lemieux's version (2002) of the full distribution accounting methodology developed initially by Juhn, Murphy and Pierce (1993).	• Within-establishment wage dispersion rises with size because large employers have a more diverse workforce. • Between-establishment wage dispersion decreases with employer size because smaller establishments are technologically more diversified and hence exhibit greater diversity in average workforce skills • Smaller establishments are found to rely more on incentive-based pay mechanisms, particularly in countries with a low trade union coverage rate.
Lallemand, Plasman & Rycx 2005	This study examines the magnitude and determinants of the establishment-size wagepremium in five European countries (Belgium, Denmark, Ireland, Italy, and Spain)	1995 European Structure of Earnings Survey	OLS	• Existence of positive and significant establishment-size wage premium in all countries controlling for human capital variables, occupations and gender. • The magnitude of the elasticity between establishment-size and workers' wages fluctuates considerably across countries. • Results show the existence of a significant and negative relationship between the size wage elasticity and three collective bargaining characteristics, i.e. the degree of centralization, the degree of coordination, and the trade union density.
Millimet 2006	The study tests the proposition that the size–	1994 WorkerRepresentation and Participation Survey	OLS, MLE and stochasticfrontier	• A 20% of the difference in observed wages across large and small employers is attributable

	wage premiumpartially reflects the heterogeneous job search skills(or labor market information more generally) ofindividuels.	(WRPS) for USA	model of Aigner et al. (1977)	to differences in unobservable job search skills.
Lane, Salmon & Spletzer 2007	This paper tests the magnitude of occupation and establishment wage differentials, the sorting of high-wage occupations into high-wage establishments.	micro data from the Occupational Employment Statistics (OES) program at the Bureau of Labor Statistics (BLS), USA 1996 1997	simple regression-based wage decomposition	• After controlling for detailed occupation, 21 percent of wage variation can be explained merely by knowing the individual’s particular establishment. • Controlling for skill explains part of the estimated establishment wage differentials.
Heyman 2007	The study aims to answer whether the size–wage premium is really arelationship between employer age and wages	Swedish Level ofLiving Surveys (LNU) in 1991, Swedish Establishment. Survey (APU). employment data from Statistics Sweden	Spline function model Greene (1997) and OLS	• Inclusion of firm age does not affect the impact of firm size on wages.
Feng 2009	This study examines the size wage premium controlling of training status and individual and job level unobservedHeterogeneities	NationalLongitudinal Study of Youth's 1979 Cohorts (NLSY79) for USA	OLS, Panel data methods	• Wage increases associated with receiving on-the-job training are less in large establishments than in small ones. • One cannot easily rationalize the existence of size-wage premium in a competitive setting, using differences in training and other unobserved and possibly size-dependent worker heterogeneities. • One has to look beyond the perfect competition paradigm and consider monopsonistic

				competition models

Non-Random Selection and returns to unmeasured ability

As explained in the introduction that workers are not randomly distributed across the size spectrum. If unobservable differences in productivity affect the allocation, neglect of these could impart a bias in the estimate of the effect of employer size on wages. One of the most promising explanations turned out to be self selection of workers. It is possible that workers apply themselves selectively to small or big firms or, alternatively, the small and big firms have different recruitment strategies for different pools of workers. Therefore, the discussion on the effect of size on wage is incomplete without considering the nonrandom allocation of workers. There are studies that focus on the role of observables and unobservable characteristics of workers in the explanation of wage premium paid by large employers. Heckman two step estimation procedures have been very famous in this type of questions to distinguish the rewards to observable and unobservable skills of workers in different size of establishments. Few studies are outlined below:

Idson and Feaster (1990)	This paper studies the nonrandom selection of workers into establishments of different sizes	May 1979 Current Population Survey (CPS)	Heckman two-step methodology and wage decomposition	• The nonrandom assignment of workers acts to diminish the wage gap between small and large firms. • A significant positive selection bias in small firms and negative selection bias in large firms. • Large employers attract better educated workers, who are willing to retrain; whereas small firms attract predominantly those who have a strong preference for independence and do not like to conform to stricter work organization necessary in larger firms.
Main and Reilly (1993)	Same as above	Economic and Social Research Council's 'Social Change and Economic Life Initiative' in 1986	Heckman two-step methodology	• No evidence of non-random sorting of workers across plant size is detected. • The unobserved traits that would raise men's wages also made it more likely that they would be located in firms in the

				smaller size groups. The mean wages of workers in a small size category are thus higher than the mean wages that would have prevailed if workers had been randomly allocated to size categories.
Silva 2004	This study tests thepossibility of different returns to observable human capital variables as well as examines the role played by unmeasured skills in driving the allocation of workers across firms of different sizes.	Longitudinaladministrative sourceby the Ministryof Employment (1992-1998) for Italy	Panel data methods, GMM	• The observed skills; namely, education, age, and tenure have high returns in large firms, while the opposite is true for high skilled occupations and for the gender gap. • The price of non-observed skills is reduced as firm size increases. This finding is consistent with explanations based on the premise that large employers have more difficulty monitoring workers, which therefore leads them to monitor less closely.
Ferre & Lluis 2004	This study tests the returns to unmeasured ability between large and small firms.	Longitudinal data of Survey of Labour and Income Dynamics (SLID) for Canada	non linearinstrumental variable estimations	• The returns to unmeasured ability are significantly greater in medium size firms relative to small firms but are not significantly greater in large firms relative to medium or small firms. • In large firms monitoring hypothesis dominates the ability sorting hypothesis in which large firms choose workers based on observable characteristics and pay them wage premium to avoid monitoring costs.
Lluis 2008	This study examines differential pricing of skills	LFS for canada and the	Heckman two-step	• Selectivity effects on wages are present and are similar in U.S. and Canada with

	in USA and Canada	March CPS for USA.	methodology	evidence of positive selection in both small and large firms. • There are substantial and significant country differences in the returns to education by firm size.
Winter-Ebmer (1995)	This study examines whether layoff risk explain the firm- size wageDifferential	Austrian social security recordsin May 1991	Probit, OLS	• This study introduces the effect of job insecurity on wages. Employment stability in a firm is considered to be a selection device, where workers, who are subject to greater layoff risk, receive lower wages. Workers with low human capital sort themselves into smaller and more unstable firms and they receive lower wages.
Gibbons and Katz (1992)	This study tests that the forces that cause sorting by measured human capital cause similar sorting by unmeasured human capital.	1984 and 1986 CPS Displaced Workers Surveys (DWS).	Panel data methods	• The measured differentials simply reflect unmeasured differences in workers' productive abilities. • If matching is important then endogenous job-change decisions can create important self-selection biases even in the first-differenced estimates of industry wage differentials. • The wage change experienced by a typical industry switcher would closely resemble the difference in the relevant industry differentials estimated in a cross-section.

1.4 Conclusion

The above survey of literature presents different explanations of size-wage differential; however, there is no consensus on the reasons of this differential. Perhaps, no single theory can provide a complete explanation of inter-industry wage differences because different theories are of greatest importance in different sectors of the labor market. One way or the other, every model lacks something and one cannot make a precise conclusion. Unmeasured-ability models do not explain correlations between industries that pay high average wages and industries that earn large profits, have high capital-to-labor ratios, and are populated by large firms. Efficiency-wage models do not differentiate the high correlation of the industry wage premium across occupations. And rent-sharing models do not motivate the observed similarity of the industry wage structure across countries with very different market systems. This dissertation is not presenting another theory of the size-wage explanation but keeping in view the data and time limitations, it will be an addition in the efforts of identifying the wage differential in the presence of nonrandom allocation of workers into employers of different size. Earlier studies focused on the measurable skills of the workforce. But if one does not take into account the nonrandom sorting, then, the estimates can be biased. Various studies, done recently, focused on unmeasured or unobservable human capital as omitted variable that causes the size-wage effect to exist.

We proceed by presenting some descriptive statistics and cross section estimation using regression and propensity score methods. But these methods are confronted to the same problems of endogeneity and selection bias. The dataset available to us for this dissertation is matched employer-employee cross section dataset. Thus, we are faced with the endogeneity problem in general and nonrandom selection into employers of different size in particular. Therefore, it is inevitable to explore data potentials. It will be followed by regression and propensity score methods.

Chapter-2

DESCRIPTIVE STATISTICS

2.1 Introduction

This chapter presents a detailed description of the data. The results in the subsequent chapters are computed using ECMOSS from 1992 to 2006 as described in the introduction chapter. Descriptive statistics for ECMOSS 1992 are presented here in detail while description of other surveys can be found in the Appendix-A. The main focus of this chapter is to describe the construction of variables and to highlight main features of the data.Few conditions or restrictions have been put to make things normal or drop the outliers.Main review is done for 1992 dataset as two chapters are based on this year survey. Descriptive statistics for other chapters are provided in Appendix-A

2.2 Data

2.2.1 Introduction and Brief History

The data for this dissertation come from INSEE surveys called ECMOSS (Labor Cost and Wage Structure Survey). As mentioned earlier, the data is drawn from two sets of surveys called

“The Cost of Labor” (ECMO) and “Wage Structure Survey” (ESS), jointly called ECMOSS (Labor Cost and Wage Structure Survey) for the private nonagricultural sector collected by INSEE.

The survey on the cost of labor (ECMO) is designed to examine the average annual cost of labor per employee (that includes wages, social charges, expenses of training, etc.) and the average hourly cost of labor per hour actually worked (excluding holidays, sick leaves, strikes, training periods, etc.) . It provides the detailed structure of these costs by sector, by economic activity, region where the establishment is based and by the size of the establishment. It aims to evaluate the total cost of an hour actually worked, across the European Union and for a given sector to compare the price competitiveness.

The survey on the structure of wages (ESS) aims to provide to Eurostat the individual data for a sample of employees on wages, its components (basic wage, various kinds of bonuses, overtime payments), the determinants of wages (education, qualification, professional experience etc) and the workplace characteristics (business sector, firm size, region, actual duration of work, occupation, nature of the employment contract). Its objective is to compare the wages (hourly and annual) and its components among European Union for a given profession, for a given level of qualification or in a particular sector.

Surveys on the structure of wages (ESS) were conducted in 1966, 1972, 1974, 1978, 1986, 1992, 1994 and 2002. Two broad regulations were made concerning these surveys. The first regulation was made in 1999 for the frequency of these two surveys (ECMO & ESS) that these are quadrennial. Second regulation was made in 2000 for the list of variables to be included in ESS. Variables related to employees fall within three broad themes: identification, individual characteristics and elements of remuneration. Lately, INSEE merged three surveys ESS, ECMO and ACEMO and launched first ECMOSS in 2006 on the 2005 data. Since then, ECMOSS is an annual survey. For this dissertation, we have used surveys for the years 1992, 1994, 2002, 2005 and 2006. No compiled data is available from 2006 until now[28].

[28] *These surveys are provided by Reseau Quetelet, Maurice Halbwachs Centre upon official request.*

2.2.2 Sample

The sample consists of 14,000 establishments and approximately 140,000 employees from the nonagricultural private sector. Employers respond to a questionnaire describing many workplace characteristics and give information about a random sample of their employees. This data set has many observations and a large variety of employer and employee characteristics which can be used as interesting instruments for the work on the cost of labor and structure of wages.

This is a very rich database consisting of socio economic characteristics of workers along with characteristics of establishments. There is no other data set that provides at the same time such information on the size of the establishment, its principal activity, its geographic location, its wage structure, the composition of its wages. Moreover, one can find detailed information on the education, profession, industrial distribution, age, nationality, family situation and number of dependent children of the workers.

2.3 Construction and Definition of Variables

Below, the main variables used in the estimations are described. Wherever possible, some variables have been constructed and some are modified as required. Table 2.1 shows the definition of the continuous and categorical variables. This is followed by summary statistics of all variables used for estimation in the subsequent chapters.

Table 2.1 Description of variables

Variable	Description	Categories
Wage/ salary/ remuneration	In each survey, two measures of wages are available; the gross wage and the net wage. The net wage is the standard contractual wage to which complements are added. The difference may	Hourly wage is computed from total number of working hours in a year. [29]

[29]*There is one variable called 'ratq' in the data which is about annual gross remuneration. This is used along with annual number of hours to create hourly wage. Missing observations are excluded. First and last percentile is dropped to exclude outliers. The minimum wage varies from 29 to 236 francs (Minimum 4 euro to maximum 35 Euros per hour).*

	include individual bonuses, bonuses related to group performance and those related to establishment or firm performance, such as mandatory French profit-sharing schemes (*participation*). We thus expect gross earnings to be more affected by profits than basic wages. Distinguishing the two measures of wages can tell us how the wage components are related to size. Therefore, in each chapter all the estimations are performed twice one for each measure of size and results are compared to analyze how the compensations and allowances vary across different size categories of employers. This plays an important role in measuring sources and magnitude of the size wage premium for France in the reference years.	
Working Hours	Total number of working hours in one year is available in the data through which we can form monthly, annual or hourly wage.[30]	
Establishment Size	Establishment size or the number of employees working in the establishment at all locations is categorized in three groups (small, medium and large) for simplicity. The robustness of the threshold is verified in each estimation method employed. [31]	Small = 1-49 employees Medium = 50-199 employees Large = 200 and more employees
Industry	Industries have been classified into three main groups: manufacturing, trade and services. Three categories were formed out of 12 categories. Industries are classified by INSEE based on NAP level 15A.[32]	Indus=1 if Trade, Indus=2 if Manufacturing and Indus=3 if Services

[30]*In France, the legal length of the working week is 35 hours in all types of companies. The working day may not exceed 10 hours. The maximum working day may be extended to 12 hours under a collective agreement. In principle, no more than 48 hours a week may be worked, 44 hours per week on average over a period of 12 consecutive weeks (up to a maximum of 46 hours, under conditions). In the data, total numbers of working hours in one year are available through which one can form monthly, annual or hourly wage. Around 43 hours in one week are included for estimations. The variable for annual hours is called 'hran'. The observations with zero hours are dropped.*

[31]*In all surveys, except in 1992, we have size as categorical variable due to which it was not possible to perform different variations in size.*

[32] *The Classification of Activities and Products (NAP) was in force in France from 1973 to 1992. In 1993, the NAF (French Classification of Activities) and the CPF (French Classification of Products) replaced the "activities" version and the "products" version respectively of the NAP 73. Since 1st January 2008 a revised version of the NAF (NAF rev.2) has been in force; it supersedes the first revised NAF Rév. 1 in force from 2003 until 2007. NAF rev. 1 replaced the NAF which had been in force since 1 January 1993. In 2005 onwards, NAF revision 1 is used.*

Region	France is divided into 27 Regions, 21 regions are in Metropolitan France, and remaining regions are overseas. Each region is further classified into department with different size and number depending on geographical location.[33]	Names of 21 regions are: Limousin, Franche-Comté, Auvergne, Champagne-Ardenne, Basse-Normandie, Bourgogne, Poitou-Charentes, Alsace, Haute-Normandie, Picardie, Languedoc Roussillon, Lorraine, Centre, Midi-Pyrénées, Bretagne, Aquitaine, Pays de la Loire, Nord-Pas-de-Calais, Provence-Alpes-Côte d'Azur, Rhône-Alpes, Île-de-France
Union	The unionisms and its features are described in detail in the introduction. But, unfortunately, we do not have data related to number of union representatives in one plant, number of representation by central union federations and the concluded agreements or negotiations. There are three variables, two on the presence of unions and staff representative and the third is on the salary agreements in the survey year. But the variables on the presence of personal delegate and union delegate do not ensure that all individuals are union members. The variable on the agreement could be more relevant if we could have the information on the outcome of negotiation as any agreement in an industry covers all employees.[34]	1. Presence of staff representative= yes/no 2. Presence of union representative= yes/no 3. Presence of any salary negotiation in survey year=yes/no
Status	Status: In 2002, 2005 and 2006 surveys, one variable 'status' is available, that classifies management or non-management cadre of employees.	Status=1 for management Status=0 for non- management

[33] *There are two variables corresponding to regions in the data; firstly, we have one variable related to administrative regions of France called 'Zeat' which includes eight categories; Ile de France, Bassin Parisien, North, East, West, South West, Central East and Mediterranean. This variable is available in all surveys but regions are heterogeneous and we cannot draw conclusions based on the administrative region when working with individual data as many regions are merged into one broad geographical region. Secondly, there is another variable called 'dep' which gives information of all 95 departments where establishments can be found. This is very relevant to serve our purpose but this variable is only available in the survey of 1992.*

[34] *As mentioned earlier in the introduction that minimum wages (per occupation) are set by the industry-level bargaining that covers all the firms so this could be used to estimate the additional effect of a supplementary firm agreement, rather than the usual union differential between collective and individual bargaining.*

Employment contract	There are two main types of employment contracts in France, the permanent (CDI) and fixed term (CDD) contracts. All other types of contracts are grouped in 'other contracts'.	CDI=1 CDD=0
Gender	Gender: a dummy variable for male or female is created. Base category is female.	Male = 1,Female = 0
Age	Individual between working age 25-60 are included in the sample. Four dummy variables for age have also been used for descriptive statistics. (25-30, 31-40, 41-50, 51-60). In the last chapter (chapter-6) birth years of employees are created from age variable for making cohorts.	
Education	For the French data, the educational variables are based exclusively on degree attainment. We used classification of CEREQ[35]. Since education; the information on education is available as categorical variables which further reclassified based on CEREQ classifications. The observations with non-declared education are dropped from the sample.	Edu=1 No degree Edu=2 Before Bac without degree Edu=3 CAP/BEP Edu=4 Bac professional and technical Edu=5 Bac general Edu=6 Bac +2 Edu=7 Bac+3 and plus
Experience	Experience represents the total working experience in the labor market in years. It is the total experience with the current employer and outside the current employ[36]. This is continuous experience and does not give information on the interruptions.	
Tenure	Tenure is defined in terms of the length, in years, of the current employment relationship.	
Family Situation	Three dummy variables are created, married (marié), single and others (widowed, divorced).	Single=1

[35] *Cereq (Centre d'études et de recherches sur les qualifications) is a French public administrative institution that conducts research in the areas of labor market, skills and vocational training.*

[36] *For French data Abowed et al. (2001) calculated potential experience as age minus school-leaving age (18) the same definition is given in Table 14 in CEREQ-DEP-INSEE 1990 [Céreq, Dep, Insee (1990), « Bilan Formation-Emploi 1986 », Insee-Résultats, n° 75, 150 p.] to calculate potential experience. We compared the available experience in the data file with the one computed this way and there was no difference.*

	This variable is available in 1992 survey only. Being married does not correspond to people living together or in a civil union. This may be included in the 'other marital status' type.	Married=2 Other (widowed, divorced)=3
Number of dependent children	We have a continuous variable on number of dependent children less than eighteen years of age. The ages of the kids are not given. This is available in 1992 survey only	
Profession	The French occupations are a recode of the 'Profession et Categories Socio-professionelle (PCS)' codes common to all INSEE surveys.For professions PCS 1992 level-4 was used. But all of the categories are similar to the PCS 2003 level 4. [37]	Management and high intellectual professionals=1 High skilled white collar=2 Low skilled white collar=3 Blue collar =4
Survey Years	Survey Years: The available surveys for this study are for the years 1992, 1994, 2002, 2005 and 2006. After 2006 no latest data set is available and for 1994 limited variables are available, therefore, it is not used for estimations. Moreover, the third and fourth chapter only uses 1992 surveys due to the availability of most of the information related to employers and employee which helped to construct instruments. [38]	

[37] *The new version of the nomenclature of professional categories and professions (PCS-2003) is the result of renovation work done on the nomenclature in force since 1982. Chief executive officers are excluded from the data as we are estimating hourly wages and chief executive officers are earning abnormal salaries compared to blue collar workers. Therefore, they are excluded from the data. Similarly the agricultural workers are excluded for the same reason.*

[38] *If the same kind of information is available for later surveys and for the recent years, then the results could be compared. This is left for further research in this field. Due to time constraint and the availability of this data this comparison remains undone.*

2.4 Summary statistics

The ECMOSS survey for the year 1992 consists of 14,000 establishments and approximately 140,000 employees from the non-agricultural private sector. Descriptive statistics of all qualitative variables and distribution across three size groups is presented in Appendix A. For each survey first, the distribution of various individual and employer related characteristics by establishment size and on the whole is presented and it is followed by mean hourly wage against those attributes. Lastly, the comparison of gender mean hourly wage difference from 1992 to 2006 is provided.[39] The information related to workers' and employers' characteristics that is used for this dissertation is the size of the establishment, gross and basic hourly wage, annual working hours, educational level, professional level, sector, experience, tenure, family situation, employment contract, region, nationality, age and number of dependent children. The variables have been cleaned up by various changes. Outliers in the formation of individual's hourly wage are dropped. Weekly hours are limited to 43 hours per week. Missing values in the education, family situation, contract, experience and tenure are dropped. Population with working age between 25-60 years of age is included and the rest of the observations are dropped. The final cleaned number of observations is 74,696. Number of dependent children is an important information for the employees as it can influence many decisions. This variable is used for instrumentation but not included in the descriptive statistics.[40] Summary statistics of all variables is presented below.

[39] *The variable on education is reclassified (CEREQ classification) and the formation of region variable excludes one overseas region 'corse' that was previously added in the earlier version of thesis. It discards 66 observations. All the estimations are performed with the new changes and only the tables in the appendix-A follow the earlier construction of variables. For each variable the lower level is kept as base category but this does not change the number of observations. Total number of observations with changes is 74,696 in 1992 after excluding corse region while earlier it was 74,762 as reflected in Appendix-A. Earlier, Corse was made part of cote d'azur. For subsequent surveys, no change is made in the formation of variables.*

[40] *There are around 30thousands missing observations against this variable. Therefore to save the data this is not included in the description.*

Table 2.2 Summary statistics

	All		Small establishment		Medium establishment		Large establishment	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Hourly Gross wage	71,57	32,76	66,80	32,45	66,95	30,56	78,64	32,83
Establishment Size								
Small establishment	0,41	0,49						
Mediumestablishment	0,19	0,40						
Large establishment	0,40	0,49						
Gender	0,60	0,49	0,54	0,50	0,58	0,49	0,66	0,47
Experience	20,48	9,48	19,63	9,52	20,25	9,49	21,46	9,35
Tenure	11,25	8,94	8,46	7,74	10,27	8,61	14,55	9,16
Industry								
Trade	0,10	0,30	0,17	0,38	0,08	0,28	0,04	0,20
Manufacturing	0,38	0,49	0,25	0,43	0,26	0,44	0,57	0,49
Services	0,52	0,50	0,57	0,49	0,66	0,47	0,39	0,49
Education								
No degree	0,18	0,39	0,20	0,40	0,21	0,41	0,15	0,36
Before Bac without degree	0,17	0,38	0,14	0,35	0,17	0,38	0,20	0,40
CAP/BEP	0,34	0,47	0,34	0,47	0,34	0,47	0,35	0,48
Bac professional and technical	0,07	0,25	0,07	0,26	0,06	0,23	0,07	0,26
Bac general	0,05	0,22	0,06	0,25	0,05	0,22	0,04	0,20
Bac +2	0,13	0,34	0,13	0,34	0,14	0,35	0,13	0,34
Bac+3 and plus	0,05	0,22	0,05	0,21	0,04	0,19	0,06	0,24
Marital Status								
Single	0,23	0,42	0,26	0,44	0,24	0,43	0,20	0,40
Married	0,69	0,46	0,66	0,47	0,67	0,47	0,73	0,44
Other family status (divorced, widowed etc)	0,08	0,27	0,08	0,27	0,09	0,28	0,07	0,25
Profession								
Management and High Intellectual professionals	0,12	0,33	0,13	0,33	0,11	0,31	0,13	0,34
High Skilled White Collar	0,28	0,45	0,26	0,44	0,25	0,44	0,31	0,46
Low Skilled White Collar	0,24	0,43	0,30	0,46	0,25	0,43	0,18	0,39
Blue collar	0,35	0,48	0,31	0,46	0,39	0,49	0,38	0,49
Contract	0,95	0,22	0,94	0,24	0,92	0,27	0,97	0,17
Region								
Limousin	0,02	0,13	0,02	0,13	0,02	0,13	0,02	0,13
Franche-Comté	0,02	0,14	0,02	0,15	0,02	0,15	0,01	0,12
Auvergne	0,02	0,14	0,02	0,14	0,01	0,11	0,02	0,14
Champagne-Ardenne	0,02	0,14	0,02	0,14	0,02	0,14	0,02	0,15
Basse-Normandie	0,02	0,14	0,02	0,15	0,02	0,13	0,02	0,14
Bourgogne	0,04	0,18	0,03	0,18	0,04	0,20	0,03	0,18
Poitou-Charentes	0,03	0,16	0,03	0,16	0,03	0,17	0,03	0,16
Alsace	0,05	0,22	0,04	0,20	0,07	0,26	0,05	0,22
Haute-Normandie	0,04	0,19	0,04	0,18	0,03	0,18	0,04	0,19
Picardie	0,03	0,18	0,03	0,17	0,04	0,20	0,04	0,19
Languedoc-	0,02	0,13	0,02	0,15	0,01	0,12	0,01	0,12

Roussillon								
Lorraine	0,05	0,21	0,03	0,18	0,04	0,20	0,06	0,24
Centre	0,04	0,20	0,04	0,19	0,04	0,19	0,05	0,21
Midi-Pyrénées	0,04	0,20	0,05	0,22	0,03	0,18	0,04	0,19
Bretagne	0,04	0,20	0,04	0,21	0,05	0,21	0,04	0,19
Aquitaine	0,05	0,23	0,06	0,24	0,06	0,24	0,05	0,21
Pays de la Loire	0,05	0,22	0,05	0,22	0,05	0,23	0,05	0,23
Nord-Pas-de-Calais	0,07	0,25	0,06	0,23	0,06	0,25	0,08	0,27
Provence-Alpes-Côte d'Azur	0,07	0,25	0,08	0,27	0,08	0,27	0,04	0,20
Rhône-Alpes	0,11	0,31	0,11	0,31	0,13	0,33	0,10	0,30
Île-de-France	0,18	0,38	0,18	0,39	0,13	0,34	0,20	0,40
No. of Obs.	74696		30286		14514		29896	

Source : Author's calculations.

2.4.1 Main Observations from descriptive statistics (1992-2006)

- Number of observations after removing outliers in various surveys are 74,696 (1992), 30,216 (2002), 51,272 (2005) and 53,508 (2006).
- Mean hourly wage in various surveys is 71francs (1992)[41], 17euros (2002), 20euros (2005) and 21euros (2006).
- Mean hourly wage increases with employer size. Mean hourly wage difference by employer size becomes more important when size threshold limit changes to 50 or to 200 employees.
- Male population compared to female in the sample is around 60 % in all surveys. The difference is more in large compared to medium and small size establishments. The wages on average for male workers are higher compared to female workers and the difference is more in large size employers compared to medium and small size establishments. The highest proportion of female is working in the small and medium size and for male workers; the highest proportion is associated with the medium and large size employers.
- Mean hourly wage for a married worker is higher compared to single or other family status. The proportion of married workers is more important in large employers compared to small and medium. Mean hourly wage for a single female worker is higher compared to a married female worker. The reverse is true for male workers.

[41] *This is equivalent to 10.82 euros based on conversion from French francs to euros.*

- The largest sector is services sector and the smallest sector is trade. In large size establishments, the largest sector is manufacturing sector (except in the survey 2005 where the largest sector is service sector across establishments), whereas, in small and medium size establishments, the largest sector is services sector.
- Among sectors, mean hourly wage is higher in the manufacturing sector compared to services and trade except in the year 2002 where mean hourly wage in the services sector is higher compared to manufacturing and trade. In large establishments, the highest mean hourly wage is in the manufacturing sector for 1992 and 2006. In the survey year 2005, the mean hourly wage in the manufacturing and services is similar and in the year 2002 mean hourly wage in the services sector is highest across establishments.
- The level of education has improved on average over the survey years. In the year 2002 to 2006, a majority of the population in the sample has acquired higher educational level unlike in 1992 where majority of population hold technical short and primary education.
  - Mean hourly wage increases with educational level and remains higher for large size establishments. Small establishments have more capacity to absorb low educated people than large employers.
  - In small size establishments, major professions are blue collar and low skilled white collar, while, in large size establishments the main professions are blue collar and high skilled white collar. In the survey for 2005 and 2006, majority of population in the sample is associated with management and high intellectual professions followed by high skilled white collar jobs across establishments.
- Mean hourly wage depends on the level of profession, for higher level of professions mean hourly wage is higher. For all professions mean hourly wage in the large size establishments is higher.
- Ile de France is the biggest region where the largest proportion of establishments is based. It is followed by Rhone Alpes and Mediterranean.
- Among all regions, mean hourly wage in large size establishments is higher. Mean hourly wage difference between small and medium is low across regions.
- Average age of the population in the sample is 39 to 42 years across surveys. The majority of population is between 31 to 40 years of age and secondly between 41 to 50

years of age. In each age group mean hourly wage increases with employer size and also as age increases mean hourly wage increases.

- Mean tenure is 11-12 years across surveys. A majority of people in the sample has completed between 0-5 years in the current job, whereas, in large size establishments the largest sample of workers have completed 11-20 years in the current job.
- Majority of contracts in all surveys and in across establishments are fixed term employment contracts. Mean hourly wage of the same contract type in large is higher compared to medium and small.
- Average total labor market experience is 20 years in the data
- In region Ile the France mean hourly wage is higher in all industries compared to other regions. Higher mean hourly wage in large regions across sectors and higher mean hourly wage in manufacturing sector across regions.
- Higher mean hourly wage with higher educational and higher professional levels. Among professional categories other than high intellectual professions, the mean hourly wages across educational levels do not vary and stay almost similar for all levels of education.
- Mean hourly wage against level of education and sector show that for particular level of education, mean hourly wage is higher in the manufacturing sector compared to trade and services and the difference is prominent for higher educational levels. On the other hand, in the survey for the year 2002 mean hourly wage is higher in the services sector for lower levels of education and mean hourly wages are similar for higher educational levels in the manufacturing and services sector.
- The mean hourly wage difference across regions is more important for high intellectual professions compared to other types of professional groups. In other words, mean hourly wage for higher level of professions depends on regions but for lower level of professions it does not vary among regions.
- The difference in the mean hourly wage across sector exists for all professional groups with higher mean hourly wage in the manufacturing sector. The difference in the mean hourly wage across sector is more important for lower levels of professions.

Summary statistics and frequency distribution of variables related to unions is presented in Table 2.3 and 2.4. Union representative (délégué syndical) must be set in companies or workplaces with 50 employees or more and staff representative (délégué du personnel) must be set in companies or workplaces with 11 or more employees. Therefore, we may expect large establishments to have both type of representatives and this is evident from the descriptive table below. Overall, 41% of the individual responded that there was some sort of salary negotiation in the survey year.

Table 2.3 Summary Statistics on unions

	All		Small		Medium		Large	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Personal Delegate	0,65	0,48	0,34	0,47	0,72	0,45	0,95	0,22
Union Delegate	0,51	0,50	0,10	0,31	0,50	0,50	0,92	0,27
Negociation	0,41	0,49	0,15	0,35	0,42	0,49	0,69	0,46
Obs	**69,733**		**28,310**		**13,758**		**27,665**	

Table 2.4 Descriptive statistics on unions

Union Variables	**All**		**Small**		**Medium**		**Large**	
Personal Delegate	Number	Percent	Number	Percent	Number	Percent	Number	Percent
No	24,092	34.55	18,791	66.38	3,823	27.79	1,478	5.34
Yes	45,641	65.45	9,519	33.62	9,935	72.21	26,187	94.66
Total	69,733	100.00	28,31	100.00	13,758	100.00	27,665	100.00
Union Delegate								
No	34,431	49.38	25,346	89.53	6,918	50.28	2,167	7.83
Yes	35,302	50.62	2,964	10.47	6,84	49.72	25,498	92.17
Total	69,733	100.00	28,31	100.00	13,758	100.00	27,665	100.00
Negociation								
No	40,833	58.56	24,173	85.39	7,95	57.78	8,71	31.48
Yes	28,9	41.44	4,137	14.61	5,808	42.22	18,955	68.52
Total	**69,733**	**100.00**	**28,31**	**100.00**	**13,758**	**100.00**	**27,665**	**100.00**

2.5 Gender Wage Difference

In the descriptive statistics it is observed that the difference in the mean hourly wage between male and female workers is quite prominent with higher mean hourly wage for male workers; therefore, it will be interesting to see the distribution and returns to different characteristics of male and female by establishment size categories to identify the source of this difference. In this section distribution of gender by establishment size, mean hourly wage by establishment size and finally the comparison of mean hourly wage difference by establishment size is presented.

The difference in the mean hourly wage of male and female is computed and a comparison by variables and by establishment size is performed from 1992-2006. Table-A-12 presents the mean hourly wage difference across categories of variables and by establishment size. Gender mean hourly wage difference exists in all variables. There is positive difference in all cases. In some categories it is larger while in other it is small. Gender mean hourly wage difference is large in large size employers. As size increases, gender wage difference increases. In educational categories the highest mean hourly wage difference is in higher education levels. In type of contract, difference is higher in other types of contracts. Mean tenure category between 21-30 years contributes larger for wage difference. Higher age and higher level professions have more discrimination. Among professions gender mean hourly difference exists in all categories and larger in high intellectual professions. Among industries overall mean hourly wage difference is larger in trade. In 1992 it is larger in both trade and manufacturing, in 2006 it is larger in services.

On the whole, the higher gender difference in the mean hourly wage is observed in large employers, for higher educational levels, higher professional level, in the trade and service sector, for marital status as married, for higher experience, more age and higher tenure levels. The reasons of this prominent gender mean hourly wage difference can be the choices of female workers for less investment in human capital, less mobility, less flexibility in the career and employment related issues, preference for specific sector, workplace or management and/or more career breaks. There is a need to analyze it further and identify the sources of this wage difference in the presence of non-random selection of workers across employer sizes.

2.6 An Overview of the Mean Gross and Basic Hourly Wage by Employer Size

This section provides an overview of the mean wage by size. Three establishment categories are formed to group the establishment into small (1-49 workers), medium (50-199) and large (200 and more) scale employers. Table-2.5, below, shows the mean wage by three size categories in each cross sectional data. The comparison between the two measures of wages is presented to see whether the difference of two measures matters by size categories.

Table-2.5 Mean Gross and Basic Wage by Size of employer (three categories) and by Year

	1992		2002		2005		2006	
Size of employer	**Gross**	**Basic**	**Gross**	**Basic**	**Gross**	**Basic**	**Gross**	**Basic**
1-49	66,79	57,82	15,55	13,19	18,49	13,97	19,15	14,43
50-199	66,95	56,15	16,89	14,08	20,57	15,44	21,57	16,11
200+	78,64	62,3	18,43	15,04	21,42	16,45	22,19	16,82
Total	71,56	59,29	17,05	14,16	20,48	15,58	21,23	16,01

Note: In the year 1992 figures are available in Francs; while the rest of the figures are in Euros.

As mentioned above, the threshold limit for size is also important for French labor market, and it becomes important for certain legislation when size grows to 11 or 50. In general, mean wage is increasing by size. In the year 1992, the mean gross wage is the highest for large size, while, between medium and small it is almost the same. The mean basic wage is more in the small category as compared to medium size; while it again rises for large size employer. In the year 2002-2006, the mean wage is monotonically increasing by size for gross and basic wage. In the last two years, the difference between medium and large is very small. This might be due to the revision of SMIC and the implementation of Aubry Act law.

Second, if we look at the difference of gross and basic wage, which depicts the contribution of bonuses, allowances, over time payments, then one can see that the difference increases as size increases for 1992 and 2002. This suggests that large employer offer more benefits and may be employer size wage differential is based on the difference of gross and basic wage. This will be explained in detail in the chapters.

2.7 Conclusion

The detailed description of data shows some interesting facts. This motivates to explore the data more precisely, keeping in view objectives highlighted in the introduction part of this dissertation. Among different size of establishments, there are different roles of different variables. One is more important than the other which is clear in the distribution and mean hourly wage comparisons. In order to identify the sources and magnitudes of the size-wage gap and to predict the effect of establishment size on wage, controlling for observable and unobservable characteristics of workers and employers, we have to disentangle the combined effects.

Few conditions or restrictions have been advanced to drop the outliers. The difference of initial wage in different categories of establishment size can be thought of as difference in the composition of workforce by level of education, by experience, by tenure, by sector, by profession and by age etc. There is a tendency in big establishments to have more educated people, more experienced, having more tenure etc. Large establishments base their decisions in selecting workers on observable individual characteristics, therefore, the distribution and mean hourly wage for observable individual specific characteristics is employer size specific.

The detaileddescription of all variables and cross tabulation provides some interesting facts of the distribution of individual and employer characteristics and the mean hourly wage against those characteristics. Some variables play more important role than the other but we have to disentangle all effects of individual and employer characteristics to study the effect of establishment size on wage. As people are not the same in small or large size establishments nor does the employers' behaviors in selecting and evaluating workers is the same. The self selection of both agents, employer and employees, bring them together. We have to see how the effects will change when we take into account all the information into the ordinary least square estimation. We cannot draw conclusions using the information on data description on the behaviors of firms in rewarding workers and how establishment size is related to wage in the presence of self-selection. Thus, we have to move further and test various methods to identify the sources of the wage premium paid by large employers controlling for individual and employer characteristics.

From the gender distribution and mean hourly wage comparisons, the following question arises from the mean hourly wage difference: Why the mean hourly wage is higher for male workers and why the difference is increasing by establishment size? We have to answer whether the gender wage difference is employer size specific? Is it profession specific, industry specific or whether the returns to workers characteristics experience, education, tenure, age or family situation creates the wage discrimination?

The dissertation aims to answer many questions that are raised after exploration of the data. Nonetheless, the rich dataset is not sufficient to encompass all the unobservable factors that can affect the employer size and wage relationship.

Chapter– 3

THE ESTABLISHMENT SIZE-WAGE PREMIUM: AN ANALYSIS USING PROPENSITY SCORE MATCHING

Abstract

In this chapter ordinary least square (OLS) and propensity score matching (PSM) methods are applied. The objective is to estimate the magnitude of the size-wage impact controlling for observable individual and employer characteristics and to study the wage differential of people working in large size establishments with the matched workers working in small size establishments. OLS estimates with gross hourly wage show around 7% of the wage differential with respect to dummy for large size establishment, whereas, for net hourly wage the premium almost disappearsat 0.8%.Matching conditional on propensity score of the observed covariates reduces wages difference between people working in large and small size of establishments. A pre-matching wage difference of 17% between large and small and a post-matching wage difference of 9% for gross hourly wage is found. For basic hourly wage, the PSM shows mean wage difference of 8% and post-matching difference of 2.5%. The wage difference between large and small size establishments is mainly due to being male, to working in the manufacturing sector and in blue collar jobs. The results are robust for different matching algorithms. The employer size wage premium is more the result of the employer's compensation policy than the wage difference itself. Results show that preferences of large employers for workers with high earning capacity and distinct features of large size establishments determine the wage premium which reflects employer's heterogeneity in rewarding similar workers differently.

3.1 Introduction

In the last chapter of descriptive statistics, higher mean wage by employer size is observed. A detail description of data reveals that both worker and employer characteristics cause higher mean wage in large establishments. This chapter presents mainly two pieces of analysis. Firstly, to start with, Ordinary Least Square (OLS) estimation is performed in order to estimate the magnitude of employer size-wage premium (henceforth ESWP) controlling for observable individual and employer characteristics. Secondly, matching technique through propensity score is applied to study the wage differential of people working in large size establishments with the 'matched' workers in small size establishments. The underline objective is to compare the effect of working in large size establishments for those in large size establishments (treatment effect on the treated) with the average impact of working in large size establishments (average treatment effect). In doing so, we have to see the importance of personal, job and workplace characteristics in both size groups.

First step in evaluating the effect of employer size on hourly wage with cross section data is the standard OLS regression method. The standard Mincer (1974) wage equation is used to estimate the effect of employer size on hourly wages controlling for observable individual and employer characteristics. The random assignment of workers across establishments could explain the true causal effect of working in large size of establishment, but practically randomization is not feasible. If there are systematic differences in characteristics across people working in large or small size establishments that are likely to influence earnings, then failure to take into account those characteristics will bias any estimate of the effect of size on wages because there is a self-selection by workers as well as by establishments. In assessing the expected effect for individuals who work in large establishments, we have to evaluate how workers' earnings compare with what they would have received, had they not been working in large size establishments on average. In other words, we have to measure the 'counterfactual' outcome.

The matching methods are used to analyze the average treatment effects as an alternative to the regression method. The most common estimates with matching methods are the "average effect of the treatment on the treated" (ATT), which is the effect for those who participate in the program or are considered as treated, and the "average treatment effect" (ATE), which is the effect on all individuals (treatment and control group) (Imbens 2004). The difference and the choice between matching method and regression has been discussed many timesin the literature (Behrman, Cheng and Todd 2004, Smith and Tod 2005, Angrist and Pischke 2008, Dehejia and Wahba 1999 and 2002). The following arguments are often provided to compare both methods:

- Matching is a semi-parametric approach and does not assume a linear functional form in parameters unlike regression methods because the true relationship between explanatory variables and the outcome variable may be very nonlinear. Matching avoids this problem.
- It is possible that only treatment observations are found over certain ranges of x (explanatory variables), and only control observations over other ranges. OLS

extrapolates the results to unknown or missing observations, whereas, non-parametric methods restrict the analysis to only those areas that are similar. Compared to OLS matching is "better" as it compares only comparable individuals.

- To estimate treatment effects, the common support is a mandatory condition and matching methods are based on the common support condition. It is possible and usually observed that the treatment and control groups are different on average. The matching techniques reduce the differences between both groups since matching techniques seek to find the best match by re-weighting the individuals in the control group.

The matching method used for this study is Propensity Score Matching (PSM) proposed by Rosenbaum and Rubin (1983) which is defined to be the probability of treatment assignment conditional on observed baseline covariates. This solves the problem of dimensionality. In PSM we match on the predicted probability of going to large size as a function of observed variables rather than matching directly on observed variables. Once we have the distribution of propensity score for people working in large and small size of establishments we can compare the two densities to get a clear sense of the common support problem. (Rosenbaum and Rubin, 1983; Dehejia and Wahba, 2002; Heckman et al., 1998; Caliendo and Kopeinig, 2008; Galdo, Smith and Black 2007; Smith and Todd, 2005).

Rosenbaum and Robin (1983) defined two conditions,the Conditional Independence Assumption (CIA) and Common support condition. The CIA is the key identifying assumption underpinning the matching methodology. The CIA requires that conditional on observable characteristics; potential non-treatment outcomes are independent of treatment participation. The plausibility of the CIA depends on the informational richness of the data since the set of *X*’s should contain all the variables thought to influence both participation (working in large employer size) and the outcome (earnings) in the absence of participation. Hence, after adjusting for observable differences the mean of untreated (potential) outcomes is the same as for those receiving treatment. This allows non-participants’ outcomes to be used to infer participants’ counterfactual outcomes. However, this is only valid if there are non-participants for all participants’ values of *X* (this is known as the common support condition).

In order for PSM to work, however, one must be able to control for all characteristics affecting both selection into an establishment and the resulting outcome ‘wages’. This requires very informative data. ECMOSS 1992 data set, a matched employer-employee dataset, is used to implement this methodology. This is a rich dataset that include a wide range of information related to employees and employer characteristics that can help to capture the selection factors and factors that determine wage. If data are available to make the CIA plausible then matching is feasible. In the absence of CIA both methods, OLS and PSM, deal with selection on observables. In this scenario the two methods are complementary and best used in combination.

Controlling for differences in observable characteristics does nothing to alleviate the selection bias problem associated with the endogenous choice of size. Without addressing the issue of sample selection, the estimated treatment effect will be biased. Therefore, we focus on the ATT and report ATE as less credible estimates. However, it is worth mentioning that controlling for wide range of observable characteristics can go some way towards minimizing the bias associated with unobservable factors. The rich nature of the available data can help us to control a wide range of individual and employer characteristics in order to compute the wage differential between large and small size establishments.

The remainder of this chapter is set out as follows. Section two presents regression estimates; it is followed by empirical implementation of propensity score matching in section three. In section four, dynamics of wage differential from 1992-2006 are presented. Section five concludes.

3.2 Regression Estimates of Employer Size-Wage Premium

Table 3.1 presents the wage equation estimates for all population in the sample and by gender. Only the coefficients for establishment size are shown, while detailed estimates of all explanatory variables are shown in the *Appendix-BTable B-1*.[42]Based on OLS estimates the gross

[42]*Size is used as a dummy variable in both OLS and matching. Matching is performed through stata’s program ‘psmatch2’ that only uses the dummy treatment variable. There are other advancements in this field where one can do multi-treatment and continuous treatments. This is left for further work. In the descriptive statistics it was evident that the mean wage difference becomes more important when an establishment size grows to 200 and more workers. Therefore, the large size is defined as establishments with 200 and more employees.*

hourly employer size-wage premium (ESWP) controlling for observable characteristics is 7% for the whole population. The net wage premium after excluding the bonuses, overtime payments and other allowances is 0.8%. This reflects that the ESWP is basically the result of the difference in the compensation structure that large employers pay to their workers. Gender comparison reveals a size premium of 9% among males in large size establishments compared to males in small size groups. On the other hand for females the ESWP is 5%. We can say that wage difference among male in different employers sizes is large and plays more important role in determining the overall size-wage premium. This may be related to the employer’s composition structures of the workforce as men are over represented in large and women in small size establishments. Further, for net hourly wages, we observe small differences in wages for male sample in large establishments compared to smaller ones, and no differences in wage for female workers in different sizes.

Table 3.1 OLS estimation of the employer size-wage impact

Dependent variable log of hourly wage	Gross	Net	Gross	Net	Gross	Net
	All Population		**Male**		**Female**	
Size[a]	0.072***	0.008***	0.088***	0.018***	0.053***	-0.006**
	(0.002)	*(0.002)*	*(0.003)*	*(0.003)*	*(0.003)*	*(0.003)*
Worker's Characteristics[b]	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***
Establishment's Chracteristics[c]	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***
Constant	3.294***	3.218***	3.390***	3.278***	3.311***	3.244***
	(0.012)	*(0.011)*	*(0.016)*	*(0.016)*	*(0.016)*	*(0.015)*
Observations	74,762	74,453	44,699	44,516	30,063	29,937
R-squared	0.66	0.66	0.65	0.67	0.63	0.62
Pseudo R-squared	0.65	0.66	0.65	0.67	0.63	0.62

*Notes: The sample includes full time jobs, private sector, non agriculture and non chief executive officers. The dependent variable is the logarithm of the hourly wage rate. This is semi-log model where dependent variable is in log form and all the variables on the right hand side are without log. Robust standard errors of size coefficients are between parentheses. *** p<0.01, ** p<0.05, * p<0.1*
***a:** size is defined as number of people working in all establishments, it's a dummy with 1-199 equal zero and 200 and more equals one.*
***b:**Worker's characteristics include experience, tenure, marital status, education, profession. Educational categories are based on CEREQ classification and Professions are coded from PCS codes.*
***c**:Employer's characteristics include the sector, type of employment contract, region where establishment is based. Industries are classified as per NAP15,*
Source; ECMOSS 1992, author's calculations

Other coefficients of the control variables show the same signs and significance as those reported by previous studies. Male workers earn around 15% higher wages compared to female. One more year of experience increases the wage by 2.4%, one more year in the current job (tenure) increases it by 1.8%. Returns to education increases by level from 6% for the base category to 39% for the highest level. Further, regressions estimates also confirm the higher wages for higher professional levels, higher wage in manufacturing sector compared to trade or services sector, for permanent employment contract compared to temporary contract, for married compared to other marital status and in region Ile de France compared to other regions.

The detailed analysis of OLS across gender, industries (*Appendix-BTable B-2*) and professions (*Appendix BTable B-3)* confirms that the wage premium is higher in the large manufacturing establishments, for blue collar jobs and for male workers. OLS analysis across industries reveals that size-wage premium for gross wage is highest in the manufacturing sector (10.5%) followed by 6% in the services sector and 2% in the trade sector. Among professional groups the wage premium between large and small establishments in the higher level professions is smaller compared to lower level professions. A blue collar worker in a large size establishment earns a wage which is 14% higher compared to a blue collar worker in a small size establishment. This is in line with Silva's (2004) findings that for high skilled occupations wage difference is smaller between large and small and opposite is true for low skilled occupations.

After controlling for the observable individual and employer characteristics to a wide range as used by previous studies, we find a significant size- wage premium for the gross hourly wage, whereas, regression with net hourly wage almost removes the size-wage difference. As noted above, the regression-adjusted estimate of the wage premium paid by large size establishments involves inferring beyond the common support and cannot be interpreted as the causal effect of size on wages. Therefore, matching method is used which is generally preferred to standard regression methods as mentioned above. Results corresponding to matching method are presented in the following section.

3.3 Empirical Implementation of Propensity Score Matching

The first step in PSM analysis is the estimation of propensity score. Normally, a logit or probit estimation is used for this purpose, given that treatment is typically dichotomous (i.e., 1 for the treated and 0 for untreated units).One of the key issues in characterizing the propensity score is the specification of the selection model, i.e., the identification of the variables that determine the participation. There is no comprehensive list of clearly relevant variables to assure that the matched comparison group provides an unbiased impact estimate. For each evaluation it is important to consider what factors make the comparison units distinct from treated units. The choice of matching method also depends on the information richness of data. ECMOSS-1992 survey includes a wide range of individual characteristics including gender, experience, tenure, educational levels, professional distribution, nationality and many others. On the other hand, it includes information on employers including industrial distribution, region where employer is based, type of employment contract etc. These variables side by side affect selection into an establishment and wage. The same variables as used in the OLS estimation above are utilized to compute the propensity score. Results for the probit model are presented in *Appendix- BTable B-4.*

3.3.1 Propensity Score graph

An important step in investigating the validity or performance of the propensity score matching estimation is to verify the common support or overlap condition. We assume that the probability of going to large size establishments conditional on observed characteristics lies between 0 and 1 (implying participation is not perfectly predicted, that is, $0<P$ (size=1| X) <1). This assumption is critical to estimation as it ensures that units with the same X values have a positive probability of belonging to both large size and small size establishments.

Checking the overlap or area of common support between people working in large (treatment) and those working in small establishments (control groups) can be done with straightforward methods. One obvious approach is through visual inspection of the propensity score distributions for both the treatment and control group. Figure-1 shows the distribution of propensity score among treatment and control group before matching. In the control group the predicted probability of belonging to large size establishments ranges from 0.0068 to 0.9168 with a mean of 0.317.In the treatment group the predicted probability ranges from 0.0118 to 0.9519 with amean of 0.401. Thus, the zone in which there is no common support given by control group is above 0.9168. Further, enforcing ‘common support’ at the extreme cases result in the loss of 38 observations of the treated and 5 observations of the control group.[43]

[43]*There is a stata command ‘common’ which implements the trimming of extreme cases proposed by dehejia and wahba (1999). The option ‘common’ imposes a common support by dropping treatment observations whose propensity score is higher than the maximum or less than the minimum propensity score of the controls.*

Figure-1 Propensity score distributionoftreatment and control group

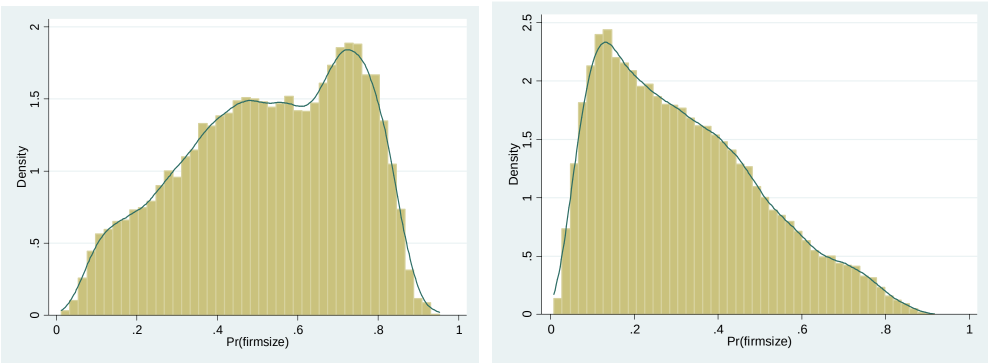
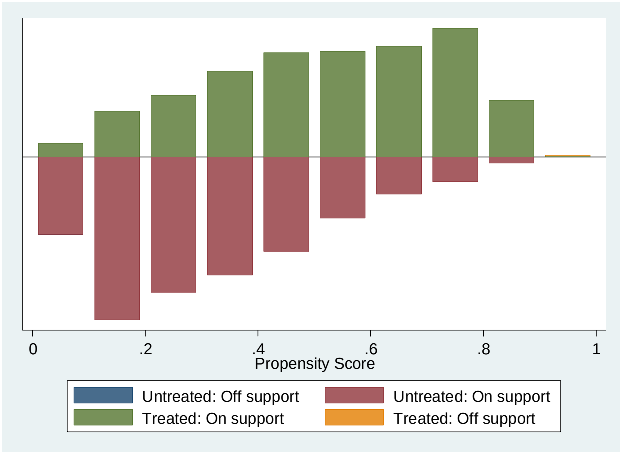


Figure a Figure b

Figure a: Treated (propensity score of large establishment workers)
Figure b: Control (propensity score of small establishment worker)
Source: Author's calculations

In the next step, we will examine the overlap condition of propensity score before and after matching of both groups. In order to show the overlap of the propensities both densities of treatment and control are shown in one graph. Figure-2 shows distribution of propensity scores before and after matching. Visual inspection suggests that the densities of propensity scores are more similar after matching. The plot also reveals a clear overlapping of the distribution. This implies that common support is given. The histogram shows that non-treated scores are bunched in the lower quartile of distribution and for treated scores are bunched in the upper quartile of the distribution but nevertheless, there is a common support of treated and non-treated throughout the distribution as shown by figure 2.

Figure 2 Testing Common support condition



Source: Author's calculations

Once the propensity score model is estimated and the score is computed for each unit, the next step consists of performing the actual matching after choosing a matching algorithm. Matching algorithms differ not only in the way the neighborhood for each treated individual is defined, but also with respect to the weights assigned to these neighbors. The matching quality depends on the closeness of the match or distance measure to determine whether an individual is a good match. The nearest neighbor matching nearly always estimates the ATT, as it matches control individuals to the treated group and discards controls who are not selected as matches.

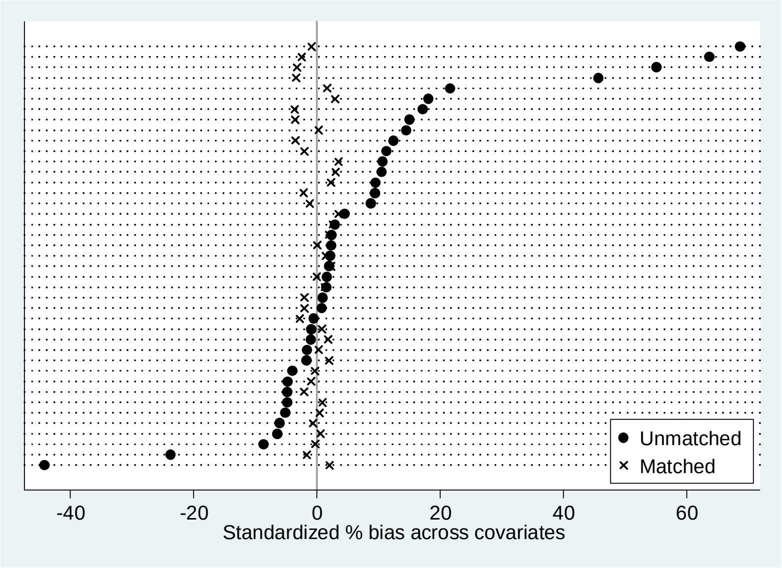
The nearest-neighbor propensity score matching method is used for this study. Nearest neighbor matching only uses individuals close to the area of common support. In contrast, the sub-classification and weighting methods generally use all individuals, regardless of the overlap of the distributions. When using those methods it may be beneficial to explicitly restrict the analysis to those individuals to the region of common support (as in Heckman et al., 1997; Dehejia and Wahba, 1999). The nearest-neighbor matching has two variations, i.e., matching with replacement and matching without replacement. Matching with replacement is used for this study to allow an untreated individual to be used more than once as a match. Allowing the non-treated to be used more than once as comparators improves the performance of the match (Dehijia and Wahba 1998). Matching with replacement involves a trade-off between bias and variance. If we allow replacement, the average quality of matching will increase and the bias will decrease. This in turn avoids the number of distinct non-participants to be used to construct the counterfactual outcome and thereby increases the variance of the estimator (Smith and Todd, 2005). The advantage of nearest neighbor matching is to make the match as good as possible and to minimize the bias across the treatment and comparison groups. Initial results are computed and interpreted using basic matching estimator while in the robustness checks different algorithms are compared. In large samples all PSM estimators should yield the same results because with growing sample size they all become closer to comparing only exact matches (Smith, 2000).

3.3.2 Test of covariate balancing

Next step in assessing the quality of matching is to perform tests that check whether the propensity score adequately balances the characteristics between treatment and comparison group units (Rosenbaum and Rubin 1985, Leuven and Sianesi 2003, Diprete and Gangl 2004). Formally, the objective of these tests is to verify that treatment is independent of unit characteristics after conditioning on observed characteristics (as estimated in the propensity score model): *Appendix-B Table B-5* shows covariate balancing test for gross hourly wage.The quality of match depends on the distance of the marginal distributions of relevant characteristics in both groups (Rosenbaum and Rubin 1985). To test this, a bias before and after matching is calculated for each variable and the change in this bias is stated. The standardized bias (formulae from Rosenbaum and Rubin 1985) should be less than 5% after matching.[44]Overall, the quality of the match is good based on the distance of the standardized bias of the covariates before and after matching. The mean standardized bias pre-match is 13% and post-match is1.8%. The standardized bias for each variable after matching tends to range from –3.6% to +3.6%. Figure-3 shows the distance before and after matching and it is clear that the distance of the marginal distributions of relevant characteristics in both groups is reduced after matching.

[44]*For a given covariate, the standardized difference after matching is defined as the difference of the sample means in the treated and matched non-treated subsamples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (Sianesi, 2001).*

Figure 3 Covariate Balancing



Source: Author's calculations

In examining the results of after-matching balancing tests, we notice that any differences in the covariate means between the two groups in the matched sample have been either eliminated or reduced. We may conclude that through PSM it was possible to generate a control group which is similar enough to the treatment group to be used for the ATT estimation.

3.3.3 Employer Size -Wage Effect through Propensity Score Matching

After assessing the quality of matching based on propensity score graphs and covariate balance, the next step is to analyze the output table of the matching method.[45]Keeping in view the employer's and worker's heterogeneity, it cannot be assumed that CIA is satisfied because PSM only allows us to take into account the observables characteristics whereas there are unobservable factors that can also affect the employer size-wage relationship. Therefore, ATEs are less credible estimates and they are reported in the notes of each table. Results are presented in Table 3.2 for gross and net hourly wage. Of the 74,762 sample, 29,896 are treated (workers in

[45]*The stata program 'pasmatch2'gives output of the mean wage and the difference of mean wage between large establishment and small establishment worker conditioned on propensity score. It provides the treatment effect on the treated (ATT), treatment effect on the untreated (ATU) and average treatment effect (ATE).*

large size establishments) and 44,866 are non treated (workers in small size establishments). 43 observations, 38 treated and 5 untreated are discarded by imposing 'common option'. 15,245 control cases are used to match with the treated. The mean match weight is 1.92 for non-members.59% of the control cases have weight of 1 as they are matched to the single treated case. The largest weight is 35 and in 191 cases a non-member is used as a match to 10 and more members.

Table 3.2 Average Treatment effect on treated using Propensity Score Matching[46]

Employer size wage difference	**Gross hourly wage**		**Net hourly wage**	
	pre-match	**post-match (ATT)**	**pre-match**	**post-match (ATT)**
Mean hourly log wage in large size establishments	4.293	4.293	4.056	4.056
Number of observations	*29896*	*15503*	*29851*	*15469*
Mean hourly log wage in small size establishments	4.117	4.201	3.966	4.030
Number of observations	*44866*	*15244*	*44602*	*15257*
% differential	**17.67**	**9.14**	**8.94**	**2.50**
bootstrap standard errors	0.003	0.005	0.003	0.006

Notes: For Gross Hourly Wage: The propensity score is estimated using a probit of treatment status. Standard errors for the ATT are computed using a bootstrap with 50 replications. ATE is 0.068 and ATU is 0.053.
For the basic hourly 15257 control cases are used to match with the treated. Mean match weight for the non treated is 1.95. Mean propensity score is 0.4021437. ATE is 0.006 and ATU is -0.006.

The table above shows the pre-matching and post-matching mean hourly log wage and their percentage differential in the large and small size establishments. This indicates that 'unmatched' mean hourly wages of employees working in large size establishments are 17 % higher compare to small size establishments. After comparing the matched treated with matched untreated the difference of mean wage is 9 % compare to 7% in the regression estimation in the above section. This shows that OLS under predicts the employer size and wage relationship. The PSM wage differential pre and post matching reflects the treatment effect on the treated. It shows that employees working in large size establishment earn 9% higher wages. The difference of the log hourly wage is reduced after matching which tells that in the private sector the observable differences in personal and workplace characteristics account for nearly half of the wage differential (52.9%) in the gross wage measure. It shows that observable characteristics of

[46] *Matching is implemented using the Stata module psmatch2 (Leuven and Sianesi, 2003).*

workers play an important role in explaining employer size wage difference. As Garen (1985) mentioned that monitoring costs rise with firm size. As a result, large firms acquire less accurate information about the abilities of their workers, and thus will rely less heavily on their own evaluation of workers than do small firms and more on other indicators of ability such as schooling. Similarly, Silva (2004) found that the observed skills; namely, education, age, and tenure have high returns in large firms. This explanation is consistent with our results that large employers choose workers based on observed quality. But after matching there remains a significant size difference which remains unexplained. The size effect may become clearer when we add variables related to profits, compensation practices, market power and unionization etc. But we do not have data on profit and market power, however, results are computed using net log hourly wage. PSM estimation with net hourly wage shows a 8% pre-matching mean difference in log wage between large and smallestablishments which is even less than after matching wage difference with gross hourly wage. The difference between 17 % and 8 % for pre matching in both measures of wage is the compensation/incentives packages and overtime paid hours offered by large employers. After matching with net hourly wage the difference in mean hourly wage between large and small size reduces to 2.5 %. It reveals that using net hourly wage the observable characteristics of employees and employer account for almost all of the employer size wage difference. The gross wage difference before and after matching reduces to smaller amount compared to net wage difference. This confirms that the ESWP is dominant for the employers' compensation practices compared to wages. This is similar to the findings of Melow (1982) who found that firm size impacts are greater on the measure of compensation than on wages.

This shows that a major part of the size premium is paid in terms of fringe benefits depending on employer size. Two things are evident from the analysis above, the high earning capacity of workers working with large size employers and secondly, employer size impact on compensation practices compared to the level of wages. In order to find support for theoretical explanations, it is important to repeat the analysis by sectors, gender and professions to distinguish whether we find some conclusion in support of compensation differentials, unionization or efficiency wage hypothesis. Before doing so, first we test the robustness of the results in the preceding section to be sure that results are accurately computed.

3.3.4 Sensitivity Analysis

When interpreting the results, it is important to evaluate the robustness of the estimations by changing the matching algorithms or by altering the parameters of a given algorithm. Robustness checks help to increase the reliability of the results by showing that the estimations do not depend crucially on the particular methodology chosen. To make sure that these findings are not driven by the selection of a particular strategy, coefficients are estimated using different matching algorithms. The results of first robustness check are shown in Table-3.3.The matching algorithms used are Nearest Neighbor, Kernel Matching, Kernel Normal and Caliper. Only the percentage mean wage differential between large and small size is reported for comparisons. The impact found on wage after matching does not appear to depend critically on the algorithm used since both the value of the coefficients and its significance are very similar using different alternatives.

Table 3.3 ESWE by using different Matching Algorithms

	Gross hourly wage	
Matching algorithms	**pre-match (Mean Wage % differential)**	**post-match (ATT)**
Nearest Neighbour (5)	0,177	0,090
Nearest Neighbour (1)	0,176	0,089
Nearest Neighbour (10)	0,177	0,090
Caliper 0,002	0,176	0,091
Caliper 0,001	0,18	0,09
Nearest Neighbour (1)caliper 0,001	0,177	0,092
Nearest Neighbour (10)caliper 0,001	0,176	0,090
Kernel Matching	0,176	0,092
Kernel Normal Matching	0,176	0,091
Matching without replacement	0,176	0,120

Secondly, in order to test that whether the results are not driven by few observations, we dropped more than 2% of the sum of weight in the control matched group and re-estimated the effects. The new sample is 74,510, in which 29,896 are treated as before and 44,614 are non treated cases[47]. We re-estimated the nearest neighbor matching model; now the wage differential before matching is 17.7% and after matching this reduces to 8.5%. Results are not different from the earlier case. The mean weight match is 1.97. This shows that results donot depend on some (few) observations.

Table 3.4 Robustness check for PSM

Employer Size Wage Difference	**Gross hourly wage**		**Net hourly wage**	
	pre-match	**post-match (ATT)**	**pre-match**	**post-match (ATT)**
Mean hourly log wage in large size establishments	4.293	4.293	4.056	4.055
Number of observations	*29,896*	*15,338*	*29,851*	*15,180*
Mean hourly log wage in small size establishments	4.116	4.207	3.965	4.036
Number of observations	*44,614*	*15,148*	*44,252*	*14,908*
% differential	***17.7***	***8.5***	***9.0***	***1.9***
bootstrap standard errors	0.002	0.006	0.002	0.005

For gross wage: The propensity score is estimated using a probit of treatment status. Standard errors for the ATT are computed using a bootstrap with 50 replications. Mean match weight for the non treated is 1.97. Mean propensity score is 0.402. ATE is 0.066 and ATU is 0.054.

For basic wage: Mean match weight for the non treated is 2.0. Mean propensity score is 0.404. ATE is 0.004 and ATU is -0.004.

Thirdly, weighted regression of wage on the establishment size dummy is performed with a sample of treated and matched control. Weighted regression with all covariates gives the wage difference of 8%.[48]Results are presented in table B-6 in the *Appendix-B.*

Matching on covariates reduces the wage differences but there still remains a significant wage difference. In the next step PSM across different groups is performed to get further information on the effect of employer size on wage. Next section presents comparisons across groups.

[47]*It discarded 252 control cases. (More than _weight=8). 41 observations are discarded by imposing common support option. 15,148 control cases are used to match the treated.*

[48]This option is used in stata. ([fweight = _weight] if _weight! =.

3.4 ESWP across different groups

3.4.1 ESWP across gender

In this section matching method is applied to different groups, separately for male and female, for each sector and across occupations in Tables 3.5 to 3.7 to have a closer look to the employer size wage difference.

Table 3.5 shows results of PSM across gender for both measures of wage. Matching is performed separately for sample of men and women in order compare the size effect within gender, if workers work in different sizes of establishments. Matching reduces 50% of the wage difference for men and 34% for women. This reflects that observable characteristics of workers play a more important role in large size establishments for male workers but overall size-wage premium exist for both groups. The pre-matching size-wage difference of female workers is less compared to male in both measures of wage. For net hourly wage, post-matching wage difference reduces substantially for both groups. Large employers have different behaviors towards male workers. The type and nature of job may be male oriented in large size employers and this further leads to the fact that male workers get more incentive and pay package. They work more over time and benefit from wage premium by working supplementary hours. On the other hand the difference also represents women segregation into low wage work places that may be the result of lower human capital or stereotype educational and professional choices (Machin and Puhani 2003). Women face horizontal segregation where pay and promotional opportunities are less important. There may be barriers for women to join high paying occupations or high paying employers. Women are subject to career breaks and this could result in discrimination in jobs and in wages. Therefore, there is preference for male workers in the large employers where turnover can be costly. The gender wage differential is explained in detail in chapter-5.

Table 3.5 Mean employer size-wage difference across gender

Employer Size Wage Difference	Male Sample				Female Sample			
	Gross wage		Net wage		Gross wage		Net wage	
	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)
Mean log wage of large establishment worker	4,362	4,361	4,108	4,106	4,160	4,160	3,954	3,954
Number of observations	*19758*	*9091*	*19741*	*9146*	*10138*	*6199*	*10110*	*6182*
Mean log wage of small establishment worker	4,181	4,273	4,017	4,082	4,036	4,079	3,903	3,940
Number of observations	*24941*	*8971*	*24775*	*9084*	*19925*	*6076*	*19827*	*6140*
% differential	**18.11**	**8.84**	**9.10**	**2.44**	**12.34**	**8.04**	**5.10**	**1.36**
bootstrap se	0,004	0,006	0,003	0,006	0,004	0,006	0,003	0,006

Male Gross Hourly Wage: The propensity score is estimated using a probit of treatment status. Standard errors for the ATT are computed using a bootstrap with 50 replications. ATE is 0.081 and ATU is 0.075.Mean match weight for the non treated is 2.19. Mean propensity score is 0.4429. For basic hourly wage ATE is 0.019 and ATU is 0.015.Mean match weight for the non treated is 2.16. Mean propensity score is 0.444
Female Gross Hourly Wage: ATE is 0.045 and ATU is 0.027.Mean match weight for the non treated is 1.667. Mean propensity score is 0.3377. For basic hourly wage ATE is -0.013 and ATU is -0.027.

3.4.2 ESWE across industries

Table 3.6 presents a comparison of wage difference by two measures of wage across industries. Only the gross wage comparison is presented here. Results using net hourly wage can be seen in *Appendix-B Table B-7*.The highest size-wage difference is observed in the manufacturing sector. Although matching reduces wage difference of around 50% in trade and services and 55% in manufacturing,still there remains a significant wage difference after matching which is more important in manufacturing and service sector.

Many studies[49] on the industry wage differential found that wages are higher in large manufacturing plants. The reasoning is linked to the efficiency wage and unionization hypothesis. This may be due to the fact that in large manufacturing plants the cost of shirking is high and the profit depends on the production quantities, therefore, the wage premium in terms of bonuses and overtime paid hours is given to the employees. The wage rate may also be related to the relative importance of labor and capital in an industry. As Masters (1969) explained that in large

[49] *Masters (1969), Krugar and Summers (1986), Bayard and Troske (1999).*

establishments there is more division of labor and more formal rules and regimentation to manage various phases of production. Therefore, large plants require more dependable workers as small mistakes can jeopardize the performance of many others. Thus, to obtain workers ready to behave in this way large plants have to pay higher wages. Other things being equal, these higher standards should lead to higher wages inthe large plants.The net wage pre-matching wage difference in manufacturing sector is 12% compared to 21% in the gross wage. Matching reduces all the wage difference in the net wage. This again confirms that compensation structure of large employers determine the size-wage premium along with evaluation based on observable characteristics.

Table *3*.6 Mean employer size-wage difference across industries

Employer size wage difference	**Trade**		**Manufacturing**		**Services**	
	pre-match	**post-match (ATT)**	**pre-match**	**post-match (ATT)**	**pre-match**	**post-match (ATT)**
Mean log wage of working in large size establishments	4,138	4,138	4,313	4,313	4,280	4,280
Number of observations	*1187*	*955*	*17166*	*5969*	*38543*	*8111*
Mean log wage of working in small size establishments	4,071	4,101	4,094	4,214	4,138	4,210
Number of observations	*6240*	*915*	*11405*	*6076*	*27000*	*7823*
% differential	**6.70**	**3.67**	**21.85**	**9.76**	**14.28**	**7.02**
bootstrap se	0,012	0,019	0,004	0,008	0,004	0,006

**Trade= The propensity score is estimated using a probit of treatment status. Standard errors for the ATT are computed using a bootstrap with 50 replications. ATE is 0.0195 and ATU is 0.0162.Mean match weight for the non treated is 1.297. Mean propensity score is 0.1598.*
**Manufacturing= ATE is 0.099 and ATU is 0.102.Mean match weight for the non treated is 2.19. Mean propensity score is 0.4429.*
**Services= ATE is 0.055 and ATU is 0.048.Mean match weight for the non treated is 2.19. Mean propensity score is 0.4429.*

3.4.3 ESWE across occupations

Furthermore, PSM results are computed across professions to compare the pre-matching and post-matching wage difference among occupations. Results by gross hourly wage are shown in Table-3.7 and results for net hourly wage can be seen in *Appendix- B Table B-8.*It is shown in the table below that the post-matching wage difference reduces considerably for higher level professions compared to blue collar jobs. This may be due to the fact that blue collar jobs do not require higher observable characteristics compared to higher level professions (education, experience). The highest wage difference before matching is in blue collar jobs (20%) and even after matching the wage difference remains at 14%. The net wage comparison in T*able B-8 of Appendix-B* shows that the post matching wage difference is negligible among all categories except blue collar workers where post-matching size-wage difference remains at 5%. In examining the size-wage premium by worker type, it turns out that managers and professional workers receive the smallest premium, blue-collar workers receive the largest wage premium and other occupational groups receive a premium in between these two extremes. This is again consistent with previous results (Brown and Medoff, 1989 and Troske 1999). Efficiency wage models may justify the wage premium in this case as large plants are mostly manufacturing plants where monitoring each employee is not possible and the cost of low productivity is high, therefore, blue collar workers may get premium in large employers to put more worker effort and to invoke productivity. The result of higher wages for production workers in the large establishments is in line with other studies. We see different explanations to support this result. FitzRoy (2002) found that firms appear to share rents with manual workers independent of union influence while Rosen (1982) mentioned that the production worker effect possibly could be equalizing on more rigid work routines and the impersonality of the work environment in large establishments.

Table 3.7 Mean employer size-wage difference across professions

Employer Size Wage Difference	Management		High skilled white collar		Low skilled white collar		Blue collar	
	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)
Mean log wage of working in large size establishments	4,859	4,859	4,405	4,404	4,074	4,074	4,114	4,114
Number of observations	*3904*	*1885*	*9221*	*4207*	*5433*	*3603*	*11338*	*5530*
Mean log wage of working in small size establishments	4,763	4,818	4,283	4,369	3,930	4,015	3,912	3,972
Number of observations	*5418*	*18857*	*11688*	*4200*	*12621*	*3447*	*15139*	*5494*
% differential	**9.59**	**2.54**	**12.11**	**4.06**	**14.34**	**5.75**	**20.16**	**14.19**
bootstrap se	0,007	0,013	0,004	0,006	0,004	0,006	0,003	0,005

**Management= The propensity score is estimated using a probit of treatment status. Standard errors for the ATT are computed using a bootstrap with 50 replications. ATE is 0.027 and ATU is 0.029.Mean match weight for the non treated is 2.102. Mean propensity score is 0.4195.*
**High skilled white collar=ATE is 0.050 and ATU is 0.058.Mean match weight for the non treated is 2.19. Mean propensity score is 0.4418.*
**Low skilled white collar= ATE is 0.046 and ATU is 0.041.Mean match weight for the non treated is 1.570. Mean propensity score is 0.3015.*
**Blue collar= ATE is 0.1168 and ATU is 0.098.Mean match weight for the non treated is 2.06. Mean propensity score is 0.4290.*

We see that the ESWP is dominant in the blue collar jobs. It remains even after matching on observed characteristics and for net hourly wage. This demands to analyze role of unions and collective bargaining agreements in French labor market. This is explained in section 3.4.4 below.

3.4.4 Effect of unions and collective bargaining agreements

As mentioned in the introduction, French unionization system is unique compared to other European Union countries. The French industrial collective relation system does not require belonging to a union that signed one agreement to benefit from the provisions contained in it. All employees of the company or branch are covered under the agreement even if they are not unionized or if they belong to another union that did not sign the agreement. This characteristic leads to a paradox resulting in an unionization rate of around 8%. On the other hand 90% of workers are covered by an industrial collective agreement (at national, professional or company-level agreement).

Bargaining positions differ in competitive and non-competitive industries. Segal (1964) has argued that in a monopolistic industry a union can bargain for high wages with less danger of adverse employment effects. In a competitive industry it is easier for new nonunion firms to enter the industry and these firms will gain a competitive advantage if they can pay wages below the union scale. In order to answer how wages are negotiated and what is the effect of different levels of bargaining agreements on wage we need information on the firm level or sector level and industry level collective bargaining wage agreements. Unfortunately, we don't have data on firm level and industry level agreement or agreements by sectors that would help us to draw the effect of agreements on wage neither we have information on the profit of plants to link the collective bargaining agreements with market structure. The presence of a union increases the likelihood of a long-term relation between profits and wages. (Segal 1964). However, it can be said that firm level and industry level agreements vary depending on size as French law binds the firms in certain threshold limits.

It is not straightforward to see the effect of unions on wage in French labor market and to assess the bargaining power of unions. The maximum available information about unions in the data is related to three questions (three dummy variables); (1) Presence of staff representatives, (2) presence of union representatives and (3) and finally, whether there was a negotiation on salary in the survey year (1992). The third can apply to all sample population as any agreement covers all employees in that firm or establishment, whereas, union representative (délégué syndical) must be set in companies or workplaces with 50 employees or more and staff representative (délégué du personnel) must be set in companies or workplaces with 11 or more employees. As size of establishment increases the number of delegates also increases. The information on salary negotiation is not sufficient to give the answer about increase in wage or any outcome of the negotiation. Therefore, all three variables are used in different ways to determine the effect of union elements in the survey year.

Firstly, results are computed by considering covered workplaces where staff representatives are present. The number of observations in this case is 48,166.[50] In 66% of the sample there is a staff representative in the establishment but the number of staff representatives per establishment is not available as the number of staff representatives increase with size. Post matching size-wage difference in covered staff representative places is around 8%. If union wage premium is linked to union strength, then covered members (large establishment workers) are likely to receive a higher premium than covered nonmembers (small establishment workers). But given this information we cannot conclude about union strength as we do not have data on union density that could explain how much is the union density in the covered treated and covered non treated workplaces. Next, covered workplaces with union representation are segregated and PSM is performed. As the minimum requirement for a union representative to be present is 50 workers, therefore, establishments less than 50 workers are dropped. It drops 29,414 observations.78% of the sample has union representative in their workplace.[51]The post matching wage difference is 5% which is less than the whole private sector sample (9%).

It is difficult to answer to the question whether bargaining coverage results in the union wage premium because, at the sectoral level, there is a higher degree of heterogeneity of coverage depending on the employer size. In general, all workers benefit from the wage bargaining.

[50] *It drops 24892 observations after including only the covered workplaces. Further the establishments with size less than 11 are dropped to meet the minimum requirement of having staff representative.Size dummy becomes 11-199=0, 200 and more=1.*

[51]*By dropping establishments with less than 50 employees, we have size dummy as 50-199=0 (21% of the establishments) and 200+=1 (79% of the establishments).*

Table 3.8 PSM estimation in covered workplaces

Employer size wage difference	Covered Staff representation		Covered union representation	
	pre-match	post-match (ATT)	pre-match	post-match (ATT)
Mean log wage of working in large size establishments	4,304	4,303	4,310	4,309
Number of observations	27 737	10 440	27 244	5 223
Mean log wage of working in small size establishments	4,152	4,223	4,168	4,254
Number of observations	20 429	10 502	7 069	5 439
% differential	**15**	**7,9**	**14**	**5,4**
bootstrap se	0,003	0,005	0,004	0,008

In the next step variables on presence of union representative, staff representatives and presence of any salary negotiation are used in the probit model and PSM analysis is repeated. In order to use all information in one model we have to put restriction on the threshold limits of size.[52] Results are computed by introducing many variations with these variables. A principal component factor (PCF) is constructed to weight all three variables and this gives one factor with combined effect of three variables.[53] Results by using three variables and weighted factor are presented in table 3.9.

[52] *For using staff representative establishments more than 11 workers are chosen and to use union representatives, establishments more than 50 workers are included in the regression.*

[53] *Three PCFs are constructed, firstly by using all three variables, secondly by using union representative and negotiation and thirdly by using staff representative and negotiation. Although the factor does perform well with more number of observations but still the 'egen value' falls within the critical region. Wherever union representative variable is used then establishments less than 50 are excluded and wherever staff representative variable is used then establishments less than 11 are excluded from the sample .*

Table 3.9 ESWP using union presence variables

Employer size wage difference	**A+B+C**				**B+C**				**A+C**			
	Variables		Weighted factor		Variables		Weighted factor		Variables		Weighted factor	
	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)
Mean log wage in large size	4,295	4,294	4,295	4,294	4,295	4,295	4,295	4,295	4,295	4,294	4,295	4,294
Number of observations	27665	6787	27665	7022	27688	10728	27688	10910	27814	6294	27814	7237
Mean log wage in small size	4,125	4,242	4,124	4,236	4,122	4,223	4,122	4,230	4,124	4,222	4,124	4,222
Number of observations	13763	6932	13763	7164	37939	10752	37939	10864	13771	6976	13771	7352
% Differential	**17.08**	**5.24**	**17.08**	**5.74**	**17.34**	**7.23**	**17.32**	**6,55**	**17.01**	**7,1**	**17.01**	**7.17**
Bootstrap se	0,003	0,008	0,003	0,008	0,003	0,006	0,003	0,006	0,003	0,007	0,003	0,007

Note: A: presence of union delegates, B=presence of staff delegate, C= presence of any negotiation on salary.

There is no difference of using the weighted factor or using variables separately in the model. If we assume that this information represents the union system or collective bargaining system in France, then the tables above reveal that ESWP exists independent of union influence even after matching. OLS estimates for the union effect on wage with the same variations of variables are presented in Table B-9 of Appendix-B. In the covered sector by union representative workplaces, employer size magnitude is 5.8% and in the covered staff representative workplaces employer size magnitude is 7%.

Hence, it may be concluded from the above analysis that either the so called ‘unionization’ effect does not exist or the available information is not sufficient to explain the role of collective bargaining system in France.

The post-matching wage difference remains unexplained in the gross wage analysis. The size-premium is dominant in the compensation practices of employers of different sizes compared to wage. This may be result of employer heterogeneity. As suggested by efficiency wage models, if firms of different sizes differ in the ease of monitoring workers, in training costs, or in their reliance on teamwork, they may find it profitable to pay differing wages to identical workers Morissette (1993).

3.5 Dynamics of Size-Wage Differential (1992-2006)

In this section, the dynamics of the wage structure from 1992 to 2006 is presented.[54] The objective is to compare the size-wage differentials over the time using four cross sections from 1992 to 2006. Fortunately, the latest dataset is available in this type of survey but the variables are not the same: for example, the marital status, number of dependent children, departments of France or detailed regions where establishment is based are not available. **Tables A-8 and A-9** shows OLS wage equation for both measures of wages in three years (2002, 2005 and 2006). Here, only a comparison of propensity score matching is presented below in ***Table-3.9 and 3.10***. The tables show only the difference in the mean wage of the two groups and bootstrap standard errors in parenthesis. The relevant figures are the bold figures in the tables.

Table-*3*.9 PSM Gross Hourly Wage (1992-2006)

year	Unmatched	ATT	ATU	ATE
1992.	**0.177**	**0.090**	0.053	0.068
	(0.003)	(0.004)		
2002	**0.147**	**0.066**	0.045	0.053
	(0.006)	(0.011)		
2005	**0.099**	**0.036**	0.049	0.043
	(0.004)	(0.006)		
2006	**0.105**	**0.038**	0.044	0.041
	(0.004)	(0.006)		

Over the years, we see that the wage difference depending on size is decreasing. The mean wage difference between large and small establishments was 10 % in 2006 compared to 17 % in 1992. The reason could be looked for in occupational segregation, workplace flexibility and salary negotiations. The mean difference conditional on propensity score and after balancing covariates is 3 % in 2006 against 9 % in 1992.

[54] *The construction of variable education is modified in the above sections based on cereq classification but these results are based on the earlier version of thesis.*

Table- 3.10 PSM with Basic Hourly Wage (1992-2006)

Year	Unmatched	ATT	ATU	ATE
1992.	**0.089**	**0.027**	-0.005	0.008
	(0.003)	(0.004)		
2002	**0.115**	**0.045**	0.022	0.031
	(0.007)	(0.009)		
2005	**0.130**	**0.073**	0.075	0.074
	(0.004)	(0.005)		
2006	**0.127**	**0.053**	0.074	0.064
	(0.004)	(0.005)		

Mean wage differences by basic hourly wage show some interesting facts. In 2005 and 2006 mean wage differential with respect to basic hourly wage between large and small establishments is greater than the wage differential by gross hourly wage. This reveals that the difference of net wage is greater than the difference of gross wage between large and small size establishments. Earlier we observed that the compensation, pay packages and over time paid hours could explain the wage difference. But for 2005 and 2006, the role of basic contractual wage premium is more important than the role of gross wage premium. The change is due to the revision of minimum wage in 2005[55]. After matching there remains a wage premium of 6 % for the measures of basic hourly wage in 2006. In recent years, wage premium for basic is more than earlier years due to revisions of salary laws. While for 1992 and 2002, there remains a very low premium explaining that the compensation and pay packages by large employers explain the wage difference between large and small establishments Therefore, it may be concluded that wage differences between large and small establishments are decreasing over the years for gross hourly wage and increasing for the basic hourly wage.

[55]*The minimum wage in France was first called SMIG, the "guaranteed inter professional national minimum wage" (Salaire Minimum National Interprofessionnel Garanti) and was created in 1950 as a statutory law. It was replaced in 1970 by the SMIC (Salaire Minimum Interprofessionnel de Croissance). The rate of SMIC is set annually by the government. In 1998, the reduction of the statutory working week to 35 hours from 39 hours has established a guaranteed monthly wage (garantie mensuelle de rémunération) so as to maintain an unchanged amount of the monthly earnings of those already working at the minimum wage prior to the 35 hour week law, all new comers being paid at the new hourly rate. In 2002, up to five different rates were defined as more and more companies adopted the 35-hour week.*

3.6 Conclusion

The objective of this chapter was to estimate the magnitude and sources of the wage difference of employees working in large size compare to similar employees working in small size establishments. OLS and PSM are used to identify wage differential attributable to workers' and employers' characteristics.

The main findings are listed below:

- OLS shows a 7% size effect on gross hourly wage controlling for individual and employer characteristics. On the other hand PSM shows pre-matching mean gross wage difference of 17% compared to post-matching difference of 9% between large and small establishments. OLS under predicts the relationship between employer size and wage. Thus, the regression analysis with the assumption of linearity and beyond common support under predicts the wage premium paid by large establishments.
- Matching conditional on propensity score of the observed covariates reduces the difference in mean wage between people working in large size and those working in small size of establishments. This reflects that large employers value more observable characteristics. 50 % of the mean wage difference is explained by differences in observable characteristics of large establishment workers, whereas, net wage difference is entirely the difference of observed characteristics between large and small establishments.
- From both measures of wages, two evident aspects emerge : first large employers prefer workers with high earning capacity. Secondly, large size establishments determine the wage premium in the form of compensations and pay packages (bonuses, allowances, over time payments etc).We may say that the employer size wage premium is more importantly the result of the employer's compensation policy than the wage difference itself.
- The employer size-wage effect based on observable characteristics is more evident for male workers, in blue collar jobs and in large manufacturing plants.
- The employer size difference is more present in lower level professions compared to higher level professions.

- The available information related to unions and collective bargaining agreements is not sufficient to determine role of bargaining agreements in the employer size-wage gap.
- The employer size wage difference after matching remains positive and significant which is not explained by the model. The difference of gross and net wage may explain the remaining difference but which measure of compensation policy is more related to employer size is not clear.

PSM has largely been used in many types of program evaluation research. But nevertheless, this technique operates with some restrictions and limitations. As Heckman et al .1997 highlighted PSM can only recover mean effects and cannot answer questions relating to the distributional effects of the program, such as the percentage of program participants who benefit from it. Further, PSM cannot estimate the local average treatment effect (LATE), which is the mean impact of the program on those whose participation status changes due to a change in policy. Instrumental variable methods can remove selection bias due to observed and unobserved confounding but in many circumstances plausible instruments are not available. Regression and propensity score (PS) methods can address selection bias under ‘unconfoundness’ but they also assume that the functional form of the regression model or the propensity score is correctly specified.

Further work can be done to deal with the unobserved characteristics of workers that result in the non-random selection of workers across employers of different size. OLS and PSM do not solve the problem of endogeneity and causal relationship of employer size and wage. There is a need to estimate LATE through selection models that deal with non-random sorting across size groups.

Chapter-4

NONRANDOM SELECTION AND EMPLOYER SIZE-WAGE PREMIUM IN FRENCH ESTABLISHMENTS

Abstract

This chapter takes into account nonrandom selection of heterogeneous workers into employers of different sizes. Jointly estimated maximum likelihood function and switching regression models are used for this study to see the magnitude and sources of the size-wage gap in the French labor market. Negative selection into large size establishment and positive selection into small size establishment is found suggesting that the conditional wage is lower than the unconditional wage for the large size establishment's worker and vice versa. It is concluded that the unobserved factors have high rewards in small size establishments while observed factors have high rewards in large size establishments. Wage premium and effect of selection exists for gross hourly wages and for male workers.

4.1 Introduction

The problem of selection bias has got little importance in the literature of employer size-wage gap. Though it is considered as important, the difficulty in finding the relevant instruments has been a problem. If there is heterogeneity in the underlying populations then workers and firms are systematically sorted and selected among classes of jobs according to their choices. The complexity of factors affecting decisions differs among individual and firms.

In the previous chapter, it was observed that matching through propensity score reduces the wage differential between workers of large and small size establishments. But still there remains a significant wage premium paid by large employers after controlling for a wide range of employer and employees' characteristics. OLS and Matching analysis provided interesting findings but both methods do not deal with unobserved characteristics that result into nonrandom selection. This chapter discusses worker's unobserved heterogeneity that results intoself-selection of workers into establishments of different sizes. Two selection models (jointly estimated maximum likelihood function and Heckman two-step estimation procedure) are applied to take into account workers' "unobserved heterogeneity and employers' evaluation of measured and unmeasured skills of workers.

Beginning in the late 1970s and early 1980s methods for detecting and statistically correcting selection bias were developed in economics and related areas (Greene 1981, Heckman 1976, 1978, 1979). In the decades since, an extensive literature has evolved in the area of sample selection bias (Berk, 1983, Lee & Marsh 2000, Miller & Wright 1995, Stolzenberg and Relles 1997, Vella and Verbeek, 1999, Winship and Mare 1992 etc.). The methods to deal with sample selection bias are known as “sample selection” models.

The question of selection in size-wage differential analysis is not new in the empirical literature. Idson and Feaster (1990), Main and Reilly (1993), Lluis (2008) have worked on the same question for the labor markets in USA, UK and Canada respectively. Among others, there are studies highlighting firms’ sorting based on employees’ skills.[56] Garen (1985) also explained that employers of different sizes evaluate observable and unobservable skills differently.

The key econometric difficulty in explaining the causal relationship between employer size and wage comes from nonrandom selection of workers into establishments of different sizes. The objective of this chapter is to consider the selection bias in the presence of nonrandom selection in order to identify the wage premium paid by large employers. The possible source of selection bias in the estimates of size-wage effect is the potential endogeneity of choice of establishment as a result of the selection process. If, for any reason, size is correlated with unobservable factors, size is said to be an endogenous explanatory variable in the wage equation. Consistency is the minimum requirement for an estimator. OLS is consistent in the simple regression case only in the presence of Conditional Independence Assumption (CIA) which means in the absence of endogenous variables. Otherwise the model is incomplete and from the statistic point of view it is not identified. The propensity score matching does not solve the problem either in the presence of unobserved heterogeneity. Thus; it is inevitable to capture the endogenous selection to determine the sources of wage premium paid by large establishments.

Different methods are employed to consider the selection mechanism and to determine the wage differential based on the size of the employer. Firstly, a joint maximum likelihood function is estimated. Secondly, a switching regression model is estimated to see the returns of measured

[56] *Gibson (2004), Garen (1985) Farber (1999) and Lluis and Ferrer (2004).*

and unmeasured characteristics across the size categories. This is not a usual switching regression model; rather it is special kind of switching model where we estimate individual wage in three different regimes. Examples of this model can be seen in Maddala (1983) and Maddala and Nelson (1975). The methodology developed by Idson and Feaster (1990) and Main and Reilly (1992) is combined and replicated. Later, Selection bias correction based on the multinomial logit model, a two-step method introduced by Bourguignon, Fourier and Gurgand (2007), is performed for robustness checks.

This chapter is organized as follows: Section 4.2 and 4.3 present results for Heckman two-step model and FIML respectively. It is followed by the similar estimations using basic hourly wage as a dependent variable in the wage equation. The last section 4.5 concludes. First, the introduction of selection models and choice of instrumental variables for exclusion restrictions is described.

4.1.1 Selection models of employer size-wage premium

Main focus of economic studies is to explain the causal effect rather than simple association between variables. The notion of 'potential outcome' became famous after Rubin's work (Rubin 1973, 1974) in the observational studies without randomization. The causal effect of a treatment on a single individual is defined as the comparison (difference) between the value of the outcome if the unit is treated and the value of the outcome if the unit is not treated (Angrist et al. 1996). This definition of causality is applicable if we are able to observe outcome in the circumstances other than those to which individuals are actually exposed (potential outcome). We need additional assumptions in evaluating a treatment under imperfect compliance beyond randomization of treatment assignment.

Instrumental variable (IV) methods are typically used to address the omitted variable bias, measurement error and reverse causality or simultaneity problems in OLS regression. But in general it is very difficult to satisfy the exclusion restriction requirements before selecting an IV. Moreover, it is very difficult to meet the exogeneity requirement even if instruments fulfill testing requirements. In the cases where the effects are not homogeneous we have to go beyond two-stage least squares (2SLS) estimation because with heterogeneous treatment effects endogeneity creates severe problems for identification of population averages. One of the key assumptions in the heterogeneous effects is "monotonicity" which states that the instruments may have no effect on some people but all those who are affected are affected in the same way (Angrist and Pischke 2008). If an instrumental variable satisfies four assumptions (Independence, exclusion restriction, monotonicity and relevance) then the average treatment effect is called local average treatment effect (LATE) (Angrist et al., 1996).[57] LATE implies in general that one cannot consistently estimate average effects for sub-populations other than compliers (for whom we can identify effects).

Heckman‘s two-step selection model is very famous in this type of research because not only it offers a theoretical framework for modeling sample selection but is also based on what was at the time a pioneering approach for correcting selection bias. LATE is relevant to the selection models as all the effects come through the instruments. Unlike 2SLS type, a selection model involves modeling both the dependence of the outcome and the treatment on the covariates.

The treatment for selection bias is done using Greene's maximum likelihood estimator (Greene, 1995) and Heckman's classic model (Heckman, 1979). It is found that selection models are better to predict the causal relationship only if the model is correctly specified. It is not straight forward to choose one model over the other. Both Heckman and full information maximum likelihood function can be used side by side as both methods predict better the causal relationship compared to OLS and PSM in the presence of nonrandom assignment of workers.

4.1.2 Selection of instrumental variables for treatment choice

As the main objective of this chapter is to control for unobserved heterogeneity, therefore, valid instruments are required to satisfy the exclusion restriction requirement. In order to choose instruments we use ECMOSS 1992 which is the most relevant dataset available to us, as this data contains maximum information related to employer and employees characteristics.

[57] *LATE is the average effect of x on y for those whose treatment status has been changed by the instrument z.*

The first instrument used in the selection model (first step) is the ***interaction of regional size with type of industry***. The intuition behind this interaction is that as large firms and establishments are mostly found in large regions and people living in large regions would more likely be working in large firms, their choice will vary on the type of industry they want to work. This suggests that information on region size may help to explain the choice of establishment size beyond industry choice or controlling for type of industry. With this, we assume that the regional wage premium should be the same for all industries in one region and for the average region, there is no additional region impact by industry. There is an industry wage premium and similarly there is a regional wage premium but in general industry wage premium does not depend on the regional size where it is based, in other words, there is no direct region-industry interaction effect. The size of the region may depend on many factors. The regional distribution of population, of building and of commercial services depends on the regional distribution of employment in industry.In France, based on European Commission report, four regions(Ile de France, Rhône-Alpes, Provence Alpes Côte-d'Azur and Nord Pas-de-Calais) are together responsible for about 50% of the French GDP. A second group of regions, composed of Western regions (Pays de la Loire, Aquitaine, Bretagne) are responsible for slightly less than 15%. Overall, there has been a reduction in disparities to the benefit of "peripheral" regions. Western and Southern regions (Pays de la Loire, Bretagne, Aquitaine, Midi-Pyrénées, Provence Alpes Côte-d'Azur, Corse) have had rates of growth of GDP and population higher than average, some of them are experiencing significant inward migration from Ile de France and Northern France[58]. This reflects that regions are heterogeneous and it is important to control each region separately and rank them so that we capture all the regional heterogeneities.

In order to define the regional size, we have to choose among regional characteristics that take into account regional heterogeneity. The geographical areas of reference frequently used in France, are the"*airesurbaines*". They have been introduced in the middle of the 90's and applied at the time of the last Census of population (1999). As defined by the INSEE, they are units formed with "*communes*" [59] , which correspond to local labor markets and economic attractiveness. They are composed of one urban pole (at least 5000 jobs) and adjacent "*communes*", called "*péri-urbain*" ring. At least, 40% of the labor force population works in the

[58]*See for detail: European Commission report on 'WP1– Coordination of evaluation of SF 2000-2006: Task 4'*
[59]*"Communes" are the first local administrative and jurisdictional level of the French urban system*

"pole" or in the "ring". The French zoning method barely relies upon population density, but also upon job concentration and home-to-work commuting flows. This has the advantage of focusing upon the connections between urban and rural areas. Based on this definition of urbanization, we can find the rate of urbanization and the figure of urban population by department. In every region there are further departments and communes. Therefore, in one region there can be more than one urban pole or ring. In urban areas there are more employment opportunities and more facilities, people are likely to go in big cities and big urban poles. We do not have data to control for the mobility or migration between cities or between regions but we can expect that people can move to the center pole for employment from peripheral cities or rural areas.

Regarding the size of the regions, they are sufficiently big in terms of population to allow reliable conclusions and sufficiently distinct to allow differences among them. Every region has its own characteristics. Moreover in France, the collective bargaining agreements apply to all sectors uniformly and apply to all firms: this eliminates the direct interaction effect. Similarly, working conditions and living conditions are different across regions and in one region people will face the same living conditions and same working conditions that should not differ by industry if we assume that regional wage premium is the compensating premium for living conditions in large regions. Thus, we would expect the same premium for people living in Paris and working in different types of industries whereas it can be different from a person living in Brittany region.

Interactions of industry and region size are created on the basis of various regional characteristics. We have data on all departments[60] and we regrouped departments into regions. We used regional urban population, population density, surface area, ranking of regions by GDP and by rate of urbanization.[61]The estimation results in the next section show regional size ranked with respect to urban population. Table C-1.1 provides ranking of regions with respect to various categories. Table C-1.2 and C-1.3 shows that distribution of type of sector and type of establishment in each region respectively.

[60] *Department (Département) is one of the three levels of government below the national level between the region and the commune. There are 96 departments in metropolitan France and 5 overseas departments, which also are classified as regions. A department belongs to one and only one region.*

[61] *In general all of the characteristics rank regions in the same order except for surface. All the information was obtained by INSEE based on population census of 1999*

Secondly, in the 1992 dataset, ***number of dependent children*** is available as a continuous variable starting from 1 dependent child. This variable is used in consideration of the fact that individual with more dependent kids will more likely work in establishments with stable jobs, higher fringe benefits and promotional opportunities. This can result into self selection. In the literature, Main & Reilly (1993) used number of dependent children; Lluis(2008) used dummy for big city by population and interacted it with industry in the first stage of Heckman procedure. We used both types of instruments separately and together to predict the selection models. Table C-1.4 in the Appendix-C-1 shows the correlation of instrumental variables with establishment size. In order to satisfy exclusion restrictions the instruments are supposed to be strongly related with the endogenous variable and should not have any direct correlation with wage. The instrument "Number of dependent children" meets both conditions and the interaction terms also meet the conditions in most of the cases.

4.2 Heckman two-step estimation procedure

The challenge to the application of the IV approach is to find instruments that are omitted but meet the two conditions, of strength and validity. Due to this reason, a treatment effect model is often used that directly estimates the selection process. One of the earliest sample selection methods is known as the "Heckman two-step estimator" (Heckman, 1976, 1978, 1979). This is considered as the classical selection model. Heckman (1997) examined the use of the IV approach to estimate the mean effect of treatment on the treated, the mean effect of treatment on randomly selected persons, and the local average treatment effect. He paid special attention to the economic questions that were addressed by these parameters and concluded that when responses to treatment vary, the standard argument justifying the use of instrumental variables fails unless person-specific responses to treatment do not influence the decision to participate in the program being evaluated.

In this section, the question of selection is more precisely examined by estimating separate wage equations by employer size: these are presented to study the behavior of establishments through endogenous switching regression model, where the switch between different regimes is endogenous. The switching regression model (SRM) is a variant of the classical Heckman selection model. This SRM is different from the usual switching model. Instead of observing a truncated distribution for wage, we observe wage in large, medium and small sized establishments from three different distributions.[62] We will find the expected earnings of a large establishment worker who self-selected into large size and the expected earnings of a small establishment worker who self selected into small establishment.

It is hypothesized that there are differences in intercept and slope coefficients between the employers of different sizes. Since the employer size (choosing to go to an establishment of a particular size) is a decision variable by individuals, it is treated as endogenous. The econometric methodologies used by Idson and Feaster (1990) and Main and Reilly (1993) are combined and extended. The estimation procedure is explained in detail in Appendix C.2. Stata's program is used for implementation of this methodology[63].

Let us assume that there are *j* size categories (*j*= 1, 2, 3) and *i* number of workers (*i*=1, 2, 3…..n). Let K_{ij} be the maximum attainable utility for worker *i* in category *j*. We assume that K_{ij}includes wage and non-wage factors related to job. This utility function is composed of stochastic and non-stochastic components and may look like the following.

$$\boldsymbol{K_{ij} = \delta'_{ij}X_i + \varepsilon_{ij}} \qquad \textbf{\textit{(1)}}$$

Where X_i is a vector of observable individual characteristics, δ is a parameter vector of individual i in j size category. For simplicity three size groups are formed and named as small, medium and large. For example: j=1 for small (1/49 employees = 1), j=2 for medium (50/199 employees =2) and j=3 for large (200/max employees =3).We assume that individual chooses among large, medium and small size categories. We denote L, M and S for large, medium and small size category respectively. The probability that individual chooses large employer is given by

[62] *Endogenous switching regression model is used to address the issue of self-selection and the estimation of treatment effects when there is non- random allocation of subjects to treatment and control groups as is generally a case in observational studies. The endogenous switching regression model provides a way of calculating the expected level of outcome if persons are assigned to employer size other than the ones they infect entered.*

[63] *In stata 'heckman' command implements the two step procedure but for dummy endogenous variable. Since we have three size groups, therefore, we wrote the program in stats for equation 1(zgamma) after ordered probit model and subsequently three λ equations (10, 12, 14) are written and then OLS is performed with selection terms.*

$$pr_{iL} = pr(K_{iL} > K_{iM} > K_{iS}) \quad (2)$$

Further by substituting from 1 we get

$$pr_{iL} = pr\left(\delta'_{iL}X_i + \varepsilon_{iL} > \delta'_{iM}X_i + \varepsilon_{iM} > \delta'_{iS}X_i + \varepsilon_{iS}\right) \quad (3)$$

Utility of going to large size employer is higher than going to any other size category. Individual choose to go to a size category where they get maximum utility. But workers base their choice for selecting employers on the observable (X) and unobservable attributes (ε) which result in nonrandom selection. In order to get unbiased estimates we need to take into account the selection process by predicting the size of the establishment to which individual is attached. In order to incorporate the effects of selection bias, two equations, the selection equation (the ordered-probit model) and the outcome equation (OLS) are required. As employed people are being studied, therefore, the selection equation deals with the determinants of being into a particular size category. Dependent variable is the probability of belonging to large size compared to medium or small size establishments.

Our ***latent variable model*** (the selection model) is as follows:

$$Y_i^* = Z_i\gamma + \varepsilon_{1i} \quad (4)$$

If:

$Y_i^* < 0$ The individual works in small sized establishment

$0 \leq Y_i^* < \mu$ The individual works in medium sized establishment

$Y_i^* \geq \mu$ The individual works in large sized establishment

And the wage equation for the three size categories is given by:

$$W_{1i} = X_i\beta_1 + v_{1i} \quad (5)$$

$$W_{2i} = X_i\beta_2 + v_{2i} \quad (6)$$

$$\boldsymbol{W_{3i} = X_i\beta_3 + v_{3i}} \qquad (7)$$

Where $Y_{1i}^{*}$is a latent variable associated with "being employed in a particular size category ", Z contains the set of determining variables of being in a size category, γ *is* the associated parameter vector. W_{1i} is the log hourly wage for small size category, X is a matrix of wage determining variables, β is a vector of unknown parameters and $\boldsymbol{\varepsilon_{1i}}$ and $\boldsymbol{v_{1i}}$are the *i.i.d* error terms that follow a bivariate normal distribution (0,0, $\sigma_{\varepsilon 1}, \sigma_{v1}, \rho_1$). The variance of $\boldsymbol{\varepsilon}$ i is normalized to be one $\boldsymbol{\varepsilon i}$ ~ N (0, 1).

$$\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} \sim N \ \{E = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \sum = \begin{pmatrix} 1 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_{v^2} \end{pmatrix}\}$$

$\sigma_{\varepsilon v} = \rho$ where ρis interpreted as the correlation coefficient between errors in the selection equation (ε_i) and errors in the wage equation (v_i) ; and the σ_{v^2}terms are the error variances of the wage equations. The final wage equation with selection term by employer size is the following.

$$\boldsymbol{W_{ij} = X_{ij}\beta_{ij} + \theta_{ij}\lambda_{ij} + v_{ij}} \qquad (8)$$

4.2.1 Results using Heckman estimation method

Standard Mincer type wage equations are estimated in the second stage of the two step procedure keeping each category of establishment size as the reference group. Three different equations by size are estimated for gross and net hourly wage separately. Table C-2.1 in Appendix C-2 presents detailed results for three models with different instruments. Here, in Table 4.1 main results are presented with selection terms. We see that gender wage difference increases with size, a male worker in small size gets a wage higher by 14% to the female wage, in medium size establishments he gets more by 14.3% and in large he gets more by 15.7% .One more year of experience increases wage by 2.5% in small, 2% in medium and 3% in large establishments, respectively. Similarly, rewards to education are higher in large size establishments. Compared to trade sector, wages are higher in manufacturing and services sectors and the difference is bigger in large size establishments. Thus, we can say that overall, rewards for observable individual and employer characteristics are greater in large size compare to medium and small sized establishments.

The selection term is the composition of $\sigma_v, \rho_{v,\varepsilon}$ and λ_i, which represent error variance of the wage equation, correlation of the error terms (of the selection equation and wage equation) and the ratio of the probability density function to the cumulative distribution function (truncated means). Therefore, the effect of selection on wage will depend on the signs of the estimated coefficients and truncated means (Idsen and Feaster 1990).

The correlation between $\boldsymbol{v_{ij}}$ and $\boldsymbol{\varepsilon_{ij}}$ would mean that selection effect is present. If it is positive then high ability workers are more likely to work in large establishments and earn high wages compared to a random draw from the population with a comparable set of characteristics. On the other hand, if it is negative then less able people are likely to work in large establishments and get higher wages. The results show uniform negative IMRs across size group but they are only significant in the case of large employers (model-1). This means that the nonrandom selection is only present in the large size establishments.Those who self selected themselves into large establishments earn lower than population average of earnings in large establishments but higher than the earnings of small establishment worker.[64] The negative selection implies that large employers reward negatively the unmeasured skills or they do not positively assign reward to unmeasured skills. Because they do not do effort to observe unmeasured ability of workers, therefore, workers of average ability or low ability may find it preferable to work in large firms compared to small firms. Large establishments have more formal rules, regulations and uniformity due to which they require workers who are more reliable and executable. They require workers to maintain the processes and perform repetitive tasks. Thus, large employers acquire less accurate information about the abilities of their workers and depend more on observable characteristics like education and experience. This implies that those individuals who have more observed characteristics that increase their utility, they find it more profitable to work with large employers as they know that their observed characteristics will be highly valued in large establishments. As Masters (1969) pointed out that the average skills requirement in large

[64]*The size of the bias depends on the magnitude of the correlation, the error variance of the wage equation, and the severity of the truncation (the IMR is larger when the cutoff is smaller).*

plants are lower than at a small plant. Larger employers hire workers with more homogeneous skill levels because larger employers rely on standardized production technologies that call for homogeneous labor (Davis et al, 1996 and Oi 1999).

Besides, there are returns to some worker-specific abilities. To illustrate, some people can express themselves better in a particular type of environment. There may be preference for particular management or they may be preferred for particular management tasks. Especially, more qualified and high ability workers are risk averse. They are more productive when they get an environment that matches their ability. They are not productive in an environment that does not correspond to their specific ability and risk-averse nature. Lluis and Idson and Feaster named it high independence drive feature. Or we may also say that low ability workers have comparative advantage in large because they are not directly monitored. Large firms do not require high unobserved skills as there is more division of labor and work organization requires repetitive tasks especially in the production and public services.

Both Instruments generate negative selection terms only for large size establishments. The statistical insignificance of the selectivity effects may partly due the inclusion of most of the variables from the first-step equations (Kingdon, 1996). It is evident by this analysis that unobserved factors have high reward in small size establishments while observed factors have high rewards in large size establishments. This reduces the overall wage difference between large and small employers. This result is in line with the previous studies about the same question.

The detail results are presented in Table C-2.1 to C-2.3 for all population and for male and female sample separately. Heckman results for male sample reveals the same pattern of results as in the total population case. For male sample we see the evidence of nonrandom selection in the large size establishments. For female the first and third model does not show any evidence of nonrandom selection. Second model confirm the nonrandom selection in the small and large size establishments with negative selection into large and positive selection into small.[65] Overall, results confirm nonrandom assignment for male workers as selection terms are not significant in the female sample.

[65] *Results for basic hourly wage are also computed but not shown in the appendices.*

Table 4.1 Size-Wage premium- Estimation results with Heckman estimation procedure (Second-step)

Variables	**Model-1** : ***Industry type *Region Size***			**Model 2:** ***Number of Dependent Children***		
	(Small)	**(Medium)**	**(Large)**	**(Small)**	**(Medium)**	**(Large)**
Gender (base female)	0.131***	0.134***	0.146***	0.155***	0.151***	0.132***
	(0.003)	(0.004)	(0.003)	(0.005)	(0.010)	(0.005)
Experience	0.025***	0.021***	0.032***	0.015***	0.017***	0.030***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)
Before Bac without degree (base no degree)	0.047***	0.055***	0.075***	0.065***	0.046**	0.007
	(0.005)	(0.007)	(0.005)	(0.009)	(0.020)	(0.008)
CAP/BEP	0.069***	0.083***	0.112***	0.071***	0.070***	0.083***
	(0.004)	(0.006)	(0.004)	(0.006)	(0.010)	(0.006)
Bac professional and technical	0.147***	0.149***	0.187***	0.151***	0.132***	0.144***
	(0.007)	(0.010)	(0.006)	(0.009)	(0.017)	(0.009)
Bac general	0.153***	0.172***	0.202***	0.157***	0.166***	0.180***
	(0.007)	(0.010)	(0.008)	(0.009)	(0.015)	(0.010)
Bac +2	0.195***	0.209***	0.254***	0.215***	0.193***	0.149***
	(0.006)	(0.009)	(0.007)	(0.012)	(0.028)	(0.011)
Bac+3 and plus	0.303***	0.344***	0.375***	0.357***	0.332***	0.228***
	(0.010)	(0.015)	(0.010)	(0.020)	(0.046)	(0.017)
CDI Contract (base CDD)	0.007	0.003	0.139***	-0.010	0.009	0.210***
	(0.006)	(0.008)	(0.008)	(0.011)	(0.022)	(0.014)
Manufacturing (base trade)	0.034***	0.034**	0.044***	0.129***	-0.001	-0.349***
	(0.010)	(0.016)	(0.011)	(0.032)	(0.085)	(0.033)
Services	0.024***	0.031***	0.054***	0.066***	0.022	-0.142***
	(0.006)	(0.010)	(0.008)	(0.014)	(0.039)	(0.018)
PersonalCharacteristics	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes
λ	-0.009	-0.015	**-0.027***	0.122***	-0.058	**-0.561*****
	(0.013)	(0.013)	(0.015)	(0.047)	(0.082)	(0.044)
σ	0,24	0,22	0,22	0,24	0,22	0,21
ρ	***-0,04***	***-0,07***	***-0,12***	***0,52***	***-0,26***	***-2,62***
Constant	3.287***	3.346***	3.238***	3.379***	3.395***	4.234***
	(0.019)	(0.028)	(0.029)	(0.032)	(0.093)	(0.082)
Observations	30,286	14,514	29,896	17,480	8,594	18,828
R-squared	0.64	0.66	0.65	0.66	0.67	0.66
Adj. R-squared	0.64	0.65	0.65	0.66	0.66	0.66

*Notes: The sample includes full time jobs, private sector, non agriculture and non chief executive officers. The dependent variable is the logarithm of the GROSS hourly wage rate. Robust standard errors of size coefficients are between parentheses. *** p<0.01, ** p<0.05, * p<0.1.Personal characteristics include tenure, family situation and profession. Region includes 21 dummies including the base categories. Experience squared and cubic term also computed with experience variable.*

4.2.2 Sensitivity Analysis of Heckman Model

Results are computed for different groups in order to test and compare selection effects:

1. Firstly, similar (two-step) equations by employer size are estimated keeping each category of profession as a reference to compare different groups. Table C-2.4 shows three models and in each model wage equation is estimated keeping each category of profession as reference and for each size class. In each column only the selection coefficients by size class are shown. For management level professions selection effect is only present in the medium sized establishments with positive coefficient showing that management workers possess unobserved traits that increases their wages. In high skilled white collar and blue jobs we see negative selection into large showing that if they are sorted into large establishments they will get lower wages then random draw of population with similar characteristics in large. In model-2 (number of dependent children as exclusion restriction) there are uniform negative selection coefficients in large size across professional groups. Thus, for all the professions in general the unmeasured ability is not positively rewarded in large establishments and the high intellectual professionals would get higher wage regardless of the size they are attached with.
2. We saw that blue colors workers are over present in large so in the next step we tested the ESWP by excluding the blue collar workers. Table C.2.5 show results when we exclude sample of blue collar workers. Here again, only the selection terms are shown. The significance level becomes very low for all categories. In the whole population case all selection terms become insignificant. For second model we can see negative selection term. For male sample there is negative selection terms in all models and for female, in general, there is no evidence of non random selection except in large establishments using model-2. Thus, blue collar workers are over present in large but they do not determine the entire effect.
3. In the next step it is tested that if we exclude the large regions as they make up of half of country's GDP and there may be more competition among regions in some industries that

can affect wages and this can cause severe estimation problems. Table C-2.6 show results when we exclude the large region. It drops 26,593 observations. In this case there is no evidence of non-random selection. All the selection terms become insignificant but this does not prove that there are direct region-industry impacts. This test loses almost half of the data.

4. Table C-2.7 in Appendix-C-2 shows the inclusion of variables related to unions and collective bargaining system in France. If we consider only the covered sample of workplaces where there are staff representatives then we found again negative selection coefficient into large size establishments. The similar results are observed by considering the sample of workplaces where only union representatives are present. Both instruments generate similar results. Using the information on the wage bargaining structure of France in different ways does not explain anything.
5. Now in the next step employer size decision equation is estimated as a multinomial logit model. The second step is OLS. All these procedures are implemented with -selmlog- command in Stata. Bourguignon et al. (2007) methodology is used to estimate the employer size wage equation correcting for selection using the command 'selmlog' available in stata. Table C-2.8 in Appendix-C-2 shows main results using method BFG (dmf (2)), while for other methods dhl, dmf (0), dmf (1) and lee are presented in table C-.2.9 in Appendix C-2. Bourguignon et al. (2007) provided an overview of the methods available to account for selection issues in the context of the multinomial logit model. BFG conduct a set of Monte Carlo experiments and find that in many cases the approach introduced by Dubin and MacFadden (1984) is the preferred method in comparison to the most commonly used procedure proposed by Lee (1983). Selection corrections through multinomial correction with same exclusion restrictions produce similar results but selection coefficients are significant under this method except lee's method where results are not according to the one reported earlier. Overall, results through this method confirm the negative correlation of error for wage and selection equation. All IMRs in three earning estimations by employer size are significant which suggests that ordinary least squares (OLS) estimates would be biased.

The results can be more intuitive if we do similar analysis to determine the selection in each region by introducing the population dummy in each region but as regions are big and independent so comparing one to other can be very problematic. There are disparities within regions with different magnitudes. Similarly, the correlation coefficient is very high when we use number of dependent children as instruments. It will be useful to have more family background information. Nevertheless, the instruments are not perfect to predict the unobserved choices of workers. The same sets of instruments are used in the literature. It is very hard to find relevant instruments; therefore, selection models are not very famous in ESWP literature. Otherwise, selection models are very relevant in cross sectional data dealing with individual choices.

4.3 Full Information maximum likelihood (FIML) estimation

In the (last) section, the size-wage magnitude is estimated controlling for individual and employer characteristics with Heckman two-step procedure. Here we will jointly estimate the selection model and wage model (Greene, 1995, 2000). Since Heckman (1976, 1979), numerous models for detecting and statistically correcting sample selection bias have been developed. Current sample selection models typically involve the simultaneous estimation of two multiple regression models.The key assumption of maximization model is that people enter into the size category where they get higher expected rewards.In this section, we estimate the model using Full Information Maximum likelihood (FIML) method. Detail of estimation procedure is given in the Appendix C-3.

We have two models, the selection and wage equation model as earlier. The latent variable model Y_i^* is the selection model (equation-1) estimated through ordered probit for three employer size choice (small, medium and large represented by capital S) and wage model is the model of wage determinants (equation-2).

$$\boldsymbol{Y_i^* = Z_i \gamma + \varepsilon_{1i}} \qquad \textbf{(1)}$$

$$\boldsymbol{W_i = X_i \beta_i + \varepsilon_{2i}} \qquad \textit{(2)}$$

$$S_i = \begin{cases} 1(small\ size) & if\ Y_i^* < 0 \\ 2(medium\ size) & if\ 0 \le \mathrm{Y_i^*} < \mu \\ 3(large\ size) & if\ Y_i^* \ge \mu \end{cases}$$

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N\ \left\{E = \left\langle \begin{matrix} 0 \\ 0 \end{matrix} \right\rangle, \ \Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right\}$$

We have to normalize $\sigma_{11} = 1$

The first equation is selection equation as above where dependent variable is the probability of belonging to one of the size group and the second equation, equation 2, is the wage equation where the dependent variable is log of gross hourly wage. The difference is that here we jointly estimate both equations.

The likelihood function can be written as;

$$\boldsymbol{L} = \prod_{i=1}^{N} \boldsymbol{li} = \prod_{i=1}^{N} \boldsymbol{l}_{1i}^{1(Yi=1)} . \boldsymbol{l}_{2i}^{1(Yi=2)} . \boldsymbol{l}_{3i}^{1(Yi=3)}$$

The detail estimation procedure is presented in the Appendix-C. The three parts of the likelihood function are written in stata. The maximum likelihood model is used, by programming the likelihood function in STATA and then using *ml max* procedure. The correlation ρ $(\varepsilon_1, \varepsilon_2)$ between two unobservable components in the selection equation and wage equation is also estimated along with the variation of ε_2 in disturbances variance-covariance matrix.

4.3.1 Results of Maximum likelihood function

Results for maximum likelihood function are presented in ***Table-4.2***.We can see that in FIML wage equation, the effect of employer size on individual hourly wage is monotonically increasing, even after controlling for employer and employees characteristics. In FIML wage equation analysis (Model-1), results indicate that people working in the medium size establishments earn 4.7% more wages compared to small size establishments, whereas, people working in the large establishments earn as much as 15% more wages compared to small establishments. Model-2 shows employer size premium from medium to large increasing from 11.7% to 31.9% compared to small size establishments. The signs of explanatory variables in the FIML wage equation are the same and in the right direction as mentioned above. The significance level and the magnitude is almost the same in both methods of Heckman and FIML. Table-C-3.1 shows increasing level of wages by employer size (confirming the effect of size on wage), increasing returns to educational level, profession, tenure and experience. It also shows more wages for male compared to females, more wage for permanent employment contract compared to fixed term contract and more wages in region ile de France compared to the other regions.

The correlation coefficient (correlation of error terms in the selection and wage equation) is negative which shows that unobservable in the selection equation and wage equation are negatively correlated with each other. The unobserved factors that increase the likelihood of going to large size of establishment are associated with lowering wages. On the other hand the unobserved factors that decrease the probability of going to large size employers are associated with increasing wages. The paradox implies that those who self-selected themselves in small size employers earn better than the population average of the earning in small firms. In contrast, those who self-selected into large, earn lower than population average of earnings in large firms. The similar arguments justifies the negative selection into large as mentioned in the above section. The correlation coefficient in Model-1 is negative and significant for male and for females it is positive and not significant.

Thus, it is observed that large employers do not rely on unobserved ability of workers rather; they evaluate workers based on education, tenure, experience etc. Moreover, monitoring is difficult in large plants as Garen (1985) mentioned that monitoring costs rise with firm size. As a result, large firms acquire less accurate information about the abilities of their workers, and thus will rely less heavily on their own evaluation of workers than do small firms and more on other indicators of ability such as schooling. Similarly, Silva (2004) found that the observed skills; namely, education, age, and tenure have high returns in large firms.

The hypothesis of compensating wage differentials has got importance in the size-wage differential but it is not empirically proven as it is difficult to test the working conditions across professions and jobs. The working conditions and selection are considered to go side by side which can provide another explanation of sorting of low ability workers into large size establishments. Brown (1980) introduced that if workers differ in their abilities and these differences are unobserved, and if the list of job attributes is incomplete, it is likely that more capable workers would be more probably found on jobs offering favorable working conditions. Rosen (1983) further explained that workers with greater earning capacity would "spend" some of it on more on-the-job consumption. This is the fundamental reason why low paying jobs tend to be the "worst" jobs. This explanation goes against the pure equalizing difference hypothesis.

Thus, in general, the effect of size on wage is positive and monotonically increasing after controlling for employer and employees characteristics. After taking into account the self-selection mechanism, different patterns of wages are observed when employees are sorted nonrandomly into different size of employers. Comparing FIML to Heckman shows better results as significance level of all variables improves with the correlation coefficient is significant in most of the cases. Table C-3.2 presents results by using number of dependent children as exclusion restriction. All the results are same with higher correlation coefficients.

We see that for Male the FIML shows the wage premium depending on size varies from 7.6% to 23% compared to base category (small) while for female workers the size premium exists in large with difference of 5%. Thus, we may say that the wage premium by taking into account the selection bias is more important among male workers working in different size of establishments. Among female workers in different sizes of establishments the FIML show small size magnitudes and only significant in large size compared to small size. The correlation coefficient is negative for both males and females cases but for females it is not significant. Wage premium among male workers is higher in large compared to small and there is strong evidence of nonrandom assignment for male works.

The size-wage premium and non random assignment is more important in the male workers because in large establishments or plants there is more division of labor especially in the production plants there is heavy machinery and expensive capital equipment for which large plants need more 'dependable' workers as Master (1969) pointed out. To maintain such workers, they pay them higher wages. Due to safety risks women do not mostly work with heaving equipment and especially in production plants. The gender wage differential patterns are explained in detail in the next chapter.

Table-4.2 Estimation Results with Joint Model FIML

Dependent variable gross hourly Wage	**All**	**Male**	**Female**	**All**	**Male**	**Female**
Instruments	Region size*Industry			Number of dependentchildren		
Medium size (base small size)	0.046***	0.074***	0.009	0.111***	0.090***	0.159***
	(0.011)	(0.011)	(0.012)	(0.010)	(0.013)	(0.010)
Large size	0.144***	0.209***	0.052**	0.277***	0.246***	0.341***
	(0.021)	(0.021)	(0.022)	(0.020)	(0.026)	(0.018)
Workercharacteristics [a]	Yes	Yes	Yes	Yes	Yes	Yes
Employer'scharacteristics [b]	Yes	Yes	Yes	Yes	Yes	Yes
Correlation coefficient	***-0,145***	***-0,235***	***0,008***	***-0,406***	***-0,306***	***-0,573***
Test of correlation	*11.01*	*34,43*	*0,02*	*134,96*	*39,91*	*444,900*
p-value	*0,000*	*0,000*	*0,880*	*0,000*	*0,000*	*0,000*
Observations	74,696	44,654	30,042	44941	27304	17637

*Notes: The sample includes full time jobs, private sector, non agriculture and non chief executive officers. The dependent variable is the logarithm of the GROSS hourly wage rate. This is a semi-log model. Robust standard errors of size coefficients are between parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.*

a) Worker characteristics include experience, tenure, education, family situation and profession. Experience and tenure also includes squared and cubic terms.

b) Employer's characteristics include type of employment contract, type of industry and region.

4.3.2 SensitivityAnalysis

The similar estimations are performed as mentioned in the above section (4.2.2) for different groups using joint FIML model. All results are provided in Appendix C-3.

1. Results by keeping each profession as a reference category are presented in Appendix-C-3 Table C-3.3.For management level professions there is positive correlation and decreasing size wage relationship. Employer size wage gap exists for middle level and lower level professions. The high intellectual professionals will get higher wage whatever size category they are attached with. Employer size wage gap is higher in high skilled white collar jobs followed by blue collar jobs. But there is no evidence of nonrandom selection for lowest level professions in the first model. In the second model we see uniform negative and significant correlation across professions and significant and increasing employer size wage relationship. Based on the results of first model, it is observed that employer size wage relationship is strong among high skilled white collar professions and there is evidence of nonrandom assignment with negative selection effect.
2. Table C-3.4 shows results when blue collar workers are excluded from the sample. It is observed that excluding the blue collar workers do not change results. The employer size wage gap is still present and there is strong evidence of negative selection into large size establishments.
3. Excluding large regions (Table C-3.5) make correlation coefficient insignificant.
4. For unionization, it is observed that establishments with covered staff representatives or covered union representative show increasing employer size relationship as well as evidence of nonrandom selection. This is shown in Appendix-C Table C-3.6. The information on unionization available in the data is used in different ways but all variables are failed to explain the collective bargaining arrangements of French system.

The results of FIML and Heckman are in the same direction and signs for most of the variables are similar. Both of these methods, FIML and Heckman selection, were important to determine the effect of unobservable factors that result into self-selection on size-wage premium.

Heckman's solution was devised within a framework of structural equation modeling that is simple and concise and that can be used in conjunction with the standard framework of OLS regression (Guo and Fraser 2010) while maximum likelihood estimator requires more computing time, and computing speed. Kennedy (2003) argues that the Heckman two-stage model is inferior to the selection model or treatment effect model using maximum likelihood because the two-stage estimator is inefficient and introduces a measurement error problem, because an estimate of the expected value of the error term is employed in the second stage. Further, Heckman model is criticized because the standard error estimates are inconsistent and heteroskedastic due to selection. The estimates from the FIML model are consistent and asymptotically efficient under the assumption of normality and homoscedasticity of the uncensored disturbances. FIML is efficient because of using the likelihood function rather than the method of moments and, second, the estimation of ρ subject to the constraint $-1<\rho<1$. (Greene 2010[66]).

The Maximum likelihood estimator is generally preferred over Heckman's two-step method; however, the Heckman approach provides a useful way to explore the problem. Both models can be used side by side. The switching regression model has been famous in the employer-size wage selection models and with FIML it was possible to determine the size-wage magnitude with selection.

4.4 Basic HourlyWage

Results for the measure of basic hourly wage are presented in Tables 4.3 and 4.4. FIML show insignificant size-wage magnitude and no correlation of the error terms. Among male employees there is positive employer size-wage coefficient though very low compared to gross

[66] *Econometric Analysis, Fourth Edition by William Greene Chapter 19: Models with Discrete Dependent Variables.*

wage measures. For female the size coefficients are negative with positive correlation of the error terms. Results for the second model are quite normal as in the gross wage with low magnitudes. Heckman results in table 4.4 for basic wage are also not clear. The selection coefficient in large size is positive in the first model. It is observed from the net hourly wage analysis that the pattern of nonrandom selection is not clear and results are hard to interpret. The employer size effect is low or negligible. Again the results show that the employer size wage effect is mainly the effect of compensation and pay practices that vary by employer size.

Table-4.3Estimation Results with Joint Model FIML

Dependent variable, gross hourly Wage	**All Sample**	**Male**	**Female**	**All Sample**	**Male**	**Female**
Instruments	**Model-1** : ***Industry type *Region Size***			**Model 2:** ***Number of Dependent Children***		
Medium size (base small size)	-0.004	0.019**	-0.027***	0.053***	0.034***	0.147***
	(0.008)	(0.009)	(0.010)	(0.013)	(0.011)	(0.009)
Large size	0.022	0.078***	-0.041**	0.137***	0.110***	0.294***
	(0.015)	(0.018)	(0.019)	(0.025)	(0.022)	(0.015)
Worker'scharacteristics	Yes	Yes	Yes	Yes	Yes	Yes
Employer'scharacteristics	Yes	Yes	Yes	Yes	Yes	Yes
Correlation coefficient	***-0,043***	***-0,143***	***0,075***	***-0,288***	***-0,204***	***-0,618***
Test of correlation	*1,820*	*15,270*	*2,930*	*31,480*	*21,130*	*809.83*
p-value	*0,177*	*0,000*	*0,087*	*0,000*	*0,000*	*0,000*
Observations	74 453	44 516	29 937	44 775	27215	17 560

Notes: same as in Table 4.2 except that the dependent variable is basic (net)hourly wage in this case.

Table 4.4 Heckman estimation procedure (Second-step regression): Net hourly wages

Variables	Model-1 : *Industry type *Region Size*			Model 2: *Number of Dependent Children*		
	(Small)	(Medium)	(Large)	(Small)	(Medium)	(Large)
Gender (base female)	0.099***	0.093***	0.108***	0.120***	0.124***	0.091***
	(0.003)	(0.004)	(0.003)	(0.005)	(0.010)	(0.005)
Experience	0.024***	0.018***	0.026***	0.017***	0.011**	0.027***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)
Before Bac without degree (base no degree)	0.022***	0.044***	0.069***	0.046***	0.082***	0.013
	(0.005)	(0.007)	(0.005)	(0.009)	(0.020)	(0.008)
CAP/BEP	0.061***	0.074***	0.105***	0.066***	0.086***	0.084***
	(0.004)	(0.005)	(0.004)	(0.006)	(0.010)	(0.005)
Bac professional and technical	0.128***	0.135***	0.184***	0.141***	0.150***	0.151***
	(0.006)	(0.009)	(0.006)	(0.009)	(0.017)	(0.008)
Bac general	0.132***	0.144***	0.191***	0.143***	0.161***	0.177***
	(0.007)	(0.010)	(0.007)	(0.009)	(0.015)	(0.009)
Bac +2	0.187***	0.200***	0.258***	0.216***	0.246***	0.174***
	(0.006)	(0.009)	(0.006)	(0.012)	(0.028)	(0.011)
Bac+3 and plus	0.299***	0.351***	0.424***	0.360***	0.438***	0.318***
	(0.010)	(0.015)	(0.009)	(0.019)	(0.046)	(0.017)
CDI Contract(base CDD)	0.022***	0.037***	0.149***	-0.001	-0.003	0.200***
	(0.006)	(0.008)	(0.008)	(0.011)	(0.023)	(0.013)
Manufacturing (base trade)	0.039***	0.047***	0.067***	0.140***	0.178**	-0.249***
	(0.010)	(0.015)	(0.011)	(0.031)	(0.086)	(0.032)
Services	0.019***	0.036***	0.062***	0.061***	0.094**	-0.092***
	(0.005)	(0.009)	(0.007)	(0.013)	(0.039)	(0.018)
PersonalCharacteristics	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes
λ	-0.008	-0.003	0.041***	0.132***	0.122	-0.395***
	(0.013)	(0.013)	(0.014)	(0.045)	(0.082)	(0.043)
σ	0,23	0,22	0,21	0,23	0,22	0,20
ρ	-0,03	-0,01	0,20	0,58	0,56	-1,93
Constant	3.252***	3.240***	3.019***	3.325***	3.111***	3.814***
	(0.018)	(0.027)	(0.028)	(0.031)	(0.095)	(0.079)
Observations	30,102	14,434	29,851	17,383	8,550	18,803
R-squared	0.64	0.66	0.69	0.66	0.67	0.70
Adj. R-squared	0.64	0.66	0.69	0.66	0.67	0.70

Notes: same as in Table 4.1 except that the dependent variable is net hourly wage in this case.

4.5 Conclusion

This chapter attempts to explain the nonrandom selection of workers across employers of different sizes to examine the size-wage gap in French establishments. Different estimation methods including full information maximum likelihood and Heckman two-step estimation procedures are used to analyze and compare the wage outcomes by employer size. Although we were not able to get panel data or perfect instruments but the impact of selection bias is neither thrown away nor assumed to be random but is explicitly used and modeled in the equation estimating the hourly wage.

Main findings are listed below:

- The selection terms in Heckman model are mostly not significant while FIML shows strong negative correlation for all population and for male sample. Two things are evident from this analysis; either the choice of selection model contains the maximum jobs and observable characteristic that can make the selection effect not significant in some cases or the choice of instruments is not good to control for unobserved heterogeneity.
- The size premium and selection effect is strong among male workers in different size of establishments. Among female workers, the selection component is not significant for the larger group.
- Negative selection into large size establishment and positive selection into small size establishment is found suggesting that the unconditional wage is lower than conditional wage for the small size establishment worker. On the other hand, the conditional wage is lower than the unconditional wage for the employees working in large size establishment. The unobserved factors have high reward in small size establishments while observed factors have high rewards in large size establishments.
- There is strong need to test the role of unions to study the behaviors of large employers. This chapter fails to analyze the role of collective bargaining system of France with the given information. Data does not enable us to test theories of employer size wage gap,

but results are in line with other studies in the size-wage literature (Idson and Feaster 1990 and Lluis 2008).

- Generally FIML is considered more efficient compared to Heckman but the switching regression models have been popular in the employer size wage gap literature dealing with selection that allow us to observe wages in different regimes with selection effect. Therefore, we may conclude that both models are equivalent and may be used side by side especially in applied work.

Given the importance of correcting for selection bias in the size wage literature, the lack of definitive findings is of concern. More research needs to focus on the causal relationship of employer size wage gap. The question to determine the strengths and weaknesses of each method and the optimal conditions under which each method should be used remains unanswered.

Next chapter decomposes the gender wage differentials across size categories in order to know whether the size wage differential is actually a gender wage differential.

Chapter - 5

DECOMPOSITION OF GENDER WAGE DIFFERENTIAL BY EMPLOYER SIZE

Abstract

There are hundreds of papers on gender wage gap but how the size of employer can alter the compensation structure of male and female wages is not studied in detail in the size-wage literature. Heckman two step estimation procedures and standard Oaxaca (1973) Blinder (1973) wage decomposition method is used to decompose the gender wage difference across employer size in order to compare the patterns of gender wage gap in different sizes of employer in French labor market. There is obvious gender wage gap in all employer size categories and in all occupations where men and women are evaluated differently for the same characteristics. Gender wage difference increases by size of employers. Two factors are important in explaining the employer size gender wage gap: first, the women segregation into low paying workplaces; stereotype selection of jobs hinder women career development. Second, employer's behavior is discriminatory against women. In large, men get an unfair advantage over women. Women are disproportionately represented in the low paid occupations. There is prevalence of horizontal segregation that results into low wages and increases gender wage gap. A larger part of the gender wage gap remains unexplained even after adjusting for selection.

5.1 Introduction

Why do women earn lower wages than men? What are the factors that determine gender wage differentials? These questions have been discussed many times in the literature of gender wage differential. This resulted in various theoretical and empirical explanations of gender wage gap. The traditional approach in analyzing the determinants of the wage gap is to consider the role of gender differences in human capital characteristics and labor market discrimination. Key determinants of gender discrimination include gender segregation in organizational hierarchies, undervaluing of women's work, uneven division of domestic labor based on the ability of women and men to devote time to labor market work and/or women's concentration in jobs where, on one hand, pay is lower and, on the other hand, career prospects are weaker (Smith 2010).

One of the earlier theory of explaining gender wage gap is 'human capital theory' which hypothesized that investment in training and acquisition of skills depend on the anticipated returns from such investment and since women are less likely to invest in education they are less likely to get returns on these skills because of breaks from employment.[67] This theory became less applicable with the increase in the female participation in the labor market and higher educational level. A second set of literature on gender wage gap shows segregation of women into low-wage jobs. Many studies found women segregation into low paying occupations as the main source of gender wage differentials. This is called the occupational sex segregation[68]. While another aspect of segregation is firm-segregation. The inter-firm wage differentials results in gender wage differentials. Such studies drawing on matched employer-employee data reveal that female segregation into low-wage workplaces play a particularly important negative impact on their relative wages.[69] The size of the gender pay gap is related to the global characteristics of the wage structure and, in particular, to the extent of wage dispersion. As women are usually concentrated in the lower part of the wage structure, the more dispersed the structure prevailing in a country the greater the penalty for female wages. Accordingly, empirical evidence shows that gender wage gaps are generally higher in those countries with comparatively more dispersed wage structures.[70] It is observed that women are concentrated in low wage occupations but generally gender wage gap exists at all levels. According to European Commission Report of 2003, European countries are still suffering the discrimination for all positions particularly at a supervisory level despite women's higher educational levels and increasing labor force participation.

The gender wage gap in France remains at an intermediate level compared to other EuropeanUnion countries. The job characteristics and labor market segregation both contribute to the gender wage gap. Particularly, in France, the vertical segregation positively contributes to explaining the full-time gender wage gap (Matteazzi et al. (2013).There are other studies showing that women are more strongly discriminated against men in full-time than in part-time

[67] *Becker (1975), Human Capital: A Theoretical and Empirical Analysis, with SpecialReference to Education, 2nd ed., National Bureau of Economic Research.*
[68] *Velling, Johannes 1995, Groshen 1991, Dolado et al. 2004, Bayard et al. 2003, Macpherson and Hirsch 1995, Simon 2012.*
[69] *Bayard et al., 2003, Meng, 2004, Groshen 1991*
[70] *Blau and Kahn, 1992, 1996, 2003, Simón and Russell 2007, Simon, 2012.*

jobs (see Johnson and Stafford 1974). Similarly, Meurs and Ponthieux (2006) focused on the evolution of the wage gap in France between 1990 and 2002. They found that 60% of the explained part of the gender wage gap is due to the length of working hours. In other words, part-time employment is a key factor in explaining the gender wage gap. One other explanation could be the educational and professional choices of women that result into low wage. Women are less likely to go to mathematics and pure sciences field. Machin and Puhani (2003) in their project (using person-level data from Britain, France and Germany) focused on university graduates and showed that the subject of degree (diplome) matters for the gender wage gap. Wages differ by subject of degree where men are predominant in engineering and related fields and women are predominant in education and language studies. Luca (2011) found that a positive differential in College education for women is by now a common feature in OECD countries. The distribution between a first level degree (Maîtrise degree in France) and a second level degree (DEA in France) shows a greater concentration of women in first level degrees. Social Sciences, Business and Law are the preferred fields for women while the preferred fields for men are Engineering and Architecture with Business and Law. Women acquire a little more tertiary education than men but they are more concentrated in the first level of tertiary education (e.g. B.A.) than in the second level (e.g. Master).

The gender wage gap for France was 18 % in 2008 (European Structure of Earnings Survey). Based on 2011 European commission of justice report, gender wage difference in France is 17 %. The wage differential between women and men for France narrows very slowly because of its key determinants which remain stable over the years. Those may include activity profile and job status. 75% of this wage discrepancy is accounted for by differences in job characteristics, the duration of work and working hours. Without radical measures, further improvement is hardly expected (IRES 2012 draft).

The gender wage differential is also observed using ECMOSS 1992 data from France. In the previous chapter, we saw that there exists a significant wage differential between male and female employees.The FIML and switching regression model confirmed that male workers are paid more wages as compared to females. Using FIML method, the wage difference between male and female is 16%. Further, we saw in Heckman two-step method that gender wage difference is higher in large size establishments. We saw that gender wage difference increases with employer size, a male worker in small size gets 14% more than a female worker, the difference being 14.3% in medium size establishments and 15.7% in large ones. Here, we will determine which factors constitute this gender wage gap. We will examine the share of each component, explained and unexplained, in the total differential and in each size category. Thus, we will decompose the gender wage gap by employer size categories. The objective is to analyze why different size establishments pay male and female workers of similar characteristics differently. The gender wage gap is largely studied but the gender wage gap by work place is not explored in detail for French labor market. There are three types of selection: selection on the decision to work, selection for occupation and third is selection for employer size. As this study addresses the sample of employed workers only, we will particularly focus on the selection bias in the employer size and wage relationship. The other two types of selection biases are left for future work.

With the popularity of wage decomposition methodology introduced by Blinder (1973) and Oaxaca (1973, henceforth O-B) many forms of discriminations have been evaluated using this wage decomposition method including gender discrimination, wage differentials based on ethnicity or race etc. In the presence of nonrandom selection, OLS estimates are not consistent (sample selection bias). Sample selection has been shown to be a potential source of bias in several studies of earnings differentials. Wage decomposition with sample selectivity bias correction is realized by many authors[71]. Moreover, quintile regression approach is largely adopted in recent studies to observe the distribution patterns at upper and lower tails. Several papers decompose the gender wage gap across the distribution for different countries[72]. Results for gender wage decomposition in this chapter are estimated through classical O-B wage decomposition. The contribution of discrimination, human capital and selectivity in different sizes of establishments is examined following Jann (2008), Oaxaca and Ransom (1994) and Neuman and Oaxaca (2004). The dataset used is ECMOSS 1992 as used in the previous chapters.

---

[71] *Neumark 1988, Oaxaca and Ransom 1994, Neuman and Oaxaca 2004,Reimers 1983, Dolton et all 1989.*
[72] *Badel A. & Pena X. (2010) for Columbia, Albrecht et al. (2003) for Sweden, de la Rica et al. (2007) for Spain, Ganguli and Terrell (2009) for Ukraine, and Ñopo (2006) and Fernández, (2006) for Chile. Albrecht et all (2004) for Netherlands.*

The rest of the chapter is as follows; section 5.2 below presents methodology. This is followed by wage equation estimates in section three. Section four presents wage decomposition analysis and section five concludes.

5.2 Methodology

This chapter is in continuation of the previous chapter, accordingly, the same data (ECMOSS 1992) and exclusion restrictions are used for wage decomposition analysis. The switching regression model is employed to compare the expected earnings of a male worker compared to a female one in small size establishments if he self-selected into small size establishment and the expected earnings of a male employee working in large size establishments who self-selected into large size compared to his female counterpart. The usual procedure to measure the male-female wage gap is to consider the differences between the average male wage and his female counterpart in different employers' classes.

The comparison of mean wage differences motivates to explore further the gender wage gap patterns in the presence of selection bias. For this purpose, Heckman method is employed to study the gender wage gap across employer sizes. A simple two equations model of wage determination and employer size selection among employed workers illustrates the application. The Heckman two step estimation procedures is used for identifying parameters and later standards O-B decompositionis applied to the regression equations.

5.2.1 Wage Determination

Following the methodology of Neuman and Oaxaca (2004), we consider the gender wage differential by employer size. First, we determine the hourly wages of male and female by employer size and later we decompose the wage differential into explained and unexplained components. The unadjusted (without taking into account selection) and adjusted (with selection effect) wage gap is estimated in large and small establishments.[73] We make a simple two

[73] *The employer size dummy is used in this chapter unlike three size groups as in the last chapter because the difference in coefficients of small and medium was less. Moreover, the program in stata 'oaxaca' command only allows the dummy variable for decomposition. Large establishments are defined with 200 and more workers and small establishments are defined as less than 200 workers.*

equations model of wage determination and employer assignment to illustrate the Heckman two steps estimation procedure. We assume that the employer size and wage functions for individual *i* in gender group *j* be given by;

$$\boldsymbol{Y^*_{ij} = Z_{ij}\gamma_j + \varepsilon_{ij}} \qquad \textbf{\textit{(1)}}$$

$$\boldsymbol{W_{ij} = X_{ij}\beta_j + v_{ij}} \qquad \textbf{\textit{(2)}}$$

Where $Y^*_{ij}$is a latent variable associated with probability of being employed in large (or small) size establishment, Z_{ij} is a vector of determinants of employer assignment, W_{ij} is the hourly wage (in logs), X_{ij} is a vector of wage determinants, $\gamma_j$and $\beta_j$are the associated parameter vectors and εij and vij are i.i.d. error terms that follow a bivariate normal distribution $(0, 0, \sigma\varepsilon j, \sigma vj, \rho j)$.

The probability of belonging to large size establishment is given by;

$$prob\left(Y^*_{ij} > 0\right) = prob\ (\varepsilon_{ij} > -Z_{ij}\,\gamma_j)$$

$$= \Phi(Z_{ij}\,\gamma_j) \qquad (3)$$

Where $\Phi(.)$is the standard normal C.D.F. (the variance of εj is normalized to 1).
Wages are observed for those for whom $Y^*_{ij} > 0$, so that the expected wage of a large establishment worker is determined by;

$$E\left(W_{ij} \middle| Y^*_{ij} > 0\right) = X_{ij}\beta_j + E(v_{ij}|\varepsilon_{ij} > -Z_{ij}\,\gamma_j)$$

$$= X_{ij}\beta_j + \theta_j\lambda_{ij} \qquad (4)$$

Where $\theta_j = \rho_j\sigma_{vj}$ and $\lambda_{ij} = \frac{\phi\left(Z_{ij}\,\gamma_j\right)}{\Phi\,(Z_{ij}\,\gamma_j)}$ and $\phi(.)$is the standard normal density function. The expected wage for small establishment workers ($Y^*_{ij} < 0$) is determined by $\lambda_{ij} = \frac{\phi\left(Z_{ij}\,\gamma_j\right)}{1-\Phi\,(Z_{ij}\,\gamma_j)}$. The estimating equation for individuals may be expressed as

$$\mathbf{W_{ij}|\,Y^*_{ij} > 0 = X_{ij}\beta_j + \theta_j\lambda_{ij} + error} \qquad \textbf{(5)}$$

The parameters of (5) will be estimated by Heckman two-step estimation procedure separately for male and female.

5.2.2 Wage Decomposition

We denote M for males and F for females. We use the classic threefold B-O decomposition (Blinder 1973, Oaxaca 1973): Stata command ‘oaxaca’ computes decomposition (see Jann (2008) for details of the procedure). The decomposition below is formulated from the view point of women;

$$\boldsymbol{G} = \{\boldsymbol{E}(\boldsymbol{X_M}) - \boldsymbol{E}(\boldsymbol{X_F})\}\boldsymbol{\beta_F} + \boldsymbol{E}(\boldsymbol{X_F})(\boldsymbol{\beta_M} - \boldsymbol{\beta_F}) + \{\boldsymbol{E}(\boldsymbol{X_M}) - \boldsymbol{E}(\boldsymbol{X_F})\}(\boldsymbol{\beta_M} - \boldsymbol{\beta_F}) \qquad \textbf{\textit{(6)}}$$

G represents the gender wage gap on the left-hand side. This is threefold decomposition where gender wage gap is divided into three components;

G=E +C+ I

The first components, $\{\boldsymbol{E}(\boldsymbol{X_M}) - \boldsymbol{E}(\boldsymbol{X_F})\}\boldsymbol{\beta_F}$ represents the effect of endowments. This amounts to the part of differential that is due to group difference in the predictors. The group differences in the predictors are weighted by the coefficients of women i.e. the expected change of women’s mean wage if they had the same predictor levels as men. This is also called the explained component of the gender wage gap.

The second component, $\boldsymbol{E}(\boldsymbol{X_F})(\boldsymbol{\beta_M} - \boldsymbol{\beta_F})$, measures the contribution of differences in the coefficients and intercept. The difference in coefficients of both groups is weighted by women’s predictor levels, i.e. the expected change of women’s mean outcome if they had the same coefficients as men. This component represents the “discrimination component”, or the unexplained part of the gender wage gap.

Finally, the third component, $\{\boldsymbol{E(X_M)-E(X_F)}\}(\boldsymbol{\beta_M-\beta_F})$**,** is the interaction term that simultaneously measures the difference in endowments and coefficients between the two groups. The first two components are considered as most relevant in the gender wage gap literature. The decomposition from the viewpoint of men can be written as;

$$G=\{E(X_M)-E(X_F)\}\beta_M+\ E(X_M)(\beta_M-\beta_F)+\ \{E(X_M)-E(X_F)\}(\beta_M-\beta_F)$$

Now the endowment effect represents the expected change in men's hourly wage if they had female's predictor level. The coefficient effect quantifies the expected change in males' hourly wage if they had the same coefficients as those of women. Alternative decomposition method in the discrimination literature is the use of nondiscriminatory coefficient vector to determine the contribution of differences in the predictors. Oaxaca (1973) proposed an index number to estimate the unknown nondiscriminatory coefficient vector, Reimers (1983) proposed using the average coefficients over both groups, Cotton (1988) suggested to weight the coefficients by the group size and Neumark (1988) and Oaxaca and Ransom (1994) suggested to use the coefficients from a pooled regression over both groups as an estimate of the unknown nondiscriminatory vector. Stata's command 'Oaxaca pooled' implements this method.

For selectivity bias adjustment, we follow the Neuman and Oaxaca (2004) methodology (see Neuman and Oaxaca (2004) for details). This approach deducts the selection effect from the overall differential and then the standard threefold decomposition is applied. To implement this, 'Oaxaca' command in Stata is used with Heckman two steps procedure where the decomposition automatically adjust for selection. For decomposition of adjusted wage gaps, the formulation is expressed as:

$$G-\ (\theta_M\lambda_M-\theta_F\lambda_F)\ =\{E+C+I\}$$

5.3 Empirical Results

5.3.1 Wage Equation unadjusted and adjusted for selection

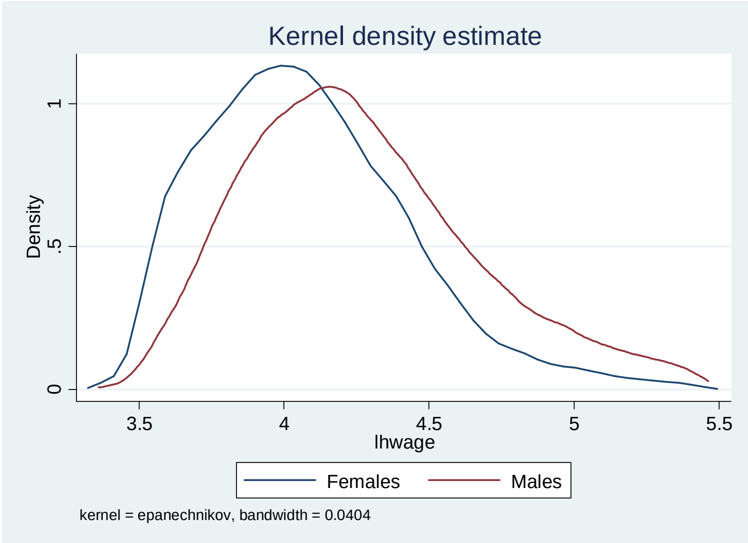
The data used for this chapter is ECMOSS 1992. For the selection model in the first step, the same instruments as in the last chapter are used[74]. Here the selection model is a probit model instead of an ordered probit as we have only two size categories of employers. Table D-1 in Appendix D-5 shows probit estimates for male and female. The adjusted and unadjusted wage equations are presented in Table D-2. Standard Mincer type equation is estimated. These results are then used to calculate the share of the endowment, discrimination and selectivity components in the wage differential by size and by gender. The log of individual hourly wage is regressed on various control variables related to individual and employer's characteristics. The results of the wage equation complement the previous studies. We see the same sign and direction of the effect of observable individual characteristics on hourly wage. Education has a strong positive effect on the wages for both male and female workers. As education increases, rewards increases and as size increases reward increases. Similarly, experience and tenure pose positive impact on wage. The selection coefficient (λ) is only significant for male sample in large size establishments. The negative sign indicates that the unobservable factors are present in both the wage equation and selection equation and both are correlated with common factors. Therefore, it was important to treat them endogenous to study the employer size effects on wage. For female sample there is no evidence of nonrandom selection. There is negative selection on unobservable in the large establishments for male workers. This implies that workers who self-selected into large establishments possess unobserved traits that depress their wages. Positive selection on observables and negative selection on unobservables and both are negatively correlated with each other.

5.3.2 Blinder-Oaxaca Wage Decomposition

Figure 5.1 describes the kernel density distribution of log wage for male and female. We see that females are concentrated more in the lower level of log distribution of hourly wage.

[74] *Results are presented using interaction of region size and type of industry as exclusion restrictions. Results for using other instrument (number of dependent children) also computed and can be requested from author.*

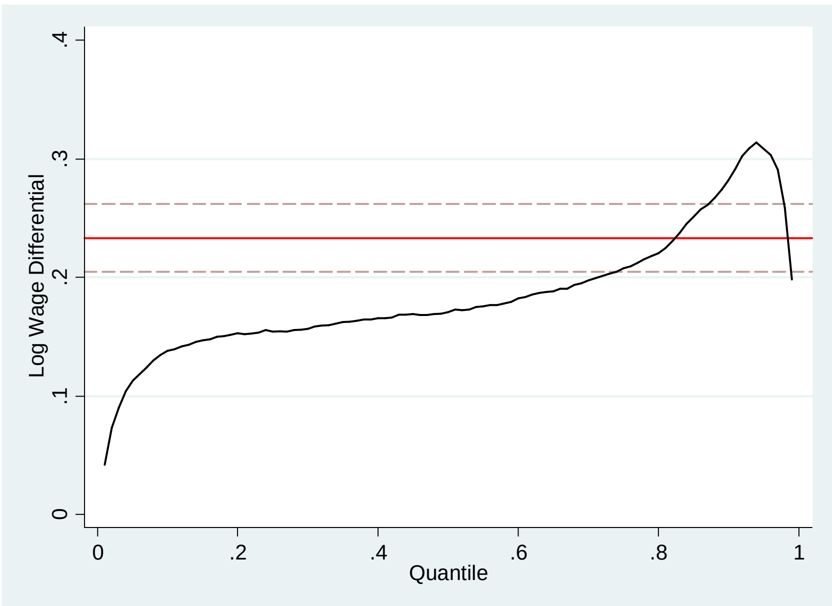
Figure 5.1 Distribution of hourly wage by gender



Source: ECMOSS 1992 author's calculations

Figure 5.2 shows the quintile distribution of gender wage differential. The gender wage gap becomes greater at the top end of the log distribution of wages. The red line shows average gender wage gap. The gender wage gap reaches at top for 90th to 95th percentile. This is glass ceiling effect or may be the sample at the extreme end is unusual.

Figure 5.2 Log gender wage differential



Source: Ecmoss 1992 author's calculations

In the table 5.1 below, the results of B-O decomposition, based on the methodology described in section two, are reported. Table 5.1 shows gender wage decomposition in large and small size establishments. First column show the mean prediction by group and their difference. The second column shows geometric means of wage of both groups and their difference.[75]The detailed table of decomposition and share of each variable in the entire component can be seen in Table D-3.

Table-5.1 Threefold decomposition of Gender wage differential by establishment size

	Large			Small		
	Mean log hourly wage	Exp(b)	%	Mean hourly log wage	Exp(b)	%
Males	4.362***	78.41		4.181***	65.43	
	(0.003)			(0.003)		
Females	4.160***	64.06		4.036***	56.62	
	(0.003)			(0.003)		
Difference	0.202***	1.22	18.31%	0.145***	1.15	13.46%
	(0.004)			(0.004)		
Endowments	0.029***	1.02	13.72%	-0.010***	0.99	-6.64%
	(0.004)			(0.003)		
Coefficients	0.126***	1.13	63.36%	0.112***	1.11	78.26%
	(0.003)			(0.003)		
Interaction	0.047***	1.04	22.91%	0.042***	1.04	28.38%
	(0.003)			(0.003)		
Total N. Obs	29,896			44,800		
Males	19,758			24,896		
Females	10,138			19,904		

Results correspond to OLS regression Appendix-D Table D-2.Results are computed using stata command 'Oaxaca'. exp (b) column is obtained through stata command 'oaxaca eform'. The % wage difference is calculated as (hourly male wage - hourly female wage)/ hourly male wage.

The decomposition output shows the mean wage prediction by gender and their difference. Two wage equations are estimated separately for male and female in one size group. The mean gross hourly wage for males is 4.36 in large size establishments (with 200 and more workers). The mean gross hourly log wage is 4.16 for women, yielding a wage difference of 0.20. The wage gap is divided into three components. The endowment part reflects the mean

[75] *Exp(b) coefficients are obtained though 'Oaxaca eform' option*

increase in women's wage if they had the same characteristics as men. The second component quantifies the change in women's wage when applying the men's coefficients to the women's characteristics. The third term measures the simultaneous effect of difference in endowments and coefficients. The second column shows that in the large size establishments the raw geometric mean of men's wage is 78.4 French francs while that of women is equal to 64 French francs which amounts to a difference of 18.31%. The difference is calculated as the ratio of the wage difference of both groups and men's average wage. The wage difference is coming through three components as explained above. The difference in endowment accounts for 13.7% of the total gender wage gap. The endowment component is 1.029 which reflects that women's wage would increase by 2.9% if they had the same characteristics as men. The coefficients component comes to 1.134. It amounts an increase of 13.4% of the women wage if we apply men's coefficients to the women characteristics. The difference in coefficients explains 63% of the gender wage gap. The interaction component explains 22% of the wage gap. It reflects the simultaneous effect of differences in endowments and coefficients. Among the endowments if women had the same experience as men, they would earn 3.6% more and if they had same tenure as men they would earn 2.5% more. For the educational variables, if women had the same BAC+2 educational levels as men they would earn 0.8% less. If they had the same highest educational level BAC+3 and more, they would earn 0.9% more. Among coefficients, if we apply men's coefficient of experience to experience of women then the later would earn 10.5% more. For tenure, women would earn 7% more. Similarly, by applying men's coefficients to the educational levels of women the difference remains less than 1%. The difference in the type of contract amounts to 11% of the wage difference. This means that if women had the similar type of employment contract as men, they would earn wages 11% higher. Overall, experience, tenure and type of employment contract show larger difference in mean log wage of male and female.

On the other hand the total gender wage gap in small size establishments (establishments with less than 200 employees) is 13% compare to 18% in large. This shows that gender wage differential is greater in large compared to small size establishments. The endowment amounts to 0.99 which reflects that if women had the same characteristics as men, they would earn 1% less. This amounts to a decrease of 6% in the total gender gap in small size establishments. On the other hand women's wage would increase 11.9% by applying men's coefficients to women characteristics. The difference in coefficients explains 78% of the total gender wag gap in small establishments. The remaining interaction component explains 28% of the total wage gap.In the small size establishments, if women had the same experience as men, they would earn 0.9% more and if they had same tenure as men they would earn 0.5% more. For the educational variables, if women had the same BAC+2 educational levels as men they would earn 0.4% less. If they had the same highest educational level BAC+3 and more, they would earn 0.3% more. Similarly, in small size establishments if we apply men's coefficient of experience to women's experience, the wage of the later would increase by 15%. For tenure, it would increase by 1.8%.

It is found that the gender wage gap is greater in large size compared to small size. This gender wage gap is explained less by the characteristics even after controlling for a wide range of individual characteristics, experience, tenure, education, profession , type of contract etc. Over four times as much of the wage gap is explained by difference in coefficients as in difference in endowments. This shows that the discrimination against women is higher compared to the difference in human capital. Men get unfair advantage against women. The total gender wage gap is more in large size establishments but the unfair advantage is more in small size establishments.

Alternatively, the twofold decomposition is computed from a pooled model over both samples to be used as reference coefficients (see for details Oaxaca and Ransom 1994). The conclusion from this model is similar to the threefold decomposition: namely that the discrimination component or the unexplained accounts for more than a half of the gender wage gap. Results are reported in Table 5.2 and detail decomposition table can be seen in Table D-4 in Appendix-D. In large size establishments, the unexplained component account for 73% of the total gender wage gap compared to 26% of the explained component. In small size establishments, more than 90% of the wage difference in unexplained.

Table-5.2 Twofold decomposition of Gender wage differential by establishment size

	Large			Small		
	Mean log hourly wage	Exp(b)	%	Mean log hourly wage	Exp(b)	%
Males	4.362*** (0.003)	78.41		4.181*** (0.003)	65.43	
Females	4.160*** (0.003)	64.06		4.036*** (0.003)	56.62	
Difference	0.202*** (0.004)	1.22	**18.31**	0.145*** (0.004)	1.15	**13.46**
Explained	0.055*** (0.004)	1.05	26.14	0.011*** (0.003)	1.01	7.49
Unexplained	0.147*** (0.003)	1.15	73.86	0.133*** (0.003)	1.14	92.51
Observations	29,896			44,800		
Males	19,758			24,896		
Females	10,138			19,904		

Results are computed using stata command 'Oaxaca pooled'. exp (b) column is obtained through stata command 'oaxaca pooled eform'. The % wage difference is calculated as (hourly male wage - hourly female wage)/ hourly male wage.

In the presence of nonrandom selection, the OLS estimates are biased. Therefore, table-5.3 below presents the decomposition results when we adjust for self-selection. The selection effect is deducted from overall differential and the standard decomposition is applied to the adjusted differential following Reimers (1983) and Neuman and Oaxaca (2004).[76] The Stata command 'Oaxaca' is compatible with Heckman two step selection models. Simultaneous selection model for male and female for $Y_{ij}^{*} > 0$ and for $Y_{ij}^{*} < 0$ are computed.[77] The results are reported in table 5.3 and detail decomposition results can be seen in Table D-5 in Appendix-D.

[76] *The same exclusion restrictions are used as in the previous chapter. The interaction variable of region size and industry type is used in the first step (probit model) and the IMR is used in the wage equation in the second step.*

[77] *Computation of 'heckman' with 'oaxaca' is not straight forward. Although it incorporates the selection of both groups (two probit models) but it can bias the standard errors (Jan 2008). Second way is to compute the selection model outside Oaxaca and then performing 'oaxaca' command with the option of 'adjust' but this option does not allow to take into account the selection of both male and female together.*

Table-5.3 Decomposition of adjusted wage differential

	Large			Small		
	Mean log hourly wage	Exp(b)	%	Mean log hourly wage	Exp(b)	%
Males	4.479***	88,126		4.177***	65,172	
	(0.012)			(0.008)		
Females	4.156***	63,833		4.037***	56,644	
	(0.014)			(0.007)		
Difference	0.322***	1,381	**27,57**	0.140***	1,151	**13,09**
	(0.018)			(0.011)		
Endowments	0.029***	1,030	8,19	-0.010***	0,990	-6,75
	(0.005)			(0.004)		
Coefficients	0.268***	1,307	84,79	0.108***	1,114	77,26
	(0.020)			(0.010)		
Interaction	0.025***	1,025	7,02	0.043***	1,043	29,49
	(0.005)			(0.003)		
λm	-0.166***			0.004		
	(0.017)			(0.014)		
λf	0.004			-0.006		
	(0.014)			(0.014)		
Observations	29,896			44,800		

Notes: Similar as in the tables above

Comparing the adjusted wage differentials to unadjusted (Table 5.1), it seems that the uncorrected wages of women are slightly biased upward, (4.160 versus selectivity corrected 4.15) and the wage gap is under estimated (0.20 versus selectivity corrected 0.32). The adjusted wage gap is similar to the unadjusted one in the small size establishments: 13%. The results are conditional to the choice of instruments and to the choice of selection model to take into account selection effect. The results are also conditional to the decomposition method as this methods takes out selection effect from total effect and decomposes the remaining wage difference while other methods make selection as additional component of the wage gap (see Neuman and Oaxaca (2004).

It is observed that the gender wage gap increases if selection bias is taken into account. The share of endowment in the total wage gap decreases to 8% and share of coefficients increases to 84%. Among small size establishment, the share of each component in the adjusted gender wage gap remains similar to the unadjusted wage gap. Only the selection coefficient for male in large size establishments is significant. Negative selection into large implies that in large size establishments the less able men are likely to enter and get higher wages or if men are selected to work in large size establishments then they would get lower wages compare to random draw of men with a comparable set of characteristics.

After observing the adjusted and unadjusted wage gaps, we saw that more than a half of the wage gap remains unexplained and shows the dominating discrimination effect that determines gender wage gap. One thing is evident: gender wage gap exists in all size groups but is larger in large size employers. On the one hand, we see segregation of women into low wage workplaces where opportunities of promotion are low while, on the other hand, there is unequal access to high-paying jobs for women. There may be a barrier to entry for women in large size establishments. Even if pay is equal, there is unequal access to high-paying jobs. In that case, discriminatory barriers in jobs become important component of gender wage gap. Moulin (2004) found for France that discriminatory barriers affect both segregation and discrimination. He found that a portion of the occupational segregation is related to a discriminatory barriers effect.

Based on human capital theory as employers anticipate that female would spend less time in labor market; they will anticipate getting less return on training and, as a consequence, they may hire less female workers or give them lower opportunities for promotion. The structural elements of the labor market in France may be less responsible for the gender pay gap as the law of minimum wage and collective bargaining agreements apply to all agents of labor market. The interruptions to work and working timings are very important elements that reduce the positive effect of higher labor force participation and educational attainment.

Further, the gender pay gap by age cohorts and by occupation can explain the type of segregation (vertical or horizontal) in explaining the gender wage gap. Table 5.4 shows that women are concentrated in the low skilled white collar occupations. The ‘female occupations’ are often referred to as the ‘five c’s’: cleaning, catering, caring, cashiering and clerical work.

Table 5.4 Distribution of professions among male and female across employer size

Profession	Large			Small		
	Female	**Male**	**Total**	**Female**	**Male**	**Total**
Management and High Intellectual professionals	824	3,080	3,904	1,608	3,808	5,416
	21.11%	78.89%	100	29.69%	70.31%	100
High Skilled White Collar	3,159	6,062	9,221	5,308	6,369	11,677
	34.26%	65.74%	100	45.46%	54.54%	100
Low Skilled White Collar	3,911	1,522	5,433	10,002	2,598	12,600
	71.99%	28.01%	100	**79.38%**	20.62%	100
Blue collar	2,244	9,094	11,338	2,986	12,121	15,107
	19.79%	80.21%	100	19.77%	80.23%	100
Total	10,138	19,758	29,896	19,904	24,896	44,800
	33.91%	66.09%	100	44.43%	55.57%	100

Table 5.5 shows that the gender wage gap across professions is higher in lower level professions and remains largely unexplained. The higher educational levels of women are offset by the coefficients component as women are underpaid against men for similar characteristics. The higher wage difference in lower level profession indicate the higher labor supplied by women in these professions which depress wages.There is predominantly horizontal segregation in our sample where females are employed in low-paid occupations from males. This results into persistent low wages because of over female labor supply into these occupations.[78] On the other hand, to some extent there is vertical segregation because women are under present in high paying occupations. There are entry barriers or men are getting unfair advantage.

[78]*Blau and Kahn, 2000. Gender Differences in Pay Francine D. Blau, Lawrence M. Kahn, NBER Working Paper No. 7732Issued in June 2000*

Table 5.5 Adjusted gender wage difference by profession in large size establishments

Mean log hourly wage and difference	Prof1	Prof2	Prof3	Prof4
Males	4.910***	4.498***	4.190***	4.284***
	(0.025)	(0.017)	(0.030)	(0.015)
Females	4.675***	4.373***	3.978***	3.803***
	(0.058)	(0.021)	(0.025)	(0.017)
Difference	**0.235*** **	**0.125*** **	**0.213*** **	**0.481*** **
	(0.063)	**(0.026)**	**(0.039)**	**(0.023)**
Endowments	0.052***	-0.023***	0.008	0.026***
	(0.011)	(0.006)	(0.005)	(0.005)
Coefficients	0.168***	0.139***	0.203***	0.441***
	(0.064)	(0.031)	(0.039)	(0.024)
Interaction	0.015	0.010	0.002	0.013**
	(0.010)	(0.009)	(0.004)	(0.006)
λm	-0.023	-0.095***	-0.068**	-0.161***
	(0.035)	(0.027)	(0.036)	(0.020)
λf	0.065	-0.046**	0.069***	0.139***
	(0.070)	(0.241)	(0.023)	(0.024)
Observations	3,904	9,221	5,433	11,338

*Prof1= Management and High Intellectual professionals, prof2=High Skilled White Collar, prof3=Low Skilled White Collar, prof4=Blue Collar.*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Further results are computed using the basic hourly wage. Results are presented in Table 5.6 and detailed results are presented in Table D-7. It is observed that the gender wage difference in large size establishments increases with the difference of gross and basic hourly wage, i.e. allowances, bonus and overtime payments. For the basic contractual wage the unadjusted wage gap is 14% compared to 18% and for the adjusted the wage gap is 16% compared to 27%. Again the share of discrimination is highest in the basic wage. The larger part of the wage gap remains unexplained.

Table 5.6 Threefold decomposition of gender wage differential in large establishments

Basic hourly log wage	**Unadjusted**			**Adjusted**		
	Mean log hourly wage	**Exp(b)**	**%**	**Mean log hourly wage**	**Exp(b)**	**%**
Males	4.108***	60.80		4.117***	61.37	
	-0.003			-0.011		
Females	3.954***	52.16		3.942***	51.51	
	-0.003			-0.013		
Difference	0.153***	1.17	**14.21**	0.175***	1.19	**16.07**
	-0.004			-0.017		
Endowments	0.030***	1.03	18.99	0.032***	1.03	17.62
	-0.004			-0.005		
Coefficients	0.089***	1.09	58.81	0.113***	1.12	65.25
	-0.003			-0.019		
Interaction	0.035***	1.04	22.21	0.031***	1.03	17.13
	-0.003			-0.004		
λm				-0.013		
				(0.015)		
λf				0.013		
				(0.014)		
Observations	29,851			29,851		
Male	19,741			19,741		
Female	10,110			10,110		

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0

5.4 Conclusion

Decomposition of wage differentials has been studied by many authors in the context of gender, race, ethnicity etc. But decomposing wage differentials by employer size has not been explored in detail. The Heckman two step estimation procedures is used for identifying parameters and later standards Oaxaca (1973) Blinder (1973) wage decomposition is applied to the regression equations. The objective is to decompose the gender wage difference across employer size in order to compare the patterns of gender wage gap in different sizes of employers. The work-place segregation is considered and the effect of differences in personal characteristics on the gender wage gap is disentangled with the effect of selection into different establishments of women and men.

The gender wage gap is greater in large compared to small size establishments but among all the cases the larger part of the gender wage gap remains unexplained. The adjusted regressions for selection increases gender wage gap in large size but the evidence of nonrandom selection is only found among male workers and no selection term is significant for women. The wage gap exists in both measures of wage, gross and basic, the wage gap increases as the difference of gross and basic wage increases.

A prominent conclusion from this analysis is the women segregation into low paying workplaces. There is stereotype in women allocation into particular jobs that results into low wages compared to men. Employers' behavior is discriminatory against women and they offer men an unfair advantage in the same job. Two factors are important in explaining the employer size gender wage gap, first is the women segregation into low paying workplaces; stereotype selection of jobs hinder women career development. Secondly, employer's behavior is discriminatory against women. Women are disproportionately represented in the low paid occupations, there is prevalence of horizontal segregation that results into low wages and increases gender wage gap. To some extent there is also vertical segregation as only women are underrepresented in high paying occupations.

Decomposition by age cohorts and by working hours of both groups is the next step. Further work should examine the different decomposition method, quintile decomposition and with different adjustments of selection correction as ‘Oaxaca’ option did not enable us to use many types of decompositions. Large part of gender wage gap remains unexplained. Further work can also be done to simultaneously take into account occupational segregation, work-place segregation and decompose gender wage differentials.

Chapter- 6

A PSEUDO PANEL ESTIMATION OF EMPLOYER SIZE-WAGE GAP IN THE PRESENCE OF SELECTION BIAS

Abstract

A cohort analysis by generating a pseudo panel data to deal with unobserved individual heterogeneity is used in many disciplines. The same methodology is used in this chapter to examine the relationship between size of the employer and wages of employees for French labor market. Successive cross sections of the matched employer-employee dataset ecmoss 2005-06 for France are used for building cohort sample and later ecmoss 2002 is added to compare the results. The chapter presents the fixed effects, between effects, random effects and first difference estimates for impact of size on wage based on the cohort analysis for workers who were born between 1928 and 1986. Four different cohort data sets are generated by taking different groups of time invariant individual characteristics. The findings consistently support the hypothesis that within a cohort, there is positive and significant effect of employer size on wage. The impact is strong in the medium scale establishments. The preferred estimation method is fixed effects as unobserved heterogeneity is the principle source of selection bias. Results can be improved by incrasing number of years.

6.1 Introduction

In the previous chapters, different methods were employed to study the relationship of employers' size and wage controlling for observable and unobservable characteristics. Nevertheless, the unobserved heterogeneity cannot be fully captured if valid instruments are not available. This chapter presents how in the absence of true panel data, we can still estimate the fixed effects through pseudo panel dataset in order to explain the relationship between employer size and wage. Further, it shows how in the size wage literature, one can analyze various panel methods using pseudo panel dataset.Cohort specific effects are controlled to deal with problem of endogeneity bias

The causal relationship between the size of employers and wages of worker cannot be explained in the presence of unobserved heterogeneity[79]. To deal with the problem of

[79]*The common problem in the cross-section empirical work when dealing with individual data is the potential endogeneity of explanatory variables on the right hand side of the equation. Ordinary least squares estimates will*

endogeneity bias in the cross section, relevant instruments are required. But if exogenous instruments are not available then one needs to estimate fixed effects and first differencing methods with panel dataset to deal with the problem of endogeneity bias. Panel data is nevertheless important to control for the individual heterogeneity and to obtain more information on the variability between variables. Panel data is useful to have more degrees of freedom, more efficiency and to identify and to measure some effects which are not detectable when using cross-sectional data or time series alone (Baltagi 1995). Pseudo panel has some advantage over the genuine panel. Firstly, there is no attrition problem since new samples of individuals are drawn each year and secondly, data can be made available over a long period of time (Hammer 2007)[80].

There is a growing literature on pseudo panel estimation methods[81]. Pseudo-panel data are constructed from repeated or non-consecutive cross sections on the same reference populations in different time periods. This dataset is constructed to controls for unobserved individual specific effects that may otherwise bias the estimated size wage premium in cross sectional regressions. By constructing a cohort data set from repeated cross sectional surveys on the labor costs and wage structure (for three years 2002, 2005, 2006), this chapter presents the magnitude of size wage impact for workers who were born between 1928 and 1986.

The basic assumption underlying the construction of a pseudo panel is that individuals can be unambiguously allocated to specific cohorts on the basis of some common invariant individual characteristic(s), which implies that all individuals within a cohort are behaving similarly over time (Deaton, 1985). The rationale behind pseudo panel is that when we aggregate all the time invariant information into cells, it homogenizes the individual effects in the same group and these do not vary between two time periods. The unobserved individual effects can further be removed by within or first-differences transformations (Gardes et al.2005).A cohort is

yield biased and inconsistent estimates in the presence of endogeneity bias. The sources of potential endogeneity bias may include omitted variables, simultaneity, and measurement error (Wooldridge, 2002). The most important source of endogeneity in size wage relationship is the sample selection bias. This means that employees are not randomly sorted into establishments of different size. There is self-selection in this relationship of employer size and wage.

80 Hammer(2007) The FE estimator in pseudo panel

81Russell & Fraas 2005, Deaton, A. 1985, Gardes et al.2005, Anil and Peter 2010, Gassner 1998, Warunsiri & McNown, 2010, Glocker & Steiner 2007, Sutton 2004 and many others.

defined as a group with fixed membership, so that an individual is a member of exactly one cohort which is the same for all periods. Examples are age cohorts or cohorts based on sex or other time invariant individual characteristics Deaton (1985).

Compared to a true panel, which is defined by the time dimension and the number of individuals, a pseudo panel can further be differentiated along two additional dimensions, namely the number of observations within a cohort and the number of cohorts. In allocating individuals to cohorts, there is a trade-off between the number of individuals within a cohort and the number of cohorts: On the one hand, if the case of a small number of cohorts is chosen to gain a larger number of observations per cohort, the heterogeneity of the individuals within a cohort arises, this may result in inefficiency in estimation due to aggregation. On the other hand, if individuals in the sample are allocated to a large number of cohorts there is a danger of small number of observations in one cell, this may result in biased estimators. Verbeek and Nijman (1993) argue that cells must contain about one hundred individuals, although the cell sizes may be smaller if the individuals grouped in each cell are sufficiently homogeneous. Thus, the challenge in construction a pseudo panel is to find the optimal choice between the numbers of cohorts on the one hand, and the cohort size on the other. Ideally, the optimal choice would yield homogeneous cohorts of sufficient size (Glocker & Steiner 2007). For this purpose, the pseudo panel data cells are created in such a way to make them homogenous and large. The cell size with at least 100 observations is used in this chapter which makes measurement error problem negligible (Verbeek and Nijman, 1993). Moreover, four different cohorts of time invariant individual characteristics are constructed and results are compared.To our knowledge, this is the first work based on pseudo panel data to analyze the employer size and wage relationship. The most recent available dataset in ECMOSS type survey is for the successive years 2005 and 2006. Firstly, pseudo panel analysis is made for these two consecutive years and later 2002 dataset is also used to construct pseudo panel and estimation is re performed.

Variables included and generated through this dataset are hourly wage, size of the establishment, gender, age of the worker, status of management and non-management, region where establishment is located, professional distribution, type of employment contract, tenure, educational level and industrial sector. The construction of variables is the same except that age variable is taken as a whole[82] and educational categories are re-formed to gain more than 100 observations per cell. The analysis revealed the expected positive and significant impact of size on gross hourly wage and higher in the medium scale establishments. However, due to short time series (two to three years) it was not possible to construct an appropriate panel data set. The results are mixed for different estimation methods and across categories of different variables.

The remaining of this chapter is as follows. The next section defines the methodology and idea of pseudo panel estimation. In section 3, the construction of cohort datasets is explained in detail. It is followed by the estimations and results interpretation for different estimation methods of panel data in Section 4. The last section concludes.

6.2 Pseudo Panel Estimation

The basic wage equation is of the following form:

$$\boldsymbol{W_{ij} = \alpha + \beta_{ij} X_{ij} + u_{ij}} \qquad \textbf{(1)}$$

Where $\boldsymbol{W_{ij}}$ is the hourly earnings of worker i in establishment j, X includes three components, 1-the size of the establishment where worker i is working, 2-vector of worker's characteristics and 3-vector of characteristics of employer. $\boldsymbol{u_{ij}}$is an error term.

In order to evaluate the role of unobservable heterogeneity, we have to estimate the time variant individual specific effects and also firm effects[83]. This can be done with panel data but in the absence of availability of true panel data one can apply the estimation techniques to pseudo panel data.

By taking into account both time and individual effects the new equation is the following: (for simplicity, j subscript is excluded and individual and time dimension are written to construct pseudo panel):

[82]*In the previous chapters employees between age 25-60 are included in the sample, whereas, for this chapter no restriction is put on age.*

[83]*We cannot estimate firm fixed effects because the information to identify firms is not available in the data. So, we will only estimate individual fixed effects.*

Starting from the error component model[84]

$$\boldsymbol{W_{it} = \alpha + \beta_1 x_{it} + v_i + \epsilon_{it}} \qquad i= 1\ldots\ldots I,\ t= 1\ldots\ldots T \qquad \textbf{(2)}$$

Here W_{it} is the log of hourly wage of worker i at time t. v_i captures unobserved individual heterogeneity. This model assumes ϵ_{it} is uncorrelated with x and v. As the repeated observations of the same individual are not available as in the case of panel data, therefore, a cohort approach proposed by Deaton (1985) is applied with repeated cross sections in order to model fixed individual effects. A group of cohorts 'c' is defined based on variables that do not change over time (year of birth, gender, education, and region) and then all observations are aggregated in each cohort and in each time period. This serves as observations to construct pseudo panel data. Pseudo panel is equivalent to the least squares dummy variable (LSDV) model in genuine panels. Taking the mean value of each cohort's sample in each time period results in:

$$\boldsymbol{\overline{W}_{ct} = \beta\overline{X}_{ct} + \overline{v}_{ct} + \overline{\varepsilon}_{ct}} \qquad c=1,\ldots.,C\ t=1,\ldots.,T \qquad \textbf{(3)}$$

In equation (3), $\boldsymbol{\overline{W}_{ct}}$ is the mean of hourly wage (average of W_{it}) over a sample of observations in cohort c at time t. $\boldsymbol{\overline{v}_{ct}}$ as defined by Deaton (1985) is the "average of the fixed effects" for those individuals in cohort c at time t; but as the sample is collected individually at different times, $\boldsymbol{\overline{v}_{ct}}$ is not "constant over time and may be correlated with the explanatory variables (Devereux, 2007). This results in a potentially different $\boldsymbol{\overline{v}_{ct}}$ value for each period unlike v_i in true panel data. Verbeek and Nijman (1993) proposed that if the sample size in each cohort is sufficiently large (cell size greater than 100 observations) then $\boldsymbol{\overline{v}_{ct}}$ can be considered as unobserved cohort fixed effect. Verbeek and Nijman (1993) also point out that there is a trade-off between the number of observations in a cohort and the number of cohorts in a panel. Since a decrease in the number of observations in a panel implies an increase in the variance of the FE-estimator, this results in a tradeoff between bias and variance. With this solution, the equation (4)

[84] *This is one way error component model. We decompose the error in two terms: A person specific error v_i and an idiosyncratic error ϵ_{it},(person specific effects do not change over time while the other remaining unobserved effects (ϵ_{it}) change over both persons and time.*

$$u_{it} = v_i + \epsilon_{it}$$

presented below can be estimated by using cohort dummies or cohort fixed effects (Warunsiri & McNown 2010).

$$\overline{W}_{ct} = \beta\overline{X}_{ct} + \overline{v}_c + \overline{\varepsilon}_{ct} \qquad \textbf{(4)}$$

Given the proposition of large cell size, it is assumed that ($\overline{v}_{ct} = \overline{v}_c$) for every t and the fixed cohort effect ($\overline{v}_c$) is treated like a fixed individual effect (v_i).

Gardes et al. (2005) suggested that the individual effects $\overline{v}_c$, represent the aggregated individual specific effects as it includes the influence of unknown explanatory variables, constant through time for the reference group and also the individual specific effects containing effects of unknown explanatory variables. As the individual specific effects are eliminated by the within and first difference operators estimated with true panel data, similarly, the operators in pseudo panel data eliminates the aggregated individual specific effects.[85]

The aggregation in pseudo panel may create systematic heteroscedasticity. Gardes et al (2005) suggested a correction of heteroscedasticity that involves weighting each observation by a heteroscedasticity factor which is a function of, but not exactly equal to, cell size. Each aggregated observation is multiplied by the square root of mean number of observation for each given cell in each year[86]. This is not an exact correction, this is an approximated correction.

6.2.1 Dependent and Independent Variables

The variables constructed and used for the pseudo panel analysis are listed below:

- **Wage:** log of gross hourly wage is used as dependent variable in all the wage equations.

[85] *The aggregate individual specific effects depend on time. Therefore, parts of the aggregated individual specific effects are eliminated that do not depend on time and on the pseudo group selection based on homogeneity. This part decreases as cell homogeneity increases.*

[86] *We take the average of number of observations (cell count) in different years. Then we take its square root and multiply that factor with all the dependent ad independent variables in that row. As the cell size is the same for all variables in one row, therefore, the same heteroskadestic factor is multiplied with each variable in one row.*

- **Size**: The information of size of the employer is reported as categorical variable in the data; three categories of size are used including first 1-49, second 50-199 employees, and last 200 and more employees.
- **Education:** There are five categories of education: first:primary and lower secondary education, second: upper secondary general, third: upper secondary technical, fourth: university level 1st and 2nd stage and last university level 3rd stage /doctoral studies. This variable construction is different from the one used for propensity score analysis using the same survey because here the objective was to keep a sufficient number of cells (more than 100 observations).
- **Generation**: Five cohorts are formed from the information of age available in the data. Full information of the age variable is used and no observation is dropped as previously working age was defined from 25-60. Following five groups are constructed for the birth cohorts based on availability of data in three surveys, First, 1928-1954; Second, 1955-1962; Third, 1963-1969; Fourth, 1970-1975 and Fifth, 1976-1986.
- **Profession**: Four groups of professions are created: First: Management and High Intellectual professionals: second: High Skilled White Collar, Third: Low Skilled White Collar and last Blue collar.
- **Region**: Region represents the geographical location of the administrative region where establishment is based. There are eight categories against a variable of INSEE called ZEAT. The eight categories are the following : first : Ile de France , second : Bassin Parisien, third : North – Pas-de-Calais, fourth : East fifth : Ouest Sixth : South-west Seventh : centre East and last : Mediterranean.
- **Industry/ Sector**: Three categories are formed for industry: first: manufacturing, second: Trade and last Services.
- **Type of contract**: three dummies, first permanent contracts called CDI; second fixed contracts called CDD and last other type of contracts.
- **Tenure** is defined as the number of years in current employment, dummy for **management**, dummy for **female**.

6.3 Construction of Pseudo Panel Dataset

There are plenty of different types of methods to be used for building pseudo panel cells. But most popular method, defined by Deaton (1985) and discussed by others, is used for this chapter. This is a usual method which is well known from a statistical point of view. There are other methods such as matching subsets based on neural network method or using individual matching to define the cells. For this chapter, we have only employed the Deaton's method for building pseudo panel cells, in the next step, different methods will be compared.

Deaton's method of building cells is criticized in recent work. As individuals grouped in a cohort are not the same for successive periods, and it results in a measurement error and inconsistent estimators. Even the number of individual changes in each survey, so the "fixed" effect obtained also changes with time for the same cohort. Firstly, measurement errors cause an inconsistency of the estimators and secondly, loss of efficiency is due to the grouping of individuals. The construction of a pseudo panel from repeated cross-sections using a neural network like the self-organizing map appears to be a means to overcome the major drawbacks attached to the classical pseudo panels (Cottrell and Gaubert (2003)[87].

In the neural network method, cells are groupedby multi-dimensional grouping methods. This technique is used for automatic grouping method which is powerful for building groups of similar individuals. In that case we do not have problem for choosing variables which may be used for pseudo penalization. This method uses multiple matching instead of individual matching. Matching one individual to other individual is more powerful than any linear method. The method of building cells is the Kohonen map, which a Self-Organizing Map or an ordered mapping[88]. This method constructs pseudo panel from repeated cross-sections using a neural network like the self-organizing map.This method constructs cohorts using factors that are stable over time in order to link reasonably the successive observations of each cohort. This method

[87] *Cottrell and Gaubert (2003) Efficient estimators: the use of neural networks to construct pseudo panels (hal-00122817, version 1 - 5 Jan 2007).*

[88] *The Self-Organizing Map (SOM), commonly also known as Kohonen network (Kohonen 1982, Kohonen 2001) is a computational method for the visualization and analysis of high-dimensional data, especially experimentally acquired information.*

includes large number of individuals to allow the use of asymptotic reasoning on the obtained estimators.

Based on the Deaton (1985) method, the pseudo panel cells are constructed by taking different groups of variables that do not change over time. This chapter utilizes the available information in the data and the time invariant variables that include: (1) gender, (2) generation (birth cohort), (3) education, (4) region where establishment is based etc. These four variables are used in different combinations to construct pseudo panel. Generally four different data sets are generated,

- First by gender, education and generation; (GEG)
- Secondly by gender, region and generation (GRG)
- Thirdly, by region and generation (RG)
- Fourth by gender and generation (GG)

Cohort stability is very important, therefore, cohort is defined in four different ways and estimation is re performed in each way. People born between 1928-1986 were included in the sample. There are two consecutive cross sections available for 2005 and 2006. Firstly a pseudo panel is constructed for consecutive years and later 2002 survey is utilized to build a three years pseudo panel dataset. Each cell represents more than 100 observations in each cohort (Verbeek & Nijman 1993, Gardes et al (2005)).

The number of cells in each pseudo panel data is the following:

- In the first pseudo dataset, the information is aggregated on gender, education and generation (cohort). Two groups of gender (male and female), five levels of education and five categories of birth cohort generates 50 cells (2*5*5) and as data is for two years (50*2) so we get the final data with 100 observations in the first pseudo panel.
- Similarly for second type of pseudo panel (gender, region and generation (GRG), we get 160 observations as regions are divided into eight categories: 160 observations = gender (2)* region (8)*generation (5) *2(years)

- For region and generation (RG): 80 observations: region (8)*generation (5)=40*2(years)= 80
- gender and generation (GG): 20 observations: gender (2)* generation (5)=10*2(years)= 20
- For three years survey, the number of observations becomes 150(GEG), 240(GRG), 120(RG) 60(GG) respectively.

The dependent and independent variables in the data are proportional variables. The cells show the proportion of employees for combined individual characteristics (cohort). Each observation is the mean of combined characteristics and help to estimate the aggregated effect of control variables on the hourly wage.

The variables on which the data is aggregated are not used in the model if they are used for aggregation. As the gender dummy is not used in the first and second group but included in the third group. Similarly, education variable is used as explanatory variable in the rest of cohort combination. There is a variable on the age of the worker which is used for making generation variable (birth year) but this variable is not included in the regressions because for the within estimation, between estimation and first difference estimation, it cannot be used to take into account age characteristics that do not change over time because the data is only available for two years. Therefore, the age variable is not used as explanatory variable but it is used to construct generation variable. One other reason of not using age variable in the estimation because of potential duplication of this information as generation is included in all pseudo panel data cells variations.

There may be an identification problem caused by difference in cohorts, difference in age and difference in periods. As suggested by Russell & Fraas (2005) all these effects are not simultaneously identified because at one time, only one dimension of time, individual and cohort exists. A linear restriction is imposed that all effects are included in the constant term.

It is very important to take into account the unobserved factors that cause potential endogeneity in the size and wage relationship. Fixed Effects (FE) are important to remove omitted variable bias because FE partially removes endogeneity by demeaning the data.

Therefore, fixed effects estimation method is the most relevant to study the size and wage relationship.

Results are presented in Table 6.1 and Table 6.2 in Appendix-E. Table 6.1 compares results by the choice of cohorts to build pseudo panel and Table-6.2 compare results by the estimation method. Results in table-6.1 and table-6.2 are the same. Table 6.1 shows the comparison of four estimation methodsand Table-6.2 depicts comparison of three ways of building cohorts. Below, each estimation method is briefly described and then, results are interpreted. We start with fixed effects/within transformation model and later we will move to other estimation methods for comparison and choosing the right method.

Results are calculated for different estimation methods including random effects, fixed effects, between effects and first difference. As stated above four different group of cohorts are generated but here in this section results for the first three are presented because with generation and gender the number of id were 20 and it was not possible to regress the explanatory variables other than size on wage.

6.4 Results and Interpretations

6.4.1 Fixed effect model/the within transformation model

Starting from error component model[89],[90]

$W_{it} = \beta_1 x_{it} + v_i + \epsilon_{it}$ (i)

After averaging this equation over time for each i (between transformation)

$\overline{W_{\iota}}= \beta_1 \overline{x_{\iota}} + v_i + \overline{\epsilon_{\iota}}$ (ii)

After subtracting the second equation from the first for each t (within transformation)

$W_{it} - \overline{W_{\iota}} = \beta_1 (x_{it} - \overline{x_{\iota}}) + \epsilon_{it} - \overline{\epsilon_{\iota}}$ (iii)

[89] *We decompose the error in two terms: A person specific error v_i and an idiosyncratic error ϵ_{it},*

$$u_{it} = v_i + \epsilon_{it}$$

[90] *The constant is omitted from the equation because it would be collinear with vi.*

This is called fixed effect (FE) estimator. In equation iii, v_i disappeared because we no longer need the assumption that v_i is uncorrelated with x_{it} . Time-constant unobserved heterogeneity is no longer a problem. Here we time demean the data. After subtracting the between variation, we are left with within variation. Therefore, this estimator is also called the within estimator.[91] The FE estimator assumes that cov (x_{it}, ϵ_{it}) =0

In the case of pseudo panel dataset, the model will take this form,

$$\overline{\boldsymbol{W}}_{\boldsymbol{ct}} = \boldsymbol{\beta}\overline{\boldsymbol{X}}_{\boldsymbol{ct}} + \overline{\boldsymbol{v}}_{\boldsymbol{c}} + \overline{\boldsymbol{\varepsilon}}_{\boldsymbol{ct}} \qquad \textbf{(5)}$$

This equation is identical to equation 4. $\overline{X}_{ct}$ includes set of C cohort. The difference between fixed effect and random effect is that RE assumes that cohort components are IID and are simply included in the$\overline{\boldsymbol{v}}_{\boldsymbol{c}}$ error term whereas, fixed effects assumes that cohort components are fixed across time and are significantly different. Thus, the type of the cohort determines the inference from fixed effect model.

The FE model eliminates $v_i$by demeaning the variables using the *within* transformation. With pseudo panel data the within transformation controls for cohort fixed effects by calculating each variable's mean value across time for each cohort, then subtracting that mean from all observations. The within transformation will follow the same formulations as in i-iii above, we have to calculate the time mean values from the cohort mean values.

$$\overline{\boldsymbol{W}}_{\boldsymbol{c}} = \boldsymbol{\alpha} + \boldsymbol{\beta}\overline{\boldsymbol{X}}_{\boldsymbol{c}} + \boldsymbol{v}_{\boldsymbol{c}} + \overline{\boldsymbol{\varepsilon}}_{\boldsymbol{c}} \qquad \textbf{(6)}$$

This is identical to equation 4 except the *t* subscript has been removed where a mean value across time as well as across cohorts is calculated. The $\boldsymbol{v}_{\boldsymbol{c}}$error term represents the unobserved fixed cohort effect and consequently is unchanged between Equations 4 and 6. The within transformation is obtained by subtracting Equation 6 from Equation 4 as follows:

[91] *It examines at how changes in the explanatory variables cause dependent variable to vary around a mean within the unit. It lets to estimate the changes in the variables over time to estimate the effects of the independent variables on the dependent variable. It is used to control for omitted variables that differ between cases but are constant over time.*

$$\boldsymbol{W_{ct} - \overline{W_c} = \alpha - \alpha + \beta\,(x_{ct} - \overline{x_c}) + v_c - v_c + \epsilon_{ct} - \bar{\epsilon}_c} \qquad \textbf{(7)}$$

The intercept term (α), as well as the cohort fixed effect (v_c), do not change over time. Consequently, they are already time means by definition. If $v_c$is assumed to sum to 0 across all cohorts, the within transformation is estimated as follows (Baltagi, 1995):

$$\boldsymbol{W_{ct} - \overline{W_c} = \beta\,(x_{ct} - \overline{x_c}) + (\epsilon_{ct} - \bar{\epsilon}_c)} \qquad \textbf{(8)}$$

The results for the fixed effect model are presented in *Table-6.1 in Appendix-E*. Thedependent variable is the log of hourly wage (averaged on the cohort combinations). The first variable of interest on the right hand side of the wage equation is the size of employer. There are three categories of employer (small, medium and large). The small category is kept as reference. The explanatory variables include; Industrial sector (three sectors, manufacturing, trade and industry, base category is manufacturing), region (eight categories, base category is ile de France), profession (four categories, base category is management and high intellectual professionals), dummy for status (management or non-management), dummy for gender and tenure.

Table-6.1- 6.2 presents the result of three different aggregated dataset; each represents pseudo dataset on different set of time invariant individual characteristics as explained in the previous section. As noted above, Table 6.1 shows the comparison of four estimation methodsand Table-6.2 depicts comparison of three ways of building cohorts.Table-6.1, first column of FE, shows that the effect of size on wage is largest in the medium size of employers.[92] For large it is positive and significant but coefficient is low compare to medium scale establishments.

In the first type of pseudo panel (gender_education_generation) employee characteristics variables are significant but the sign of region is not correct. Moreover, the large establishment size dummy is only significant in FE. In the pseudo panel with cohort (gender_region_generation), the size-wage effect is significant in all cases and the difference in

[92]*Results are computed using 'xtreg' command in stata. As the panel data is formed only for two years, therefore, we don't expect problems of autocorrelation and heteroskedasticity. For FE 'vce robust' command is used along with 'xtreg'.*

medium and large is low. Most of the control variables are significant. In the pseudo of region_generation the significance of most of the variables is low. In FE the medium size group is not significant.

So, looking at the results of fixed effects the preferred pseudo panel cohort combination is the second one (gender region generation). On the whole in the fixed effects with the variation of employees and employers characteristics, we see that the effect of size on wage is increasing but higher is higher in the medium size group. Thus, with the aggregated data, it was possible to compute the fixed effects.

The high magnitude of FE may represent the effect of latent variables as OLS does not capture the unobserved heterogeneity and this is reflected in the FE coefficients. FE increases the magnitude of impact because other methods do not truly capture the unobserved heterogeneity. The individual heterogeneity or the effect of latent variables is powerful which an unbiased effect is. But there is not so much variation in time series and cross section. More years are needed to make results intuitive.

6.4.2 The Random Effects Model

The random effect (RE) assumes that $\boldsymbol{\overline{v}_c}$ (equation3) which represents possible bias from unobserved, fixed cohort heterogeneity, is identically and independently distributed (IID) with a mean of zero and cov (x_{it}, v_i) = 0 (Baltagi, 1995, Wooldridge 2002). For most research problems there is a possibility that this variance is not equal to zero. In this case RE estimator will be biased.

In estimating the RE Model, the same variables are included as used for fixed effects estimation to represent the characteristics of workers and employers. All the cells of explanatory variables and dependent variables show the proportion of employees in that cell aggregated on the chosen cohort.

Table-6.1 second column shows RE estimates for the size-wage effect. We see the significant random effects in most of the cohort cases and larger magnitudes compare to FE. In the second and third pseudo panel group, we see most of the variables related to employer and employee characteristics are significant and in the right direction compare to the first group of pseudo panel. The magnitude of the size-wage effect is high compare to FE and the direction of the effect is the same.

The use of the random effects model relies on the assumption that significant fixed effects do not exist[93]. But, as we saw from Table-6.1 that significant fixed effects exist for the size-wage effect.[94] The high significance in the RE shows that the estimates are biased and exaggerate the impacts of explanatory variables on the wage.

The Hausman test is generally used after RE to see the difference between two estimators. The null hypothesis is that both estimation methods can be used and, therefore, it should yield coefficients that are "similar". The alternative hypothesis is that the FE estimation is preferred over RE, therefore, it should yield coefficients that are "different". The bigger the difference (the less similar are the two sets of coefficients), the bigger the Hausman statistic. A large and significant Hausman statistic means a large and significant difference, and one can reject the null that the two methods are good but the preferred estimation is the FE.

The Hausman stats are shown in the Tables 6.9 for each of the cohort pseudo data and for different variations of the size variable. The test stats are very high which suggests that fixed effects estimator should be used. So, we will reject random effects as inconsistent and rely on fixed effects instead. Hausman test is conducted separately for employer size and than by adding covariates.

In the next step the First-Differenced Model, along with the Between Transformation Model is estimated to provide additional insight into the core results of the FE model and RE model.

6.4.3 Data Transformation Models: between effects and first difference estimation

A. Between Effects Model

[93] *If significant fixed effects exit, the Random Effects Model cannot be used. One alternative to using the Random Effects Model is to use the Fixed Effects Model.*

[94] *Stata's random-effects estimator is a weighted average of fixed and between effects.*

The rationale behind using Between Effects (BE) estimator is that it averages observations over a unit and regresses average dependent variable on average independent variable to look at differences across units. Regression with between effects model is used to control for omitted variables that change over time but are constant between cases[95]. It allows using the variation between cases to estimate the effect of the omitted independent variables on the dependent variable. The between effects estimator is mostly important because it gives insight to interpret random effects estimator.

Results using BE estimation are computed using equation 6. The results of between effects are highly insignificant. The size coefficient is not significant is most of the cases. The results for the second group of pseudo panel are comparatively better for the size-wage effect.

The results are unusual because BE are very often biased by permanent latent variables. The usual case is when there is difference between cross section and time serious estimates as in this case, which means we have permanent latent variables which are eliminated in the time series but which bias the cross section estimates. Thus, when we have biased estimates of BE, then it is due to latent variables.

B. The First-Differenced Model.

First-Differenced estimator is not used for the between person comparisons. It is used for only within person changes. For testing hypothesis involving a trend the FD estimator can be used. First-differenced data are obtained by subtracting each cohort's variable values from the prior year's values as follows:

$$W_{ct} - W_{c,t-1} = \alpha - \alpha + \beta\left(x_{ct} - x_{c,t-1}\right) + v_c - v_c + \epsilon_{ct} - \epsilon_{c,t-1}$$

The First Differenced model is estimated as follows:

$$W_{ct} - W_{c,t-1} = \beta\left(x_{ct} - x_{c,t-1}\right) + \epsilon_{ct} - \epsilon_{c,t-1}$$

[95] *BE estimator does not eliminate unobserved heterogeneity but it isolates unobserved heterogeneity.*

The first differenced results are in line and comparable with FE but the significance level is slightly low compare to FE and the magnitudes of the explanatory variables is also low.

6.4.4 Results for three years (2002-2006)

Now the same estimations are re performed as done in the above section. The only difference is that now, pseudo panel data is generated for three years i. e. 2002, 2005 and 2006. Results for this section are presented in ***Tables 6.3-6.4***. A pseudo panel with one more year (with gap) improves the significant and magnitudes of the various effects. The direction remains the same with higher effect on wage in the medium size group compare to large group. In the three years case the choice of preferred cohort combination is one of the first or second as in the second group FD results for size variable are not significant while in first pseudo panel the significance of other control variables is low. On the whole the preferred method is FE and preferred pseudo is the second one.

If we compare the magnitude of the size wage impact (aggregated cohort effect of the explanatory variables on hourly wage) in three different pseudo panel data sets than the results from fixed effects are realistic compare to other effects as RE and BE are biased. FD results are lower than FE but in the right direction. Using two years cross sections, being in the medium size increases earnings 64% compare to small and by 56% in large more compare to small. In the first different estimates the change from small to medium is 42% and from small to large is 37%. This shows that earnings are higher in medium compare to large. In three years data results are in the same direction. The fixed effects show 78% higher wage in medium compare to small, and 37% higher wages in large compare to small. In FD the difference between medium and large is low compare to FE. It varies from small to medium about 28% and from small to large around 26%. Thus, overall, the results when gender-generation-region is used to aggregate the data are more rationale compare to other combinations of invariant individual characteristics.

6.4.5 Robustness Checks

The information about size of employer is available as categorical variable in the data. In order to check the elasticity of wage by size, two different variations of size are produced and all the results are re-estimated and compared by different estimation methods. Results are computed with two different variations of size variable other than categories.

1. In the first case, the size variable is modified and now the proportion of each size category in the size variable as a whole is calculated and used as log of size instead of categories[96]. We create a normal curve of the proportion of each size category among all categories and then use those proportions as continuous values. To make it clearer,the probability of each establishment of falling into a particular size category is computed through normal distribution and then it is multiplied by 100. Results corresponding to this method are presented in the preceding section.
2. In the second case, the size variable is produced by simply taking mean value of each size category and then it is used as continuous variable in the data and all results are re-performed[97].

A. Size as continuous variable- case-1 (proportion for size category)

In this section, results are presented by the same method and estimation techniques. The only difference is that size variable is modified and now the proportion of each size category in the size variable as a whole is calculated and used as log of size instead of categories. Results are calculated in same lines as in the previous section.

Here the size is used as log, so the coefficient will show the elasticity of wage by size. The coefficients show the proportion of the combined elasticity for the cohort. In all the different cohort variations, the effect of size on wage remains significant and varies between 0.3 to 0.9 depending on the cohort combination and on the choice of explanatory variables.

[96]*We have information on size as categorical variable in the data. There are eleven categories of size. Each category shows a range of employees (for example the first category includes all establishments with 1-9 workers). We have generated another size variable where first average of each category is calculated (it is named as x) and then the mean of x (xbar) is subtracted by each value of x (x-xbar). This further is divided by its standard deviation. And the resulting value is checked in the Student's t-distribution table and new list is created equal to the number of categories (11categories). This is then used as continuous variable in the data. Reference: Theil, H. A rank-invariant method of linear and polynomial regression analysis. I.*Nederl. Akad. Wetensch., Proc.***53,** *(1950) 386–392 = Indagationes Math. 12, 85–91 (1950).*

[97]T*he first category includes all establishments with 1-9 workers. We take average of each category (the first category average is 5) and then it is used as continuous variable in the data as log of size.*

The results corresponding to first three pseudo panel are presented in the *Tables-6.5-6.6*. When size is used as continuous variable by taking the proportion of each size category than significance level of most of the variables is low compare to using it as categorical variable. FE magnitude of size wage impact is low compare to other effects and the second cohort combination is better than the other two groups of pseudo panel.

When size is used as continuous variables in the first case the elasticity of wage by size varies from 0.4 to 0.9 in different estimation methods. Taking the second group of pseudo panel data set, if we double the size, wage will increases by 4% in FE, 6% in RE, 9% in BE and 7% in FD. In the first pseudo panel the elasticity remain at 5% in first three estimation methods compare to 9% in FD. In the third case of pseudo panel the BE show highest elasticity of 0.97.

If we compare the results of using size as continuous variable with the previous section where size was used as categorical variable then we see that the results of other explanatory variables other than size are quite mix. In the first pseudo panel the magnitude of explanatory variables is mix, sometime high, while other time low or approximately similar. The signs are almost the same but for some controls as in regional categories signs are different. Fixed effects are significant for profession, sector and tenure.

The same conclusion is drawn from this analysis as in the previous case. The preferred estimation is FE, and the preferred cohorts are first and second. The variable on education is highly insignificant in this case compare to the previous case. The BE show large effect of size on wage which as stated above may be caused by permanent latent variables. Therefore, we cannot trust the results of BE.

When size is further modified by taking average of each category than the magnitude is low compare to using it s proportion of size category. This is explained below.

B. Size as continuous variable- case-2 (Average of size)

In the second case, size is again used as a continuous variable. The mean value of each category of size is computed and then used as a continuous variable.

Now the same four types of pseudo panel datasets have been regenerated by combining different variables that don't change over time. The difference is that now we use log of size as explanatory variable instead of categories of size. The objective is to estimate the elasticity of wage by size. Results for different estimation methods are presented below in ***Tables 6.7-6.8.***

Log of size is positive and significant in all the estimation methods of panel data. The significance of explanatory variables is low in all cases compare to Table 6.1. Keeping in view the results from this estimation method, it may be concluded that the preferred type of pseudo panel data is the second one and preferred estimation method is FE.

In the second case where size is used as continuous variable, the magnitude of the size wage impact varies from -0.1 to 0.4 which is very low compare to the case-1. So overall, when size is used as continuous variable than in the FE doubling the size will increase wages by 4% (0.37) in the first case and by 1% (0.148) in the second case. Overall results are better by using size as categorical variable.

6.5 The preferred Estimation method and preferred cohort

In the presence of selection bias, the size-wage relationship cannot be explained as many characteristics of employer and employees are unobserved. The available dataset does not give relevant information to control for unobserved characteristics. Unobserved factors are potential source of endogeneity. Unobserved heterogeneity in the size-wage effect needs to be removed. To study the employer-employee relationship, FE estimation is the most relevant because FE partially removes endogeneity by demeaning the data but it does not remove the endogeneity biased caused through selection bias. It only captures and reduces given the observed characteristics of employer and employee. Moreover, the numbers of years are too few to make some inference from the results.

FE methods are important to remove omitted variable bias. That is why FE are preferred over other estimation methods. RE are more significant but they are biased. RE exaggerates the true effect. Thus, the coefficients do not show the true effect. BE estimation results are biased because of permanent latent variables. FD are close to FE as the data is for two years. Second proffered estimation method is FD.

The preferred cohort combination is the second one 'Gender_region_generation'. Second preference is for the first one. Both of these combinations show comparatively logical results compare to the third.

6.6 Conclusion

This chapter demonstrates how pseudo panel data can be constructed to address the lack of a true panel data. The cohort analysis is presented to study the causal relationship between size and wage. Generating an aggregated panel data set based on some constant individual or employer related characteristics show some interesting findings. Four different types of aggregated data sets are generated based on individual characteristics that do not change over time. Results in each type of data are different from the other as expected.

The findings consistently support the hypothesis that within a cohort, there is positive and significant effect of employer size on wage. The impact is strong in the medium scale establishments. As the size of employer increases, wages increase and this effect remains significant even after controlling employer and employee characteristics.

A comparison between the random effect model and the fixed effects model revealed that the fixed effects were statistically significant. Consequently, the FE Model formed the core of our analysis. Additional insights were provided by the results produced by the between transformation, and first-differenced models, which used transformed data sets. The random effects, fixed effects, between transformation models and FD all showed a significant and positive effect of size on wage.

By using any of the four estimation methods, it cannot be said that this method eliminate bias from unobserved cohort heterogeneity that changes over time. Different cohort data were generated and results were compared. The preferred pseudo panel data is the combination of gender, region and generation. The significance, direction and sign of the variables are correct. The preferred estimation method is FE. This corresponds to the fact that FE captures heterogeneity by demeaning the data. As unobserved heterogeneity is the principle source of selection bias. The RE and BE results cannot be trusted. They exaggerate the true effect and are biased. BE are biased due to permanent latent variables.

In order to perform better estimation, the cell number needs to be increased by having more repeated cross section with less distant gaps. Further, it would also be better to construct pseudo cells by different methods, for instance, neural network method for defining more homogeneous cells.

GENERAL CONCLUSION

"I may not have gone where I intended to go, but I think I have ended up where I needed to be."Douglas Adams

This dissertation aims to explore why large employers pay higher wages compared to similar workers in smaller size establishments in the presence of nonrandom assignment of workers. The main focus is to determine the magnitude of the size-wage impact in the presence of selection and to test various methods to explain employer size and wage relationship. The objectives of the dissertation are multifold. One chapter leaves us with some questions and that question becomes motivation for the next chapter. All the chapters aim to explore the employer size and wage relationship in the presence of selection bias. This is not a new question in labor economics. The econometric methods are not new either. But this is an addition to the French labor market in this literature. Moreover, the methods applied on this question have not been used,to our knowledge, before, for example, PSM, FIML, and pseudo panel etc. The institutional elements in the French labor market show that employer size is definitely an important element for policies related to employment, wage structure and for work councils. The conclusions are relevant in this literature of size-wage gap for France, first, there are few studies on the question of employer size and wage and second, the sample selection bias has been overlooked in the studies using cross sectional data. Therefore, in France, the most crucial issue is to find a good balance between the respect of the autonomy of labor and management representatives and the intervention of the

State to compensate the defects of such autonomy.[98] The dissertation is not aimed to formulate or to suggest policy changes in the employer employee relationship in France but to study the dynamics of this relationship that can help to redirect policy.

There are six main chapters among which four present results of the research work. The first chapter reviews the literature on the relationship between employer size and wages. The second chapter explores the data where descriptive statistics for four cross sections (1992, 2002, 2005, and 2006) along with descriptive statistics across gender are presented. In the third chapter, the overall wage differential is identified through OLS and propensity score matching method where employees in the big size of employers are matched with those working in small size establishments based on observable characteristics and the wage differential is analyzed. This chapter also deals with the relative importance of bonuses, overtime payments, and allowances through comparing gross and basic hourly wage. Third chapter leaves us with the problem of unsolved selection bias or nonrandom sorting of workers. Therefore, in the fourth chapter, full information maximum likelihood model is used.In this model nonrandom selection of workers is studied and the magnitude of size-wage premium is analyzed. Two equations models, selection model and wage model, are jointly analyzed controlling for maximum available information. The instruments for exclusion restrictions are utilized to predict the size of establishment. Two types of instruments: number of dependent children less than eighteen years of age and interaction of worker industry and region size where establishment is based, are utilized. Further, rewards for measured and unmeasured skills by size group are studied by using endogenous switching regression method where inverse mills ratio are computed through Heckman two step estimation procedure. Further, it is answered in the fifth chapter to the question whether the size wage differential is actually a gender wage differential and the male and female wage outcomes are compared by different size categories. Finally, cohort analysis is presented in chapter six using the latest successive cross section datasets to estimate fixed effects. The fixed effects are compared with first difference, random effects and data transformation models.

It is observed that the employer size-wage premium still prevails with different sources of data, different geographical locations and different workforce composition structure. Different

[98] *Eric Labaye, Charles Roxburgh, Clarisse Magnin, and Jan Mischke march 2012, French employment 2020: Five priorities for action, The McKinsey Global Institute (MGI)*

theoretical explanations are valid for different countries depending upon local labour market condition. Among all these studies, in general, no attention has been paid to the endogeneity problem that is to the fact that there is a correlation between the error term and the explanatory variable. There are few studies that used longitudinal data to take into account the employer and worker heterogeneity, but on the other hand, majority of cross sectional studies ignore the selection bias problem.The main unobserved factors affecting the size-wage relationship are selectivity factors that result in nonrandom allocation of workers across employer size. This in turn, makes size an endogenous variable in the wage equation and failure to take into account this fact of nonrandom selection biases the parameters. Two things are important to deal with endogeneity; first, to know the extent of problem and second to remove it. We have to clarify which causality is more important in this relationship, causality on time or causality on cross section. This work deals with causality on cross section as the data is not sufficient to cover causality on time series. Different econometric methods are used to obtain more information on causality. The matching method identifies the extent of the problem. Using selection models and wage decompositions, we are able to study the different behaviors of employers of different size. We have been able to know how small, medium and large size employers reward workers' observed and unobserved characteristics. Further, the magnitude of the size effect in the presence of selection bias and various components of the wage differential are identified. The causal inference depends on the choice of the selection model and on the choice of instruments. The results can be improved by using more relevant and valid instruments. Although endogeneity is not completely removed, it is clearly identified and methods to deal with it are highlighted.

The principle result follows: size does matter for French labor market. There is positive and significant relationship between size of employer and wages of employees. This is evident by all methods. The comparison of the two measures of wages shows a clear difference in the wage premium. Computation of results by distinguishing the two measures of wage tells us how the wage components are related to employer size. The wage premium is strongly associated with the compensation structure of large employers. The magnitude of the size wage impact becomes negligible when using basic (net) hourly wage. Gross earnings are related to profits which reflect that large employers earn more profits compared to small employers and the profits are shared with their workers. It is observed that the employer size –wage difference is more of a difference of compensation policy of employers.

The discussion on the effect of size on wage is incomplete without considering the nonrandom allocation of workers. Therefore, the selection methods were applied to compare the returns of observable and unobservable characteristics of workers in different size of establishments. The negative correlation coefficient and negative inverse mills ratios (IMR) are found using FIML and Heckman procedures showing that unobservable in the selection equation and wage equation are negatively correlated with each other. Thus, the unobserved factors have high reward in small size establishment while observed factors have high rewards in large size establishment. The average skill requirements in the large establishments are low compared to small establishments. Results support that large employers choose better workers and pay them higher wages based on observable characteristics. Similarly after controlling for observed and unobserved characteristics, the wage premium still exists and is stronger for male workers. This may be result of employer heterogeneity. As suggested by efficiency wage models, if firms of different sizes differ in their ease of monitoring workers, in their training costs, or in their reliance on teamwork, they may find it profitable to pay differing wages to identical workers (Morissette (1993). The gender wage gap is greater in large compared to small size establishments but among all of the cases the larger part of the gender wage gap remains unexplained. At large firms, wages are usually negotiated between the firm and a union, with the union attempting to apply pressure on the firm in order to gain higher wages but the conclusion does not support the union threat hypothesis as the information available on unionization is not sufficient to see the impact on employer's wage structure.

PSM and OLS are equivalent but deal with observables and do not solve the problem of selection bias. As Heckman, Smith and Clements (1997) pointed out that PSM cannot answer questions relating to the distributional effects of the program. PSM can only recover mean effects. Among all methods, selection models are superior to predict the causal relationship between employer size and wage if good instruments are available. For the heterogeneous treatment effectsthe selection models are superior to 2SLS and GMM. Generally FIML is considered more efficient compared to Heckman but the switching regression models have been popular in the employer size wage gap literature dealing with selection that allow us to observe wages in different regimes with selection effect. Therefore, both models are equivalent and may be used side by side especially in applied work. Pseudo panel data solves the problem of unavailability of true panel data but using only few years' cross section this does not give sound results. There are other methods of building pseudo panel cells. Results can be more intuitive if we compare different methods for building cohort cells.

As mentioned above, it is difficult to say that all the objectives are achieved, but many new questions are opened up and many advancements and extensions are possible to improve the quality of the present work and for further research. The selection models are most relevant for this type of question when working with cross sectional data but it is very important to get good instruments to predict the selection model. More instruments need to be searched. Multinomial selection models can be studied in detail (Bourguignon et al. 2007) to improve the results. It will also be interesting to exploit the detailed information on compensation structure elements in the big size employers and to clearly identify which of the fringe benefit is more size sensitive and to differentiate between the voluntary and involuntary nature of supplementary working hours. For matching methods, further work can be done using multi-treatment in the propensity score matching. The neural network method can be applied to build homogeneous cells and results can be compared with Deaton's method of building Pseudo panel cells.

Bibliography

Abadie, A., D. Drukker, J. Leber Herr, and G. W. Imbens, 2004. “Implementing Matching Estimators for Average Treatment Effects in STATA," *The Stata Journal*, 4(3), 290-311.

Abowd, J. M., F. Kramarz, and D. N. Margolis, 1999.“High wage workers and high wage firms,” *Econometrica*,67:251–334.

Abowd, J. M., F. Kramarz and S. Roux, 2006. “Wages, mobility, and firm performance: Advantages and insights from using matched worker-firm data”, *Economic Journal* ,116:245–85.

Abowd, J. M., F. Kramarz, P. Lengermann and S. Roux, 2005. “Persistent inter-industry wage differences: Rent sharing and opportunity costs”, *Center for Research in Economics and Statistics (CREST),* Working Paper.

Abowd, John M., Kramarz Francis, Margolis David N., and Troske, Kenneth R., 2001. "The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States," *Journal of the Japanese and International Economies*, Elsevier, vol. 15(4), pages 419-436, December.

Albrecht, J., A. van Vuuren and S. Vroman, 2009. “Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustments”, *Labour Economics*, 16(4), 383-396.

Albrecht, J., Bjorklund, A., and Vroman, S., 2003. “Is there a glass ceiling in Sweden?” *Journal of Labor Economics*, vol. 21, no. 1, 145-177.

Alchian, A.A. and H. Demsetz, 1972. "Production, information cost, and economic organization", *American Economic Review*, 62: 777-795.

Alejandro Badel and Ximena Peña, 2010. "Decomposing the Gender Wage Gap with Sample Selection Adjustment: Evidence from Colombia," *Economic Analysis Review*, vol. 25(2), pages 169-191.

Angrist, J.D., G.W. Imbens and D.B. Rubin, 1996.“Identification of Causal Effects Using Instrumental Variables”, *Journal of the American Statistical Association*, 91, 444-472.

Angrist, J.D. and J.S. Pischke, “Mostly Harmless Econometrics: An Empiricist’s Companion,” *Princeton University Press*, 2009.

Anja Heinze A., 2010. “Beyond the Mean Gender Wage Gap: Decomposition of Differences in Wage Distributions Using Quantile Regression”, *ZEW research Discussion Paper*, No. 10-043.

Badel, A., and Peña, X., 2009. “Decomposing the gender wage gap with sample selection adjustment: evidence from Colombia”, Bogotá, Colombia: Universidad de las Andes. Mimeographed document. Available at: http://ximena. pena. googlepages. com/gender. pdf.

Baltagi, B. H., “Econometric Analysis of Panel Data”, *John Wiley & Sons*, New York, 1995.

Bargain, Olivier and Melly, Blaise, 2008. "Public Sector Pay Gap in France: New Evidence Using Panel Data," *IZA Discussion Papers* 3427.

Bayard Kimberly and Troske R. Kenneth, 1999. “Examining the Employer-Size Wage Premium in the Manufacturing, Retail Trade, and Service Industries Using Employer–Employee Matched Data,” *AEA Papers and Proceedings*, VOL. 89 NO. 2.

Bayard, K., Hellerstein, J., Neumark, D. and K. Troske, 2003. “New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data,” *Journal of Labor Economics*, 21(4), 887-922.

Becker, G. S., 1975. “Human capital: A theoretical and empirical analysis, with special reference to education”, *University of Chicago Press*.

Belfield, C. R., and Wei, X., 2004.“Employer size–wage effects: evidence from matched employer–employee survey data in the UK”, *Applied Economics*, 36(3), 185-193.

Berk Richard A. 1983.“An Introduction to Sample Selection Bias in Sociological Data” *American Sociological Review*, Vol. 48, No. 3 (Jun., 1983), pp. 386-398.

Behrman, Jere R., Yingmei Cheng and Petra Todd, 2004. “Evaluating Preschool Programs when Length of Exposure to the Program Varies: A Nonparametric Approach,” *Review of Economics and Statistics*, 86:1 (February 2004), 108-132.

Blanchard, P., J.-P.Huiban and C. Mathieu, 2010. "Do sunk costs matter in exiting? An empirical investigation based on French firms", Micro-Dyn Working Paper n°20/10.

Blau, F.D.; Kahn, L.M., 1992. “The Gender Earnings Gap: Learning from International Comparisons”, *American Economic Review*, vol. 82, pp. 533-38.

------------------ 1996. “Wage Structure and Gender Earnings Differentials: an International Comparison”, *Economica*, 63.

------------------ 2003. “Understanding International Differences in the Gender Pay Gap”, *Journal of Labor Economics*, vol. 21, nº 1.

Blinder, A. S., 1973. “Wage Discrimination: Reduced Form and Structural Estimates,” *The Journal of Human Resources*, 8: 436–455.

Borjas, George J, 1987. "Self-Selection and the Earnings of Immigrants," *American Economic Review, American Economic Association*, vol. 77(4), pages 531-53, September.

Bourguignon, F., Fournier, M., & Gurgand, M., 2007. “Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons”, *Journal of Economic Surveys*, 21(1), 174-205.

Brown, C., and Medoff, J., 1989. “The employer size wage effect,” *Journal of Political Economy*,University of Chicago Press, vol. 97(5), pages 1027-59, October.

Brown, C., 1980."Equalizing differences in the labor market", *Quarterly Journal of Economics.*

Brown, C., Hamilton, J. and Medoff, J., 1990. “Employers Large and Small, *Harvard University Press.*

Bryson Alex, 2002. “The Union Membership Wage Premium: An Analysis Using Propensity Score Matching,” *ISBN* 7530 1551 X.

Bryson A, John Forthand Laroche, P., 2009. "Unions and Workplace Performance in Britain and France," *NIESR Discussion Papers 327*, National Institute of Economic andSocial Research.

Caliendo.M. and Kopeinig.S., 2008. “Some Practical Guidance for the Implementation of Propensity Score Matching,” *IZA Discussion Paper* No. 1588.

Christophe Daniel and Catherine Sofer, 1998. “Bargaining, Compensating Wage Differentials, and Dualism of the Labor Market: Theory and Evidence for France”, *Journal of Labor Economics*, Vol. 16, No. 3, pp. 546-575.

Claudio Lucifora and Dominique Meurs, 2006. "The Public Sector Pay Gap in France, Great Britain and Italy," *Review of Income and Wealth, International Association for Research in Income and Wealth*, vol. 52(1), pages 43-59, 03.

Cotton, J. 1988. “On the Decomposition of Wage Differentials”, *The Review of Economics and Statistics*, 70: 236–243.

Cottrell and Gaubert 2003.“Efficient estimators: the use of neural networks to construct pseudo panels”, hal-00122817, version 1.

Criscuolo, C., 2000. Employer Size - Wage Effect: A Critical Review and an Econometric Analysis,”*University of Siena Economics Working Paper* No. 277.

Daniela Glocker, Viktor Steiner, 2007. “Self-employment – A Way to End Unemployment? Empirical Evidence from German Pseudo-Panel Data”, *DIW Berlin*, Discussion Papers Berlin, January 2007 (German Institute for Economic Research).

Daniel Christophe and Sofer Catherine, 1998. “Bargaining, Compensating Wage Differentials, and Dualism of the Labor Market: Theory and Evidence for France”, *Journal of Labor Economics*, Vol. 16, No. 3 (July 1998), pp. 546-575

Davis, Steven J. and Haltiwanger John, 1996.“Employer Size and The Wage Structure in U.S. Manufacturing” *NBER Working Paper*, No. 5393 Issued in December 1995

Deaton, A., 1985.“Panel data from time series of cross-sections”, *Journal of Econometrics*, 30, 109-126.

Dehejia, R. H., and Wahba, S., 1999. “Causal effects in non-experimental studies: Reevaluating the evaluation of training programs”, *Journal of the American statistical Association*, 94(448), 1053-1062.

Dehejia, R.H. and Wahba, S., 2002. “Propensity score matching methods for non-experimental causal studies”, *Review of Economics and Statistics*, 84(1): 151–161.

Dela Rica, S., Dolado, J. J., and Llorens, V., 2008.“Ceilings or floors?Gender wage gaps by education in Spain”, *Journal of Population Economics*, 21(3), 751-776.

Devereux, P. J., 2007.“Improved errors-in-variables estimators for grouped data”, *Journal of Business & Economic Statistics*, 25(3), 278-287.

DiPrete, T. A., and Gangl, M., 2004."Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments", *Sociological Methodology*, 34(1), 271-310.

Doeringer, P. and M. Piore, 1971."Internal labor markets and manpower analysis" *Lexington Books* D.C. Heath, Lexington, MA. 1971.

Dolado, J. J., F. Felgueroso and J. F. Jimeno, 2004. "Where Do Women Work?: Analysis Patterns of Occupational Segregation by Gender," *Annales d´Economie et deStatistique (Special Issue on Discrimination and Unequal Outcomes)*, 71/72(3), 293-315.

Dolton, P.J., G. H. Makepeace, and W. Van Der Klaauw, 1989."Occupational Choice and Earnings Determination: The Role of Sample Selection and Non-Pecuniary Factors," *Oxford Economic Papers* 41, 573—594.

Dubin, J. A. and McFadden, D. L., 1984."An econometric analysis of residential electric appliance holdings and consumption", *Econometrica* 52: 345–362.

Erica L. Groshen, 1991. "Sources of Intra-Industry Wage Dispersion: How Much Do Employers Matter?" *The Quarterly Journal of Economics*, August 1991, pp. 869–884.

Farber, H. S., 1999. "Mobility and stability: The dynamics of job change in labor markets", *Handbook of labor economics*, 3, 2439-2483.

Fathi, F., and FitzRoy, F. 2002. "Basic Wages and Firm Characteristics: Rent-sharing in French Manufacturing,"*Review of Labor Economics and Industrial Relations*, Vol. 18, No. 4, pp. 615-631, December 2004.

Fathi, F., and FitzRoy, F., 2006. "Dynamic Monopsony: Evidence from a French Establishment Panel,"*Economica*,Economica, Vol. 73, No. 291, pp. 533-545, August 2006.

Feng, S., 2009. "Return to Training and Establishment Size: A Reexamination of the Size-Wage Puzzle,"*IZA & Princeton University Discussion Paper* No. 4143.

Fernández, P. 2006. "Determinantes del Diferencial Salarial por Género en Colombia, 1997-2003." Desarrollo y Sociedad 58(2): 165-208.

Foss, M.R., 1981."Changes in the workweek of fixed capital," *American Enterprise Institute*, Washington, DC.

Gardes François, Greg J. Duncan, Patrice Gaubert, Marc Gurgand and Christophe Starzec, 2005."Panel and Pseudo-Panel Estimation of Cross-Sectional and Time Series Elasticities of Food Consumption: The Case of U.S. and Polish Data" *Journal of Business & Economic Statistics*, Volume 23, Issue 2.

Fujiwara-Greve Takako and Greve Henrich, 2004. "The Role of expectation in job search and firm size effect on wages", *Japanese Economic Review, Vol. 55, No. 1, pp. 56-85, March 2004.*

Green, F., Machin, S., and Manning, A., 1996."The employer size-wage effect: can dynamic monopsony provide an explanation?", *Oxford Economic Papers*, 48(3), 433-455.

Ganguli, I., and Terrell, K., 2009. "Closing Gender Wage Gaps in Ukraine: Composition, Returns and the Minimum Wage".

Galdo, J., Smith, J., & Black, D., 2007. "Bandwidth selection and the estimation of treatment effects with unbalanced data".

Garen E. John, 1985. "Worker Heterogeneity, Job Screening, and Firm Size," *Journal of Political Economy*, 1985, vol. 93, issue 4, pages 715-39.

Garicano, L., LeLarge, C. and Reenen, V., 2013."Firm Size Distortions and the Productivity Distribution: Evidence from France," *NBER Working Papers* 18841, National Bureau of Economic Research, Inc.

Gassner, K. 1998."An estimation of UK telephone access demand using pseudo-panel data", Utilities Policy, vol.7, no.3, pp. 143-154.

Giaimo R., Bono F., Lo Magno G. L., 2010.“ Interpreting the Decomposition of the Gender Earnings Gap”, Department of National Accounts and Social Processes Analysis, University of Palermo, Italy.

Gibbons, R., and L. Katz, 1992. "Does unmeasured ability explain inter-industry wage differentials?" *Review of Economic Studies* 59: 515~35.

Gibson, J., 2004. “Firm size, wages, and unmeasured skills: Evidence from an international database”, Unpublished manuscript, University of Waikato.

Glocker, D., and Steiner, V. 2007. “Self-employment: a way to end unemployment? Empirical evidence from German pseudo-panel data”, IZA *Discussion Paper,* (No. 2561).

Greene, William H., 1981. “Sample selection bias as a specification error”, *Econometrica*, 49 (3), 795-798.

---------------- 1995. “LIMDEP user’s manual. Plainview”, NY: Econometric Software.

---------------- 2000-2008. “Econometric Analysis”, (International Edition), *New York: Prentice Hall*

Grlliches, Z., 1969. "Capital-skill complementarity," *Review of Economics and Statistics* 51: 465-468.

Gronau Reuben, 1974. “Wage Comparisons--A Selectivity Bias” Journal of Political Economy, Vol. 82, No. 6 (Nov. - Dec., 1974), pp. 1119-1143

Groshen, E. L., 1991a: “The Structure of the Female/Male Wage Differential: Is It Who You Are, What You Do, or Where You Work?” *Journal of Human Resources*, 26(3), 457-472.

Groshen, E. L., 1991b: “Sources of Intra-Industry Wage Dispersion: How Much Do Employers Matter?” *Quarterly Journal of Economics*, 106(3), 869-884.

Groshen, E.L. and A.B. Krueger, 1990. "The structure of supervision and pay in hospitals," *Industrial and Labor Relations Review*, 43: $134--S146.

Guo, S., and Fraser, M. W., 2010. "Propensity score analysis. Statistical methods and applications".

Hamermesh, D., 1993. "Labor demand," *Princeton University Press*, Princeton, NJ.

Heckman, James J., 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5(4), 475-492.

----------------1978. ''Dummy Endogenous Variables in a Simultaneous Equation System.'' Econometrica 46(4):931–59.

--------------- 1979. "Sample Selection Bias as a Specification Error," *Econometrica* 47, 153-163.

-------------- **1997**. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32: 441-462.

------------- **2005.** "The scientific model of causality", *Sociological Methodology, 35, 1(97).*

------------- **2008.** "Econometric Causality" *International Statistical Review, 2008, (76) 1, 1-27).*

Heckman, James J. Tobias, Justin & Vytlacil, Ed, 2003. "Simple Estimators for Treatment Parameters in a Latent Variable Framework," *Staff General Research Papers 12012*, Iowa State University, Department of Economics.

Heckman, James J. & Hidehiko Ichimura & Jeffrey Smith & Petra Todd, 1998.*"*Characterizing Selection Bias Using Experimental Data*"*, *Econometrica, Vol. 66, No. 5 (September, 1998) 1017-1098).*

Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith., 2000."The Economics and Econometrics of Active Labor Market Programs." Pg. 1865-2097 in Handbook of Labor Economics, *Elsevier*, edition 1, volume 3, number 3.

Heckman, James J., Smith Jeffrey and Clements Nancy, 1997. “Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts”, *The Review of Economic Studies*, Vol. 64, No. 4, pp. 487-535.

Heyman Fredrik, 2007. “Firm Size or Firm Age? The Effect on Wages Using Matched Employer–Employee Data,” *Research Institute of Industrial Economics (IFN), Labour*, Vol. 21, No. 2, pp. 237-263, June 2007

Hmermesh, D., 1980. “The Economics of Firm Size, Market Structure and Social Performance,” Washington DC, Federal Trade Commission.

Hu, L., 2000. “Who Gets Good Jobs? The Hiring Decisions and Compensation Structures of Large Firms”, *Industrial Relations Section Working Paper*, (436).

Hu, L., 2003.“The Hiring Decisions and Compensation Structures of Large Firms”, *Industrialand Labor Relations Review*, vol 56 (4).pp.663-81.

Idson, T.L., 1996. "Employer size and labor turnover", in: S. Polachek, ed., *Research in labor economics* (JAI Press, Greenwich, CT).

Idson, T.L. and D. Feaster, 1990. "A selectivity model with employer size differentials," *Journal of Labor Economics* 8 (part 1): 99-122.

Imbens, G. W,. 2004. “Nonparametric estimation of average treatment effects under exogeneity: A review”, *Review of Economics and statistics*, *86*(1), 4-29.

Jamet, S., 2006, "Improving Labour Market Performance in France", *OECD Economics Department Working Papers*, No. 504, OECD Publishing.doi: 10.1787/877354041783

Jann, B., 2008. "A Stata implementation of the Blinder-Oaxaca decomposition", *Stata Journal*, *8*(4), 453-479.

Johnson, G. E., and Stafford, F. P., 1974."The earnings and promotion of women faculty", *The American Economic Review*, *64*(6), 888-903.

Julia I. Lane, Laurie A. Salmon, and James R. Spletzer, 2007. "Establishment Wage Differentials," *BLS Working Paper* No. 403 (Washington, DC, U.S. Department of Labor, Bureau of Labor Statistics, March 2007).

Katharina Gassner, **1998.**"An estimation of UK telephone access demand using Pseudo-Panel data" *Utilities Policy* (November 1998), 7 (3), pg. 143-154.

Katz Lawrence F. andSummers, Lawrence H. 1989. "Can Inter-industry Wage Differentials Justify Strategic Trade Policy?," NBER Chapters, in: Trade Policies for International Competitiveness, pages 85-124 *National Bureau of Economic Research*, Inc.

Kennedy, P., 2003. "A Guide to Econometrics, (5th Edition)", *Cambridge, MA: MIT Press*

Kenneth R. Troske, 1999. "Evidence on The Employer Size-Wage Premium From Worker-Establishment Matched Data," *The Review of Economics and Statistics*, MIT Press, vol. 81(1), pages 15-26, February.

Kingdon, G., 1996. "The quality and efficiency of private and public education: a case-study of urban Indian", *Oxford Bulletin of Economics and Statistics*, 58: 57-81.

Kramarz, Francis, 2008. "Offshoring, Wages, and Employment: Evidence from Data Matching Imports, Firms, and Workers," *CREST working paper*.

Kramarz Francois, Sébastien Perez-Duarte, 2009. "Wage Structure in France, 1977-96," chapter in Book, The Structure of Wages: An International Comparison (p. 401 - 418) *National Bureau of Economic Research ,University of Chicago Press.*

Krueger, A. and L. Summers, 1989."Efficiency Wages and the Wage Structure," *NBER Working Papers 1952*, National Bureau of Economic Research, Inc.

Kruse, D., 1992. "Supervision, working conditions, and the employer size effect", *Industrial Relations* 31: 229-249.

Kumar Anil and Rossiter Peter, 2010. A Methodology for Decomposing Age, Period and Cohort Effects Using Pseudo-Panel Data to Study Children's Participation in Organised Sports, Australian Statistical Conference, Perth

Lallemand, T., and Plasman, R., 2005."The establishment-size wage premium: evidence from European countries," *IZA Discussion Paper* No. 1569.

---------------- 2007. "The establishment-size wage premium: evidence from European countries," *Empirica*, *Springer*, vol. 34(5), pages 427-451.

Lallemand Thierry, 2007. "Employer Size and the Structure of Wages: A Critical Survey" *ULB Institutional Repository* from ULB -- Universite Libre de Bruxelles.

Lane, I. J., Salmon, A. L., and Spletzer, R., 2007. "Establishment Wage Differentials",*Monthly Labor Review*, April 2007, Vol. 130, No. 4.

Lanfranchi, Joseph, Henry Ohlsson and Ali Skalli, 2002. "Compensating Wage Differentials and Shift Work Preferences", *Economics Letters*, Vol. 74, February, No. 3, pages 393-398.

Lanfranchi, Joseph, Henry Ohlsson and Ali Skalli, 2002, "Compensating Wage Differentials and Shift Work Preferences", *Economics Letters*, Vol. 74, February, No. 3, pages 393-398.

Patrice Laroche and Heidi Wechtler, 2011. "The Effects of Labor Unions on Workplace Performance: New Evidence from France," *Journal of Labor Research*, Springer, vol. 32(2), pages 157-180, June.

Lazear, E. P., 1995: “Hiring risky workers,” *NBER Working Paper* No. 5334.

Leclair M. and Petit P., 2003."The Effects of Workplace Union on the Gender Wage Gap," *Documents de Travail de la DESE* - 2003-10.

Lee, B., and Marsh, L. C., 2000. “Sample selection bias correction for missing response observations”, *Oxford Bulletin of Economics and Statistics*, 62 (2), 305-323.

Lee, L. F., 1983. “Generalized econometric models with selectivity”, *Econometrica*, 51: 507–512.

Lester, R.A., 1967. "Pay differentials by size of establishment," *Industrial and Labor Relations Review* 7: 57-67.

Leuven, E., and Sianesi, B., 2003–2006. “PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing [Computer software]”, Retrieved June 2006 from http://ideas.repec.org/c/boc/bocode/s432001.html

Lluis, S., 2008. “The Structure of Wages by Firm Size: A Comparison of Canada and the United States,”. Economics Department, 200 University West, Waterloo, Ontario. September 2008.

Lluis, S., & Ferre, A., 2004. “Should Workers Care about Firm Size?”*Industrial Relations Center*, University of Minnesota.

Luca, Flabbi, 2011. “Gender differences in education, career choices and labor market outcomes on a sample of OECD countries?”,*World development report 2012,* gender equality and development

Luojia Hu, 2003.“The Hiring Decisions and Compensation Structures of Large Firms”, *Industrial and Labor Relations Review*, Vol. 56, No. 4 (Jul., 2003), pp. 663-681.

Machado, J. A. F and J. Mata, 2005.“Counterfactual Decomposition of Changes in Wage Distribution Using Quantile Regression”, *Journal of Applied Econometrics*, 20(4) 445-465.

Machin, S., & Puhani, P. A., 2003. “Subject of degree and the gender wage differential: evidence from the UK and Germany”, Economics Letters, 79(3), 393-400.

Macpherson, D.A.; Hirsch, B.T., 1995. “Wages and gender composition: Why do women’s jobs pay less?” *Journal of Labor Economics*, vol. 13 nº3, pp.426-471.

Maddala, G S., 1983 “Limited-Dependent and Qualitative Variables in Economics”, New York: *Cambridge University Press*, pp.257-91.

Maddala, G. S. and F. D. Nelson, 1975. “Switching Regression Models with Exogenous and Endogenous Switching”, *American Statistical Association*, Proceedings of the Business and Economic Statistics, Section 423-426.

Maddala, G. S., 1977. “Self-selectivity problems in Econometric Models”, chapter Applications of Statistics, pages 351–366. *Amsterdam: North-Holland*.

Magali Beffy, Elise Coudin, Roland Rathelot, 2008. "“Who is Confronted to Insecure Labor Market Histories ? Some Evidence Based on the French Labor Market Transitions”," *Centre de Recherche en Economie et Statistique(INSEE)*,Working Papers 2008-34.

Main, B.G.M. and B. Reilly, 1993."The employer size-wage gap: evidence for Britain", *Economica* 60: 125-142.

Manning, A., “Monopsony in Motion: Imperfect Competition in Labor Markets,” *Princeton University Press* 2003.

Margolis, David N. and Salvanes, Kjell G., 2001. "Do Firms Really Share Rents with Their Workers?," *IZA Discussion Papers* 330, Institute for the Study of Labor (IZA).

Martin Segal, 1964., "The Relation Between Union Wage Impact and Market Structure." *Quarterly Journal of Economics* LXXVIII (Feb. 1964).

Masters, Stanley, 1969. “An Interindustry Analysis of Wages and Plant Size”, *Review of Economics and Statistics* 51 (August 1969) 341-345.

Matteazzi et al., 2013. “Does Part-Time Employment Widen the Gender Wage Gap? Evidence from Twelve European Countries”, *Society for the Study of Economic Inequality*, Working Papers from ECINEQ, No 293.

Mayo, J.W. and M.N. Murray, 1991. "Firm size, employment risk, and wages: further insights on a persistent puzzle," *Applied Economics* 23:1351-1360.

Mellow, Wesley, 1982, “Employer Size and Wages,” *Review of Economics and Statistics* 64 (August 1982) 495-501.

Meng, X., 2004.“Gender Earnings Gap: the Role of Firm Specific Effects”, *Labour Economics*, vol. 11, pp.555-573.

Meurs Dominique and Sophie Ponthieux, 2006. “Can the gender wage go further down?”,*Economie et Statistique*, 398: 99-129.

Miller, Paul & Mulvey, Charles, 1996."Unions, Firm Size and Wages," The Economic Record, The Economic Society of Australia, vol. 72(217), pages 138-53, June.

Miller, R. B., & Wright, D. W., 1995. “Detecting and correcting attrition bias in longitudinal family research”, *Journal of Marriage and the Family*, 57 (4), 921-929.

Millimet, Daniel L., 2005. “Job search skills, employer size and wages,” *Applied Economics Letters*, 12:2, 95 — 100.

Mincer, J., 1974.“Schooling, Experience, and Earnings”, *Columbia University Press*, New York.

Mizala, Alejandra & Romaguera, Pilar, 1998. "Wage Differentials and Occupational Wage Premia: Firm-Level Evidence for Brazil and Chile," *Review of Income and Wealth*, *International Association for Research in Income and Wealth*”, vol. 44(2), pages 239-57, June.

Moore, Henry L., 1911. “Laws of wages,” (*Macmillan, New York*) (reprinted 1967, *Augustus M. Kelley*, New York).

Morissette René, 1993. “Canadian Jobs and Firm Size: Do Smaller Firms Pay Less?” *The Canadian Journal of Economics / Revue canadienne d'Economique*, Vol. 26, No. 1(Feb., 1993), pp. 159-174.

Moulin, S., 2004.“Discriminatory barriers and the gender wage gap in France”, *Labour*, *18*(3), 443-463.

Myeong-Su Yun, 2005. "A Simple Solution to the Identification Problem in Detailed Wage Decompositions," *Economic Inquiry*, vol. 43, Oct 2005, 766 - 772.

Neumark, D., 1988. “Employers’ Discriminatory Behavior and the Estimation of Wage Discrimination”, *The Journal of Human Resources* 23: 279–295.

Neuman, Shoshana & Oaxaca, Ronald L, 1998. "Estimating Labour Market Discrimination with Selectivity Corrected Wage Equations: Methodological Considerations and an Illustration from Israel," *CEPR Discussion Papers* 1915, C.E.P.R. Discussion Papers.

Neuman, S., and R. L. Oaxaca, 2004. “Wage decomposition with Selectivity-Corrected Wage Equations :A Methodological Note,” *Journal of Economic Inequality*, 2004, vol. 2, issue 1, pages 3-10.

Ñopo, H. 2006. “The Gender Wage Gap in Chile 1992-2003 from a Matching Comparisons Perspective”, Research Department Working Paper 562. Washington, DC, United States: Inter-American Development Bank.

Oaxaca, R. L, 1973. “Male-Female Wage Differentials in Urban Labor Markets”, *International Economic Review*, 14(3), 693-709.

Oaxaca, R. L. and M. R. Ransom, 1994. “On Discrimination and the Decomposition of Wage Differentials,” *Journal of Econometrics*, 61(1), 5-21.

Oaxaca, Ronald L. and Michael Ransom, 1999. "Identification in Detailed Wage Decompositions"(with Michael R. Ransom), *Review of Economics and Statistics*, vol. 81, no. 1, 154-157.

Oi, W. Y., and Idson, T. L., 1999. “Firm Size and Wages in*Handbook of Labor* Economics,” *Elsevier Science B.V., edition 1*, volume 3, number 3.

Oi, Walter, 1983. "Heterogeneous Firms and the Organization of Production," *Economic Inquiry*, 1983, vol. 21, issue 2, pages 147-71.

Paez Paul, 2003. “The effects of firm size on wages in Colorado: a case study,” *Monthly Labor Review* July 2003.

Pedace, R., 2008. “Firm Size-Wage Premiums: Using Employer Data to Unravel the Mystery,”*Journal of Economic Issues*, Vol. 44, No. 1, pp. 163-181, 2010.

Piekkola Hannu, 2000. “Unobserved Human Capital and Firm-Size Premium,” *Discussion Papers from The Research Institute of the Finnish Economy*, No 739.

Raquel Vale Mendes, 2005. “Decomposition of Gender Wage Differentials among Portuguese Top Management Jobs,” *Escola Superior de Gestão, Instituto Politécnico do Cavado*- Portugal

Reimers, C. W., 1983. “Labor Market Discrimination Against Hispanic and Black Men”, *The Review of Economics and Statistics*, 65: 570–579.

Rosen Sherwin, 1986. “The theory of equalizing differences”, University of Chicago Handbook of Labor Economics, Volume I, Edited by O. *Ashenfelter and R. Layard;'Elsevier Science Publishers* B V, 1986

Rosenbaum Paul R. and Donald B. Rubin, 1983."The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, Vol. 70, No. 1.(Apr., 1983), pp. 41-55.

Russell Jeffrey E. and John W. Fraas, 2005. "An Application of Panel Regression to Pseudo Panel Data"*Multiple Linear Regression Viewpoints*, 2005, Vol. 31(1)

Rubin, D. B., 1973. "Matching to remove bias in observational studies", *Biometrics*, 159-183.

Rubin, D. B., 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies", *Journal of educational Psychology*, *66*(5), 688.

Russell, J. E., & Fraas, J. W., 2005.An application of panel regression to pseudo panel data. *Multiple Linear Regression Viewpoints*, *31*(1), 1-15.

Roy, A. D., 1951, "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, June 1951, 3, 135-46.

Schere, F.M., 1976. "Industrial Structure, Scale Economies, and Worker Alienation," *Essays in Industrial Organization in Honor of Joe S. Bain*, Cambridge, MA ; Ballinger, 1976.

Segal Marti, 1964. "The Relation Between Union Wage Impact and Market Structure", *The Quarterly Journal of Economics*, Vol. 78, No. 1 (Feb., 1964), pp. 96-114Published

Sherwin Rosen, 1983. "The Equilibrium Approach to Labor Markets," *NBER Working Papers,* 1165, National Bureau of Economic Research, Inc.

Shinohara, M., "Growth and cycles in the Japanese economy," *Kinokuniya Bookstore Co.*, Tokyo.1962.

Sianesi, B., 2001. "Implementing propensity score matching estimators with Stata",UK Stata Users Group, VII Meeting London, May 2001

Silva, C., 2004. "Estimating the employer size-wage premium in a panel data model with comparative advantage and non-random selection,"*NIPE Working Papers* 6/2004.

Simón, H., 2012. "The Gender Pay Gap in Europe: An International Comparison with Matched Employer-Employee Data," *Universidad de Alicante.*

Simon, H. and H. Russell, 2005."Firms and the Gender Pay Gap: A Cross-National Comparison," *Pay Inequalities and Economic Performance Working Paper* 15, London School of Economics (LSE), London.

Simón, H.; Russell, H., 2007. "The Gender Pay Gap: New Evidence from International Matched EmployerEmployee Data", unpublished

Smith A. J. and Todd E, P., 2005. "Does matching overcome LaLonde's critique of nonexperimental estimators?", *Journal of econometrics, 125*(1), 305-353.

Smith, James P., 2000. "Why is Wealth Inequality Rising?" in Finis Welch (ed.), Increasing Income Inequality in America: The Facts, Causes, and Consequences, *University of Chicago Press*, 2000, pp. 83-116.

Stafford, Frank, 1980. "Firm Size, Workplace public Goods and Worker Welfare," *The economics of firm size, Market Structure and Social Performance (Ed.)* John Siegfried. Washington DC ; Federal Trade Commission, 1980.

Stephen G. Bronars and Melissa Famulari, 1997. "Wage, Tenure, and Wage Growth Variation Within and Across Establishments," *Journal of Labor Economics*, April 1997, pp. 285–317.

Stigler George, 1962. "Information in the Labor Market," *Journal of Political Economy* 70, October 1962 ; 94-105.

Stolzenberg, R. M., and Relles, D. A.,1997. "Tools for intuition about sample selection bias and its correction", *American Sociological Review*, 62 (3), 494-507.

Sutton M., 2004.“The effects of Income on Health: Evidence from a British Pseudo-Panel, General Practice and Primary Care Community-Based Sciences”, University of Glasgow January 2004 Working Paper #10.

Terza, Joseph V., 1997. “Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects”, *Journal of Econometrics*, Volume 84, Issue 1, May 1998, Pages 129–154.

Theil, H., 1952. “On the time shape of economic microvariables and the munich business test”, *Revue de l'Institut International de Statistique*, 105 – 120.

Thomas Breda, 2010. "Are union representatives badly paid? Evidence from France," *Working Papers halshs-00564905, HAL.*

------------ 2010. "Firms' rents, workers' bargaining power and the union wage premium in France," *Working Papers halshs-00564903, HAL.*

Troske Kenneth R., 1999. “Evidence on the employer size wage premium from worker-establishment matched data,” *The review of Economics and Statistics*, February 1999, Vol. 81, No. 1, Pages 15-26.

Velling, Johannes, 1995. “Wage discrimination and occupational segregation of foreign male workers in Germany,” *ZEW Discussion Papers*, No. 95-04.

Vella, F., & Verbeek, M. 1999.“Two-step estimation of panel data models with censored endogenous variables and selection bias”, *Journal of Econometrics*, 90 (2),239-263.

Verbeek, M. and Nijman, T., 1992. “Can cohort data be treated as genuine panel data?” *Empirical Economics* 17(1), 9–23.

Verbeek, M. and Nijman, T., 1993.“Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections”, *Journal of Econometrics* 59(1-2), 125–136.

Weiss, L., 1966. “Concentration and labor earning,”*American Economic Review*, 56 March 1966 ; 96-117.

Warunsiri, Sasiwimon & McNown, Robert, 2010. "The Returns to Education in Thailand: A Pseudo-Panel Approach," World Development, *Elsevier*, vol. 38(11), pages1616-1625, November.

Winship, C.,andMare, R. D., 1992. “Models for sample selection bias”, *Annual Review of Sociology*, 18, 327-350.

Winter-Ebmer, R., 1995. "Does layoff risk explain the firm-size wage differential?," *Applied Economics Letters* 2: 211-214.

Wooldridge, J.W., 2002. “*Econometric Analysis of Cross Section and Panel Data”*.The MIT press. 2002.

Yallen, 1984. “Efficiency Wage Models of Unemployment”, *The American Economic Review*, Vol. 74, No. 2.

Zabojnik, Jan and Bernhardt, Dan, 2001. "Corporate Tournaments, Human Capital Acquisition, and the Firm Size-Wage Relation," *Review of Economic Studies*, WileyBlackwell, vol. 68(3), pages 693-716, July.

Appendix- A Chapter 2

A -1 Descriptive Statistics 1992

A-1.1 Descriptive statistics by establishment size

As most of the variables are qualitative, therefore, *Table-A-1* shows details of all categories. For continuous variables, including tenure, experience and age, dummies are introduced to see the non-linear relationship of these variables with establishment size. For estimations these variables are used as continuous. The first column in table A-1 shows results of all population and it is followed by columns of three size categories. The objective is to see the distribution of all control variables in each establishment size category.

- Gender comparisons reveal that there are more males than females. The difference between male and females ratio is more in large employers. Male and females distribution is in a ratio of 60:40. In small establishments there are 45% females and 54% males, whereas, in large establishments, there are 33% females and 66% males. It shows that female labor supply is more in small establishment size.
- A Majority of people in the sample is between 31-40 years of age.
- There is more married population (69%) in the sample compared to single or other family status. In small establishments 65% population is married followed by 67% in medium and 73% in large.
- More than 94 percent of the employees hold CDI contract (Contrat à Durée Indéterminée, long-term contract) while only 5 percent hold CDD (Contrat à Durée Déterminée , fixed term employment contract or short term).
- A Majority of people in the sample completed between 0-5 years in the current job, whereas, in large establishment size the largest sample of workers completed 11-20 years in the current job. We may say that tenure and family situation variables are important for workers to go and stay in large establishments.
- Labor market experience is an important job and pay determinant variable. We see 35% of people in the sample hold 11-20 years of labor market experience followed by 21-30 years of experience, 30%. Among the size categories, one can see the similar trend.
- In this survey, a majority of people are working in the blue collar jobs, 35%; followed by 27% in the high skilled white collar jobs, 24% in low skilled white collar and 12% in the

high intellectual professions. In small size establishments, major professions are blue collar and low skilled white collar, while, in large size establishments the main professions are blue collar and high skilled white collar.

- The largest sector is services sector with 51.5% workforce and smallest sector is trade with 10% of sample. In large size establishments, the largest sector is manufacturing sector having 57% workforce employed. In small and medium the largest category is services sector i.e. 57% and 66% of workforce respectively.
- The highest proportion of all population holds short technical education and primary education (that also includes people without degree). This trend prevails in all the employer size categories.
- Ile de France is the biggest region where around 18% of the establishments are based. It is followed by Rhone Alpes with 11% of all establishments.
- There is a variable on nationality with very low variation as 95% of the sample is French national.

Table A-1 Descriptive Statistics of all qualitative variables

Variables	All		Small		Medium		Large	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Females	30,063	40.21	13,831	45.58	6,094	41.97	10,138	33.91
Males	44,699	59.79	16,516	54.42	8,425	58.03	19,758	66.09
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
Age 25-30	16,953	22.68	8,099	26.69	3,526	24.29	5,328	17.82
Age 31-40	27,499	36.78	11,012	36.29	5,357	36.90	11,130	37.23
Age 41-50	21,558	28.84	7,944	26.18	3,983	27.43	9,631	32.22
Age 51-60	8,752	11.71	3,292	10.85	1,653	11.39	3,807	12.73
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
Single	17,343	23.20	7,887	25.99	3,492	24.05	5,964	19.95
Married	51,644	69.08	20,01	65.94	9,788	67.42	21,846	73.07
Others (widowed, divorced)	5,775	7.72	2,450	8.07	1,239	8.53	2,086	6.98
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
CDI	70,932	94.88	28,522	93.99	13,359	92.01	29,051	97.17
CDD	3,830	5.12	1,825	6.01	1,160	7.99	845	2.83
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
Tenure 0-5	28,539	38.17	15,501	51.08	6,137	42.27	6,901	23.08
Tenure 6-10	12,910	17.27	5,654	18.63	2,454	16.90	4,802	16.06
Tenure 11-20	21,191	28.34	6,667	21.97	4,071	28.04	10,453	34.96
Tenure 21-30	10,286	13.76	2,149	7.08	1,589	10.94	6,548	21.90
Tenure 31-46.5	1,836	2.46	376	1.24	268	1.85	1,192	3.99
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
Experience 0-5	2,730	3.65	1,200	3.95	503	3.46	1,027	3.44
Experience 6-10	11,452	15.32	5,479	18.05	2,356	16.23	3,617	12.10
Experience 11-20	26,346	35.24	11,045	36.40	5,237	36.07	10,064	33.66
Experience 21-30	22,435	30.01	8,321	27.42	4,205	28.96	9,909	33.14
Experience 31-40	10,858	14.52	3,886	12.81	2,002	13.79	4,970	16.62
Experience 41-49	941	1.26	416	1.37	216	1.49	309	1.03

Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
Management and High Intellectual professionals	9,322	12.47	3,855	12.70	1,563	10.77	3,904	13.06
High Skilled White Collar	20,909	27.97	7,99	26.33	3,698	25.47	9,221	30.84
Low Skilled White Collar	18,054	24.15	8,965	29.54	3,656	25.18	5,433	18.17
Blue collar	26,477	35.42	9,537	31.43	5,602	38.58	11,338	37.92
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
Manufacturing	28,571	38.22	7,68	25.31	3,725	25.66	17,166	57.42
Trade	7,648	10.23	5,256	17.32	1,205	8.30	1,187	3.97
Services	38,543	51.55	17,411	57.37	9,589	66.04	11,543	38.61
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
primary education	20,437	27.34	8,198	27.01	4,352	29.97	7,887	26.38
Secondary	9,374	12.54	4,123	13.59	1,82	12.54	3,431	11.48
Technical Short	26,015	34.80	10,446	34.42	4,974	34.26	10,595	35.44
Technical Long	5,1	6.82	2,131	7.02	837	5.76	2,132	7.13
Higher	13,836	18.51	5,449	17.96	2,536	17.47	5,851	19.57
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
Ile de France (11)	13,347	17.85	5,487	18.08	1,871	12.89	5,989	20.03
Champagne-Ardenne(21)	1,592	2.13	585	1.93	310	2.14	697	2.33
Picardie (22)	2,51	3.36	853	2.81	582	4.01	1,075	3.60
Haute-Normandie (23)	2,666	3.57	1,066	3.51	481	3.31	1,119	3.74
Centre (24)	3,149	4.21	1,18	3.89	544	3.75	1,425	4.77
Basse-Normandie (25)	1,597	2.14	741	2.44	240	1.65	616	2.06
Bourgogne (26)	2,64	3.53	1,003	3.31	637	4.39	1	3.34
Nord (31)	5,067	6.78	1,709	5.63	932	6.42	2,426	8.11
Lorraine (41)	3,424	4.58	1,02	3.36	632	4.35	1,772	5.93
Alsace (42)	3,848	5.15	1,247	4.11	1,025	7.06	1,576	5.27
Franche-Comte (43)	1,439	1.92	666	2.19	343	2.36	430	1.44
Pays de la Loire (52)	3,924	5.25	1,531	5.04	793	5.46	1,6	5.35
Bretagne (53)	3,198	4.28	1,359	4.48	702	4.84	1,137	3.80
Poitou-Charentes (54)	2,027	2.71	818	2.70	445	3.06	764	2.56

Aquitaine (72)	4,036	5.40	1,804	5.94	874	6.02	1,358	4.54
Midi-Pyrenees (73)	3,121	4.17	1,570	5.17	474	3.26	1,077	3.60
Limousin (74)	1,253	1.68	521	1.72	243	1.67	489	1.64
Rhone-Alpes (82)	8,237	11.02	3,365	11.09	1,865	12.85	3,007	10.06
Auvergne (83)	1,392	1.86	591	1.95	186	1.28	615	2.06
Languedoc-Roussillon (91)	1,326	1.77	709	2.34	199	1.37	418	1.40
Provence-Alpes-Cote d’Azur (93)	4,969	6.65	2,522	8.31	1,141	7.86	1,306	4.37
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00
Other Nationality	3,740	5.00	1,646	5.42	772	5.32	1,322	4.42
French	71,022	95.00	28,701	94.58	13,747	94.68	28,574	95.58
Total	74,762	100.00	30,347	100.00	14,519	100.00	29,896	100.00

A-1.2 Descriptive Statistics by Mean Hourly Wage

Table A-2 shows descriptive statistics in the same lines as in the table A-1 above except that now the comparison among establishments is performed by mean hourly wage not by the distribution among categories of employer size. Hourly wage is computed by utilizing the total working hours in a year and total gross wage in a year. The highest mean hourly wage is highlighted by yellow color. In all cases mean hourly wage of all variables is highest in large size establishments.

Main observations are summarized below.

- Descriptive statistics by mean hourly wage show that wages on average for males are higher compared to females and the difference is more in large size employers compared to small size establishments. One other notable thing is that the wage difference among male workers depending upon size of employer increases more compared to female employees.
- Returns to educations are higher with higher levels of education as expected. Mean hourly wage against secondary education is more compared to technical short across employer size groups. Among all categories, mean hourly wage for different levels of education is higher in large size employers compared to medium or small size employer.
- For type of contracts, as there are more permanent type contracts compared to fixed term contracts, therefore, we would expect higher average wage for permanent contracts. This is evident from the table A-2 that mean hourly wage is higher for permanent contracts and higher in large employers compared to other size category, while, for fixed term contracts, mean hourly wage in small is more compared to medium and large although the difference is small.
- Mean hourly wage increases with age and with tenure. The difference between medium and small remains negligible across categories of age and tenure but remains low compared to large size category.
- Among experience categories, the highest mean hourly wage is observed from 21-40 years of experience in all population and across size categories with highest mean

hourly wage in large size establishments. No difference in the mean hourly wage between medium and small size establishments is observed.

- Being single results in lower wage compared to other situations of marital status. Mean hourly wage is higher for married workers compared to other categories. Among size categories, mean hourly wage remains similar against marital status situations in the medium and small size groups and below the mean hourly wage in large size establishments.
- Mean hourly wage increases by level of profession. Mean hourly wages in all professions are high in large employers compared to medium and small.
- The average wage in the manufacturing sector is highest, 74francs, followed by services sector, 71france; however, the difference is small. In small size employers highest wages are paid in services sector unlike in large establishments where mean hourly wage is highest in the manufacturing sector. In medium size establishments, mean hourly wage is similar in the services and manufacturing sector. Overall, mean hourly wage in all sectors remains high in large size compared to other size categories with highest difference in the manufacturing sector compared to other sector.
- We observe higher mean hourly wage in large regions compared to other regions. As Ile de France is the largest region, therefore, mean hourly wage is highest in this region across all size categories. Mean hourly wage difference among regional categories between small and medium size establishments is low or either no difference or in few cases it is negative meaning higher wage in small compared to medium.
- Table A-2 also shows higher mean hourly wage by employer size for being a French national compared to any other nationality.

Table A-2 Descriptive Statistics by Mean hourly wage

	All			Small			Medium			Large		
Gender	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Female	63.19	26.38	30063	60,92	26,52	13831	60,29	25,23	6094	68,03	26,2	10138
Male	77.19	35.32	44699	71,71	35,94	16516	71,76	33,06	8425	84,08	34,52	19758
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Education	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
primary education	58.57	22.82	20437	54,88	23,55	8198	55,07	20,76	4352	64,33	21,94	7887
Secondary	72.99	31.74	9374	69,53	32,27	4123	68,31	29,47	1820	79,62	31,16	3431
Technical Short	65.16	23.64	26015	59,62	22,98	10446	61,45	21,64	4974	72,35	23,31	10595
Technical Long	79.74	32.04	5100	73,55	30,64	2131	77,83	32,44	837	86,67	31,9	2132
Higher	98.80	42.67	13836	93,73	42,4	5449	93,55	40,83	2536	105,8	42,72	5851
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Contract	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
CDI	72.28	32.77	70932	67,25	32,4	28522	67,85	30,66	13359	79,26	32,83	29051
CDD	58.15	29.28	3830	59,57	32,23	1825	56,52	27,21	1160	57,33	24,9	845
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Tenure	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Tenure 0-5	65.00	32.46	28539	63	31,57	15501	61,2	29,45	6137	72,89	35,56	6901
.6-10	70.51	32.75	12910	66,42	31,59	5654	66,71	30,41	2454	77,27	34,11	4802
.11-20	73.66	30.56	21191	70,68	32,17	6667	69,82	29,36	4071	77,05	29,57	10453
21-30	83.28	32.35	10286	79,7	34,19	2149	78,59	31,16	1589	85,6	31,79	6548
31-46.5	90.87	35.39	1836	85,83	40,48	376	87,95	35,75	268	93,11	33,34	1192
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Age	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Age 25-30	60.31	23.31	16953	57,07	22,24	8099	57,79	22,33	3526	66,89	24,18	5328

Age 31-40	70.98	31.15	27499	67,24	31,64	11012	66,81	29,24	5357	76,68	30,68	11130
Age 41-50	78.58	36.03	21558	73,99	37,06	7944	73,81	34,72	3983	84,35	34,85	9631
Age 51-60	77.87	38.16	8752	71,84	37,91	3292	70,4	34,02	1653	86,33	38,42	3807
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Experience	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Experience 0-5	72.12	26.73	2730	68,59	26,48	1200	69,02	28	503	77,77	25,43	1027
.6-10	63.90	28.51	11452	59,77	26,62	5479	61,85	27,91	2356	71,5	30,12	3617
.11-20	70.22	32.65	26346	66,24	32,42	11045	65,52	29,7	5237	77,04	33,21	10064
21-30	75.49	33.94	22435	71,01	34,62	8321	70,57	32,18	4205	81,34	33,18	9909
31-40	75.14	34.50	10858	69,34	35,27	3886	69,25	32,17	2002	82,04	33,51	4970
41-49	65.54	30.17	941	60,54	27,99	416	60,39	25,05	216	75,86	33,54	309
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Family Situation	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Single	65.65	28.15	17343	61,91	27,33	7887	61,93	26,48	3492	72,79	28,8	5964
Married	73.81	34.01	51644	68,75	34,02	20010	68,96	31,62	9788	80,61	33,86	21846
Others (widowed, divorced)	69.19	31.73	5775	66,51	32,68	2450	65,2	31,01	1239	74,72	30,23	2086
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Profession	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Management and High Intellectual professionals	128.40	40.73	9322	123,23	41,33	3855	125,15	39,92	1563	134,82	39,56	3904
High Skilled White Collar	79.26	22.33	20909	74,72	22,96	7990	76,82	21,11	3698	84,18	21,22	9221
Low Skilled White Collar	55.11	15.94	18054	52,25	15,32	8965	53,71	14,18	3656	60,76	16,59	5433
Blue collar	56.68	16.96	26477	51	14,38	9537	52,83	15,04	5602	63,35	17,51	11338
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Sector/Industry	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Manufacturing	74.07	32.83	28571	63,87	29,95	7680	67,52	30,83	3725	80,06	33,08	17166
Trade	64.37	31.12	7648	63,71	31	5256	63,41	29,51	1205	68,25	32,95	1187

Services	71.12	32.78	38543	69,01	33,71	17411	67,17	30,55	9589	77,59	32,21	11543
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Region	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Ile de France (11)	89.01	40.45	13347	83,13	39,15	5487	85,22	40,06	1871	95,58	40,77	5989
Champagne-Ardenne(21)	65.55	26.69	1592	60,07	27,07	585	63,57	26,7	310	71,03	25,29	697
Picardie (22)	66.23	27.86	2510	61,33	27,23	853	59,91	25,45	582	73,54	27,93	1075
Haute-Normandie (23)	70.08	30.15	2666	65,44	29,41	1066	63,74	25,93	481	77,23	31,07	1119
Centre (24)	66.54	28.09	3149	61,06	28,09	1180	62,57	25,63	544	72,59	27,81	1425
Basse-Normandie (25)	63.16	27.35	1597	57,97	25,96	741	58,67	24,96	240	71,15	28,01	616
Bourgogne (26)	64.61	26.77	2640	61,35	28,07	1003	60,18	25,39	637	70,7	25,14	1000
Nord (31)	66.44	28.47	5067	63,96	31,03	1709	62,09	29,5	932	69,85	25,64	2426
Lorraine (41)	68.41	27.98	3424	63,17	29,3	1020	61,56	22,58	632	73,86	27,84	1772
Alsace (42)	72.51	31.00	3848	67,95	31,87	1247	70,09	28,47	1025	77,68	31,13	1576
Franche-Comte (43)	60.60	24.30	1439	61,74	26,59	666	55,6	18,42	343	62,84	24,19	430
Pays de la Loire (52)	65.46	27.97	3924	61,99	29,24	1531	63,31	26,89	793	69,84	26,64	1600
Bretagne (53)	63.57	27.23	3198	61,81	29,75	1359	58,82	23,96	702	68,6	25,12	1137
Poitou-Charentes (54)	66.26	27.47	2027	59,71	28,2	818	62,51	25,6	445	75,46	25,14	764
Aquitaine (72)	70.09	31.93	4036	62,77	29,24	1804	67,27	31,63	874	81,62	32,29	1358
Midi-Pyrenees (73)	65.29	30.31	3121	60,03	28,65	1570	62,05	25,85	474	74,38	32,35	1077
Limousin (74)	61.30	23.61	1253	58,46	24,9	521	57,44	22,89	243	66,24	21,67	489
Rhone-Alpes (82)	71.96	31.77	8237	66,19	30,94	3365	67,83	28,77	1865	80,97	32,43	3007
Auvergne (83)	64.02	27.43	1392	57,42	27	591	68,39	34,91	186	69,03	23,75	615
Languedoc-Roussillon (91)	69.06	31.03	1326	65,65	32,8	709	64,06	23,41	199	77,21	29,58	418
Provence-Alpes-Cote d'Azur (93)	71.46	31.90	4969	66,81	30,52	2522	69,88	33,17	1141	81,82	31,04	1306
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896
Nationality	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.

Other	61.77	29.48	3740	59,01	28,62	1646	59,44	28,79	772	66,57	30,32	1322
French	72.07	32.84	71022	67,24	32,59	28701	67,37	30,6	13747	79,2	32,83	28574
Total	71.56	32.75	74762	66,79	32,44	30347	66,95	30,55	14519	78,64	32,83	29896

A-2 Descriptive Statistics 2002

A-2.1 Summary Statistics

Summary statistics and descriptive statistics are presented in the same way as shown for the 1992 survey. The variables on labor market experience, family situation, number of children and region are not available in the subsequent surveys. There is a variable called 'zeat' on the eight administrative regions which is available in all surveys. For type of contract there are three categories unlike in 1992 where there are only two types of employment contract. Another variable called 'status' is available in the subsequent surveys that is a dummy variable for management cadre and non management cadre. The categories for education, sector/industry, education and profession are formed in the same way as described above. Size of establishment is available as categorical variable unlike in 1992 where it is a continuous variable. Hourly wage is computed in the same manner by taking gross annual wage and annual working hours.

Table-A-3 belowshows the summary statistics of all variables. It is followed by descriptive statistics in Table-A-4. Total number of observations after removing outliers in the data are 30, 216. Mean hourly wage of the population in the sample is 17euros, mean tenure is 12 years and mean age is 41years. The description of categorical variables by establishment size is described below in Table A-4.

Table A-3 Summary Statistics of all Variables

Variable	No. of Obs.	Mean	Std. Dev.	Min	Max
Hourly Wage	30216	17.05	10.58	1.87	100
Size	30216	1.99	0.80	1	3
Education	30216	3.30	1.45	1	5
Contract	30216	1.05	0.28	1	3
Tenure	30216	12.66	10.48	0.08	45.5
Age	30216	41.14	9.48	25	60
Profession	30216	2.56	1.17	1	4
Sector	30216	2.06	0.92	1	3
Region	30216	3.40	2.52	1	8
Status	30216	0.24	0.43	0	1
Gender	30216	0.63	0.48	0	1
Nationality	30216	0.93	0.26	0	1

A-2.2 Descriptive statistics by Size of Establishment

Descriptive statics by percentage distribution of population in the sample across size of establishments are shown below in table A-4. The first column reports results of all population and it is followed by columns of three establishment size categories. Main observations are summarized below:

- Like in the previous survey, male and females distribution is in a ratio of 62:37 meaning that there are more males than females. A female composition in different sizes of establishments is 38%, 37% and 35% respectively in small, medium and large.
- 29% of population in the sample has attained higher level of education followed 28% of technical short education. In small size establishment the highest percentage of population has acquired technical short education, while, in medium size establishments the highest percentage of population has attained technical short and higher education. In large size establishments the first highest percentage is coming through higher education. In large establishments, the highly educated workforce makes up 33% of total workforce followed by 28% in medium and 26% in small.
- 96% of the employment contracts are permanent contracts.
- A Majority of people in the sample is between 31-40 years of age. This is similar in medium and small. In large the percentage of people between 41-50 is 31% which is highest among other age groups.
- A Majority of people in the sample completed between 0-5 years in the current job. This trend prevails across all size groups.
- Like in the 1992 survey, a majority of people in the year 2002 are working in the blue collar jobs, 32%; followed by 28% in the high skilled white collar jobs, 24% in high intellectual professions and 16% low skilled white collar. As 29% of the workforce is highly educated, therefore, we can expect the change in the professional distribution where more than 50% of the population in the sample is working in the higher level of professions. This also reflects the change in the labor force structure compared to last ten years.

- The services sector remains the largest sector of economy with 45% share, against 39% contribution from manufacturing sector and 15% from trade sector. In large size establishments, the largest sector is manufacturing sector having 49% workforce employed. In small and medium the largest category is services sector i.e. 44% and 48% of workforce respectively.
- Ile de France is the biggest region where around 38% of the establishments are based. The other big regions are Bassin Prissiene 13% and centre-est with 12% contribution in total population of sample. It is worth to mention that the regional categories are not the same in this survey. Here, the administrative regions information is given unlike detail departments' information in survey for the year 1992.
- 92% of the sample is French national.
- There is one additional dummy variable called management status, which shows that 75% of the workforce is doing non-management cadre jobs and 24% is doing the management cadre jobs.

Table A-4 Descriptive Statistics of all qualitative variables

Variables	All		Small		Medium		Large	
Gender	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Females	11,295	37.38	3,768	38.55	3,369	37.96	4,158	35.95
Males	18,921	62.62	6,007	61.45	5,506	62.04	7,408	64.05
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Education	Number	Percent	Number	Percent	Number	Percent	Number	Percent
primary education	6,028	20	1,890	19.34	1,885	21.24	2,253	19.48
Secondary	1,708	5.65	535	5.47	499	5.62	674	5.83
Technical Short	8,520	28.2	2,944	30.12	2,552	28.75	3,024	26.15
Technical Long	4,984	16.49	1,774	18.15	1,412	15.91	1,798	15.55
Higher	8,976	29.71	2,632	26.93	2,527	28.47	3,817	33.00
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Contract	Number	Percent	Number	Percent	Number	Percent	Number	Percent
CDI	29,092	96.28	9,431	96.48	8,547	96.30	11,114	96.09
CDD	681	2.25	250	2.56	232	2.61	199	1.72
Other	443	1.47	94	0.96	96	1.08	253	2.19
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Tenure	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Tenure 0-5	11,241	37	4,320	44.19	3,357	37.83	3,564	30.81
.6-10	4,705	16	1,712	17.51	1,393	15.70	1,600	13.83
.11-20	7,200	23.83	2,171	22.21	2,128	23.98	2,901	25.08
21-30	4,871	16	1,172	11.99	1,382	15.57	2,317	20.03
31-45.5	2,199	7.28	400	4.09	615	6.93	1,184	10.24
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Age	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Age 25-30	5,130	16.98	1,817	18.59	1,535	17.30	1,778	15.37
Age 31-40	9,686	32.06	3,347	34.24	2,912	32.81	3,427	29.63
Age 41-50	9,107	30.14	2,821	28.86	2,618	29.50	3,668	31.71

Age 51-60	6,293	20.83	1,79	18.31	1,81	20.39	2,693	23.28
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Profession	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Management and High Intellectual professionals	7,389	24.45	2,009	20.55	2,184	24.61	3,196	27.63
High Skilled White Collar	8,363	28	2,789	28.53	2,388	26.91	3,186	27.55
Low Skilled White Collar	4,736	16	1,88	19.23	1,399	15.76	1,457	12.60
Blue collar	9,728	32	3,097	31.68	2,904	32.72	3,727	32.22
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Industry/Sector	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Manufacturing	11,838	39.18	2,848	29.14	3,226	36.35	5,764	49.84
Trade	4,629	15.32	2,588	26.48	1,356	15.28	685	5.92
Services	13,749	45.5	4,339	44.39	4,293	48.37	5,117	44.24
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Region	Number	Percent	Number	Percent	Number	Percent	Number	Percent
ile de France	11,754	38.9	3,646	37.30	3,724	41.96	4,384	37.90
Bassin Parisien	4,043	13	1,239	12.68	1,215	13.69	1,589	13.74
NORD – PAS-DE-CALAIS	1,61	5.33	474	4.85	436	4.91	700	6.05
EST	2,056	6.8	693	7.09	587	6.61	776	6.71
OUEST	2,674	8.85	903	9.24	725	8.17	1,046	9.04
SUD-OUEST	2,17	7.18	831	8.50	608	6.85	731	6.32
CENTRE-EST	3,805	12.59	1,287	13.17	1,003	11.30	1,515	13.10
MÉDITERRANÉE	2,104	6.96	702	7.18	577	6.50	825	7.13
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Status	Number	Percent	Number	Percent	Number	Percent	Number	Percent
NC	22,827	75.55	7,766	79.45	6,691	75.39	8,37	72.37
CD	7,389	24.45	2,009	20.55	2,184	24.61	3,196	27.63
Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00
Nationality	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Other Nationality	2,24	7.41	676	6.92	655	7.38	909	7.86
French	27,976	92.59	9,099	93.08	8,22	92.62	10,657	92.14

Total	30,216	100	9,775	100.00	8,875	100.00	11,566	100.00

A-2.3 Descriptive Statistics by Mean Hourly Wage

Table A-5 below shows descriptive statistics by mean hourly wage across the categories of establishments to compared the average wage in different establishment size groups, professions, sectors and other worker and employer related variables. It will be followed by cross tabulation of selected variables. Overall, the trends in the mean hourly wage are similar to the previous survey with few exceptions. Main observations are summarized below:

- Mean hourly wage is higher for male, increasing level of education, tenure, growing level of age, in region ile de France, for high level of professions, in the services sector, other type of employment contracts, management status and for French nationals.
- The comparison of mean hourly wage across establishment's size categories shows that wages for male and female are higher in large. The gender wage difference is higher in large compared to medium and small.
- Mean hourly wage increases with educational levels and remains higher in large for all categories.
- For contract mean hourly wage is higher for CDI contracts in large and medium while in small mean hourly wage is higher for other types of employment contract (that may include seasonal contract).
- Mean hourly wage increases with tenure and age in all establishment size groups.
- For all professions, sectors and regions mean hourly wage is higher in large establishment size group. The highest mean hourly wage in survey 2002 is coming through services sector in small and medium size establishments unlike 1992 where mean hourly wage in manufacturing sector was highest. In large establishments mean hourly wage in the services sector is 19euros against 18euros in the manufacturing. So, we can say that wage in both sectors go side by side.
- Among regions mean hourly wage is higher in region ile de France and in Mediterranean.
- Mean hourly wage by management status (cadre or non-cadre) is higher in large establishments.
- Finally, we observe higher mean hourly wage for French nationals.

Table A-5 Descriptive Statistics by Mean hourly wage

Variables	All			Small			Medium			Large		
Gender	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Females	14,79	8,31	11295	13,76	7,87	3768	14,87	8,46	3369	15,67	8,46	4158
Males	18,39	11,52	18921	16,67	11,69	6007	18,13	11,45	5506	19,98	11,22	7408
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Education	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
primary education	12,58	7,00	6028	11,82	7,58	1890	12,39	6,50	1885	13,37	6,82	2253
Secondary	14,87	8,35	1708	13,01	7,69	535	14,65	8,42	499	16,51	8,50	674
Technical Short	14,31	7,31	8520	13,18	7,44	2944	14,29	7,48	2552	15,44	6,85	3024
Technical Long	17,32	9,27	4984	16,05	9,23	1774	17,29	9,04	1412	18,59	9,33	1798
Higher	22,90	13,29	8976	21,06	13,70	2632	23,11	13,44	2527	24,05	12,75	3817
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Contract	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
CDI	17,17	10,56	29092	15,59	10,40	9431	17,04	10,53	8547	18,61	10,51	11114
CDD	12,38	9,54	681	13,08	10,22	250	12,70	10,05	232	11,14	7,80	199
Other	16,04	11,84	443	17,60	16,86	94	14,10	9,56	96	16,20	10,21	253
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Tenure	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Tenure 0-5	15,03	10,09	11241	13,75	9,72	4320	15,42	10,55	3357	16,22	9,91	3564
.6-10	16,49	10,46	4705	15,41	10,13	1712	16,90	11,25	1393	17,28	10,00	1600
.11-20	18,44	11,01	7200	17,12	11,01	2171	17,74	10,38	2128	19,93	11,29	2901
21-30	18,78	10,18	4871	18,15	10,75	1172	17,87	9,39	1382	19,65	10,27	2317
31-45.5	20,14	10,66	2199	19,37	12,23	400	19,82	10,62	615	20,57	10,09	1184
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Age	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Age 25-30	13,17	6,56	5130	12,29	6,92	1817	13,01	6,21	1535	14,21	6,33	1778
Age 31-40	16,20	9,53	9686	14,78	9,07	3347	16,33	9,98	2912	17,48	9,40	3427
Age 41-50	18,10	10,86	9107	16,72	11,10	2821	17,81	10,69	2618	19,36	10,64	3668
Age 51-60	19,99	12,98	6293	18,45	13,47	1790	19,77	12,78	1810	21,16	12,67	2693
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Profession	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Management and High Intellectual professionals	27,87	13,32	7389	26,76	14,18	2009	28,10	13,27	2184	28,42	12,73	3196

High Skilled White Collar	16,60	6,76	8363	15,56	7,24	2789	16,49	6,42	2388	17,60	6,41	3186
Low Skilled White Collar	12,01	5,49	4736	11,51	5,92	1880	11,75	5,00	1399	12,90	5,26	1457
Blue collar	11,66	5,35	9728	10,72	5,78	3097	11,28	4,70	2904	12,73	5,27	3727
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Industry/Sector	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Manufacturing	16,29	9,53	11838	13,49	8,29	2848	15,39	9,28	3226	18,19	9,83	5764
Trade	14,69	10,26	4629	14,44	10,48	2588	14,86	9,97	1356	15,26	9,95	685
Services	18,49	11,31	13749	17,56	11,38	4339	18,67	11,28	4293	19,13	11,23	5117
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Region	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
ile de France	19,98	12,17	11754	18,31	12,32	3646	20,05	11,91	3724	21,31	12,11	4384
Bassin Parisien	14,79	8,46	4043	13,46	8,24	1239	14,41	8,20	1215	16,12	8,64	1589
NORD – PAS-DE-CALAIS	14,39	8,04	1610	12,65	6,81	474	13,77	8,64	436	15,96	8,15	700
EST	15,51	9,71	2056	15,19	10,82	693	14,27	8,39	587	16,73	9,45	776
OUEST	14,27	8,77	2674	13,11	8,93	903	14,22	8,45	725	15,32	8,74	1046
SUD-OUEST	15,09	8,55	2170	13,76	8,85	831	14,50	8,70	608	17,08	7,70	731
CENTRE-EST	15,85	9,46	3805	14,58	8,99	1287	15,41	9,67	1003	17,21	9,55	1515
MÉDITERRANÉE	16,25	9,14	2104	14,23	8,11	702	15,25	8,85	577	18,66	9,62	825
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Status	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
NC	13,54	6,37	22827	12,65	6,74	7766	13,24	5,95	6691	14,61	6,19	8370
CD	27,87	13,32	7389	26,76	14,18	2009	28,10	13,27	2184	28,42	12,73	3196
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566
Nationality	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Other Nationality	14,41	9,43	2240	13,42	9,56	676	13,97	9,34	655	15,45	9,29	909
French	17,26	10,64	27976	15,71	10,53	9099	17,13	10,59	8220	18,68	10,58	10657
Total	17,05	10,58	30216	15,55	10,49	9775	16,89	10,54	8875	18,43	10,52	11566

A-3 Descriptive Statistics 2005

A-3.1 Summary Statistics

Summary statistics and descriptive statistics are presented in the same way as shown in the preceding sections. All the variables are available in the ecmoss survey of 2005 as used in the year 2002 survey except for the nationality variable. Table-A-6 presents summary statistics of all variables. The variables have been cleaned by various changes, for instance, the observations are dropped for non declared education (22042), person with zero tenure or not declared any (3500), working hours more than 2280 annual hours (699), outliers in hourly wage (1280), age group below 25 or above 60 (7360), and type of contract not declared (4362).

The categories for education, sector/industry, education, establishment size and profession are formed in the same way as described above. Hourly wage is computed in the same manner by taking gross annual wage and annual working hours. Total number of observations after removing outliers in the data are 51, 272. Mean hourly wage of the population in the sample is 20 euros, mean tenure is 12 years and mean age is 41years. The description of categorical variables by establishment size is described below in Table A-7.

Table A-6 Summary Statistics of all Variables

Variable	No. of Obs.	Mean	Std. Dev.	Min	Max
Hourly Wage	51272	20,48	11,64	7,44	99,78
Size	51272	2,21	0,80	1	3
Education	51272	3,75	1,33	1	5
Contract	51272	1,04	0,22	1	3
Tenure	51272	12,79	10,38	0,08	45,5
Age	51272	41,82	9,42	25	60
Profession	51272	2,29	1,13	1	4
Sector	51272	2,10	0,97	1	3
Region	51272	3,87	2,40	1	8
Status	51272	1,62	0,49	1	2
Gender	51272	0,59	0,49	0	1

A-3.2 Descriptive Statistics by Size of Establishment

The first column of table A-7 reports results of all population and it is followed by columns of three establishment size categories. The objective is to see distribution in each establishment size category.

Main observations are summarized below:

- Gender comparison reveals that there are more male in the population of this survey for the year 2005 compared to females. There is approximately 59% males' population in the survey against 41% of females. The proportion of gender distribution is similar across establishment sizes
- The largest age group is 31-40 making 32% of the sample followed by 41-50 making 31% of the sample.
- 96% employment contracts are fixed term contracts in whole sample population and across establishments.
- A majority of population in the sample has completed between 0-5 years in the current job. This is stronger in small size establishments with 43% of the population, followed by 39% in medium and 32% in large. In large size establishments 21% of the population in the sample has completed 21-30 years with current employer. This reflects the fact that shirking is costly is large establishments, therefore, large employer may attract workers with different packages in order to reduce quit rates.
- Non-management cadre sample is around 62% of the workforce. In medium and large the management to non management cadre status ratio is 40:60, whereas, in small it becomes 32:67.
- In this survey 32% of population in the sample is associated with management and high intellectual professions. This trend is similar across establishments unlike in the previous surveys where blue collar workers made up the largest population in the sample. Second highest population in the sample is the high skilled white collar professions around 27%. This is similar across establishments. With this we would expect the highest percentage of higher educated people, which is true, as 42% of sample is with higher educational level. In large the percentage of the highly educated employees is highest, i.e. 46%,

followed by 41% in medium size establishments. In small establishments the population with higher educational level is 36%.

- The sectoral distribution is in line with the previous surveys. The largest sector is services and smallest sector is trade. In large establishments, we see the similar pattern unlike 1992, where largest sector in large establishments was manufacturing.
- Ile de France is the biggest region where 24% of the sample works and establishments are based. It is followed by 17% of basin Parisian. This pattern is similar across employer sizes.

Table A-7 Descriptive Statistics of all qualitative variables

Variables	**All**		**Small**		**Medium**		**Large**	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Females	21,225	41.40	5,089	41.19	5,498	39.49	10,638	42.56
Males	30,047	58.60	7,265	58.81	8,425	60.51	14,357	57.44
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00
Age 25-30	7,188	14.02	1,769	14.32	1,988	14.28	3,431	13.73
Age 31-40	16,446	32.08	3,988	32.28	4,581	32.90	7,877	31.51
Age 41-50	16,225	31.64	3,874	31.36	4,239	30.45	8,112	32.45
Age 51-60	11,413	22.26	2,723	22.04	3,115	22.37	5,575	22.30
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00
CDI	49,271	96.10	11,879	96.16	13,36	95.96	24,032	96.15
CDD	1,816	3.54	465	3.76	552	3.96	799	3.20
Other	185	0.36	10	0.08	11	0.08	164	0.66
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00
Tenure 0-5	18,91	36.88	5,375	43.51	5,508	39.56	8,027	32.11
.6-10	8,325	16.24	2,236	18.10	2,278	16.36	3,811	15.25
.11-20	11,304	22.05	2,600	21.05	3,137	22.53	5,567	22.27
21-30	9,065	17.68	1,565	12.67	2,059	14.79	5,441	21.77
31-45.5	3,668	7.15	578	4.68	941	6.76	2,149	8.60
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00
CD	19,571	38.17	4,045	32.74	5,764	41.40	9,762	39.06
NC	31,701	61.83	8,309	67.26	8,159	58.60	15,233	60.94
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00
Management and High Intellectual professionals	16,558	32.29	3,218	26.05	4,95	35.55	8,390	33.57
High Skilled White Collar	14,322	27.93	3,643	29.49	3,398	24.41	7,281	29.13
Low Skilled White Collar	9,493	18.51	2,495	20.20	2,349	16.87	4,649	18.60
Blue collar	10,899	21.26	2,998	24.27	3,226	23.17	4,675	18.70
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00

Manufacturing	21,636	42.20	4,72	38.21	5,984	42.98	10,932	43.74
Trade	2,704	5.27	764	6.18	1,084	7.79	856	3.42
Services	26,932	52.53	6,87	55.61	6,855	49.24	13,207	52.84
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00
primary education	5,364	10.46	1,499	12.13	1,708	12.27	2,157	8.63
Secondary	2,733	5.33	656	5.31	743	5.34	1,334	5.34
Technical Short	13,412	26.16	3,395	27.48	3,534	25.38	6,483	25.94
Technical Long	7,82	15.25	2,268	18.36	2,167	15.56	3,385	13.54
Higher	21,943	42.80	4,536	36.72	5,771	41.45	11,636	46.55
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00
ile de France	12,188	23.77	2,174	17.60	3,481	25.00	6,533	26.14
Bassin Parisien	8,771	17.11	1,93	15.62	2,286	16.42	4,555	18.22
NORD – PAS-DE-CALAIS	3,47	6.77	914	7.40	960	6.90	1,596	6.39
EST	5,452	10.63	1,484	12.01	1,291	9.27	2,677	10.71
OUEST	6,287	12.26	1,716	13.89	1,899	13.64	2,672	10.69
SUD-OUEST	4,625	9.02	1,247	10.09	1,19	8.55	2,188	8.75
CENTRE-EST	6,296	12.28	1,721	13.93	1,688	12.12	2,887	11.55
MÉDITERRANÉE	4,183	8.16	1,168	9.45	1,128	8.10	1,887	7.55
Total	51,272	100.00	12,354	100.00	13,923	100.00	24,995	100.00

A-3.3 Descriptive statistics by Mean Hourly Wage

In the next step, descriptive statistics by mean hourly wage are presented in Table A-8 to compare the average wage in different establishment size groups, professions, sectors and other worker related and employer related variables. It will be followed by cross tabulation of selected variables. Main observations are summarized below:

- Mean hourly wage is higher for male, increasing level of education, tenure, age, in region ile de France, for high level professions, in manufacturing sector, for CDI contracts and for management status.
- Among establishment size groups, in all cases mean hourly wage of all variables is highest in large size establishments. In some cases mean hourly wage difference between medium and large is very low or does not exist.
- Overall gender wage difference is higher in medium and large size establishments compared to small size establishments.
- Mean hourly wage increases with educational levels and higher for large size establishment. Mean hourly wage for higher educational level is similar in medium and large size establishments.
- For professional categories mean hourly wage is higher in medium and large compared to small size establishments except in blue collar jobs where mean hourly wage in large establishments is higher among all categories of size.
- Overall, we see higher mean hourly wage in the manufacturing sector followed by trade sector. In small size establishments the highest mean hourly wage is in services sector, in medium mean hourly wage in trade and services is similar, whereas, in large size establishments, mean hourly wage in trade and manufacturing is almost similar. The detail table of all industries with respect to six categories of establishment size is presented in *Appendix C table-8* where there are three sub-categories of the trade sector with the mean hourly wage in the large size employers varying between 18 to 29euros. On the other hand, as manufacturing sector is a very large sector and the mean hourly wage in the large size employer with 200 and more workers varies between 17 to 38euros. On the whole, the average hourly wage in both sectors become equal by including all sub-categories

- Mean hourly wage in region ile de France is highest among all regions. Among establishments, mean hourly wage in large is higher or equal to medium across regions.

Table A-8 Descriptive Statistics by Mean hourly wage

Variables	All			small			medium			large		
Gender	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.	Mean	Std. Dev.	Freq.
Females	17,19	8,62	21225	15,83	7,82	5089	17,24	8,89	5498	17,82	8,77	10638
Males	22,81	12,87	30047	20,36	12,23	7265	22,73	12,93	8425	24,09	12,97	14357
Total	20,48	11,64	5127	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995
Education	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
primary education	13,93	6,71	5364	13,16	6,20	1499	13,52	6,50	1708	14,78	7,10	2157
Secondary	16,48	8,71	2733	15,01	8,61	656	16,36	8,50	743	17,27	8,79	1334
Technical Short	15,94	7,49	13412	15,36	7,45	3395	16,15	8,15	3534	16,13	7,11	6483
Technical Long	19,38	10,13	7820	18,37	10,29	2268	20,01	10,59	2167	19,66	9,67	3385
Higher	25,75	13,11	21943	23,16	12,78	4536	26,10	13,14	5771	26,59	13,10	11636
Total	20,48	11,64	51272	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995
Contract	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
CDI	20,67	11,68	49271	18,64	10,91	11879	20,76	11,89	13360	21,62	11,79	24032
CDD	16,38	9,98	1816	14,95	8,93	465	15,98	8,62	552	17,49	11,24	799
Other	12,05	3,67	185	10,59	1,96	10	14,15	6,96	11	12,00	3,40	164
Total	20,48	11,64	51272	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995
Tenure	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Tenure 0-5	18,72	11,21	18910	16,42	9,75	5375	19,00	11,60	5508	20,06	11,61	8027
.6-10	20,27	11,52	8325	18,70	10,61	2236	20,41	11,55	2278	21,12	11,91	3811
.11-20	21,76	12,27	11304	20,21	11,91	2600	21,84	12,13	3137	22,44	12,46	5567
21-30	21,46	10,99	9065	20,83	11,19	1565	21,50	11,19	2059	21,62	10,86	5441
31-45.5	23,71	12,17	3668	22,92	11,87	578	23,78	12,59	941	23,89	12,05	2149

Total	20,48	11,64	51272	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995
Age	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Age 25-30	16,06	7,39	7188	14,17	6,26	1769	16,14	7,62	1988	16,99	7,60	3431
Age 31-40	19,53	10,43	16446	17,58	9,57	3988	19,59	10,73	4581	20,48	10,53	7877
Age 41-50	21,21	12,12	16225	19,22	11,38	3874	21,28	12,16	4239	22,13	12,32	8112
Age 51-60	23,61	13,59	11413	21,61	12,97	2723	23,85	13,82	3115	24,45	13,66	5575
Total	20,48	11,64	51272	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995
Profession	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Management and High Intellectual professionals	30,80	12,98	16558	28,97	13,14	3218	30,16	12,61	4950	31,88	13,02	8390
High Skilled White Collar	18,88	7,74	14322	18,37	8,34	3643	19,18	8,55	3398	19,00	6,98	7281
Low Skilled White Collar	12,96	4,46	9493	12,72	4,64	2495	12,83	4,58	2349	13,15	4,29	4649
Blue collar	13,47	4,69	10899	12,21	3,72	2998	12,93	4,59	3226	14,65	5,03	4675
Total	20,48	11,64	51272	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995
Industry/Sector	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
Manufacturing	21,95	12,00	21636	17,96	10,42	4720	20,91	11,85	5984	24,25	12,18	10932
Trade	20,87	12,10	2704	17,68	10,43	764	19,84	11,92	1084	25,01	12,59	856
Services	19,27	11,15	26932	18,95	11,19	6870	20,38	11,76	6855	18,85	10,76	13207
Total	20,48	11,64	51272	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995
Region	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
ile de France	25,43	13,82	12188	23,72	14,25	2174	25,44	13,47	3481	25,99	13,80	6533
Bassin Parisien	20,67	11,81	8771	17,11	9,55	1930	21,39	12,20	2286	21,83	12,18	4555
NORD – PAS-DE-CALAIS	18,55	9,80	3470	17,33	9,41	914	19,35	10,63	960	18,78	9,42	1596
EST	18,11	9,60	5452	17,42	9,63	1484	18,32	9,81	1291	18,40	9,47	2677

OUEST	17,50	9,48	6287	16,97	9,70	1716	16,90	9,30	1899	18,27	9,41	2672
SUD-OUEST	17,95	9,62	4625	17,01	9,27	1247	17,22	9,47	1190	18,87	9,81	2188
CENTRE-EST	19,82	10,61	6296	18,30	10,03	1721	19,65	11,18	1688	20,83	10,50	2887
MÉDITERRANÉE	18,65	9,98	4183	17,44	9,60	1168	18,54	10,36	1128	19,47	9,90	1887
Total	20,48	11,64	51272	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995
Status	Mean	Std. Dev.	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.	Mean	Std, Dev,	Freq.
NC	30,01	12,88	19571	28,24	13,04	4045	29,59	12,64	5764	31,00	12,86	9762
CD	14,60	5,10	31701	13,75	4,91	8309	14,19	5,22	8159	15,28	5,06	15233
Total	20,48	11,64	51272	18,49	10,87	12354	20,57	11,82	13923	21,42	11,79	24995

A-4 Descriptive Statistics 2006

A-4.1 Summary Statistics

This section presents descriptive statistics and cross tabulation using the cross section for the year 2006. All the variables are available in this survey as used in the 2005 survey. Table-2.33 presents summary statistics of all variables. The variables have been cleaned by various changes, for instance, the observations are dropped for non declared education (23408), person with zero tenure or not declared any (3150), working hours more than 2280 annual hours (694), outliers in hourly wage (1385), and below 25 years of age and more than 60 years of age (5006).

The categories for education, sector/industry, education, establishment size and profession are formed in the same way as described above. Hourly wage is computed in the same manner by taking gross annual wage and annual working hours. Total number of observations after removing outliers in the data is 53,508. Mean hourly wage of the population in the sample is 21 euros, mean tenure is 13 years and mean age is 42 years. The description of categorical variables by establishment size is described below in Table A-9.

Table A-9 Summary Statistics of all Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Hourly Wage	53508	21,23	12,03	7,47	99,95
Size	53508	2,21	0,83	1	3
Education	53508	3,79	1,32	1	5
Contract	53508	1,04	0,22	1	3
Tenure	53508	13,05	10,28	0,08	44,08
Age	53508	42,03	9,46	25	60
Profession	53508	2,25	1,12	1	4
Sector	53508	2,11	0,96	1	3
Region	53508	3,89	2,39	1	8
Status	53508	1,64	0,48	1	2
Gender	53508	0,58	0,49	0	1

A-4.2 Descriptive statistics by Size of Establishment

*Table-A-10*shows distribution of employed workers across categories of variables used related to employer and employees characteristics. The first column reports results of all population and it is followed by columns of three establishment size categories. The objective is to see distribution in each establishment size category. Main observations are summarized below.

- Male population in the sample is around 58% in the surveys. The difference is more in medium size establishments compared to small and large.
- 43% of the population in the sample has acquired higher education. In large size establishments, 48% employees are highly education followed by 42% in medium and 37% in small. After higher education, the second highest percentage among educational categories is coming from technical short education which is 23 to 26% in different sizes of establishments.
- 96% employment contracts are fixed term contracts in all categories

- A majority of population in the sample is associated with the current employer from 0-5 years (34%) followed by 11-20 years (22%). In large size establishments around 42% of the population in the sample has completed 11-30 years with current employer. This again confirms that large employer may attract workers with different packages in order to reduce quit rates.
- Majority of sample is between 31-50 years of age in all establishment size categories and in the sample of whole population.
- In this survey 33% of the population is associated with management and high intellectual professions followed by high skilled white collar jobs. In small the largest percentage is high skilled white collar and in medium and large, the largest percentages of professions are management and high intellectual professions.
- The largest sector is services and smallest sector is trade. More than 50% of the population is associated with services sector who are working in small and large size establishments.
- Ile de France is the biggest region where 23% of the sample works and establishments are based. It is followed by 17% of establishments in region Basin Parisian. This pattern is similar across establishment sizes.
- The sample population doing with non-management cadre is around 64%. In small the highest non-management status is observed (70%) compared to 59% in medium and 62% in large.

A-10 Descriptive Statistics of all qualitative variables

Variables	All		Small		Medium		Large	
Gender	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Females	22,314	41.70	5,755	41.12	5,426	38.39	11,133	43.86
Males	31,194	58.30	8,239	58.88	8,707	61.61	14,248	56.14
Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00
Education	Number	Percent	Number	Percent	Number	Percent	Number	Percent
primary education	5,248	9.81	1,566	11.19	1,594	11.28	2,088	8.23
Secondary	2,853	5.33	779	5.57	661	4.68	1,413	5.57
Technical Short	13,439	25.12	3,723	26.60	3,647	25.80	6,069	23.91
Technical Long	8,508	15.90	2,693	19.24	2,285	16.17	3,530	13.91
Higher	23,460	43.84	5,233	37.39	5,946	42.07	12,281	48.39
Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00
Contract	Number	Percent	Number	Percent	Number	Percent	Number	Percent
CDI	51,478	96.21	13,467	96.23	13,721	97.08	24,290	95.70
CDD	1,844	3.45	515	3.68	400	2.83	929	3.66
Other	186	0.35	12	0.09	12	0.08	162	0.64
Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00
Tenure	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Tenure 0-5	18,233	34.08	5,708	40.79	5,006	35.42	7,519	29.62
.6-10	10,153	18.97	2,854	20.39	2,790	19.74	4,509	17.77
.11-20	12,067	22.55	3,051	21.80	3,199	22.63	5,817	22.92
21-30	9,008	16.83	1,700	12.15	2,085	14.75	5,223	20.58
31-45.5	4,047	7.56	681	4.87	1,053	7.45	2,313	9.11
Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00
Age	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Age 25-30	7,348	13.73	2,035	14.54	1,884	13.33	3,429	13.51
Age 31-40	16,830	31.45	4,561	32.59	4,445	31.45	7,824	30.83
Age 41-50	17,023	31.81	4,331	30.95	4,492	31.78	8,200	32.31
Age 51-60	12,307	23.00	3,067	21.92	3,312	23.43	5,928	23.36

Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00
Profession	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Management and High Intellectual professionals	17,902	33.46	3,639	26.00	5,160	36.51	9,103	35.87
High Skilled White Collar	15,381	28.75	4,316	30.84	3,551	25.13	7,514	29.60
Low Skilled White Collar	9,386	17.54	2,679	19.14	2,201	15.57	4,506	17.75
Blue collar	10,839	20.26	3,360	24.01	3,221	22.79	4,258	16.78
Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00
Industry/Sector	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Manufacturing	22,19	41.47	5,126	36.63	6,284	44.46	10,78	42.47
Trade	3,125	5.84	961	6.87	1,249	8.84	915	3.61
Services	28,193	52.69	7,907	56.50	6,600	46.70	13,686	53.92
Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00
Region	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Ile de France	12,535	23.43	2,736	19.55	3,750	26.53	6,049	23.83
Bassin Parisien	8,875	16.59	2,135	15.26	2,156	15.26	4,584	18.06
NORD – PAS-DE-CALAIS	3,741	6.99	1,085	7.75	1,074	7.60	1,582	6.23
EST	5,917	11.06	1,689	12.07	1,322	9.35	2,906	11.45
OUEST	6,781	12.67	1,911	13.66	1,853	13.11	3,017	11.89
SUD-OUEST	4,864	9.09	1,424	10.18	1,177	8.33	2,263	8.92
CENTRE-EST	6,528	12.20	1,735	12.40	1,705	12.06	3,088	12.17
MÉDITERRANÉE	4,267	7.97	1,279	9.14	1,096	7.75	1,892	7.45
Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00
Status	Number	Percent	Number	Percent	Number	Percent	Number	Percent
NC	19,414	36.28	4,141	29.59	5,763	40.78	9,510	37.47
CD	34,094	63.72	9,853	70.41	8,370	59.22	15,871	62.53
Total	53,508	100.00	13,994	100.00	14,133	100.00	25,381	100.00

A-4.3 Descriptive statistics by Mean Hourly Wage

In the next step, descriptive statistics by mean hourly wage are presented in table a-11 below to compare the average wage in different establishment size groups, professions, sectors and other worker related and employer related variables. It will be followed by cross tabulation of selected variables. Main observations are summarized below:

- Descriptive statistics by mean hourly wage follow the same patterns of the mean hourly wage for the individual and employer related characteristics as observed in the preceding section. As mentioned above we see higher mean hourly wage for male workers compared to females, increasing level of education, tenure, age, in region ile de France, for high level professions, in manufacturing sector, for CDI contracts and for management status.
- Gender wage difference is higher in large compared to small and medium
- Mean hourly wage increases with educational levels. Mean hourly wages across categories of level of education is similar in medium and large but higher compared to small size establishments
- For tenure mean hourly wage remains equal in medium and large
- For age groups and type of contracts, mean hourly wage in large is higher
- For professional categories mean hourly wage is higher for higher level of professions in all size groups and across establishments. The mean hourly wage in large size establishments for all professions is higher compared to small and medium
- Mean hourly wage is higher in the manufacturing sector followed by trade sector; in large the same pattern in observed, while in small size establishments, mean hourly wage is similar in manufacturing and services sector. In the medium size establishments, mean hourly wage does not differ by sectors.
- The same results for regions are observed where region ile de france is offering highest mean hourly wage compared to other regions. Overall mean hourly wage in large regions is higher. Highest mean hourly wage wage in region ile de france is in medium scale establishments.

Table A-11 Descriptive Statistics by mean hourly wage

Variables	**All**		**small**		**medium**		**large**	
Gender	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Females	17,87	9,11	16,56	8,74	18,05	9,71	18,45	8,93
Males	23,64	13,23	20,96	12,81	23,77	13,63	25,11	12,99
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86
Education	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,
primary education	14,55	7,22	13,55	6,83	14,25	7,09	15,52	7,48
Secondary	16,63	7,70	15,58	7,68	16,80	8,11	17,14	7,46
Technical Short	16,35	7,70	15,39	7,77	16,68	8,22	16,73	7,26
Technical Long	20,30	10,60	18,98	10,77	20,90	11,06	20,91	10,07
Higher	26,43	13,53	24,12	13,55	27,32	14,22	26,98	13,07
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86
Contract	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,
CDI	21,42	12,06	19,30	11,58	21,74	12,60	22,41	11,85
CDD	16,91	10,55	15,45	8,57	15,88	10,43	18,17	11,41
Other	12,78	6,66	17,91	21,82	12,95	3,71	12,38	3,92
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86
Tenure	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,
Tenure 0-5	19,06	11,06	16,74	9,80	19,67	11,70	20,41	11,25
.6-10	20,51	11,51	18,86	11,11	20,80	11,81	21,38	11,46
.11-20	22,83	13,02	21,17	12,91	23,26	13,50	23,45	12,73
21-30	22,88	11,99	22,39	12,39	23,02	12,94	22,99	11,46
31-45.5	24,43	12,63	23,42	12,85	24,67	13,35	24,61	12,21
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86
Age	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,
Age 25-30	16,16	6,72	14,24	5,63	16,21	7,05	17,28	6,87
Age 31-40	19,98	10,27	17,77	9,41	20,27	10,64	21,11	10,33
Age 41-50	22,25	12,82	20,22	12,28	22,61	13,57	23,13	12,55

Age 51-60	24,57	14,19	22,95	14,33	24,96	14,64	25,18	13,80
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86
Profession	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,
Management and High Intellectual professionals	31,55	13,14	30,26	13,75	31,42	13,34	32,14	12,75
High Skilled White Collar	19,31	8,37	18,88	9,25	19,81	9,55	19,32	7,14
Low Skilled White Collar	13,28	4,78	12,77	4,45	13,27	5,34	13,60	4,65
Blue collar	13,80	4,40	12,56	3,60	13,40	4,11	15,09	4,82
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86
Industry/Sector	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,
Manufacturing	22,90	12,48	19,07	11,27	21,71	12,13	25,43	12,66
Trade	20,73	12,12	17,60	10,52	21,07	12,89	23,56	11,85
Services	19,97	11,49	19,40	11,78	21,54	12,94	19,55	10,48
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86
Region	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,
ile de France	26,18	14,15	24,11	14,93	27,32	14,69	26,41	13,33
Bassin Parisien	21,17	11,80	17,93	10,38	21,45	12,15	22,55	11,97
NORD – PAS-DE-CALAIS	19,18	10,28	17,42	8,76	19,34	10,80	20,29	10,72
EST	19,05	10,21	18,88	10,85	18,81	10,33	19,25	9,77
OUEST	18,70	10,35	17,48	10,21	18,07	10,06	19,85	10,49
SUD-OUEST	18,66	10,14	16,97	9,32	18,24	10,54	19,94	10,26
CENTRE-EST	20,32	10,94	18,40	10,68	20,32	11,44	21,40	10,66
MÉDITERRANÉE	20,01	11,28	18,35	10,01	19,08	10,37	21,67	12,33
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86
Status	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,	Mean	Std, Dev,
NC	31,42	13,34	30,43	13,97	31,22	13,61	31,96	12,86
CD	15,44	5,77	14,41	5,52	14,93	5,62	16,34	5,86
Total	21,23	12,03	19,15	11,52	21,57	12,58	22,19	11,86

Table-A-12 Comparison of Mean Gender Wage difference by Size 1992-2006

Variables	1992			2002			2005			2006		
	Small	Medium	Large	Small	Medium	Large	Small	Medium	Large	Small	Medium	Large
Education												
primary education	9.50	11.57	14.18	0.96	1.41	2.88	1.78	2.94	4.12	2.70	3.18	3.97
Secondary	17.86	15.97	18.96	1.14	2.09	4.25	2.85	3.44	4.98	3.18	4.31	2.83
Technical Short	6.36	8.69	12.94	1.32	1.57	2.70	2.50	3.55	3.81	2.13	3.49	4.61
Technical Long	22.05	21.05	19.61	4.07	3.67	4.35	5.63	5.56	4.96	5.41	6.07	5.54
Higher	26.15	24.03	31.10	7.75	7.10	6.35	8.30	7.83	8.36	8.16	7.69	8.85
Total	10.79	11.47	16.05	2.91	3.26	4.31	4.53	5.49	6.27	4.40	5.72	6.66
Contract												
CDI	10.82	11.78	15.77	2.78	3.25	4.17	4.52	5.51	6.14	4.33	5.63	6.54
CDD	8.16	6.07	8.01	3.67	3.07	2.23	3.42	2.27	7.07	4.23	4.93	6.54
Other				10.78	0.24	4.97	-2.29	-2.13	0.19	20.24	-0.16	0.59
Total	10.79	11.47	16.05	2.91	3.26	4.31	4.53	5.49	6.27	4.40	5.72	6.66
Tenure												
Tenure 0-5	9.99	10.14	14.84	2.69	3.28	3.67	3.75	4.99	6.01	3.58	4.54	6.00
.6-10	11.59	10.76	16.50	3.03	3.89	3.39	4.86	5.93	5.45	3.65	5.10	5.94
.11-20	11.64	11.99	14.45	3.45	3.52	4.66	4.55	5.52	6.94	4.29	6.65	7.46
21-30	11.42	12.02	15.58	2.95	2.55	4.71	4.81	4.52	5.83	6.05	6.40	6.47
31-45.5	8.46	17.82	14.79	3.22	2.93	3.87	5.48	7.72	6.75	6.93	6.67	7.18
Total	10.79	11.47	16.05	2.91	3.26	4.31	4.53	5.49	6.27	4.40	5.72	6.66
Age												
Age 25-30	4.55	5.11	7.31	1.53	1.20	1.69	1.46	1.54	2.68	1.13	1.93	2.94
Age 31-40	9.87	9.76	13.81	2.33	2.94	2.80	3.76	3.91	4.80	2.93	3.60	5.10
Age 41-50	16.55	17.02	18.69	2.78	3.52	4.86	4.68	6.57	6.75	4.91	6.54	7.56
Age 51-60	14.68	16.03	23.20	5.20	4.71	6.59	6.62	8.08	9.21	7.55	8.87	8.90
Total	10.79	11.47	16.05	2.91	3.26	4.31	4.53	5.49	6.27	4.40	5.72	6.66
Experience												
Experience 0-5	11.59	10.19	10.20									

Experience 6-10	6.24	4.99	10.44									
Experience 11-20	8.87	9.89	14.40									
Experience 21-30	14.11	14.54	17.57									
Experience 31-40	15.72	16.69	20.39									
Experience 41-49	9.13	13.92	19.07									
Total	10.79	11.47	16.05									
Family Situation												
Single	0.74	2.64	4.60									
Married	14.83	14.15	19.05									
Others (widowed, divorced)	10.73	13.62	15.29									
Total	10.79	11.47	16.05									
Profession												
Management and High Intellectual professionals	15.98	12.27	20.33	7.02	5.38	5.25	6.56	5.25	4.57	6.00	4.95	4.41
High Skilled White Collar	7.30	6.69	9.17	1.96	1.68	2.43	3.09	3.11	3.03	3.86	4.04	3.11
Low Skilled White Collar	3.91	4.40	5.41	0.70	0.03	0.04	1.01	1.44	1.26	1.16	1.85	1.82
Blue collar	9.64	12.19	16.06	1.51	2.26	3.76	2.16	2.68	2.58	2.13	2.52	2.90
Total	10.79	11.47	16.05	2.91	3.26	4.31	4.53	5.49	6.27	4.40	5.72	6.66
Industry/Sector												
Manufacturing	8.90	15.31	16.42	2.46	3.96	4.32	4.16	5.09	5.04	3.87	5.18	4.85
Trade	14.34	14.51	22.48	4.38	4.30	5.34	5.51	6.30	6.19	3.85	6.53	6.44
Services	13.21	10.05	14.99	4.30	4.20	4.88	5.54	5.96	4.96	5.43	6.57	5.44
Total	10.79	11.47	16.05	2.91	3.26	4.31	4.53	5.49	6.27	4.40	5.72	6.66
Region												
Ile de France (11)	12.79	14.21	18.74									
Champagne-Ardenne(21)	8.02	2.60	15.69									
Picardie (22)	10.98	12.36	16.70									
Haute-Normandie (23)	10.06	9.62	25.87									
Centre (24)	8.35	12.63	18.24									
Basse-Normandie (25)	6.11	9.23	12.53									

Bourgogne (26)	10.30	11.26	15.55									
Nord (31)	10.89	10.83	16.18									
Lorraine (41)	13.86	13.01	15.31									
Alsace (42)	17.85	14.69	15.84									
Franche-Comte (43)	8.96	10.47	14.14									
Pays de la Loire (52)	14.10	9.49	16.91									
Bretagne (53)	10.72	13.17	9.23									
Poitou-Charentes (54)	7.60	10.08	9.19									
Aquitaine (72)	11.81	11.96	24.92									
Midi-Pyrenees (73)	9.41	12.03	25.04									
Limousin (74)	2.56	4.05	15.71									
Rhone-Alpes (82)	11.33	11.71	20.25									
Auvergne (83)	7.26	11.00	9.59									
Languedoc-Roussillon (91)	14.27	10.84	15.46									
Provence-Alpes-Cote d'Azur (93)	10.83	10.69	15.40									
Total	10.79	11.47	16.05									
Nationality												
Other Nationality	7.35	12.13	9.95	0.14	0.91	4.39						
French	11.18	11.72	16.55	3.16	3.51	4.34						
Total	10.79	11.47	16.05	2.91	3.26	4.31						
Region												
ile de France				3.97	4.32	4.77	6.67	5.94	6.63	5.22	6.02	6.20
Bassin Parisien				2.74	2.90	5.48	3.87	7.83	7.70	3.91	7.31	7.77
NORD				2.28	3.21	4.40	3.16	5.07	4.31	3.58	5.28	5.84
EST				3.63	2.66	5.67	4.48	5.85	5.18	5.64	5.61	5.92
OUEST				2.43	4.34	4.54	4.53	4.33	6.10	4.48	4.96	6.42
SUD-OUEST				1.95	2.75	3.83	4.31	3.08	5.65	3.14	4.50	6.87
CENTRE-EST				3.28	4.66	4.64	5.23	6.67	5.84	5.38	5.62	6.61
MÉDITERRANÉE				2.49	3.84	6.19	4.19	5.84	6.88	4.54	6.32	8.20

Total		2.91	3.26	4.31	4.53	5.49	6.27	4.40	5.72	6.66
Status										
NC		0.78	1.13	2.32	5.97	5.52	4.67	5.98	5.22	4.57
CD		7.02	5.38	5.25	0.97	1.46	1.60	1.10	1.77	2.14
Total		2.91	3.26	4.31	4.53	5.49	6.27	4.40	5.72	6.66

Appendix - B Chapter3

Table-B-1 Regression Estimates of the ESWP

List of Explanatory variables	All population		Male		Female	
	Gross Hourly Wage	Net Hourly Wage	Gross Hourly Wage	Net Hourly Wage	Gross Hourly Wage	Net Hourly Wage
Establishment size dummy	0.072***	0.008***	0.088***	0.018***	0.053***	-0.006**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Gender (base female)	0.137***	0.101***				
	(0.002)	(0.002)				
Experience	0.024***	0.022***	0.028***	0.025***	0.021***	0.021***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Experience sq.	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experience cube	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.018***	0.008***	0.019***	0.008***	0.016***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tenure Sq.	-0.001***	-0.000***	-0.001***	-0.000***	-0.000***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure cube	0.000***	0.000***	0.000***	0.000***	0.000***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Family Status (Base single)						
Married	0.030***	0.022***	0.051***	0.039***	-0.001	-0.005
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Other family status (divorced, widowed etc)	0.021***	0.016***	0.026***	0.022***	0.011**	0.002
	(0.004)	(0.004)	(0.006)	(0.006)	(0.005)	(0.005)
Level of Education (base no degree)						
Before Bac without degree	0.063***	0.042***	0.066***	0.049***	0.050***	0.025***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
CAP/BEP	0.089***	0.079***	0.082***	0.082***	0.079***	0.060***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
Bac professional and technical	0.164***	0.149***	0.159***	0.148***	0.151***	0.134***

	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
Bac general	0.177***	0.157***	0.165***	0.153***	0.170***	0.142***
	(0.005)	(0.005)	(0.007)	(0.007)	(0.006)	(0.006)
Bac +2	0.225***	0.214***	0.206***	0.209***	0.226***	0.205***
	(0.004)	(0.004)	(0.006)	(0.005)	(0.006)	(0.006)
Bac+3 and plus	0.336***	0.349***	0.315***	0.343***	0.361***	0.355***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.012)	(0.012)
Professions (base blue collar)						
Management and High Intellectual professionals	0.635***	0.674***	0.634***	0.679***	0.675***	0.680***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.009)	(0.009)
High Skilled White Collar	0.267***	0.279***	0.250***	0.268***	0.335***	0.324***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)
Low Skilled White Collar	0.053***	0.062***	0.002	0.019***	0.125***	0.115***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)	(0.003)
CDI Contract(base CDD)	0.027***	0.055***	0.043***	0.075***	0.011*	0.037***
	(0.005)	(0.005)	(0.008)	(0.007)	(0.006)	(0.006)
Industry (base Trade)						
Manufacturing	0.043***	0.037***	0.029***	0.018***	0.069***	0.072***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)
Services	0.039***	0.031***	0.030***	0.022***	0.048***	0.040***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Region (Base regional category (Franche-Comte (43)**))**						
Ile de France (11)	0.182***	0.167***	0.167***	0.158***	0.197***	0.178***
	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
Champagne-Ardenne(21)	0.053***	0.026***	0.065***	0.033***	0.023**	0.012
	(0.008)	(0.008)	(0.010)	(0.010)	(0.012)	(0.012)
Picardie (22)	0.067***	0.043***	0.077***	0.045***	0.053***	0.042***
	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)	(0.010)
Haute-Normandie (23)	0.076***	0.040***	0.093***	0.043***	0.049***	0.035***
	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)	(0.009)
Centre (24)	0.043***	0.041***	0.051***	0.051***	0.027***	0.022**

	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)	(0.010)
Basse-Normandie (25)	0.016**	0.016**	0.013	0.012	0.019*	0.022**
	(0.008)	(0.007)	(0.010)	(0.011)	(0.011)	(0.010)
Bourgogne (26)	0.026***	0.019***	0.034***	0.029***	0.012	0.003
	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)	(0.009)
Nord (31)	0.015**	0.001	0.018**	-0.002	0.010	0.010
	(0.006)	(0.006)	(0.008)	(0.009)	(0.009)	(0.009)
Lorraine (41)	0.077***	0.038***	0.090***	0.038***	0.046***	0.040***
	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)	(0.010)
Alsace (42)	0.092***	0.073***	0.113***	0.096***	0.059***	0.039***
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
Pays de la Loire (52)	0.015**	0.008	0.017**	0.016*	0.011	-0.005
	(0.006)	(0.006)	(0.008)	(0.009)	(0.009)	(0.009)
Bretagne (53)	0.006	-0.003	0.000	-0.004	0.015	-0.001
	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)
Poitou-Charentes (54)	0.048***	0.030***	0.042***	0.018*	0.060***	0.050***
	(0.007)	(0.007)	(0.010)	(0.010)	(0.011)	(0.010)
Aquitaine (72)	0.051***	0.033***	0.068***	0.053***	0.022**	0.004
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
Midi-Pyrenees (73)	0.021***	0.044***	0.035***	0.061***	0.008	0.026***
	(0.007)	(0.007)	(0.009)	(0.010)	(0.010)	(0.009)
Limousin (74)	-0.028***	-0.011	-0.031***	-0.004	-0.021*	-0.020*
	(0.008)	(0.008)	(0.010)	(0.010)	(0.012)	(0.011)
Rhone-Alpes (82)	0.077***	0.065***	0.088***	0.078***	0.057***	0.046***
	(0.006)	(0.006)	(0.008)	(0.008)	(0.009)	(0.008)
Auvergne (83)	0.001	0.005	-0.003	0.009	0.003	-0.003
	(0.008)	(0.008)	(0.010)	(0.011)	(0.011)	(0.011)
Languedoc-Roussillon (91)	0.013	-0.010	0.031**	-0.011	-0.012	-0.012
	(0.009)	(0.009)	(0.012)	(0.012)	(0.012)	(0.012)
Provence-Alpes-Cote d’Azur (93)	0.102***	0.087***	0.115***	0.098***	0.078***	0.069***
	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)
Constant	3.294***	3.218***	3.390***	3.278***	3.311***	3.244***
	(0.012)	(0.011)	(0.016)	(0.016)	(0.016)	(0.015)
Observations	74,762	74,453	44,699	44,516	30,063	29,937
R-squared	0.66	0.66	0.65	0.67	0.63	0.62
Pseudo R-squared	0.65	0.66	0.65	0.67	0.63	0.62

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Dependent variable is log of hourly wage, gross or net (basic).

Table-B-2 Regression Estimates across Industry

List of Explanatory variables	Trade		Manufacturing		Services	
	Gross Hourly Wage	Net Hourly Wage	Gross Hourly Wage	Net Hourly Wage	Gross Hourly Wage	Net Hourly Wage
Establishment size dummy	0.024***	-0.019**	0.100***	0.017***	0.059***	0.004*
	(0.008)	(0.008)	(0.003)	(0.003)	(0.003)	(0.003)
Gender (base female)	0.132***	0.101***	0.158***	0.107***	0.120***	0.087***
	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)
Experience	0.017***	0.015***	0.028***	0.024***	0.023***	0.023***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
Experience sq.	-0.000	-0.000	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experience cube	0.000	-0.000	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.022***	0.012***	0.016***	0.006***	0.019***	0.009***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Tenure Sq.	-0.001***	-0.001***	-0.001***	-0.000***	-0.000***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure cube	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married (base single)	0.018***	0.013**	0.030***	0.024***	0.032***	0.022***
	(0.007)	(0.007)	(0.004)	(0.003)	(0.003)	(0.003)
Other family status (divorced, widowed etc)	0.024**	0.021*	0.025***	0.021***	0.019***	0.011**
	(0.011)	(0.011)	(0.007)	(0.006)	(0.005)	(0.005)
Before Bac without degree (base no degree)	0.008	0.003	0.051***	0.037***	0.081***	0.051***
	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
CAP/BEP	0.026***	0.027***	0.092***	0.095***	0.096***	0.074***
	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
Bac professional and technical	0.105***	0.088***	0.161***	0.168***	0.171***	0.142***
	(0.013)	(0.013)	(0.007)	(0.006)	(0.006)	(0.006)
Bac general	0.110***	0.108***	0.179***	0.176***	0.188***	0.153***

	(0.014)	(0.014)	(0.009)	(0.008)	(0.006)	(0.006)
Bac +2	0.187***	0.176***	0.230***	0.236***	0.232***	0.211***
	(0.014)	(0.014)	(0.007)	(0.007)	(0.005)	(0.005)
Bac+3 and plus	0.218***	0.229***	0.350***	0.394***	0.350***	0.344***
	(0.023)	(0.023)	(0.010)	(0.010)	(0.009)	(0.009)
Management and High Intellectual professionals (base blue collar)	0.693***	0.706***	0.620***	0.687***	0.624***	0.638***
	(0.013)	(0.013)	(0.007)	(0.007)	(0.006)	(0.006)
High Skilled White Collar	0.324***	0.302***	0.265***	0.291***	0.248***	0.247***
	(0.008)	(0.009)	(0.004)	(0.004)	(0.004)	(0.004)
Low Skilled White Collar	0.053***	0.043***	0.100***	0.123***	0.034***	0.029***
	(0.007)	(0.007)	(0.005)	(0.005)	(0.003)	(0.003)
CDI Contract(base CDD)	0.043***	0.048***	0.015	0.029***	0.019***	0.051***
	(0.014)	(0.013)	(0.011)	(0.009)	(0.006)	(0.006)
Ile de France (11) (Base regional category (Franche-Comte (43))	0.204***	0.205***	0.178***	0.150***	0.174***	0.162***
	(0.016)	(0.016)	(0.009)	(0.009)	(0.008)	(0.009)
Champagne-Ardenne(21)	-0.017	0.023	0.030**	-0.007	0.070***	0.040***
	(0.025)	(0.024)	(0.012)	(0.011)	(0.011)	(0.011)
Picardie (22)	0.012	0.045*	0.116***	0.054***	0.020**	0.025**
	(0.024)	(0.025)	(0.011)	(0.010)	(0.010)	(0.010)
Haute-Normandie (23)	0.052***	0.080***	0.112***	0.055***	0.047***	0.013
	(0.020)	(0.020)	(0.011)	(0.010)	(0.010)	(0.010)
Centre (24)	0.018	0.032*	0.042***	0.015	0.046***	0.059***
	(0.019)	(0.019)	(0.010)	(0.010)	(0.010)	(0.010)
Basse-Normandie (25)	-0.003	0.002	0.046***	0.026**	-0.004	0.011
	(0.019)	(0.019)	(0.012)	(0.011)	(0.011)	(0.011)
Bourgogne (26)	-0.021	0.018	0.019*	-0.014	0.032***	0.032***
	(0.019)	(0.018)	(0.011)	(0.010)	(0.009)	(0.010)
Nord (31)	-0.000	-0.013	0.027***	-0.008	0.008	0.009
	(0.017)	(0.017)	(0.010)	(0.010)	(0.009)	(0.009)
Lorraine (41)	0.017	-0.001	0.100***	0.024**	0.057***	0.048***
	(0.023)	(0.021)	(0.010)	(0.010)	(0.010)	(0.010)

Alsace (42)	0.064***	0.059***	0.158***	0.092***	0.049***	0.056***
	(0.019)	(0.019)	(0.011)	(0.010)	(0.009)	(0.010)
Pays de la Loire (52)	0.039**	0.033**	0.023**	0.010	0.002	-0.004
	(0.018)	(0.017)	(0.010)	(0.009)	(0.009)	(0.010)
Bretagne (53)	-0.017	-0.007	0.023**	0.011	-0.006	-0.015*
	(0.019)	(0.019)	(0.010)	(0.010)	(0.009)	(0.009)
Poitou-Charentes (54)	-0.036	0.009	0.056***	0.014	0.045***	0.030***
	(0.022)	(0.023)	(0.013)	(0.012)	(0.010)	(0.010)
Aquitaine (72)	-0.031*	-0.004	0.129***	0.094***	0.015*	-0.002
	(0.019)	(0.018)	(0.011)	(0.011)	(0.009)	(0.009)
Midi-Pyrenees (73)	0.030	0.079***	0.045***	0.054***	0.001	0.024**
	(0.020)	(0.020)	(0.011)	(0.011)	(0.009)	(0.010)
Limousin (74)	-0.000	0.015	-0.026**	-0.027**	-0.036***	-0.008
	(0.027)	(0.027)	(0.012)	(0.011)	(0.010)	(0.011)
Rhone-Alpes (82)	0.041**	0.066***	0.101***	0.064***	0.059***	0.058***
	(0.017)	(0.017)	(0.010)	(0.009)	(0.008)	(0.009)
Auvergne (83)	-0.083***	-0.059**	0.028**	0.007	-0.010	0.012
	(0.022)	(0.023)	(0.012)	(0.012)	(0.011)	(0.011)
Languedoc-Roussillon (91)	-0.035	-0.020	0.043***	-0.007	-0.002	-0.016
	(0.024)	(0.024)	(0.016)	(0.016)	(0.012)	(0.012)
Provence-Alpes-Cote d'Azur (93)	0.033*	0.083***	0.134***	0.100***	0.091***	0.075***
	(0.017)	(0.017)	(0.011)	(0.011)	(0.009)	(0.009)
Constant	3.390***	3.312***	3.290***	3.255***	3.366***	3.282***
	(0.034)	(0.033)	(0.019)	(0.018)	(0.015)	(0.015)
Observations	7,648	7,612	28,571	28,491	38,543	38,350
R-squared	0.65	0.64	0.67	0.71	0.65	0.64
Pseudo R-squared	0.65	0.64	0.67	0.71	0.65	0.64

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.Dependent variable is log of hourly wage, gross or net (basic).

Table-B-3 Regression Estimates across Professions

List of Explanatory variables	Management and High Intellectual professionals		High Skilled White Collar		Low Skilled White Collar		Blue collar	
	Gross Hourly Wage	Net Hourly Wage	Gross Hourly Wage	Net Hourly Wage	Gross Hourly Wage	Net Hourly Wage	Gross Hourly Wage	Net Hourly Wage
Establishment size dummy	0.020***	-0.014*	0.045***	-0.018***	0.053***	-0.006*	0.132***	0.049***
	(0.007)	(0.007)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Gender (base female)	0.114***	0.099***	0.108***	0.084***	0.083***	0.052***	0.204***	0.131***
	(0.007)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Experience	0.048***	0.045***	0.017***	0.014***	0.012***	0.013***	0.005*	0.002
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Experience sq.	-0.001***	-0.001***	-0.000*	-0.000	-0.000**	-0.000**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experience cube	0.000***	0.000***	-0.000	-0.000	0.000	0.000	-0.000*	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.013***	0.004	0.015***	0.005***	0.018***	0.008***	0.022***	0.012***
	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tenure Sq.	-0.001***	-0.000**	-0.001***	-0.000*	-0.000*	0.000	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure cube	0.000***	0.000***	0.000***	0.000**	0.000	-0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married (base single)	0.066***	0.063***	0.025***	0.014***	0.012***	0.003	0.041***	0.033***
	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Other family status (divorced, widowed etc)	0.043***	0.027*	0.010	0.011	0.011*	-0.000	0.038***	0.033***
	(0.014)	(0.015)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)
Before Bac without degree (base no degree)	-0.048**	-0.049**	0.025***	0.004	0.048***	0.029***	0.066***	0.045***
	(0.019)	(0.019)	(0.008)	(0.009)	(0.005)	(0.005)	(0.004)	(0.003)
CAP/BEP	-0.077***	-0.056***	0.021***	0.015**	0.076***	0.057***	0.096***	0.093***

	(0.018)	(0.018)	(0.008)	(0.008)	(0.005)	(0.005)	(0.003)	(0.003)
Bac professional and technical	0.020	0.032*	0.096***	0.086***	0.169***	0.146***	0.199***	0.185***
	(0.018)	(0.018)	(0.008)	(0.009)	(0.007)	(0.007)	(0.010)	(0.009)
Bac general	0.057***	0.060***	0.125***	0.111***	0.152***	0.132***	0.159***	0.142***
	(0.019)	(0.019)	(0.010)	(0.010)	(0.007)	(0.007)	(0.015)	(0.013)
Bac +2	0.119***	0.146***	0.166***	0.164***	0.217***	0.194***	0.281***	0.257***
	(0.017)	(0.017)	(0.008)	(0.008)	(0.009)	(0.008)	(0.021)	(0.018)
Bac+3 and plus	0.258***	0.305***	0.321***	0.311***	0.291***	0.256***	0.045	0.026
	(0.017)	(0.017)	(0.016)	(0.016)	(0.037)	(0.032)	(0.088)	(0.072)
CDI Contract(base CDD)	0.114***	0.121***	0.015	0.024**	0.028***	0.054***	-0.016**	0.036***
	(0.022)	(0.024)	(0.012)	(0.012)	(0.007)	(0.006)	(0.007)	(0.006)
Manufacturing (base trade)	0.046***	0.057***	0.029***	0.041***	0.094***	0.105***	0.037***	0.006
	(0.012)	(0.012)	(0.007)	(0.007)	(0.006)	(0.006)	(0.005)	(0.005)
Services	0.036***	0.014	0.006	0.012*	0.057***	0.046***	0.051***	0.045***
	(0.011)	(0.011)	(0.007)	(0.007)	(0.004)	(0.004)	(0.005)	(0.005)
Ile de France (11) (Base regional category (Franche-Comte (43))	0.175***	0.168***	0.172***	0.151***	0.212***	0.197***	0.164***	0.157***
	(0.033)	(0.034)	(0.012)	(0.013)	(0.010)	(0.010)	(0.009)	(0.009)
Champagne-Ardenne(21)	0.064	0.064	0.022	0.022	0.025*	0.023*	0.061***	0.014
	(0.040)	(0.044)	(0.017)	(0.017)	(0.014)	(0.014)	(0.011)	(0.010)
Picardie (22)	0.067*	0.074**	0.049***	0.032**	0.043***	0.052***	0.079***	0.037***
	(0.036)	(0.037)	(0.015)	(0.016)	(0.013)	(0.013)	(0.009)	(0.009)
Haute-Normandie (23)	0.085**	0.082**	0.101***	0.053***	0.056***	0.038***	0.059***	0.014
	(0.036)	(0.037)	(0.014)	(0.015)	(0.012)	(0.011)	(0.010)	(0.009)
Centre (24)	0.066*	0.082**	0.047***	0.031**	0.025**	0.032***	0.037***	0.034***
	(0.036)	(0.038)	(0.014)	(0.015)	(0.012)	(0.012)	(0.009)	(0.009)
Basse-Normandie (25)	0.069	0.052	0.051***	0.035**	-0.019	0.005	0.003	0.005
	(0.042)	(0.045)	(0.016)	(0.017)	(0.013)	(0.012)	(0.010)	(0.010)
Bourgogne (26)	0.061	0.091**	0.025*	0.014	-0.001	0.012	0.020**	-0.002

	(0.037)	(0.039)	(0.015)	(0.015)	(0.012)	(0.011)	(0.009)	(0.009)
Nord (31)	0.072**	0.070*	0.020	-0.004	0.002	0.012	-0.001	-0.020**
	(0.035)	(0.036)	(0.013)	(0.014)	(0.011)	(0.010)	(0.009)	(0.008)
Lorraine (41)	0.035	0.004	0.098***	0.044***	0.070***	0.073***	0.061***	0.014*
	(0.040)	(0.040)	(0.015)	(0.015)	(0.012)	(0.011)	(0.009)	(0.009)
Alsace (42)	0.036	0.040	0.078***	0.057***	0.059***	0.052***	0.136***	0.108***
	(0.035)	(0.036)	(0.013)	(0.014)	(0.011)	(0.011)	(0.010)	(0.010)
Pays de la Loire (52)	0.079**	0.045	0.027**	0.014	0.003	-0.001	-0.008	-0.002
	(0.035)	(0.037)	(0.013)	(0.014)	(0.011)	(0.011)	(0.009)	(0.008)
Bretagne (53)	0.041	0.021	0.011	-0.009	0.010	0.003	-0.010	-0.010
	(0.037)	(0.039)	(0.013)	(0.014)	(0.011)	(0.010)	(0.009)	(0.009)
Poitou-Charentes (54)	0.061	0.054	0.058***	0.031*	0.053***	0.052***	0.032***	0.007
	(0.040)	(0.042)	(0.016)	(0.017)	(0.012)	(0.012)	(0.010)	(0.009)
Aquitaine (72)	0.050	0.053	0.053***	0.019	0.021**	0.009	0.072***	0.060***
	(0.035)	(0.037)	(0.013)	(0.014)	(0.011)	(0.010)	(0.010)	(0.010)
Midi-Pyrenees (73)	0.035	0.038	0.014	0.016	0.041***	0.064***	0.020**	0.059***
	(0.036)	(0.038)	(0.014)	(0.015)	(0.012)	(0.011)	(0.010)	(0.009)
Limousin (74)	0.043	0.087*	-0.022	-0.023	0.005	0.010	-0.067***	-0.028***
	(0.045)	(0.048)	(0.016)	(0.017)	(0.013)	(0.012)	(0.011)	(0.010)
Rhone-Alpes (82)	0.079**	0.081**	0.054***	0.033**	0.066***	0.061***	0.091***	0.081***
	(0.033)	(0.034)	(0.012)	(0.013)	(0.010)	(0.010)	(0.008)	(0.008)
Auvergne (83)	0.041	0.061	0.002	0.000	-0.009	0.007	-0.014	-0.012
	(0.046)	(0.048)	(0.016)	(0.017)	(0.013)	(0.012)	(0.011)	(0.010)
Languedoc-Roussillon (91)	-0.052	-0.074*	0.004	-0.036**	0.031**	0.029**	0.031**	0.013
	(0.044)	(0.044)	(0.016)	(0.016)	(0.014)	(0.014)	(0.014)	(0.013)
Provence-Alpes-Cote d'Azur (93)	0.072**	0.072**	0.088***	0.068***	0.092***	0.084***	0.122***	0.103***
	(0.034)	(0.035)	(0.013)	(0.014)	(0.011)	(0.010)	(0.009)	(0.009)
Constant	3.781***	3.696***	3.729***	3.666***	3.447***	3.371***	3.388***	3.346***
	(0.047)	(0.048)	(0.023)	(0.023)	(0.020)	(0.018)	(0.020)	(0.018)
Observations	9,322	9,252	20,909	20,856	18,054	17,973	26,477	26,372
R-squared	0.29	0.26	0.27	0.19	0.36	0.27	0.41	0.27
Pseudo R-squared	0.29	0.26	0.26	0.19	0.35	0.26	0.41	0.27

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.Dependent variable is log of hourly wage, gross or net (basic).

Table B-4 Result for Probit Estimates and marginal effects (All population Gross Hourly Wage)

Dependent Variable Large Size Establishment (0,1)	Probit Estimates	Marginal Effects
Gender (base female)	0.091***	0.012***
	(0.012)	(0.002)
Experience	-0.051***	-0.007***
	(0.008)	(0.002)
Experience sq.	0.002***	0.000***
	(0.000)	(0.000)
Experience cube	-0.000***	-0.000***
	(0.000)	(0.000)
Tenure	0.069***	0.009***
	(0.004)	(0.001)
Tenure Sq.	0.000	0.000
	(0.000)	(0.000)
Tenure cube	-0.000***	-0.000***
	(0.000)	(0.000)
Married (base single)	0.047***	0.006***
	(0.013)	(0.002)
Other family status (divorced, widowed etc)	0.046**	0.006*
	(0.023)	(0.003)
Before Bac without degree (base no degree)	0.272***	0.043***
	(0.018)	(0.005)
CAP/BEP	0.123***	0.017***
	(0.016)	(0.003)
Bac professional and technical	0.244***	0.037***
	(0.024)	(0.005)
Bac general	0.130***	0.018***
	(0.028)	(0.004)
Bac +2	0.406***	0.070***
	(0.022)	(0.007)
Bac+3 and plus	0.707***	0.146***
	(0.032)	(0.013)
Management and High Intellectual professionals (base blue collar)	-0.196***	-0.022***
	(0.022)	(0.003)
High Skilled White Collar	0.006	0.001
	(0.015)	(0.002)
Low Skilled White Collar	0.035**	0.005**
	(0.016)	(0.002)
CDI Contract(base CDD)	-0.131***	-0.015***
	(0.027)	(0.004)
Manufacturing (base trade)	1.141***	0.292***

	(0.020)	(0.016)
Services	0.417***	0.072***
	(0.020)	(0.007)
Ile de France (11) (Base regional category (Franche-Comte (43))	0.470***	0.084***
	(0.041)	(0.008)
Champagne-Ardenne(21)	0.319***	0.052***
	(0.051)	(0.009)
Picardie (22)	0.230***	0.035***
	(0.047)	(0.007)
Haute-Normandie (23)	0.256***	0.040***
	(0.047)	(0.008)
Centre (24)	0.330***	0.054***
	(0.045)	(0.008)
Basse-Normandie (25)	0.203***	0.030***
	(0.052)	(0.008)
Bourgogne (26)	0.161***	0.023***
	(0.046)	(0.007)
Nord (31)	0.422***	0.073***
	(0.043)	(0.008)
Lorraine (41)	0.461***	0.082***
	(0.044)	(0.009)
Alsace (42)	0.298***	0.048***
	(0.045)	(0.007)
Pays de la Loire (52)	0.144***	0.021***
	(0.044)	(0.006)
Bretagne (53)	0.200***	0.030***
	(0.046)	(0.007)
Poitou-Charentes (54)	0.217***	0.033***
	(0.049)	(0.008)
Aquitaine (72)	0.078*	0.011*
	(0.044)	(0.006)
Midi-Pyrenees (73)	0.194***	0.029***
	(0.046)	(0.007)
Limousin (74)	0.134**	0.019**
	(0.056)	(0.008)
Rhone-Alpes (82)	0.125***	0.018***
	(0.042)	(0.006)
Auvergne (83)	0.252***	0.039***
	(0.052)	(0.009)
Languedoc-Roussillon (91)	0.113**	0.016**
	(0.054)	(0.008)
Provence-Alpes-Cote d’Azur (93)	0.011	0.001
	(0.044)	(0.006)
Constant	-1.505***	
	(0.068)	
Observations	74,762	74,762

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table-B-5 Imbalance in Means between Treated and Matched Controls and Standardized Differences (%)

Covariates	Treated	Control group Pre-match	Control group Post-match	%bias pre-match	%bias post-match
Gender (base female)	0,66	0,58	0,65	17,4	1,7
Experience	21,65	20,04	21,79	17,2	-3,6
Experience sq.	554,23	490,24	562,48	15,2	-3,6
Experience cube	15798,00	13696,00	16176,00	12,7	-3,5
Tenure	14,38	8,60	14,75	67,3	-2,5
Tenure Sq.	291,36	136,45	303,68	58,5	-3,3
Tenure cube	6847,90	2806,20	7232,00	48,7	-3,4
Married (base single)	0,73	0,67	0,73	14,5	0,3
Other family status (divorced, widowed etc)	0,07	0,08	0,07	-4	-1
Before Bac without degree (base no degree)	0,20	0,15	0,20	11,4	-2
CAP/BEP	0,35	0,34	0,35	1,9	0
Bac professional and technical	0,07	0,07	0,07	2,2	2,3
Bac general	0,04	0,06	0,04	-8,4	-0,3
Bac +2	0,13	0,14	0,13	-0,9	1,8
Bac+3 and plus	0,06	0,04	0,06	9,4	2,3
Management and High Intellectual professionals (base blue collar)	0,13	0,11	0,12	6	2,7
High Skilled White Collar	0,28	0,23	0,29	12,3	3,6
Low Skilled White Collar	0,18	0,27	0,19	-21,8	-1,6
CDI Contract(base CDD)	0,97	0,93	0,97	18,6	2,8
Manufacturing (base trade)	0,58	0,28	0,58	62,8	-0,9
Services	0,37	0,56	0,38	-38,4	2
Ile de France (11) (Base regional category (Franche-Comte (43))	0,21	0,19	0,21	6,4	-2,1
Champagne-Ardenne(21)	0,03	0,02	0,02	3,6	0,1
Picardie (22)	0,04	0,03	0,03	3,8	1,5
Haute-Normandie (23)	0,03	0,03	0,04	1,6	1,4
Centre (24)	0,05	0,04	0,04	5,4	3,5
Basse-Normandie (25)	0,02	0,02	0,02	-2,8	0,8
Bourgogne (26)	0,04	0,03	0,03	0,4	2
Nord (31)	0,08	0,06	0,08	8,1	-1,2
Lorraine (41)	0,05	0,03	0,05	9,7	3,2
Alsace (42)	0,05	0,05	0,06	1,2	-2,1
Pays de la Loire (52)	0,06	0,05	0,06	0,8	-2
Bretagne (53)	0,04	0,05	0,04	-4,7	-0,3
Poitou-Charentes (54)	0,02	0,03	0,03	-3,7	0,3
Aquitaine (72)	0,05	0,06	0,04	-6,2	0,6
Midi-Pyrenees (73)	0,04	0,04	0,04	-2,4	-2,2
Limousin (74)	0,01	0,01	0,02	-0,4	-2,9
Rhone-Alpes (82)	0,10	0,11	0,10	-5,2	0,5
Auvergne (83)	0,02	0,02	0,02	1,5	2

Languedoc-Roussillon (91)	0,02	0,03	0,01	-5,8	0,8
Provence-Alpes-Cote d'Azur (93)	0,04	0,07	0,04	-12,5	-0,5
mean standardized bias before matching					13
mean standardized bias after matching					1,8

Standardized Differences raw and matched

Sample	Pseudo R2	LR chi2	p>chi2	MeanBias	MedBias
Raw	0,166	16714,54	0,000	13	6
Matched	0,168	16940,08	0,000	1,8	2

Notes: This table is complementary to the table B-5

Table-B-6 weighted regression

	(_weight!=.)	([fweight = _weight] if _weight!=.)	([fweight = _weight] if _weight!=.)	([fweight = _weight])	if _weight!=.)
Dummy for Size	0.074***	-0.031***	0.077***	0.077***	0.073***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.003)
With covariates	No	No	Yes	Yes	Yes
Constant	4.171***	4.201***	3.330***	3.330***	3.346***
	(0.003)	(0.002)	(0.010)	(0.010)	(0.016)
Observations	30.748	74.719	74.719	74.719	30.748
R-squared	0.010	0.002	0.648	0.648	0.641

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Note: Weighted regression shows only the coefficients for Size effect while covariates inclusion or exclusion is indicated by "yes" or "no".

Table B-7 PSM across Industries (Net hourly wage)

Mean Log hourly wage difference	**Trade**		**Manufacturing**		**Services**	
	pre-match	**post-match (ATT)**	**pre-match**	**post-match (ATT)**	**pre-match**	**post-match (ATT)**
Mean log wage of working in large size establishments	3,94	3,94	4,06	4,06	4,06	4,06
N	*1182*	*944*	*17143*	*5997*	*11526*	*8093*
Mean log wage of working in small size establishments	3,93	3,95	3,94	4,04	3,99	4,05
N	*6058*	*917*	*11348*	*6075*	*26824*	*7791*
% differential	**0,42**	**-1,64**	**12,06**	**2,12**	**7,11**	**0,73**
bootstrap se	0,01	0,18	0,00	0,01	0,00	0,01

**Trade= The propensity score is estimated using a probit of treatment status. Standard errors for the ATT are computed using a bootstrap with 50 replications. ATE is 0.0195 and ATU is -0.006.Mean match weight for the non treated is 1.28. Mean propensity score is 0.1632.*
**Manufacturing= ATE is 0.020 and ATU is 0.018.Mean match weight for the non treated is 2.81. Mean propensity score is 0.601.*
**Services= ATE is -0.008 and ATU is -0.0146.Mean match weight for the non treated is 1.479. Mean propensity score is 0.300.*

Table B-8 PSM across Professions (Net hourly wage)

Mean Log hourly wage difference	**Management**		**High skilled white collar**		**Low skilled white collar**		**Blue collar**	
	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)	pre-match	post-match (ATT)
Mean log wage of working in large size establishments	4,69	4,69	4,17	4,17	3,85	3,85	3,84	3,84
N	*3897*	*1802*	*9223*	*4227*	*5415*	*3589*	*11316*	*5529*
Mean log wage of working in small size establishments	4,62	4,69	4,13	4,19	3,79	3,85	3,76	3,79
N	*5355*	*1795*	*11633*	*4200*	*12558*	*3495*	*15056*	*5456*
% differential	**6,72**	**-0,36**	**4,01**	**-1,92**	**6,29**	**0,77**	**8,80**	**5,21**
bootstrap se	0,01	0,01	0,00	0,01	0,00	0,01	0,00	0,00

**Management= The propensity score is estimated using a probit of treatment status. Standard errors for the ATT are computed using a bootstrap with 50 replications. ATE is -0.015 and ATU is -0.024.Mean match weight for the non treated is 2.17. Mean propensity score is 0.421.*
**High skilled white collar= ATE is -0.014 and ATU is -0.010.Mean match weight for the non treated is 2.19. Mean propensity score is 0.443.*
**Low skilled white collar= ATE is -0.012 and ATU is -0.020.Mean match weight for the non treated is 1.544. Mean propensity score is 0.3019.*
**Blue collar= ATE is 0.039 and ATU is 0.030.Mean match weight for the non treated is 2.07. Mean propensity score is 0.429.*

Table B-9 OLS estimates of Size-Wage impact using variables related to unions

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Firmsize [a]	0.061***	0.043***	0.046***	0.044***	0.048***	0.055***	0.053***	0.057***	0.068***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
delperson		0.021***				0.021***			
		(0.004)				(0.002)			
delunion		0.043***		0.051***					
		(0.004)		(0.003)					
nego	0.031***	0.010***		0.012***		0.021***			
	(0.002)	(0.003)		(0.002)		(0.002)			
Union weight			0.023***		0.021***		0.018***		
			(0.001)		(0.001)		(0.001)		
Worker's Characteristics[b]	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***
Establishment's Chracteristics[c]	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***	Yes***
Constant	3.291***	3.223***	3.275***	3.234***	3.270***	3.290***	3.316***	3.209***	3.262***
	(0.012)	(0.016)	(0.016)	(0.016)	(0.016)	(0.012)	(0.012)	(0.018)	(0.015)
Observations	70.570	41.428	41.428	41.428	41.428	68.260	65.283	34.313	48.166
R-squared	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.64	0.65
Pseudo R-squared	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.64	0.65

1. Model 1 include only dummy on negotiation and no restriction on sample is set.
2. Model 2 includes all three variables on unions. Agreem, Staff Rep., Union Rep keeping Establishment Threshold 50 and more
3. Model 3 is run using PCF using all variables to build union weights keeping Establishment Threshold 50 and more
4. Model 4 negociation and presence of union representative keeping Establishment Threshold 50 and more
5. Model 5 shows union weights on two negociation and presence of union representative
6. Model 6 negociation and presence of staff representative keeping Establishment Threshold 11 and more
7. Model 7 shows union weights on two negociation and presence of staff representative keeping Establishment Threshold 11 and more
8. Sample of establishments with more than 50 workers where union representative is present
9. Sample of establishments with more than 11 workers where staff representative is present.

Appendix - C Chapter4

Table C-1.1 Ranking of regions by characteristics

Serial No.	By urban population	By surface area	By pop density (hab./km²)	By GDP rank	By Rate of Urbanization
1	Limousin	Alsace	Limousin	Limousin	Limousin
2	Franche-Comté	Île-de-France	Auvergne	Franche-Comté	Basse-Normandie
3	Auvergne	Haute-Normandie	Bourgogne	Auvergne	Poitou-Charentes
4	Champagne-Ardenne	Nord-Pas-de-Calais	Champagne-Ardenne	Basse-Normandie	Midi-Pyrenees
5	Basse-Normandie	Franche-Comté	Midi-Pyrénées	Champagne-Ardenne	Auvergne
6	Bourgogne	Limousin	Centre	Poitou-Charentes	Bourgogne
7	Poitou-Charentes	Basse-Normandie	Poitou-Charentes	Bourgogne	Franche-comte
8	Alsace	Picardie	Franche-Comté	Picardie	Champagne-Ardenne
9	Haute-Normandie	Lorraine	Aquitaine	Haute-Normandie	Picardie
10	Picardie	Champagne-Ardenne	Basse-Normandie	Alsace	Pays de la Loire
11	Languedoc-Roussillon	Poitou-Charentes	Languedoc-Roussillon	Languedoc-Roussillon	Centre
12	Lorraine	Auvergne	Picardie	Lorraine	Bretagne
13	Centre	Bretagne	Lorraine	Centre	Aquitaine
14	Midi-Pyrénées	Languedoc-Roussillon	Pays de la Loire	Midi-Pyrénées	Haute-Normandie
15	Bretagne	Provence-Alpes-Côte d'Azur	Bretagne	Bretagne	Languedoc-Roussillon
16	Aquitaine	Bourgogne	Rhône-Alpes	Aquitaine	Lorraine
17	Pays de la Loire	Pays de la Loire	Provence-Alpes-Côte d'Azur	Pays de la Loire	Rhone-Alpes
18	Nord-Pas-de-Calais	Centre	Haute-Normandie	Nord-Pas-de-Calais	Alsace
19	Provence-Alpes-Côte d'Azur	Aquitaine	Alsace	Provence-Alpes-Côte d'Azur	Provence-Alpes-Cote d’Azur

20	Rhône-Alpes	Rhône-Alpes	Nord-Pas-de-Calais	Rhône-Alpes	Nord
21	Île-de-France	Midi-Pyrénées	Île-de-France	Île-de-France	Ile de France

Source, ranking is based on the population census 1999. Source INSEE

Table C-1.2 Distribution of type of sectors in a region

Region	Type of Sector Trade	Manufacturing	Services	Total
Limousin	111	484	658	1,253
	8.86%	38.63%	52.51%	100.00
Franche-Comté	192	517	730	1,439
	13.34%	35.93%	50.73%	100.00
Auvergne	142	638	612	1,392
	10.20%	45.83%	43.97%	100.00
Champagne-Ardenne	107	626	859	1,592
	6.72%	39.32%	53.96%	100.00
Basse-Normandie	221	661	715	1,597
	13.84%	41.39%	44.77%	100.00
Bourgogne	260	946	1,434	2,640
	9.85%	35.83%	54.32%	100.00
Poitou-Charentes	160	548	1,319	2,027
	7.89%	27.04%	65.07%	100.00
Alsace	316	1,288	2,244	3,848
	8.21%	33.47%	58.32%	100.00
Haute-Normandie	222	1,126	1,318	2,666
	8.33%	42.24%	49.44%	100.00
Picardie	187	1,286	1,037	2,510
	7.45%	51.24%	41.31%	100.00
Languedoc-Roussillon	149	275	902	1,326
	11.24%	20.74%	68.02%	100.00
Lorraine	221	1,643	1,560	3,424
	6.45%	47.98%	45.56%	100.00
Centre	290	1,456	1,403	3,149
	9.21%	46.24%	44.55%	100.00
Midi-Pyrénées	308	1,041	1,772	3,121
	9.87%	33.35%	56.78%	100.00
Bretagne	287	921	1,990	3,198
	8.97%	28.80%	62.23%	100.00
Aquitaine	414	1,300	2,322	4,036
	10.26%	32.21%	57.53%	100.00
Pays de la Loire	458	1,964	1,502	3,924
	11.67%	50.05%	38.28%	100.00
Nord-Pas-de-Calais	545	1,991	2,531	5,067
	10.76%	39.29%	49.95%	100.00
Provence-Alpes-Côte d'Azur	579	1,226	3,098	4,903
	11.81%	25.01%	63.19%	100.00
Rhône-Alpes	751	3,385	4,101	8,237

	9.12%	41.10%	49.79%	100.00
Île-de-France	1,721	5,214	6,412	13,347
	12.89%	39.06%	48.04%	100.00
Total	7,641	28,536	38,519	74,696
	10.23	38.20	51.57	100.00

Note: For each region the frequency distribution and row percentage of type of industry is given which shows how the three sectors are distributed in one region.

Table C-1.3 Distribution of size of establishment with respect to size of region

Region	Establishment Size Small	Medium	Large	Total
Limousin	521	243	489	1,253
	41.58%	19.39%	39.03%	100.00
Franche-Comté	666	343	430	1,439
	46.28%	23.84%	29.88%	100.00
Auvergne	591	186	615	1,392
	42.46%	13.36%	44.18%	100.00
Champagne-Ardenne	585	310	697	1,592
	36.75%	19.47%	43.78%	100.00
Basse-Normandie	741	240	616	1,597
	46.40%	15.03%	38.57%	100.00
Bourgogne	1,003	637	1000	2,640
	37.99%	24.13%	37.88%	100.00
Poitou-Charentes	818	445	764	2,027
	40.36%	21.95%	37.69%	100.00
Alsace	1,247	1,025	1,576	3,848
	32.41%	26.64%	40.96%	100.00
Haute-Normandie	1,066	481	1,119	2,666
	39.98%	18.04%	41.97%	100.00
Picardie	853	582	1,075	2,510
	33.98%	23.19%	42.83%	100.00
Languedoc-Roussillon	709	199	418	1,326
	53.47%	15.01%	31.52%	100.00
Lorraine	1,020	632	1,772	3,424
	29.79%	18.46%	51.75%	100.00
Centre	1,180	544	1,425	3,149
	37.47%	17.28%	45.25%	100.00
Midi-Pyrénées	1,570	474	1,077	3,121
	50.30%	15.19%	34.51%	100.00
Bretagne	1,359	702	1,137	3,198
	42.50%	21.95%	35.55%	100.00
Aquitaine	1,804	874	1,358	4,036
	44.70%	21.66%	33.65%	100.00
Pays de la Loire	1,531	793	1,600	3,924
	39.02%	20.21%	40.77%	100.00
Nord-Pas-de-Calais	1,709	932	2,426	5,067
	33.73%	18.39%	47.88%	100.00
Provence-Alpes-Côte d'Azur	2,461	1,136	1,306	4,903
	50.19%	23.17%	26.64%	100.00

Rhône-Alpes	3,365	1,865	3,007	8,237
	40.85%	22.64%	36.51%	100.00
Île-de-France	5,487	1,871	5,989	13,347
	41.11%	14.02%	44.87%	100.00
Total	30,286	14,514	29,896	74,696
	40.55	19.43	40.02	100.00

Note: For each region the frequency distribution and row percentage of establishments by size is given which shows how the three size groups are distributed in one region.

Table C.1.4: Correlation of establishment size with Large region by industry and number of dependent children [a]

	Model-1	Model-2	Model-3
	Cross-Industry Effects of living in a Large region on Choosing to Work in a Large establishment. [b]	*Effect of having more dependent children on choosing to work in large establishment*	*Combined effects of Model1 and Model 2 on choosing to work in large establishment*
Number of Dependent Children		0.035***	0.034***
		(0.007)	(0.007)
Industry type *Region Size (Base Industry Trade and Base region Limousin with lowest population)			
Franche-Comté * Manufacturing	-0.787***		-1.051***
	(0.182)		(0.236)
Franche-Comté * Services	-1.748***		-1.980***
	(0.178)		(0.232)
Auvergne* Manufacturing	0.510***		0.536**
	(0.189)		(0.245)
Auvergne* Services	-0.480***		-0.538**
	(0.186)		(0.242)
Champagne-Ardenne* Manufacturing	1.219***		1.090***
	(0.253)		(0.332)
Champagne-Ardenne* Services	0.661***		0.613*
	(0.249)		(0.328)
Basse-Normandie* Manufacturing	0.270		0.219
	(0.186)		(0.242)
Basse-Normandie* Services	-0.245		-0.306
	(0.184)		(0.241)
Bourgogne* Manufacturing	0.768***		0.803***
	(0.175)		(0.230)
Bourgogne* Services	0.070		0.110
	(0.170)		(0.225)
Poitou-Charentes* Manufacturing	0.237		0.187
	(0.198)		(0.259)
Poitou-Charentes* Services	0.004		-0.013
	(0.191)		(0.252)
Alsace* Manufacturing	0.448***		0.459**
	(0.167)		(0.219)
Alsace* Services	0.076		0.038
	(0.163)		(0.215)

Haute-Normandie* Manufacturing	0.654***		0.613***
	(0.175)		(0.227)
Haute-Normandie* Services	-0.091		-0.132
	(0.172)		(0.225)
Picardie* Manufacturing	0.881***		0.882***
	(0.189)		(0.244)
Picardie* Services	0.040		0.056
	(0.186)		(0.243)
Languedoc-Roussillon* Manufacturing	0.353*		0.348
	(0.194)		(0.253)
Languedoc-Roussillon* Services	-0.136		-0.193
	(0.181)		(0.239)
Lorraine* Manufacturing	1.087***		1.105***
	(0.176)		(0.229)
Lorraine* Services	0.143		0.205
	(0.173)		(0.226)
Centre* Manufacturing	0.952***		0.981***
	(0.177)		(0.233)
Centre* Services	0.100		0.136
	(0.174)		(0.231)
Midi-Pyrénées* Manufacturing	0.140		0.243
	(0.177)		(0.233)
Midi-Pyrénées* Services	-0.223		-0.241
	(0.173)		(0.229)
Bretagne* Manufacturing	0.230		0.324
	(0.175)		(0.234)
Bretagne* Services	0.188		0.400*
	(0.171)		(0.230)
Aquitaine* Manufacturing	0.341**		0.399*
	(0.167)		(0.217)
Aquitaine* Services	-0.076		-0.091
	(0.163)		(0.214)
Pays de la Loire* Manufacturing	0.349**		0.428**
	(0.164)		(0.214)
Pays de la Loire* Services	-0.353**		-0.290
	(0.163)		(0.213)
Nord-Pas-de-Calais* Manufacturing	-0.088		-0.152
	(0.162)		(0.211)
Nord-Pas-de-Calais* Services	-0.768***		-0.857***
	(0.158)		(0.208)
Provence-Alpes-Côte d'Azur* Manufacturing	-0.361**		-0.250

	(0.162)		(0.213)
Provence-Alpes-Côte d'Azur* Services	-0.771***		-0.724***
	(0.158)		(0.209)
Rhône-Alpes* Manufacturing	0.289*		0.256
	(0.158)		(0.208)
Rhône-Alpes* Services	-0.229		-0.284
	(0.156)		(0.206)
Île-de-France* Manufacturing	0.040		0.029
	(0.154)		(0.203)
Île-de-France* Services	-0.499***		-0.519***
	(0.152)		(0.201)
cut1	0.704***	0.778***	0.854***
	(0.147)	(0.096)	(0.203)
cut2	1.284***	1.349***	1.432***
	(0.147)	(0.096)	(0.203)
Number of Observations	74,696	44,902	44,902
Test of joint significance			
chi2	15647.89	9297.83	9850.27
Prob > chi2	0.000	0.000	0.000

*Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*

a. *Dependent variable is size of establishment; it includes all the workers characteristics (gender, experience, tenure, family status, education, profession), and employer characteristics (industry, type of contract, region).*
b. *Regions are ranked (low to high) with respect to urban population (based on population census 1999) Source INSEE. Establishmnet size is defined as small (1-49 workers), Medium (50-199 workers) and Large (more than 200 workers).*

Appendix C.2 Program for Heckman

The Heckman two step estimation procedures is generally followed when we need to correct for the selection bias that is associated with estimating separate wage equations by establishment size. An ordered probit model is estimated where the dependent variable is a dummy variable indicating different size of establishments. The relevant characteristics of the respondents available in the data set are considered as the determinants of the size of establishments.

For the purpose of simplicity we consider three size categories: small, medium and large. We assume that our establishment size participation function for size category is given by:

$$Y_i^* = Z_i \gamma_i + \varepsilon_i \qquad (1)$$

If:

$Y_i^* < 0$ The individual works in small sized establishment

$0 \leq \mathrm{Y}_\mathrm{i}^* < \mu$ The individual works in medium sized establishment

$Y_i^* \geq \mu$ The individual works in large sized establishment

And the wage equation for the three size categories is given by:

$$W_{1i} = X_i \beta_1 + v_{1i} \qquad (2)$$

$$W_{2i} = X_i \beta_2 + v_{2i} \qquad (3)$$

$$W_{3i} = X_i \beta_3 + v_{3i} \qquad (4)$$

Where $Y_{1i}^*$is a latent variable associated with "being employed in size category 1", Z contains the set of determining variables of being in a size category, γ *is* the associated parameter vector. $W_{1i}$is the log hourly wage for small size category, X is a matrix of wage determining variables, β is a vector of unknown parameters and $\boldsymbol{\varepsilon_{1i}}$ and $\boldsymbol{v_{1i}}$are the *i.i.d* error terms that follow a bivariate normal distribution (0,0, $\sigma_{\varepsilon 1}, \sigma_{v1}, \rho_1$). The variance of $\boldsymbol{\varepsilon}$i is normalized to be one $\boldsymbol{\varepsilon i}$ ~ N (0, 1).

$$\begin{pmatrix}\varepsilon_i \\ v_i\end{pmatrix} \backsim N \ \{E = \begin{pmatrix}0 \\ 0\end{pmatrix}, \quad \sum = \begin{pmatrix}1 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_{v^2}\end{pmatrix}\}$$

$\sigma_{\varepsilon v} = \rho$ where ρis interpreted as the correlation coefficient between errors in the selection equation (ε_i) and errors in the wage equation (v_i); and the σ_{v^2}terms are the error variances of the wage equations.

The likelihood function for the ordered probit model for three categories of size, used in thisstudy, is given by:

$$L = \Pi_{Y=1}\Phi(-Z\gamma)\Pi_{Y=2}[\Phi(\mu - Z\gamma) - \Phi(-Z\gamma)]\Pi_{Y=3}\,[1 - \Phi(\mu - Z\gamma)] \qquad (5)$$

The maximum likelihood estimates for μ and the γvector are then used to construct the truncated means. These constructed variables are then inserted into the wage equation and OLS estimation is performed.

The probability of being employed in size categories is given by:

$$pr(\varepsilon_{1i} \leq -Z_i\gamma) = \Phi(-Z_i\gamma) \qquad (6)$$

$$pr(-Z_i\gamma < \varepsilon_{2i} \leq \mu - Z_i\gamma) = \Phi(\mu - Z_i\gamma) - \Phi(-Z_i\gamma) \qquad (7)$$

$$pr(\varepsilon_{3i} \geq \mu - Z_i\gamma) = 1 - \Phi(\mu - Z_i\gamma) \qquad (8)$$

Where Φ (.) is the cumulative distribution functions of a standard normal distribution.

Because of the selection problem (the failure to observe W1 when $0 \leq Y_i^* < \mu$ or $Y_i^* \geq \mu$ and the failure to observe W2 when$Y_i^* < 0$ or so on), we need to write these outcomes in a selection equation format. Taking expectations of the outcome equations, we can find the expected earnings for an employee working in large size who self-selected into large size and similarly, we can find the expected earnings for a small firm worker who self selected into small firm.

The expected wages of a worker for whom $Y_i^* < 0$ and who self selected into small size establishment is given by:

$$E[W_1] = X'_1\beta_1 + E[v_1|\varepsilon_{1i} < -Z\gamma] \quad (9)$$

$$= X'_1\beta_1 + \theta_1\lambda_{1i}$$

Where $\theta_1 = \sigma_{v1}, \rho_1$ and λ_{1i} is defined as the ratio of the probability density function to the cumulative distribution function of a distribution. It is written as:

$$\boldsymbol{\lambda_{1i} = -\emptyset(Z\gamma)/[1 - \Phi(Z\gamma)]} \quad (10)$$

The expected wages of a worker observed to be in medium size establishment is given by

$$E[W_2] = X'_2\beta_2 + E[v_2| - Z\gamma \leq \varepsilon_{2i} < \mu - Z\gamma] \quad (11)$$

$$= X'_2\beta_2 + \theta_2\lambda_{2i}$$

$$\boldsymbol{\lambda_2 = \{[\emptyset(-Z\gamma) - \emptyset(\mu - Z\gamma)] / [\Phi(\mu - Z\gamma) - \Phi(-Z\gamma)]\}} \quad (12)$$

The expected wages of a worker observed to be in large size establishment is given by

$$E[W_3] = X'_3\beta_3 + E[v_3|\varepsilon_{3i} \geq \mu - Z\gamma] \quad (13)$$

$$= X'_3\beta_3 + \theta_3\lambda_{3i}$$

$$\boldsymbol{\lambda_3 = \{ \emptyset(\mu - Z\gamma)/[1 - \Phi(\mu - Z\gamma)]\}} \quad (14)$$

The estimating equation for those who are working in small size category is given by

$$\boldsymbol{W_{ij} = X_{ij}\beta_{ij} + \theta_{ij}\lambda_{ij} + error_{ij}} \quad (15)$$

Where W is log of hourly wage of worker, j is the size category, small, medium or large and (i) denotes worker. The parameters of 18 will be estimated through Heckman two step estimation procedures separately for the measures of gross hourly wage and Basic hourly wage and for males and females.

Table C.2.1 Heckman estimation procedure (Second-step regression): Gross hourly wages

Variables	(Small)	(Medium)	(Large)	(Small)	(Medium)	(Large)	(Small)	(Medium)	(Large)
	Interactions			Number of dependent children			Both		
Gender	0.131***	0.134***	0.146***	0.155***	0.151***	0.132***	0.147***	0.154***	0.163***
	(0.003)	(0.004)	(0.003)	(0.005)	(0.010)	(0.005)	(0.004)	(0.006)	(0.004)
Experience	0.025***	0.021***	0.032***	0.015***	0.017***	0.030***	0.016***	0.016***	0.023***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)	(0.003)	(0.005)	(0.003)
Exp. squared	-0.001***	-0.001***	-0.001***	-0.000	-0.000	-0.001***	-0.000*	-0.000	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Exp. cube	0.000***	0.000***	0.000***	-0.000	0.000	0.000**	-0.000	-0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.012***	0.013***	0.010***	0.019***	0.011*	-0.017***	0.013***	0.013***	0.008***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)	(0.002)	(0.001)	(0.002)	(0.001)
Tenure. squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000*	0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married (base single)	0.029***	0.033***	0.028***	0.010	0.020*	0.022***	0.011	0.019*	0.016**
	(0.004)	(0.005)	(0.004)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)	(0.007)
Other family status (divorced, widowed etc)	0.019***	0.023***	0.019***	0.007	0.018	0.018*	0.008	0.017	0.013
	(0.006)	(0.008)	(0.006)	(0.009)	(0.013)	(0.009)	(0.009)	(0.013)	(0.009)
Before Bac without degree (base no degree)	0.047***	0.055***	0.075***	0.065***	0.046**	0.007	0.047***	0.054***	0.074***
	(0.005)	(0.007)	(0.005)	(0.009)	(0.020)	(0.008)	(0.007)	(0.009)	(0.006)
CAP/BEP	0.069***	0.083***	0.112***	0.071***	0.070***	0.083***	0.065***	0.073***	0.107***
	(0.004)	(0.006)	(0.004)	(0.006)	(0.010)	(0.006)	(0.005)	(0.007)	(0.005)
Bac professional and technical	0.147***	0.149***	0.187***	0.151***	0.132***	0.144***	0.140***	0.137***	0.186***
	(0.007)	(0.010)	(0.006)	(0.009)	(0.017)	(0.009)	(0.009)	(0.012)	(0.008)
Bac general	0.153***	0.172***	0.202***	0.157***	0.166***	0.180***	0.150***	0.169***	0.201***
	(0.007)	(0.010)	(0.008)	(0.009)	(0.015)	(0.010)	(0.009)	(0.013)	(0.009)
Bac +2	0.195***	0.209***	0.254***	0.215***	0.193***	0.149***	0.190***	0.205***	0.242***

	(0.006)	(0.009)	(0.007)	(0.012)	(0.028)	(0.011)	(0.008)	(0.012)	(0.008)
Bac+3 and plus	0.303***	0.344***	0.375***	0.357***	0.332***	0.228***	0.315***	0.352***	0.381***
	(0.010)	(0.015)	(0.010)	(0.020)	(0.046)	(0.017)	(0.013)	(0.019)	(0.012)
Management and High Intellectual professionals (base blue collar)	0.709***	0.663***	0.562***	0.704***	0.687***	0.641***	0.724***	0.678***	0.572***
	(0.006)	(0.009)	(0.006)	(0.010)	(0.022)	(0.009)	(0.008)	(0.011)	(0.007)
High Skilled White Collar	0.323***	0.294***	0.214***	0.318***	0.298***	0.229***	0.323***	0.296***	0.212***
	(0.004)	(0.006)	(0.004)	(0.006)	(0.009)	(0.005)	(0.006)	(0.007)	(0.005)
Low Skilled White Collar	0.087***	0.063***	0.029***	0.088***	0.080***	0.054***	0.092***	0.078***	0.035***
	(0.004)	(0.006)	(0.004)	(0.006)	(0.009)	(0.005)	(0.006)	(0.007)	(0.005)
Type of Contract	0.007	0.003	0.139***	-0.010	0.009	0.210***	0.008	-0.001	0.137***
	(0.006)	(0.008)	(0.008)	(0.011)	(0.022)	(0.014)	(0.009)	(0.011)	(0.012)
Manufacturing (base trade)	0.034***	0.034**	0.044***	0.129***	-0.001	-0.349***	0.045***	0.039**	0.011
	(0.010)	(0.016)	(0.011)	(0.032)	(0.085)	(0.033)	(0.013)	(0.019)	(0.013)
Services	0.024***	0.031***	0.054***	0.066***	0.022	-0.142***	0.032***	0.040***	0.046***
	(0.006)	(0.010)	(0.008)	(0.014)	(0.039)	(0.018)	(0.007)	(0.012)	(0.010)
Franche-Comté (Base region Limousin)	0.062***	0.010	-0.012	0.045**	0.013	0.019	0.050***	0.012	-0.010
	(0.014)	(0.019)	(0.015)	(0.018)	(0.025)	(0.018)	(0.018)	(0.024)	(0.018)
Auvergne	-0.010	0.053**	0.046***	-0.015	0.074***	0.057***	-0.018	0.078***	0.069***
	(0.014)	(0.022)	(0.013)	(0.018)	(0.028)	(0.017)	(0.018)	(0.028)	(0.017)
Champagne-Ardenne	0.053***	0.120***	0.069***	0.058***	0.132***	0.001	0.041**	0.139***	0.067***
	(0.014)	(0.019)	(0.013)	(0.019)	(0.029)	(0.017)	(0.018)	(0.024)	(0.016)
Basse-Normandie	0.032**	0.020	0.062***	0.037**	0.025	0.053***	0.031*	0.027	0.079***
	(0.014)	(0.020)	(0.013)	(0.018)	(0.027)	(0.017)	(0.017)	(0.026)	(0.016)
Bourgogne	0.040***	0.055***	0.050***	0.043**	0.078***	0.002	0.031*	0.083***	0.045***
	(0.013)	(0.017)	(0.012)	(0.017)	(0.024)	(0.016)	(0.016)	(0.022)	(0.015)
Poitou-Charentes	0.029**	0.062***	0.122***	0.052***	0.047*	0.064***	0.038**	0.053**	0.121***
	(0.013)	(0.018)	(0.013)	(0.018)	(0.027)	(0.016)	(0.017)	(0.023)	(0.016)
Alsace	0.084***	0.133***	0.123***	0.107***	0.140***	0.030*	0.083***	0.150***	0.120***
	(0.013)	(0.016)	(0.012)	(0.018)	(0.031)	(0.017)	(0.016)	(0.021)	(0.015)
Haute-Normandie	0.102***	0.084***	0.106***	0.114***	0.111***	0.065***	0.103***	0.116***	0.108***
	(0.013)	(0.018)	(0.012)	(0.017)	(0.025)	(0.015)	(0.016)	(0.023)	(0.015)

Picardie	0.064***	0.074***	0.118***	0.084***	0.078***	0.043***	0.066***	0.087***	0.111***
	(0.013)	(0.017)	(0.012)	(0.018)	(0.028)	(0.016)	(0.017)	(0.022)	(0.015)
Languedoc-Roussillon	0.033**	0.066***	0.048***	0.033*	0.081***	0.128***	0.043**	0.075***	0.081***
	(0.014)	(0.021)	(0.015)	(0.018)	(0.030)	(0.019)	(0.017)	(0.028)	(0.019)
Lorraine	0.086***	0.087***	0.105***	0.120***	0.083**	-0.018	0.088***	0.099***	0.094***
	(0.013)	(0.017)	(0.012)	(0.020)	(0.038)	(0.018)	(0.017)	(0.022)	(0.015)
Centre	0.060***	0.080***	0.063***	0.070***	0.066**	-0.026	0.048***	0.077***	0.056***
	(0.013)	(0.017)	(0.012)	(0.018)	(0.031)	(0.016)	(0.016)	(0.022)	(0.015)
Midi-Pyrénées	0.037***	0.068***	0.062***	0.035**	0.094***	0.036**	0.030*	0.096***	0.055***
	(0.012)	(0.018)	(0.012)	(0.016)	(0.023)	(0.015)	(0.016)	(0.023)	(0.015)
Bretagne	0.041***	0.017	0.028**	0.039**	0.026	0.002	0.030*	0.030	0.039***
	(0.012)	(0.017)	(0.012)	(0.016)	(0.023)	(0.015)	(0.016)	(0.022)	(0.015)
Aquitaine	0.037***	0.082***	0.129***	0.045***	0.091***	0.114***	0.041***	0.092***	0.133***
	(0.012)	(0.016)	(0.012)	(0.015)	(0.021)	(0.015)	(0.015)	(0.021)	(0.015)
Pays de la Loire	0.041***	0.036**	0.036***	0.055***	0.051**	0.001	0.047***	0.055***	0.036**
	(0.012)	(0.016)	(0.011)	(0.016)	(0.022)	(0.015)	(0.015)	(0.021)	(0.014)
Nord-Pas-de-Calais	0.041***	0.027*	0.039***	0.073***	0.032	-0.081***	0.043***	0.046**	0.029**
	(0.012)	(0.017)	(0.011)	(0.019)	(0.036)	(0.017)	(0.016)	(0.021)	(0.014)
Provence-Alpes-Côte d'Azur	0.120***	0.127***	0.139***	0.131***	0.137***	0.122***	0.127***	0.138***	0.140***
	(0.012)	(0.016)	(0.012)	(0.015)	(0.021)	(0.015)	(0.015)	(0.020)	(0.015)
Rhône-Alpes	0.091***	0.102***	0.108***	0.092***	0.114***	0.062***	0.081***	0.118***	0.104***
	(0.011)	(0.015)	(0.011)	(0.015)	(0.022)	(0.014)	(0.014)	(0.020)	(0.014)
Île-de-France	0.240***	0.209***	0.175***	0.271***	0.209***	0.066***	0.244***	0.221***	0.168***
	(0.011)	(0.016)	(0.011)	(0.017)	(0.033)	(0.016)	(0.015)	(0.020)	(0.014)
lambda1	**-0.009**	**-0.015**	**-0.027***	**0.122*****	**-0.058**	**-0.561*****	**-0.003**	**-0.019**	**-0.080*****
	(0.013)	**(0.013)**	**(0.015)**	**(0.047)**	**(0.082)**	**(0.044)**	**(0.016)**	**(0.016)**	**(0.017)**
Sigma	0,24	0,22	0,22	0,24	0,22	0,21	0,24	0,22	0,21
Rho	-0,04	-0,07	-0,12	0,52	-0,26	-2,62	-0,01	-0,09	-0,37
Constant	3.287***	3.346***	3.238***	3.379***	3.395***	4.234***	3.337***	3.356***	3.390***
	(0.019)	(0.028)	(0.029)	(0.032)	(0.093)	(0.082)	(0.029)	(0.043)	(0.040)
Observations	30,286	14,514	29,896	17,480	8,594	18,828	17,480	8,594	18,828
R-squared	0.64	0.66	0.65	0.66	0.67	0.66	0.66	0.67	0.66
Adj. R-squared	0.64	0.65	0.65	0.66	0.66	0.66	0.66	0.66	0.66

*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.Dependent variable is gross hourly wage.*

Table C.2.2 Heckman estimation procedure (Second-step regression) Gross hourly wages (Male Sample)

VARIABLES	(small)	(medium)	(large)	(small)	(medium)	(large)	(small)	(medium)	(large)
		Interactions		Number of dependent children				Both	
Experience	0.031***	0.025***	0.043***	0.028***	0.027***	0.035***	0.028***	0.026***	0.031***
	(0.003)	(0.004)	(0.003)	(0.005)	(0.007)	(0.004)	(0.005)	(0.007)	(0.004)
Exp. squared	-0.001***	-0.001***	-0.002***	-0.001***	-0.001**	-0.001***	-0.001***	-0.001**	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Exp. cube	0.000***	0.000**	0.000***	0.000*	0.000	0.000***	0.000*	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.011***	0.015***	0.001	0.013***	0.011	-0.017***	0.011***	0.014***	-0.001
	(0.001)	(0.002)	(0.001)	(0.003)	(0.011)	(0.003)	(0.002)	(0.002)	(0.002)
Tenure. squared	-0.000***	-0.000***	0.000	-0.000***	-0.000	0.000***	-0.000***	-0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married (base single)	0.052***	0.048***	0.043***	0.020*	0.041**	0.040***	0.021*	0.040**	0.036***
	(0.005)	(0.007)	(0.005)	(0.011)	(0.016)	(0.010)	(0.011)	(0.016)	(0.010)
Other family status (divorced, widowed etc)	0.016*	0.032***	0.024***	-0.012	0.038	0.048***	-0.009	0.034	0.030**
	(0.009)	(0.012)	(0.008)	(0.016)	(0.025)	(0.014)	(0.015)	(0.021)	(0.014)
Before Bac without degree (base no degree)	0.045***	0.066***	0.055***	0.048***	0.050	0.021*	0.041***	0.061***	0.066***
	(0.008)	(0.010)	(0.007)	(0.014)	(0.038)	(0.011)	(0.010)	(0.013)	(0.008)
CAP/BEP	0.062***	0.084***	0.088***	0.055***	0.064***	0.063***	0.051***	0.070***	0.089***
	(0.006)	(0.008)	(0.005)	(0.009)	(0.023)	(0.008)	(0.008)	(0.010)	(0.007)
Bac professional and technical	0.153***	0.142***	0.146***	0.152***	0.107**	0.096***	0.143***	0.120***	0.150***
	(0.010)	(0.014)	(0.009)	(0.018)	(0.047)	(0.014)	(0.013)	(0.018)	(0.011)
Bac general	0.140***	0.166***	0.174***	0.146***	0.146***	0.150***	0.142***	0.153***	0.179***
	(0.011)	(0.015)	(0.010)	(0.015)	(0.031)	(0.014)	(0.014)	(0.020)	(0.013)
Bac +2	0.179***	0.202***	0.196***	0.195***	0.185***	0.135***	0.186***	0.199***	0.194***
	(0.010)	(0.014)	(0.009)	(0.018)	(0.050)	(0.015)	(0.013)	(0.017)	(0.011)
Bac+3 and plus	0.296***	0.322***	0.297***	0.316***	0.303***	0.214***	0.301***	0.326***	0.315***

	(0.014)	(0.021)	(0.013)	(0.028)	(0.085)	(0.023)	(0.017)	(0.025)	(0.015)
Management and High Intellectual professionals (base blue collar)	0.710***	0.657***	0.579***	0.720***	0.682***	0.621***	0.725***	0.674***	0.588***
	(0.008)	(0.011)	(0.007)	(0.012)	(0.029)	(0.010)	(0.010)	(0.014)	(0.009)
High Skilled White Collar	0.314***	0.277***	0.197***	0.319***	0.284***	0.199***	0.320***	0.283***	0.194***
	(0.005)	(0.007)	(0.004)	(0.007)	(0.010)	(0.005)	(0.007)	(0.009)	(0.005)
Low Skilled White Collar	0.042***	0.010	-0.044***	0.049***	0.017	-0.061***	0.046***	0.022*	-0.042***
	(0.007)	(0.009)	(0.007)	(0.011)	(0.021)	(0.009)	(0.010)	(0.013)	(0.008)
Type of Contract	0.011	0.010	0.221***	0.023	0.010	0.247***	0.028**	0.003	0.215***
	(0.009)	(0.011)	(0.012)	(0.015)	(0.030)	(0.020)	(0.014)	(0.017)	(0.019)
Manufacturing (base trade)	0.009	0.071***	-0.049***	0.044	0.016	-0.291***	0.017	0.057**	-0.050***
	(0.014)	(0.022)	(0.017)	(0.041)	(0.145)	(0.044)	(0.018)	(0.027)	(0.019)
Services	-0.001	0.050***	0.011	0.018	0.025	-0.129***	0.005	0.046***	0.018
	(0.008)	(0.014)	(0.012)	(0.020)	(0.074)	(0.027)	(0.011)	(0.017)	(0.014)
Franche-Comté (Base region Limousin)	0.089***	0.013	-0.034*	0.070***	0.003	-0.002	0.072***	0.002	-0.028
	(0.019)	(0.026)	(0.018)	(0.024)	(0.034)	(0.022)	(0.024)	(0.034)	(0.023)
Auvergne	-0.003	0.052*	0.001	0.003	0.064	0.001	-0.002	0.072*	0.030
	(0.021)	(0.029)	(0.016)	(0.027)	(0.046)	(0.022)	(0.026)	(0.037)	(0.021)
Champagne-Ardenne	0.082***	0.133***	0.043***	0.077***	0.123**	-0.018	0.069***	0.135***	0.038*
	(0.020)	(0.027)	(0.016)	(0.027)	(0.053)	(0.022)	(0.025)	(0.034)	(0.020)
Basse-Normandie	0.060***	0.009	0.048***	0.060**	-0.003	0.044**	0.058**	-0.001	0.059***
	(0.019)	(0.027)	(0.017)	(0.024)	(0.036)	(0.021)	(0.024)	(0.035)	(0.021)
Bourgogne	0.055***	0.067***	0.029**	0.054**	0.080*	-0.026	0.048**	0.091***	0.022
	(0.018)	(0.024)	(0.015)	(0.025)	(0.048)	(0.020)	(0.023)	(0.031)	(0.019)
Poitou-Charentes	0.039**	0.068***	0.096***	0.059**	0.035	0.050**	0.055**	0.042	0.086***
	(0.018)	(0.024)	(0.016)	(0.024)	(0.039)	(0.020)	(0.023)	(0.031)	(0.020)
Alsace	0.133***	0.154***	0.109***	0.132***	0.147***	0.066***	0.125***	0.157***	0.111***
	(0.017)	(0.023)	(0.014)	(0.024)	(0.045)	(0.020)	(0.022)	(0.029)	(0.018)
Haute-Normandie	0.138***	0.076***	0.116***	0.139***	0.090**	0.090***	0.135***	0.096***	0.118***
	(0.017)	(0.025)	(0.015)	(0.022)	(0.038)	(0.019)	(0.022)	(0.032)	(0.018)
Picardie	0.100***	0.087***	0.082***	0.099***	0.073	0.030	0.093***	0.084***	0.076***
	(0.018)	(0.024)	(0.015)	(0.025)	(0.049)	(0.021)	(0.023)	(0.031)	(0.019)
Languedoc-Roussillon	0.070***	0.079***	0.077***	0.071***	0.079*	0.141***	0.074***	0.073*	0.109***
	(0.019)	(0.031)	(0.019)	(0.024)	(0.045)	(0.024)	(0.024)	(0.040)	(0.024)
Lorraine	0.124***	0.111***	0.061***	0.144***	0.097	-0.015	0.132***	0.115***	0.058***
	(0.018)	(0.024)	(0.015)	(0.029)	(0.070)	(0.023)	(0.023)	(0.031)	(0.018)

Centre	0.080***	0.112***	0.028*	0.082***	0.085	-0.051**	0.072***	0.101***	0.021
	(0.017)	(0.024)	(0.014)	(0.028)	(0.065)	(0.023)	(0.023)	(0.031)	(0.019)
Midi-Pyrénées	0.063***	0.076***	0.088***	0.057***	0.091***	0.074***	0.057***	0.091***	0.075***
	(0.016)	(0.025)	(0.015)	(0.021)	(0.031)	(0.019)	(0.021)	(0.031)	(0.019)
Bretagne	0.068***	0.028	-0.004	0.061***	0.039	0.002	0.061***	0.039	0.004
	(0.016)	(0.023)	(0.015)	(0.021)	(0.030)	(0.019)	(0.021)	(0.030)	(0.019)
Aquitaine	0.064***	0.077***	0.150***	0.073***	0.084***	0.134***	0.072***	0.086***	0.147***
	(0.016)	(0.023)	(0.014)	(0.021)	(0.030)	(0.018)	(0.021)	(0.029)	(0.018)
Pays de la Loire	0.068***	0.034	0.020	0.077***	0.044	-0.011	0.074***	0.050*	0.014
	(0.016)	(0.023)	(0.014)	(0.021)	(0.034)	(0.018)	(0.021)	(0.029)	(0.017)
Nord-Pas-de-Calais	0.065***	0.039*	0.009	0.084***	0.036	-0.069***	0.074***	0.050*	-0.005
	(0.016)	(0.023)	(0.013)	(0.025)	(0.058)	(0.021)	(0.021)	(0.029)	(0.017)
Provence-Alpes-Côte d'Azur	0.159***	0.138***	0.136***	0.157***	0.131***	0.111***	0.155***	0.134***	0.129***
	(0.016)	(0.022)	(0.014)	(0.020)	(0.030)	(0.018)	(0.020)	(0.028)	(0.018)
Rhône-Alpes	0.122***	0.120***	0.095***	0.116***	0.111***	0.042**	0.110***	0.120***	0.085***
	(0.015)	(0.021)	(0.013)	(0.021)	(0.041)	(0.018)	(0.020)	(0.028)	(0.017)
Île-de-France	0.256***	0.213***	0.120***	0.271***	0.195***	0.045**	0.261***	0.209***	0.115***
	(0.015)	(0.022)	(0.013)	(0.024)	(0.058)	(0.020)	(0.020)	(0.029)	(0.017)
lambda1	**-0.021**	**0.020**	**-0.161*** **	**0.020**	**-0.035**	**-0.466*** **	**-0.018**	**0.003**	**-0.187*** **
	(0.018)	(0.019)	(0.021)	(0.057)	(0.132)	(0.055)	(0.022)	(0.022)	(0.024)
Sigma	0,25	0,23	0,22	0,24	0,23	0,22	0,24	0,23	0,22
Rho	-0,09	0,09	-0,73	0,08	-0,15	-2,14	-0,07	0,01	-0,86
Constant	3.361***	3.394***	3.522***	3.398***	3.446***	4.178***	3.391***	3.399***	3.640***
	(0.026)	(0.040)	(0.041)	(0.043)	(0.173)	(0.107)	(0.041)	(0.063)	(0.057)
Observations	16,472	8,424	19,758	9,485	5,076	12,717	9,485	5,076	12,717
R-squared	0.65	0.65	0.63	0.68	0.65	0.64	0.68	0.65	0.64
Adj. R-squared	0.65	0.64	0.63	0.68	0.65	0.63	0.68	0.65	0.63

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table C.2.3 Heckman estimation procedure (Second-step regression) Gross hourly wages (Female Sample)

VARIABLES	(small)	(medium) Interactions	(large)	(small)	(medium) Number of dependent children	(large)	(small)	(medium) Both	(large)
Experience	0.020***	0.017***	0.026***	0.004	0.010	0.022***	0.005	0.010	0.016***
	(0.003)	(0.004)	(0.003)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)
Exp. squared	-0.001***	-0.001***	-0.001***	0.000	-0.000	-0.000*	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Exp. cube	0.000*	0.000*	0.000***	-0.000**	-0.000	0.000	-0.000**	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.013***	0.012***	0.013***	0.020***	0.012**	-0.006*	0.016***	0.013***	0.012***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.006)	(0.003)	(0.001)	(0.002)	(0.002)
Tenure. squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	0.000***	-0.000***	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married (base single)	-0.001	0.013*	-0.008	0.002	-0.003	-0.010	0.003	-0.003	-0.013
	(0.005)	(0.007)	(0.005)	(0.009)	(0.013)	(0.010)	(0.009)	(0.013)	(0.010)
Other family status (divorced, widowed etc)	0.014*	0.012	0.006	0.024**	-0.000	-0.015	0.021*	0.000	-0.002
	(0.007)	(0.010)	(0.008)	(0.012)	(0.016)	(0.012)	(0.012)	(0.016)	(0.012)
Before Bac without degree (base no degree)	0.045***	0.045***	0.059***	0.062***	0.048***	0.013	0.052***	0.049***	0.052***
	(0.007)	(0.010)	(0.007)	(0.012)	(0.018)	(0.011)	(0.009)	(0.012)	(0.009)
CAP/BEP	0.065***	0.069***	0.101***	0.072***	0.068***	0.091***	0.072***	0.068***	0.092***
	(0.006)	(0.009)	(0.007)	(0.008)	(0.011)	(0.009)	(0.008)	(0.011)	(0.009)
Bac professional and technical	0.133***	0.154***	0.178***	0.124***	0.147***	0.197***	0.129***	0.146***	0.177***
	(0.009)	(0.014)	(0.010)	(0.012)	(0.018)	(0.013)	(0.011)	(0.017)	(0.013)
Bac general	0.151***	0.163***	0.197***	0.151***	0.167***	0.196***	0.152***	0.167***	0.189***
	(0.009)	(0.014)	(0.011)	(0.012)	(0.018)	(0.014)	(0.012)	(0.018)	(0.014)
Bac +2	0.197***	0.208***	0.272***	0.208***	0.205***	0.190***	0.190***	0.208***	0.261***
	(0.009)	(0.013)	(0.010)	(0.016)	(0.028)	(0.016)	(0.011)	(0.016)	(0.012)
Bac+3 and plus	0.306***	0.419***	0.435***	0.361***	0.437***	0.377***	0.337***	0.442***	0.473***

	(0.014)	(0.023)	(0.016)	(0.025)	(0.043)	(0.026)	(0.020)	(0.031)	(0.022)
Management and High Intellectual professionals (base blue collar)	0.725***	0.705***	0.585***	0.690***	0.714***	0.703***	0.722***	0.708***	0.589***
	(0.011)	(0.016)	(0.012)	(0.025)	(0.044)	(0.024)	(0.016)	(0.022)	(0.016)
High Skilled White Collar	0.358***	0.360***	0.296***	0.320***	0.356***	0.352***	0.337***	0.352***	0.295***
	(0.008)	(0.011)	(0.008)	(0.015)	(0.025)	(0.013)	(0.011)	(0.014)	(0.010)
Low Skilled White Collar	0.132***	0.136***	0.121***	0.099***	0.145***	0.189***	0.117***	0.141***	0.125***
	(0.007)	(0.009)	(0.007)	(0.015)	(0.024)	(0.013)	(0.010)	(0.012)	(0.009)
Type of Contract	0.001	-0.009	0.071***	-0.028*	-0.006	0.133***	-0.011	-0.009	0.062***
	(0.008)	(0.011)	(0.011)	(0.016)	(0.026)	(0.019)	(0.011)	(0.015)	(0.016)
Manufacturing (base trade)	0.060***	0.031	0.102***	0.139***	0.045	-0.211***	0.073***	0.056**	0.067***
	(0.014)	(0.021)	(0.013)	(0.045)	(0.085)	(0.046)	(0.017)	(0.024)	(0.016)
Services	0.050***	0.026**	0.058***	0.082***	0.042	-0.069***	0.059***	0.045***	0.052***
	(0.007)	(0.012)	(0.009)	(0.017)	(0.035)	(0.023)	(0.009)	(0.015)	(0.012)
Franche-Comté (Base region Limousin)	0.015	0.016	0.021	-0.001	0.032	0.057*	0.006	0.031	0.020
	(0.020)	(0.026)	(0.024)	(0.026)	(0.035)	(0.030)	(0.025)	(0.033)	(0.030)
Auvergne	-0.040*	0.079**	0.100***	-0.066**	0.093**	0.169***	-0.055**	0.093**	0.121***
	(0.020)	(0.034)	(0.023)	(0.027)	(0.046)	(0.030)	(0.026)	(0.043)	(0.029)
Champagne-Ardenne	0.008	0.118***	0.047**	-0.007	0.156***	0.039	-0.008	0.155***	0.046
	(0.021)	(0.029)	(0.023)	(0.027)	(0.036)	(0.030)	(0.027)	(0.036)	(0.030)
Basse-Normandie	-0.013	0.035	0.105***	-0.010	0.062	0.094***	-0.017	0.062	0.125***
	(0.020)	(0.030)	(0.021)	(0.026)	(0.039)	(0.026)	(0.025)	(0.038)	(0.026)
Bourgogne	0.002	0.054**	0.045**	-0.013	0.088***	0.068**	-0.011	0.087***	0.056**
	(0.019)	(0.023)	(0.021)	(0.024)	(0.030)	(0.027)	(0.024)	(0.030)	(0.027)
Poitou-Charentes	0.012	0.060**	0.177***	0.015	0.083**	0.127***	0.004	0.085**	0.177***
	(0.020)	(0.026)	(0.021)	(0.026)	(0.036)	(0.027)	(0.025)	(0.033)	(0.025)
Alsace	0.016	0.105***	0.133***	0.037	0.139***	0.019	0.014	0.142***	0.114***
	(0.019)	(0.023)	(0.019)	(0.028)	(0.040)	(0.028)	(0.024)	(0.030)	(0.024)
Haute-Normandie	0.043**	0.086***	0.092***	0.055**	0.130***	0.053**	0.045*	0.133***	0.092***
	(0.019)	(0.024)	(0.020)	(0.025)	(0.033)	(0.025)	(0.024)	(0.031)	(0.025)
Picardie	0.009	0.072***	0.157***	0.028	0.100***	0.115***	0.018	0.101***	0.153***
	(0.020)	(0.024)	(0.021)	(0.026)	(0.033)	(0.026)	(0.025)	(0.031)	(0.026)
Languedoc-Roussillon	-0.024	0.049*	0.034	-0.022	0.080**	0.082***	-0.015	0.078**	0.053*
	(0.020)	(0.029)	(0.023)	(0.026)	(0.038)	(0.030)	(0.025)	(0.037)	(0.030)
Lorraine	0.023	0.070***	0.118***	0.027	0.087**	0.037	0.009	0.089***	0.103***
	(0.019)	(0.025)	(0.020)	(0.027)	(0.037)	(0.027)	(0.025)	(0.031)	(0.025)

Centre	0.018	0.038	0.086***	-0.000	0.043	0.066***	-0.001	0.044	0.073***
	(0.018)	(0.025)	(0.020)	(0.023)	(0.032)	(0.025)	(0.024)	(0.032)	(0.025)
Midi-Pyrénées	-0.004	0.058**	0.063***	-0.008	0.103***	0.013	-0.016	0.104***	0.049**
	(0.018)	(0.025)	(0.019)	(0.024)	(0.034)	(0.025)	(0.023)	(0.032)	(0.024)
Bretagne	0.001	0.003	0.098***	-0.003	0.017	0.025	-0.022	0.020	0.103***
	(0.019)	(0.024)	(0.019)	(0.026)	(0.038)	(0.027)	(0.024)	(0.031)	(0.024)
Aquitaine	-0.007	0.077***	0.090***	-0.007	0.094***	0.070***	-0.010	0.094***	0.087***
	(0.017)	(0.022)	(0.019)	(0.023)	(0.029)	(0.024)	(0.022)	(0.029)	(0.024)
Pays de la Loire	-0.005	0.049**	0.065***	0.000	0.072**	0.048**	-0.003	0.072**	0.066***
	(0.018)	(0.023)	(0.019)	(0.023)	(0.029)	(0.024)	(0.023)	(0.029)	(0.024)
Nord-Pas-de-Calais	0.004	0.017	0.075***	0.013	0.041	-0.029	-0.012	0.045	0.071***
	(0.018)	(0.024)	(0.019)	(0.028)	(0.042)	(0.028)	(0.023)	(0.030)	(0.024)
Provence-Alpes-Côte d'Azur	0.059***	0.106***	0.144***	0.079***	0.141***	0.135***	0.075***	0.141***	0.152***
	(0.017)	(0.022)	(0.019)	(0.022)	(0.029)	(0.024)	(0.022)	(0.029)	(0.024)
Rhône-Alpes	0.042**	0.085***	0.116***	0.029	0.124***	0.113***	0.031	0.123***	0.110***
	(0.017)	(0.021)	(0.018)	(0.021)	(0.028)	(0.023)	(0.021)	(0.027)	(0.023)
Île-de-France	0.206***	0.215***	0.240***	0.224***	0.243***	0.153***	0.205***	0.246***	0.234***
	(0.017)	(0.022)	(0.018)	(0.024)	(0.036)	(0.026)	(0.021)	(0.028)	(0.022)
lambda1	**0.009**	**-0.025**	**0.025**	**0.123***	**-0.031**	**-0.413***	**0.018**	**-0.019**	**-0.012**
	(0.018)	(0.017)	(0.018)	(0.070)	(0.086)	(0.066)	(0.022)	(0.019)	(0.020)
Sigma	0,23	0,21	0,20	0,22	0,21	0,20	0,22	0,21	0,20
rho	0,04	-0,12	0,12	0,55	-0,15	-2,10	0,08	-0,09	-0,06
Constant	3.348***	3.372***	3.184***	3.510***	3.386***	3.923***	3.444***	3.379***	3.324***
	(0.027)	(0.036)	(0.036)	(0.058)	(0.076)	(0.107)	(0.040)	(0.056)	(0.051)
Observations	13,814	6,090	10,138	7,995	3,518	6,111	7,995	3,518	6,111
R-squared	0.60	0.65	0.64	0.60	0.65	0.64	0.60	0.65	0.64
Adj. R-squared	0.60	0.65	0.64	0.60	0.64	0.64	0.60	0.64	0.64

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table C.2.4 Heckman estimation procedure (Second-step regression) Gross hourly wages (Results across professions)

Professions	**Model 1 (Interactions)**				**Model 2 (Number of dependent children)**				**Model 3 (Both)**			
	Prof1	Prof2	Prof3	Prof4	Prof1	Prof2	Prof3	Prof4	Prof1	Prof2	Prof3	Prof4
λ1	0.067	0.022	**-0.052****	**0.078*****	0.018	**0.162****	-0.003	0.102	0.04	0.044	-0.033	**0.075*****
	(0.048)	(0.026)	(0.021)	(0.016)	(0.140)	(0.081)	(0.081)	(0.064)	(0.055)	(0.032)	(0.023)	(0.021)
σ1	0.301	0.248	0.218	0.207	0.297	0.242	0.213	0.209	0.297	0.242	0.213	0.208
ρ1	0.223	0.089	-0.239	0.376	0.061	0.669	-0.014	0.489	0.135	0.182	-0.155	0.360
Observations	3,855	7,99	8,965	9,537	2,317	4,716	5,126	5,356	2,317	4,716	5,126	5,356
λ2	**0.104****	-0.018	0.014	0.017	**-1.085*****	-0.34	-0.001	0.126	**0.135****	-0.009	-0.002	0.011
	(0.049)	(0.025)	(0.021)	(0.017)	(0.303)	(0.275)	(0.096)	(0.120)	(0.055)	(0.031)	(0.024)	(0.020)
σ2	0.278	0.228	0.201	0.204	0.275	0.226	0.204	0.200	0.276	0.226	0.204	0.200
ρ2	0.374	-0.079	0.070	0.083	-3.948	-1.507	-0.005	0.629	0.489	-0.040	-0.010	0.055
Observations	1,563	3,698	3,656	5,602	957	2,240	2,111	3,290	957	2,240	2,111	3,290
λ3	-0.004	**-0.067*****	**0.052****	**-0.096*****	**-0.359*****	**-0.26*****	**-0.301*****	**-0.404*****	**-0.091***	**-0.129*****	0.024	**-0.135*****
	(0.044)	(0.023)	(0.024)	(0.023)	(0.096)	(0.065)	(0.087)	(0.072)	(0.048)	(0.028)	(0.026)	(0.028)
σ3	0.244	0.203	0.200	0.207	0.239	0.203	0.198	0.204	0.239	0.203	0.198	0.204
ρ3	-0.016	-0.331	0.260	-0.465	-1.504	-1.282	-1.523	-1.977	-0.380	-0.636	0.121	-0.660
Observations	3,904	9,221	5,433	11,338	2,431	5,825	3,308	7,264	2,431	5,825	3,308	7,264

Notes:λ1 denotes selection terms for small size establishment, λ2 denotes the selection effect for medium size establishment, λ3 denotes the selection effect for large size establishment. Prof1= Management and High Intellectual professionals, prof2=High Skilled White Collar, prof3=Low Skilled White Collar, prof4=Blue Collar. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table C.2.5 Heckman estimation procedure (Second-step regression) Gross hourly wages (Excluding sample of blue collar workers)

	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	All Sample			Male			Female		
λ1	-0.004	0.09	0.006	-0.016	0.026	-0.006	-0.021	0.087	0.008
	(0.018)	(0.057)	(0.021)	(0.028)	(0.088)	(0.034)	(0.021)	(0.072)	(0.024)
σ1	0.250	0.245	0.245	0.271	0.265	0.265	0.231	0.227	0.227
ρ1	-0.016	0.367	0.024	-0.059	0.098	-0.023	-0.091	0.383	0.035
Observations	20.810	12.159	12.159	8.696	5.087	5.087	12.114	7.072	7.072
λ2	0.016	-0.210**	0.010	0.015	-0.621**	0.012	0.039*	-0.048	0.029
	(0.017)	(0.092)	(0.020)	(0.028)	(0.260)	(0.033)	(0.021)	(0.093)	(0.024)
σ2	0.231	0.231	0.231	0.247	0.246	0.246	0.213	0.214	0.214
ρ2	0.069	-0.908	0.043	0.061	-2.523	0.049	0.183	-0.224	0.136
Observations	8.917	5.308	5.308	4.095	2.521	2.521	4.822	2.787	2.787
λ3	-0.024	-0.47***	-0.083***	-0.102***	-0.339***	-0.155***	0.025	-0.284***	-0.007
	(0.017)	(0.047)	(0.019)	(0.025)	(0.061)	(0.029)	(0.022)	(0.076)	(0.024)
σ3	0.217	0.214	0.215	0.221	0.218	0.218	0.205	0.202	0.202
ρ3	-0.110	-2.197	-0.387	-0.463	-1.556	-0.711	0.122	-1.408	-0.035
Observations	18.558	11.564	11.564	10.664	6.851	6.851	7.894	4.713	4.713

Notes : Model 1 (Interactions), Model 2 (Number of dependent children), Model 3 (Both IV together)
*λ1 denotes selection terms for small size establishment, λ2 denotes the selection effect for medium size establishment, λ3 denotes the selection effect for large size establishment. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.*

Table C.2.6 Heckman estimation procedure (Second-step regression) Gross hourly wages (Excluding large regions)

	Model-1	**Model-2**
λ1	-0.020	-0.000
	(0.014)	(0.013)
σ1	0,23	0,23
ρ1	-0,09	0,00
Observations	18,973	24,799
λ2	-0.016	-0.008
	(0.014)	(0.013)
σ2	0,21	0,22
ρ2	-0,07	-0,04
Observations	9,642	12,643
λ3	-0.003	-0.009
	(0.016)	(0.015)
σ3	0,21	0,21
ρ3	-0,01	-0,04
Observations	19,594	23,907

*Notes : Model 1 (18 region,excluding ile de France, rhone alpes, cote d'azur), Model 2 (20 regions excluding ile de France) λ1 denotes selection terms for small size establishment, λ2 denotes the selection effect for medium size establishment, λ3 denotes the selection effect for large size establishment. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.*

Table C.2.7 Heckman estimation procedure (Second-step regression) Gross hourly wages (unionization)

	Covered staff representative workplaces		**Covered union representative workplaces**		**Salary negotiation in wage equation**		**All three variables**	
	Model-1	Model-2	Model-1	Model-2	Model-1	Model-2	Model-1	Model-2
λ1	0.007	0.144	0.002	0.501	-0.009	0.116**	-0.006	0.080***
	(0.018)	(0.110)	(0.029)	(0.350)	(0.014)	(0.047)	(0.013)	(0.031)
σ1	0,22	0,22	0,23	0,23	0,24	0,23	0,22	0,22
ρ1	0,03	0,64	0,01	2,21	-0,04	0,49	-0,03	0,36
Observations	10,291	6,234	3,134	1,938	28,687	16,503	13,758	8,136
λ2	-0.002	-0.111	0.026	-0.183	-0.007	-0.073	-0.026**	0.137
	(0.015)	(0.133)	(0.019)	(0.249)	(0.014)	(0.079)	(0.013)	(0.145)
σ2	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,22
ρ2	-0,01	-0,50	0,12	-0,85	-0,03	-0,33	-0,12	0,63
Observations	10,377	6,253	7,068	4,304	13,894	8,216	10,44	6,523
λ3	-0.107***	-0.645***	-0.067***	-0.713***	-0.028*	-0.560***	-0.085***	-0.490***
	(0.017)	(0.043)	(0.016)	(0.049)	(0.016)	(0.046)	(0.016)	(0.047)
σ3	0,22	0,21	0,21	0,21	0,22	0,22	0,21	0,21
ρ3	-0,50	-3,04	-0,31	-3,36	-0,13	-2,60	-0,40	-2,34
Observations	27,737	17,561	27,244	17,303	27,927	17,572	17,225	10,889

Model 1 (Interactions), Model 2 (Number of dependent children),

Table C-2.8 Multinomial selection model of employer size choice using BFG (dmf(2))

	(Small)	(Medium) Interactions	(Large)	(Small)	(Medium) Number of dependent children	(Large)
Gender (base female)	0.130***	0.134***	0.151***	0.146***	0.148***	0.161***
	(0.003)	(0.004)	(0.004)	(0.006)	(0.009)	(0.006)
Before Bac without degree (base no degree)	0.044***	0.054***	0.092***	0.042***	0.044***	0.073***
	(0.005)	(0.007)	(0.005)	(0.012)	(0.012)	(0.010)
CAP/BEP	0.067***	0.081***	0.114***	0.059***	0.080***	0.104***
	(0.004)	(0.005)	(0.004)	(0.007)	(0.011)	(0.008)
Bac professional and technical	0.144***	0.147***	0.194***	0.129***	0.157***	0.182***
	(0.007)	(0.010)	(0.007)	(0.012)	(0.023)	(0.014)
Bac general	0.152***	0.170***	0.204***	0.141***	0.190***	0.198***
	(0.007)	(0.012)	(0.009)	(0.014)	(0.024)	(0.013)
Bac +2	0.190***	0.206***	0.278***	0.177***	0.207***	0.240***
	(0.006)	(0.010)	(0.007)	(0.014)	(0.019)	(0.018)
Bac+3 and plus	0.297***	0.340***	0.419***	0.292***	0.371***	0.386***
	(0.010)	(0.017)	(0.013)	(0.022)	(0.035)	(0.032)
CDI Contract(base CDD)	0.008	-0.001	0.108***	0.001	0.044	0.113***
	(0.009)	(0.012)	(0.012)	(0.023)	(0.037)	(0.020)
Manufacturing (base trade)	0.018**	0.023	0.087***	0.015	0.009	-0.024
	(0.009)	(0.015)	(0.013)	(0.037)	(0.043)	(0.042)
Services	0.011	0.023**	0.060***	0.027	-0.028	-0.003
	(0.008)	(0.011)	(0.010)	(0.029)	(0.043)	(0.022)
Personal characteristics	Yes	Yes	Yes	Yes	Yes	Yes
_m1	-0.067***	-0.115*	-0.601***	-0.090	0.220	-0.356***
	(0.017)	(0.068)	(0.051)	(0.059)	(0.256)	(0.104)

_m2	-0.103***	-0.027	-0.278***	-0.012	-0.174*	-0.317*
	(0.038)	(0.021)	(0.052)	(0.168)	(0.095)	(0.171)
_m3	-0.176***	-0.172***	-0.208***	-0.251**	-0.064	-0.264***
	(0.032)	(0.047)	(0.016)	(0.109)	(0.168)	(0.041)
Sigma2	0.068***	0.060***	0.165***	0.072***	0.074*	0.120***
	(0.004)	(0.006)	(0.015)	(0.017)	(0.039)	(0.024)
rho1	-0.256***	-0.471*	-1.480***	-0.333*	0.808	-1.026***
	(0.062)	(0.259)	(0.076)	(0.193)	(0.701)	(0.280)
rho2	-0.394***	-0.109	-0.685***	-0.044	-0.638***	-0.914**
	(0.140)	(0.083)	(0.128)	(0.542)	(0.222)	(0.412)
rho3	-0.677***	-0.703***	-0.513***	-0.935***	-0.233	-0.761***
	(0.105)	(0.161)	(0.031)	(0.322)	(0.541)	(0.145)
Constant	3.251***	3.260***	2.975***	3.325***	3.660***	3.307***
	(0.025)	(0.063)	(0.035)	(0.067)	(0.288)	(0.119)
Observations	30,286	14,514	29,896	17,480	8,594	18,828

Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.Results are computed through stata command ‘selmlog’.

Table C-2.9 Multinomial selection model of employer size choice using comparison of various selmlog methods

VARIABLES	Dhl(1)			Selmlog command using (dmf(0)) Dubin-McFadden			Selmlog command using (dmf(1)) Dubin-McFadden			Selmlog command using BFG (Lee)		
	Small	Medium	Large	Small	Medium	Large	Small	Medium	Large	Small	Medium	Large
Gender (base female)	0.130***	0.135***	0.153***	0.131***	0.133***	0.146***	0.130***	0.134***	0.149***	0.131***	0.136***	0.145***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Before Bac without degree (base no degree)	0.045***	0.060***	0.095***	0.048***	0.053***	0.072***	0.045***	0.055***	0.098***	0.048***	0.059***	0.073***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)	(0.006)	(0.005)	(0.006)	(0.006)
CAP/BEP	0.068***	0.086***	0.121***	0.068***	0.083***	0.106***	0.067***	0.081***	0.119***	0.069***	0.085***	0.110***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
Bac professional and technical	0.147***	0.155***	0.204***	0.145***	0.150***	0.174***	0.145***	0.146***	0.200***	0.148***	0.153***	0.184***
	(0.007)	(0.010)	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.010)	(0.008)	(0.006)	(0.010)	(0.007)
Bac general	0.153***	0.175***	0.209***	0.150***	0.172***	0.190***	0.152***	0.169***	0.203***	0.153***	0.174***	0.199***
	(0.007)	(0.012)	(0.009)	(0.008)	(0.012)	(0.009)	(0.008)	(0.012)	(0.009)	(0.007)	(0.010)	(0.008)
Bac +2	0.192***	0.216***	0.283***	0.195***	0.207***	0.246***	0.191***	0.207***	0.284***	0.196***	0.215***	0.250***
	(0.007)	(0.010)	(0.008)	(0.007)	(0.010)	(0.008)	(0.007)	(0.010)	(0.008)	(0.006)	(0.008)	(0.007)
Bac+3 and plus	0.300***	0.358***	0.426***	0.300***	0.342***	0.352***	0.298***	0.340***	0.412***	0.306***	0.355***	0.367***
	(0.010)	(0.017)	(0.012)	(0.012)	(0.017)	(0.012)	(0.012)	(0.017)	(0.014)	(0.011)	(0.015)	(0.013)
CDI Contract(base CDD)	0.011	0.005	0.125***	0.000	0.007	0.114***	0.007	-0.005	0.093***	0.006	0.002	0.136***
	(0.008)	(0.010)	(0.012)	(0.008)	(0.011)	(0.014)	(0.008)	(0.012)	(0.014)	(0.007)	(0.010)	(0.011)
Manufacturing (base trade)	0.024***	0.053***	0.116***	0.037***	0.026*	0.033***	0.020**	0.025*	0.101***	0.039***	0.051***	0.036***
	(0.009)	(0.009)	(0.013)	(0.008)	(0.015)	(0.013)	(0.009)	(0.015)	(0.016)	(0.009)	(0.007)	(0.011)
Services	0.016**	0.033***	0.071***	0.032***	0.023**	0.075***	0.011	0.026**	0.060***	0.026***	0.035***	0.053***
	(0.007)	(0.008)	(0.008)	(0.006)	(0.011)	(0.010)	(0.008)	(0.011)	(0.011)	(0.006)	(0.008)	(0.007)

_m1					0.042**	-0.122***	-0.035***	-0.130**	-0.467***	-0.002		
					(0.021)	(0.039)	(0.010)	(0.057)	(0.040)	(0.013)		
_m2				0.047**		0.174***	-0.061	-0.009	-0.108**		0.012	
				(0.023)		(0.042)	(0.039)	(0.011)	(0.049)		(0.021)	
_m3				-0.032**	-0.034**		-0.119***	-0.153***	-0.086***			0.038**
				(0.016)	(0.017)		(0.020)	(0.039)	(0.009)			(0.015)
_m11	-0.048**											
	(0.024)											
_m21		0.070										
		(0.055)										
_m31			-0.176***									
			(0.032)									
Sigma2				0.059***	0.052***	0.082***	0.071***	0.087***	0.246***	0.057***	0.050***	0.049***
				(0.002)	(0.002)	(0.018)	(0.006)	(0.021)	(0.031)	(0.001)	(0.002)	(0.001)
rho1					0.234**	-0.547***	-0.167***	-0.568***	-1.208***	-0.009		
					(0.113)	(0.128)	(0.048)	(0.205)	(0.040)	(0.055)		
rho2				0.245**		0.776***	-0.295*	-0.038	-0.280**		0.055	
				(0.118)		(0.113)	(0.174)	(0.052)	(0.129)		(0.090)	
rho3				-0.168**	-0.193**		-0.575***	-0.668***	-0.222***			0.172***
				(0.082)	(0.089)		(0.081)	(0.117)	(0.028)			(0.066)
Constant	3.323***	3.311***	3.219***	3.311***	3.369***	3.284***	3.249***	3.206***	2.943***	3.290***	3.347***	3.260***
	(0.024)	(0.034)	(0.021)	(0.025)	(0.040)	(0.028)	(0.029)	(0.064)	(0.042)	(0.018)	(0.033)	(0.034)

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.Results are computed through stata command ‘selmlog’.

Appendix C-3Program for FIML

We simplify the model in the following way:

Our latent variable model is as follows:

$$\boldsymbol{Y_i^*} = \boldsymbol{Z_i \gamma} + \boldsymbol{\varepsilon_{1i}}$$

Where Y_i^* is unobservable, Z contains the set of determining variables, γ *is* an unknown parameter vector and $\boldsymbol{\varepsilon_{1i}}$~ N (0, 1).

If:

$Y_i^* < 0$ The individual works in small sized establishment

$0 \leq Y_i^* < \mu$ The individual works in medium sized establishment

$Y_i^* \geq \mu$ The individual works in large sized establishment

And the wage equation is

$$\boldsymbol{W_i} = \boldsymbol{X_i \beta_i} + \boldsymbol{\varepsilon_{2i}}$$

We define θ as set of all parameters;

$$\theta = \{\beta_{wage}, \gamma_{size}; \sigma_{12}, \sigma_{22}\}$$

And X_i as set of all variables;

$X_i = \{X_{size}, X_{wage}\}$

Note that Xsize is included in Xwage and constants are included in both Xsize and Xwage. This model is the case of two equations with correlated random terms, ordered probit and linear equations.

$$S_{1,i} = I(S_i = 1);\ S_{2,i} = I(S_i = 2);\ S_{3,i} = I(S_i = 3)$$

$$S_i = \begin{cases} 1(small\ size) & if\ Y_i^* < 0 \\ 2(medium\ size) & if\ 0 \leq Y_i^* < \mu \\ 3(large\ size) & if\ Y_i^* \geq \mu \end{cases}$$

$$\binom{\varepsilon_1}{\varepsilon_2} \backsim N \ \{E = \binom{0}{0}, \quad \sum = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\}$$

We have to normalize $\sigma_{11} = 1$

The likelihood function can be written as;

$$\boldsymbol{L = \prod_{i=1}^{N} li = \prod_{i=1}^{N} l_{1i}^{1(Yi=1)} . l_{2i}^{1(Yi=2)} . l_{3i}^{1(Yi=3)}}$$

Where the parts of likelihood function can be expressed as the following:

$$\boldsymbol{l_{1i} = Lsize = 1\ (\theta, X_i) = p(Z_i\gamma + \varepsilon_{1i}} < 0;\ \boldsymbol{W_i = X_i\beta_i + \varepsilon_{2i})=}$$

$$= (P(Z_i\gamma + \varepsilon_{1i} < 0 |\ W_i)) . f(W_i) =$$

$$= (P(\varepsilon_{1i} < -Z_i\gamma |\ W_i)) . \frac{1}{\sqrt{\sigma_{22}}}\ \phi\ \left(\frac{W_i - X_i\beta_i}{\sqrt{\sigma_{22}}}\right) =$$

$$= \Phi \left[\frac{-Z_i\gamma - \frac{\sigma_{21}}{\sigma_{22}}(W_i - X_i\beta_i)}{\sqrt{1 - \frac{\sigma_{21}^2}{\sigma_{22}}}}\right] . \frac{1}{\sqrt{\sigma_{22}}}\ \phi\ \left(\frac{W_i - X_i\beta_i}{\sqrt{\sigma_{22}}}\right)$$

$$\boldsymbol{l_{2i} = Lsize = 2\ (\theta, X_i) = p(\ 0 < Z_i\gamma + \varepsilon_{1i} < \mu;\ W_i = X_i\beta_i + \varepsilon_{2i})=}$$

$$= (P(Z_i\gamma + \varepsilon_{1i} < \mu |\ W_i) - P(Z_i\gamma + \varepsilon_{1i} < 0 |\ W_i)) . f(W_i) =$$

$$= (P(\varepsilon_{1i} < \mu - Z_i\gamma |\ W_i) - P(\varepsilon_{1i} < -Z_i\gamma |\ W_i)) . \frac{1}{\sqrt{\sigma_{22}}}\ \phi\ \left(\frac{W_i - X_i\beta_i}{\sqrt{\sigma_{22}}}\right) =$$

$$= \Phi \left[\frac{\mu - Z_i\gamma - \frac{\sigma_{21}}{\sigma_{22}}(W_i - X_i\beta_i)}{\sqrt{1 - \frac{\sigma_{21}^2}{\sigma_{22}}}}\right] - \Phi \left[\frac{-Z_i\gamma - \frac{\sigma_{21}}{\sigma_{22}}(W_i - X_i\beta_i)}{\sqrt{1 - \frac{\sigma_{21}^2}{\sigma_{22}}}}\right] . \frac{1}{\sqrt{\sigma_{22}}}\ \phi\ \left(\frac{W_i - X_i\beta_i}{\sqrt{\sigma_{22}}}\right)$$

$\boldsymbol{l_{3i} = Lsize = 3\ (\theta, X_i) = p(\mu < Z_i\,\gamma + \ \varepsilon_{1i}\ ;\ W_i = \ X_i\beta_i + \ \varepsilon_{2i})}$=

$= (P(\,-\varepsilon_{1i} < Z_i\,\gamma - \mu|\ W_i))\,.\,f(W_i)=$

$$\Phi\left[\frac{Z_i\,\gamma - \mu - \frac{\sigma_{21}}{\sigma_{22}}(W_i - X_i\beta_i)}{\sqrt{1-\frac{\sigma_{21}^2}{\sigma_{22}}}}\right]\cdot\frac{1}{\sqrt{\sigma_{22}}}\,\phi\ \left(\frac{W_i - X_i\beta_i}{\sqrt{\sigma_{22}}}\right)$$

Where ϕ () is a standard normal probability density function, and Φ () is the cumulative distribution functions of a standard normal distribution. To determine the above probabilities, the conditional normal distribution theorem is used, which states the following:

If vector of random variables (x_1, x_2) have a joint multivariate normal distribution

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\ \left\{E = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\ ,\ \sum = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right\}$$

Then the conditional distribution of $x_1$givenx_2 is normal with the following parameters:

$$\{x_1|x_2\} \sim N\ \{\mu_{1.2}, \Sigma_{11.2}\}$$

Where

$$\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

And then

$$f(x_1, x_2) = f_{1.2}(x_1|x_2)f_2(x_2)$$

The maximum likelihood model is used, by programming the likelihood function in STATA and then using *ml max* procedure.

In order to estimate the parameters of the covariance matrix:$\{\sigma_{12}, \sigma_{22}\}$,the Cholesky decomposition is followed.

Cholesky decomposition states that: if A is a symmetric and positive definite matrix, then there exists a triangular operator matrix L such that A= L.$L^{/}$. Where L is a lower triangular matrix with strictly positive diagonal entries.

Therefore, two parameters l_{21}, l_{22} in the Cholesky matrix are estimated:

$$L = \begin{pmatrix} 1 & 0 \\ l_{21} & l_{22} \end{pmatrix}$$

Covariance matrix can be expressed as the following (Cholesky decomposition, thus, guarantees the matrix to by symmetric and positive definite):

$$\Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = L.L^{/} = \begin{pmatrix} 1 & l_{21} \\ l_{21} & l_{21}^2 + l_{22}^2 \end{pmatrix}$$

Table C-3.1 Joint Model FIML (All sample and across gender)for gross hourly wage

Dependent variable Log of gross hourly wage	All sample		Male		Female	
Variables	Wage Equationln(Hourly Wage)	Size of Establishment (Ordered Probit)	Wage Equationln(Hourly Wage)	Size of Establishment (Ordered Probit)	Wage Equationln(Hourly Wage)	Size of Establishment (Ordered Probit)
Medium size (base small size)	0.046***		0.074***		0.009	
	(0.011)		(0.011)		(0.012)	
Large size	0.144***		0.209***		0.052**	
	(0.021)		(0.021)		(0.022)	
Gender (base female)	0.135***	0.094***				
	(0.002)	(0.010)				
Before Bac without degree (base no degree)	0.056***	0.239***	0.054***	0.292***	0.049***	0.172***
	(0.003)	(0.015)	(0.004)	(0.020)	(0.004)	(0.024)
CAP/BEP	0.088***	0.082***	0.076***	0.149***	0.079***	0.033
	(0.003)	(0.014)	(0.004)	(0.017)	(0.004)	(0.023)
Bac professional and technical	0.160***	0.167***	0.145***	0.329***	0.151***	-0.000
	(0.004)	(0.022)	(0.006)	(0.029)	(0.006)	(0.033)
Bac general	0.175***	0.063***	0.157***	0.182***	0.169***	-0.046
	(0.005)	(0.024)	(0.007)	(0.034)	(0.006)	(0.035)
Bac +2	0.215***	0.332***	0.188***	0.372***	0.223***	0.301***
	(0.004)	(0.020)	(0.006)	(0.027)	(0.006)	(0.030)
Bac+3 and plus	0.318***	0.559***	0.284***	0.663***	0.356***	0.414***
	(0.007)	(0.029)	(0.008)	(0.035)	(0.010)	(0.051)
Experience	0.016***	-0.041***	0.018***	-0.044***	0.015***	-0.031***
	(0.000)	(0.007)	(0.001)	(0.009)	(0.001)	(0.010)

Experience sq.	-0.000***	0.001***	-0.000***	0.001***	-0.000***	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experience cube		-0.000***		-0.000***		-0.000*
		(0.000)		(0.000)		(0.000)
Tenure	0.011***	0.070***	0.008***	0.082***	0.013***	0.058***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)
Tenure Sq.	-0.000***	-0.001***	-0.000***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married (base single)	0.031***	0.042***	0.051***	0.051***	0.000	0.015
	(0.002)	(0.012)	(0.003)	(0.016)	(0.003)	(0.018)
Other family status (divorced, widowed etc)	0.021***	0.064***	0.026***	0.010	0.011**	0.089***
	(0.004)	(0.019)	(0.006)	(0.029)	(0.005)	(0.026)
Management and High Intellectual professionals (base blue collar)	0.641***	-0.209***	0.641***	-0.177***	0.677***	-0.403***
	(0.004)	(0.019)	(0.005)	(0.023)	(0.007)	(0.038)
High Skilled White Collar	0.268***	-0.041***	0.250***	-0.004	0.336***	-0.213***
	(0.003)	(0.014)	(0.003)	(0.016)	(0.005)	(0.027)
Low Skilled White Collar	0.054***	-0.028**	-0.004	0.138***	0.127***	-0.226***
	(0.003)	(0.014)	(0.004)	(0.022)	(0.004)	(0.023)
CDI Contract(base CDD)	0.042***	-0.252***	0.063***	-0.220***	0.015***	-0.258***
	(0.004)	(0.021)	(0.006)	(0.030)	(0.006)	(0.029)
Manufacturing (base trade)	0.018**	0.790***	-0.015*	0.976***	0.069***	0.544**
	(0.008)	(0.136)	(0.009)	(0.182)	(0.009)	(0.211)
Services	0.026***	0.797***	0.006	0.924***	0.047***	0.714***
	(0.004)	(0.133)	(0.006)	(0.182)	(0.005)	(0.193)
Franche-Comté (Base region Limousin)	0.029***	1.060***	0.032***	1.148***	0.021	1.084***
	(0.009)	(0.152)	(0.012)	(0.206)	(0.013)	(0.222)
Auvergne	0.028***	0.028	0.020*	0.412*	0.025*	-0.319
	(0.009)	(0.165)	(0.012)	(0.222)	(0.014)	(0.247)
Champagne-Ardenne	0.076***	-0.729***	0.084***	-0.913***	0.044***	-0.523*
	(0.009)	(0.214)	(0.011)	(0.342)	(0.014)	(0.284)
Basse-Normandie	0.043***	-0.000	0.044***	0.293	0.040***	-0.256
	(0.009)	(0.154)	(0.012)	(0.208)	(0.013)	(0.228)
Bourgogne	0.051***	-0.241	0.054***	-0.056	0.032***	-0.338
	(0.008)	(0.152)	(0.011)	(0.212)	(0.012)	(0.216)
Poitou-Charentes	0.073***	-0.101	0.069***	-0.110	0.081***	0.033
	(0.008)	(0.165)	(0.011)	(0.230)	(0.013)	(0.235)
Alsace	0.112***	-0.005	0.133***	0.144	0.080***	-0.083

	(0.008)	(0.144)	(0.010)	(0.197)	(0.012)	(0.211)
Haute-Normandie	0.101***	-0.120	0.120***	0.073	0.071***	-0.305
	(0.008)	(0.154)	(0.011)	(0.208)	(0.012)	(0.229)
Picardie	0.090***	-0.274*	0.098***	-0.242	0.073***	-0.201
	(0.008)	(0.165)	(0.011)	(0.223)	(0.012)	(0.243)
Languedoc-Roussillon	0.044***	-0.220	0.068***	-0.152	0.010	-0.206
	(0.009)	(0.168)	(0.012)	(0.227)	(0.013)	(0.249)
Lorraine	0.096***	-0.225	0.105***	-0.037	0.067***	-0.401*
	(0.008)	(0.156)	(0.010)	(0.211)	(0.012)	(0.231)
Centre	0.066***	-0.278*	0.069***	-0.093	0.048***	-0.392*
	(0.008)	(0.151)	(0.010)	(0.205)	(0.012)	(0.223)
Midi-Pyrénées	0.051***	0.019	0.071***	0.106	0.030**	-0.024
	(0.008)	(0.146)	(0.010)	(0.199)	(0.012)	(0.214)
Bretagne	0.032***	-0.191	0.032***	-0.456**	0.036***	-0.019
	(0.008)	(0.149)	(0.010)	(0.223)	(0.012)	(0.209)
Aquitaine	0.080***	-0.177	0.101***	-0.223	0.043***	-0.072
	(0.008)	(0.142)	(0.010)	(0.197)	(0.011)	(0.203)
Pays de la Loire	0.041***	0.078	0.045***	0.008	0.032***	0.232
	(0.008)	(0.138)	(0.010)	(0.190)	(0.011)	(0.199)
Nord-Pas-de-Calais	0.036***	0.668***	0.040***	0.525***	0.031***	0.843***
	(0.008)	(0.135)	(0.010)	(0.186)	(0.012)	(0.194)
Provence-Alpes-Côte d'Azur	0.131***	0.414***	0.147***	0.422**	0.098***	0.495**
	(0.007)	(0.136)	(0.010)	(0.187)	(0.011)	(0.195)
Rhône-Alpes	0.103***	0.029	0.114***	0.175	0.077***	-0.082
	(0.007)	(0.134)	(0.009)	(0.183)	(0.011)	(0.195)
Île-de-France	0.203***	0.496***	0.185***	0.657***	0.218***	0.389**
	(0.007)	(0.128)	(0.009)	(0.177)	(0.011)	(0.185)
Region Size*Industry Type (Base Industry Trade and Base region Limousin with lowest population)		-0.733***		-0.809***		-0.737***
		(0.169)		(0.224)		(0.264)
Franche-Comté * Manufacturing		-1.679***		-1.835***		-1.649***
		(0.166)		(0.225)		(0.241)
Franche-Comté * Services		0.583***		0.221		0.856***
		(0.180)		(0.238)		(0.290)
Auvergne* Manufacturing		-0.406**		-0.615**		-0.231
		(0.178)		(0.240)		(0.266)
Auvergne* Services		1.260***		1.341***		1.348***
		(0.226)		(0.352)		(0.324)
Champagne-Ardenne*		0.726***		1.103***		0.248

Manufacturing			
	(0.223)	(0.350)	(0.299)
Champagne-Ardenne* Services	0.324*	-0.114	1.011***
	(0.169)	(0.223)	(0.270)
Basse-Normandie* Manufacturing	-0.211	-0.539**	0.097
	(0.166)	(0.227)	(0.245)
Basse-Normandie* Services	0.797***	0.525**	1.111***
	(0.165)	(0.226)	(0.255)
Bourgogne* Manufacturing	0.116	0.164	-0.009
	(0.161)	(0.225)	(0.231)
Bourgogne* Services	0.288	0.240	0.297
	(0.180)	(0.245)	(0.278)
Poitou-Charentes* Manufacturing	0.064	0.054	-0.028
	(0.174)	(0.242)	(0.250)
Poitou-Charentes* Services	0.518***	0.363*	0.635**
	(0.158)	(0.210)	(0.246)
Alsace* Manufacturing	0.093	-0.181	0.316
	(0.153)	(0.209)	(0.224)
Alsace* Services	0.698***	0.390*	1.239***
	(0.167)	(0.220)	(0.265)
Haute-Normandie* Manufacturing	-0.072	-0.263	0.129
	(0.163)	(0.221)	(0.243)
Haute-Normandie* Services	0.953***	0.838***	1.103***
	(0.177)	(0.234)	(0.278)
Picardie* Manufacturing	0.072	0.083	-0.046
	(0.174)	(0.236)	(0.256)
Picardie* Services	0.416**	0.282	0.570*
	(0.191)	(0.249)	(0.308)
Languedoc-Roussillon* Manufacturing	-0.090	-0.232	-0.010
	(0.178)	(0.242)	(0.263)
Languedoc-Roussillon* Services	1.138***	0.882***	1.497***
	(0.168)	(0.223)	(0.266)
Lorraine* Manufacturing	0.188	0.045	0.326
	(0.165)	(0.223)	(0.245)
Lorraine* Services	0.983***	0.745***	1.255***
	(0.164)	(0.217)	(0.257)
Centre* Manufacturing	0.143	0.161	-0.017
	(0.161)	(0.218)	(0.239)
Centre* Services	0.167	0.010	0.408

		(0.160)		(0.213)		(0.252)
Midi-Pyrénées* Manufacturing		-0.215		-0.411*		-0.095
		(0.156)		(0.213)		(0.228)
Midi-Pyrénées* Services		0.271*		0.419*		0.477*
		(0.163)		(0.235)		(0.248)
Bretagne* Manufacturing		0.214		0.385		0.156
		(0.158)		(0.234)		(0.223)
Bretagne* Services		0.455***		0.484**		0.359
		(0.157)		(0.209)		(0.245)
Aquitaine* Manufacturing		-0.020		-0.040		-0.058
		(0.151)		(0.209)		(0.216)
Aquitaine* Services		0.357**		0.340*		0.434*
		(0.150)		(0.202)		(0.233)
Pays de la Loire* Manufacturing		-0.342**		-0.198		-0.544**
		(0.148)		(0.204)		(0.215)
Pays de la Loire* Services		-0.041		0.055		-0.135
		(0.148)		(0.198)		(0.231)
Nord-Pas-de-Calais* Manufacturing		-0.717***		-0.606***		-0.833***
		(0.145)		(0.199)		(0.208)
Nord-Pas-de-Calais* Services		-0.272*		-0.291		-0.355
		(0.151)		(0.200)		(0.240)
Provence-Alpes-Côte d'Azur* Manufacturing		-0.706***		-0.694***		-0.774***
		(0.145)		(0.199)		(0.209)
Provence-Alpes-Côte d'Azur* Services		0.341**		0.168		0.528**
		(0.146)		(0.194)		(0.227)
Rhône-Alpes* Manufacturing		-0.193		-0.260		-0.148
		(0.142)		(0.194)		(0.208)
Rhône-Alpes* Services		0.051		-0.189		0.424*
		(0.140)		(0.188)		(0.216)
Île-de-France* Manufacturing		-0.489***		-0.580***		-0.424**
		(0.137)		(0.188)		(0.198)
Constant	3.303***	-0.705***	3.411***	-0.912***	3.327***	-0.362*
	(0.010)	(0.133)	(0.013)	(0.183)	(0.015)	(0.192)
Mu		0.580***		0.577***		0.597***
		(0.004)		(0.006)		(0.007)
Random Components 121		-0.034***		-0.057***		0.002
		(0.010)		(0.010)		(0.011)

122		0.229***		0.234***		0.215***
		(0.001)		(0.001)		(0.001)
Covariance matrix estimated		$\varepsilon_1 \varepsilon_2$		$\varepsilon_1 \varepsilon_2$		$\varepsilon_1 \varepsilon_2$
	ε_1	1	ε_1	1	ε_1	1
	ε_2	-.033 .053	ε_2	-.056 .057	ε_2	.001 .046
Correlation coefficient		**-0.145**		**-0.235**		**0.008**
Test of correlation	Chi square	11.01	Chi square	34.43	Chi square	0.02
	Prob>chi2	0.000	Prob>chi2	0.000	Prob>chi2	0.880
Observations	74,696	74,696	44,654	44,654	30,042	30,042

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table C-3.2 Joint Model FIML (All sample and across gender)for gross hourly wage (IV Number of dependent children)

	All sample		Male		Female	
Variables	**Wage Equationln(Hourly Wage)**	**Size of Establishment (Ordered Probit)**	**Wage Equationln(Hourly Wage)**	**Size of Establishment (Ordered Probit)**	**Wage Equationln(Hourly Wage)**	**Size of Establishment (Ordered Probit)**
Medium size (base small size)	0.111***		0.090***		0.159***	
	(0.010)		(0.013)		(0.010)	
Large size	0.277***		0.246***		0.341***	
	(0.020)		(0.026)		(0.018)	
Gender (base female)	0.149***	0.091***				
	(0.003)	(0.014)				
Before Bac without degree (base no degree)	0.049***	0.233***	0.055***	0.282***	0.037***	0.161***
	(0.004)	(0.020)	(0.006)	(0.026)	(0.007)	(0.032)
CAP/BEP	0.080***	0.086***	0.069***	0.169***	0.080***	0.005
	(0.004)	(0.018)	(0.005)	(0.022)	(0.006)	(0.030)
Bac professional and technical	0.152***	0.144***	0.140***	0.337***	0.156***	-0.056
	(0.006)	(0.028)	(0.008)	(0.037)	(0.009)	(0.043)
Bac general	0.173***	0.082***	0.160***	0.186***	0.171***	-0.020
	(0.006)	(0.031)	(0.009)	(0.043)	(0.009)	(0.045)
Bac +2	0.201***	0.313***	0.189***	0.370***	0.195***	0.259***
	(0.006)	(0.026)	(0.008)	(0.035)	(0.008)	(0.039)
Bac+3 and plus	0.319***	0.530***	0.300***	0.635***	0.365***	0.366***
	(0.008)	(0.037)	(0.010)	(0.044)	(0.015)	(0.072)

Experience	0.017***	-0.022**	0.019***	-0.015	0.016***	-0.034**
	(0.001)	(0.011)	(0.001)	(0.015)	(0.001)	(0.015)
Experience sq.	-0.000***	0.000	-0.000***	-0.000	-0.000***	0.001
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Experience cube		-0.000		0.000		-0.000
		(0.000)		(0.000)		(0.000)
Tenure	0.008***	0.074***	0.007***	0.084***	0.008***	0.063***
	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.004)
Tenure Sq.	-0.000***	-0.001***	-0.000***	-0.001***	-0.000**	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married (base single)	0.014***	-0.025	0.029***	-0.021	-0.003	-0.027
	(0.005)	(0.024)	(0.007)	(0.035)	(0.007)	(0.033)
Other family status (divorced, widowed etc)	0.011*	-0.021	0.015	-0.106**	0.003	0.031
	(0.006)	(0.031)	(0.010)	(0.048)	(0.008)	(0.041)
Management and High Intellectual professionals (base blue collar)	0.659***	-0.230***	0.653***	-0.193***	0.720***	-0.460***
	(0.005)	(0.025)	(0.006)	(0.029)	(0.011)	(0.052)
High Skilled White Collar	0.267***	-0.050***	0.250***	-0.016	0.348***	-0.234***
	(0.003)	(0.017)	(0.004)	(0.020)	(0.007)	(0.035)
Low Skilled White Collar	0.061***	-0.043**	-0.002	0.131***	0.147***	-0.241***
	(0.004)	(0.018)	(0.006)	(0.029)	(0.006)	(0.029)
CDI Contract(base CDD)	0.047***	-0.248***	0.066***	-0.204***	0.031***	-0.245***
	(0.006)	(0.030)	(0.009)	(0.045)	(0.009)	(0.041)
Manufacturing (base trade)	-0.025***	1.064***	-0.022**	1.127***	-0.029***	0.984***
	(0.008)	(0.022)	(0.011)	(0.029)	(0.009)	(0.035)
Services	0.014***	0.479***	0.006	0.571***	0.019***	0.379***
	(0.005)	(0.021)	(0.007)	(0.029)	(0.006)	(0.030)
Franche-Comté (Base region Limousin)	0.160***	0.406***	0.155***	0.417***	0.158***	0.421***
	(0.009)	(0.042)	(0.012)	(0.057)	(0.013)	(0.063)
Auvergne	0.038***	0.292***	0.053***	0.356***	0.011	0.176*

	(0.011)	(0.055)	(0.014)	(0.070)	(0.018)	(0.090)
Champagne-Ardenne	0.053***	0.288***	0.066***	0.307***	0.034**	0.262***
	(0.010)	(0.050)	(0.013)	(0.065)	(0.016)	(0.077)
Basse-Normandie	0.075***	0.216***	0.097***	0.181***	0.034**	0.298***
	(0.010)	(0.049)	(0.013)	(0.065)	(0.015)	(0.075)
Bourgogne	0.020**	0.351***	0.036***	0.470***	0.002	0.145*
	(0.010)	(0.048)	(0.013)	(0.064)	(0.015)	(0.075)
Poitou-Charentes	0.019*	0.133**	0.021	0.064	0.009	0.241***
	(0.011)	(0.055)	(0.015)	(0.075)	(0.017)	(0.082)
Alsace	0.017*	0.222***	0.026*	0.307***	0.007	0.109
	(0.010)	(0.050)	(0.013)	(0.066)	(0.015)	(0.075)
Haute-Normandie	-0.007	0.439***	0.012	0.405***	-0.039***	0.493***
	(0.009)	(0.045)	(0.012)	(0.059)	(0.014)	(0.069)
Picardie	0.053***	0.460***	0.079***	0.506***	0.003	0.374***
	(0.010)	(0.048)	(0.013)	(0.062)	(0.016)	(0.075)
Languedoc-Roussillon	0.075***	0.367***	0.109***	0.292***	0.011	0.500***
	(0.010)	(0.047)	(0.013)	(0.062)	(0.015)	(0.071)
Lorraine	0.012	0.169***	0.022*	0.162***	-0.004	0.192***
	(0.009)	(0.046)	(0.012)	(0.062)	(0.014)	(0.070)
Centre	-0.001	0.186***	0.010	0.026	-0.027*	0.403***
	(0.010)	(0.048)	(0.013)	(0.064)	(0.015)	(0.071)
Midi-Pyrénées	0.038***	0.244***	0.042***	0.198***	0.031*	0.323***
	(0.011)	(0.051)	(0.014)	(0.068)	(0.016)	(0.077)
Bretagne	0.054***	0.125***	0.083***	0.081	0.004	0.214***
	(0.009)	(0.046)	(0.012)	(0.062)	(0.014)	(0.069)
Aquitaine	0.016	0.130***	0.041***	0.022	-0.019	0.268***
	(0.010)	(0.049)	(0.013)	(0.066)	(0.015)	(0.072)
Pays de la Loire	-0.031***	0.069	-0.027*	0.020	-0.037**	0.132
	(0.012)	(0.058)	(0.015)	(0.076)	(0.019)	(0.092)
Nord-Pas-de-Calais	0.066***	0.201***	0.081***	0.256***	0.045***	0.127*

	(0.009)	(0.043)	(0.012)	(0.058)	(0.013)	(0.065)
Provence-Alpes-Côte d'Azur	0.006	0.123**	0.008	0.230***	0.009	-0.011
	(0.012)	(0.059)	(0.016)	(0.078)	(0.018)	(0.090)
Rhône-Alpes	0.040***	-0.083	0.060***	-0.159**	0.004	0.038
	(0.012)	(0.059)	(0.016)	(0.080)	(0.018)	(0.087)
Île-de-France	0.106***	0.103**	0.118***	0.068	0.079***	0.188***
	(0.009)	(0.046)	(0.012)	(0.061)	(0.014)	(0.069)
nbenf		0.029***		0.022***		0.041***
		(0.006)		(0.008)		(0.011)
Constant	3.321***	-0.877***	3.443***	-1.075***	3.292***	-0.444***
	(0.014)	(0.089)	(0.019)	(0.126)	(0.020)	(0.126)
Mu		0.570***		0.570***		0.585***
		(0.006)		(0.007)		(0.009)
Random Components						
121		-0.098***		-0.074***		-0.139***
		(0.010)		(0.012)		(0.009)
122		0.220***		0.229***		0.198***
		(0.002)		(0.002)		(0.002)
Covariance matrix estimated		$\varepsilon_1 \varepsilon_2$		$\varepsilon_1 \varepsilon_2$		$\varepsilon_1 \varepsilon_2$
	ε_1	1	ε_1	1	ε_1	1
	ε_2	-.097 .058	ε_2	-.073 .057	ε_2	-.138 .058
Correlation Coeficient		**-0,406**		**-0,306**		**-0,573**
Test of correlation	Chi square	134.96	Chi square	39.91	Chi square	444.90
	Prob>chi2 =	0.0000	Prob>chi2 =	0.0000	Prob>chi2 =	0.0000

Observations	44,941	44,941	27,304	27,304	17,637	17,637

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table C-3.3 FIML Gross hourly wages (Results across professions)

	Prof1	**Prof2**	**Prof3**	**Prof4**	**Prof1**	**Prof2**	**Prof3**	**Prof4**	**Prof1**	**Prof2**	**Prof3**	**Prof4**
	Model 1				**Model 2**				**Model 3**			
Medium size (base small size)	-0.051**	0.081***	0.026	0.013	0.057**	0.081***	0.143***	0.134***	0,022	0.092***	0,015	0,017
	(0.024)	(0.013)	(0.023)	(0.012)	(0,028)	(0,017)	(0,014)	(0,013)	(0,033)	(0,014)	(0,015)	(0,016)
Large size	-0.084*	0.183***	0.094**	0.115***	0.137***	0.189***	0.312***	0.363***	0,069	0.210***	0.075***	0.124***
	(0.043)	(0.025)	(0.041)	(0.023)	(0,052)	(0,032)	(0,025)	(0,025)	(0,060)	(0,026)	(0,027)	(0,032)
Correlation coefficient	**0,186**	**-0,281**	**-0,093**	**0,062**	**-0,206**	**-0,295**	**-0,541**	**-0,469**	**-0,086**	**-0,340**	**-0,048**	**0,036**
Test of correlation	6.35	32.23	0.97	1.47	5,370	22,840	171,230	110,190	0,640	47,630	0,570	0,250
p-value	0,0117	0,000	0,323	0.224	0,020	0,000	0,000	0,000	0,424	0,000	0,449	0,618
Observations	9320	20898	18033	26445	5705	12781	10545	15910	5,705	12781	10545	15910

*Prof1= Management and High Intellectual professionals, prof2=High Skilled White Collar, prof3=Low Skilled White Collar, prof4=Blue Collar. Model 1 (Interactions), Model 2 (Number of dependent children),Model 3 (*Model1+Model 2)Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table C-3.4 FIML Gross hourly wages (Results by excluding blue collar workers)

	Model 1			Model 2			Model 3		
	All sample	**Male**	**Female**	**All sample**	**Male**	**Female**	**All sample**	**Male**	**Female**
Medium size (base small size)	0.041***	0.046***	0,011	0.096***	0.056***	0.163***	0.061***	0.066***	0,005
	(0,013)	(0,012)	(0,016)	(0,014)	(0,015)	(0,012)	(0,013)	(0,013)	(0,016)
Large size	0.104***	0.123***	0.047*	0.216***	0.149***	0.333***	0.147***	0.169***	0,041
	(0,024)	(0,023)	(0,028)	(0,025)	(0,028)	(0,020)	(0,024)	(0,024)	(0,028)
Correlation coefficient	**-0,134**	**-0,174**	**0,002**	**-0,350**	**-0,224**	**-0,563**	**-0,220**	**-0,265**	**0,025**
Test of correlation	7,380	16,130	0,000	56,610	18,630	327,580	21,530	35,960	0,160
p-value	0,007	0,000	0,969	0,000	0,000	0,000	0,000	0,000	0,694
Observations	48 285	23 455	24 830	29 031	14459	14572	29031	14 459	14 572

*Model 1 (Interactions), Model 2 (Number of dependent children), Model 3 (*Model1+Model 2*)*. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table C-3.5FIML Gross hourly wages (Excluding large regions)

	Model 1	Model 2
Medium size (base small size)	0.024**	0.017**
	(0.010)	(0.009)
Large size	0.106***	0.094***
	(0.018)	(0.016)
Correlation coefficient	**-0,02**	**0,00**
Test of correlation	0,22	0
	0,636	0,989
Observations	48 209	61,349

Notes : Model 1 (18 region, excluding ile de France, rhone alpes, cote d'azur), Model 2 (20 regions excluding ile de France)
*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.*

Table C-3.6FIML Gross hourly wages (Unionization)

	Model 1		Model 2	
	Covered union representatives	Covered staff representatives	Covered union representatives	Covered staff representatives
Medium size (base small size)	0.007	0.024***	0.021**	0.033***
	(0.008)	(0.007)	(0.009)	(0.009)
Large size	0.088***	0.108***	0.131***	0.135***
	(0.015)	(0.013)	(0.019)	(0.018)
Correlation coefficient	-0,09	-0,08	-0,17	-0,13
Test of correlation	9,72	8,29	24,99	13,76
p-value	0,0018	0,004	0	0,0002
Observations	37446	48405	23545	30048

*Model 1: Interaction of region size and industry as IV, Model 2: number of dependent children as IVRobust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.*

Appendix - D Chapter5

Table D-1 Probit model of choice of employer size

Dependent variable size dummy	**Males**	**Females**
Experience	-0.068***	-0.040***
	(0.010)	(0.011)
Exp. squared	0.002***	0.001**
	(0.000)	(0.001)
Exp. cube	-0.000***	-0.000**
	(0.000)	(0.000)
Tenure	0.097***	0.069***
	(0.003)	(0.003)
Tenure. squared	-0.001***	-0.001***
	(0.000)	(0.000)
Married (base single)	0.043**	0.037*
	(0.018)	(0.021)
Other family status (divorced, widowed etc)	-0.015	0.079**
	(0.034)	(0.031)
Before Bac without degree (base no degree)	0.353***	0.181***
	(0.023)	(0.028)
CAP/BEP	0.213***	0.042
	(0.020)	(0.027)
Bac professional and technical	0.407***	0.077**
	(0.033)	(0.039)
Bac general	0.287***	-0.020
	(0.039)	(0.040)
Bac +2	0.457***	0.335***
	(0.031)	(0.034)
Bac+3 and plus	0.828***	0.531***
	(0.040)	(0.057)
Management and High Intellectual professionals (base blue collar)	-0.162***	-0.308***
	(0.027)	(0.043)
High Skilled White Collar	0.021	-0.081***
	(0.018)	(0.031)
Low Skilled White Collar	0.229***	-0.116***
	(0.025)	(0.027)
Type of Contract	-0.019	-0.223***
	(0.039)	(0.037)
Manufacturing (base trade)	1.086***	0.278
	(0.258)	(0.253)
Services	1.008***	0.515**
	(0.257)	(0.234)
Franche-Comté (Base region Limousin)	1.259***	1.059***
	(0.280)	(0.259)

Auvergne	-0.028	-0.387
	(0.348)	(0.309)
Champagne-Ardenne	-4.571***	-0.575
	(0.263)	(0.360)
Basse-Normandie	0.638**	-0.306
	(0.285)	(0.285)
Bourgogne	-0.372	-1.480***
	(0.339)	(0.435)
Poitou-Charentes	0.173	-0.008
	(0.314)	(0.277)
Alsace	-0.932**	-1.219***
	(0.388)	(0.364)
Haute-Normandie	-0.135	-1.282***
	(0.305)	(0.458)
Picardie	-0.111	-0.335
	(0.326)	(0.294)
Languedoc-Roussillon	-1.011**	-4.806***
	(0.469)	(0.228)
Lorraine	-4.408***	-4.696***
	(0.254)	(0.222)
Centre	-0.110	-0.523*
	(0.300)	(0.275)
Midi-Pyrénées	0.453*	0.045
	(0.275)	(0.254)
Bretagne	-0.325	-0.593**
	(0.311)	(0.284)
Aquitaine	-0.424	-0.659**
	(0.303)	(0.270)
Pays de la Loire	-0.107	0.076
	(0.278)	(0.243)
Nord-Pas-de-Calais	0.519*	0.804***
	(0.266)	(0.233)
Provence-Alpes-Côte d'Azur	0.490*	0.362
	(0.267)	(0.235)
Rhône-Alpes	0.197	-0.403*
	(0.264)	(0.243)
Île-de-France	0.820***	0.354
	(0.254)	(0.225)
Industry type *Region Size (Base Industry Trade and Base region Limousin with lowest population)		
Franche-Comté * Manufacturing	-0.850***	-0.618**
	(0.297)	(0.304)
Franche-Comté * Services	-2.482***	-1.951***
	(0.308)	(0.287)
Auvergne* Manufacturing	0.759**	1.269***
	(0.360)	(0.348)
Auvergne* Services	-0.242	-0.069
	(0.364)	(0.331)
Champagne-Ardenne* Manufacturing	5.114***	1.571***
	(0.279)	(0.396)
Champagne-Ardenne* Services	4.616***	0.244

	(0.277)	(0.377)
Basse-Normandie* Manufacturing	-0.455	1.248***
	(0.298)	(0.322)
Basse-Normandie* Services	-0.925***	0.294
	(0.302)	(0.303)
Bourgogne* Manufacturing	0.890**	2.408***
	(0.350)	(0.459)
Bourgogne* Services	0.335	0.768*
	(0.349)	(0.446)
Poitou-Charentes* Manufacturing	0.042	0.412
	(0.328)	(0.320)
Poitou-Charentes* Services	-0.416	0.071
	(0.325)	(0.293)
Alsace* Manufacturing	1.443***	1.846***
	(0.397)	(0.391)
Alsace* Services	0.657*	1.399***
	(0.396)	(0.375)
Haute-Normandie* Manufacturing	0.632**	2.430***
	(0.316)	(0.481)
Haute-Normandie* Services	-0.022	1.027**
	(0.317)	(0.468)
Picardie* Manufacturing	0.697**	1.366***
	(0.336)	(0.329)
Picardie* Services	-0.253	-0.162
	(0.338)	(0.312)
Languedoc-Roussillon* Manufacturing	1.299***	5.415***
	(0.482)	(0.303)
Languedoc-Roussillon* Services	0.659	4.684***
	(0.479)	(0.249)
Lorraine* Manufacturing	5.343***	5.948***
	(0.266)	(0.266)
Lorraine* Services	4.275***	4.544***
	(0.266)	(0.242)
Centre* Manufacturing	0.831***	1.663***
	(0.310)	(0.309)
Centre* Services	0.025	-0.030
	(0.310)	(0.293)
Midi-Pyrénées* Manufacturing	-0.198	0.599**
	(0.287)	(0.293)
Midi-Pyrénées* Services	-0.843***	-0.075
	(0.288)	(0.270)
Bretagne* Manufacturing	0.298	1.186***
	(0.322)	(0.320)
Bretagne* Services	0.134	0.754**
	(0.321)	(0.298)
Aquitaine* Manufacturing	0.706**	1.080***
	(0.314)	(0.308)
Aquitaine* Services	0.019	0.479*
	(0.314)	(0.284)
Pays de la Loire* Manufacturing	0.466	0.625**
	(0.288)	(0.279)

Pays de la Loire* Services	-0.178	-0.376
	(0.290)	(0.260)
Nord-Pas-de-Calais* Manufacturing	0.141	0.110
	(0.277)	(0.273)
Nord-Pas-de-Calais* Services	-0.745***	-0.787***
	(0.276)	(0.250)
Provence-Alpes-Côte d'Azur* Manufacturing	-0.307	0.089
	(0.278)	(0.278)
Provence-Alpes-Côte d'Azur* Services	-0.955***	-0.796***
	(0.277)	(0.251)
Rhône-Alpes* Manufacturing	0.182	0.995***
	(0.274)	(0.277)
Rhône-Alpes* Services	-0.473*	0.048
	(0.274)	(0.258)
Île-de-France* Manufacturing	-0.293	0.650**
	(0.264)	(0.259)
Île-de-France* Services	-0.715***	-0.280
	(0.263)	(0.240)
Constant	-1.766***	-0.939***
	(0.261)	(0.234)
Observations	44,654	30,042

Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table- D-2 Adjusted and unadjusted wage estimates across gender

Dependent variable: log of gross hourly wage	**Large Establishments**				**Small Establishments**			
	Unadjusted (OLS)		**Adjusted (Heckman second-step)**		**Unadjusted (OLS)**		**Adjusted Heckman second-step)**	
	Large Size				**Small Size**			
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Experience	0.037***	0.027***	0.045***	0.027***	0.028***	0.019***	0.028***	0.019***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Tenure	0.010***	0.012***	-0.001	0.012***	0.013***	0.013***	0.013***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Married (base single)	0.049***	-0.008	0.044***	-0.008	0.052***	0.004	0.052***	0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Other family status (divorced, widowed etc)	0.026***	0.005	0.027***	0.005	0.023***	0.014**	0.023***	0.014**
	(0.008)	(0.008)	(0.009)	(0.008)	(0.007)	(0.006)	(0.007)	(0.006)
Before Bac without degree (base no degree)	0.082***	0.057***	0.048***	0.057***	0.054***	0.046***	0.053***	0.047***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)
CAP/BEP	0.102***	0.101***	0.082***	0.101***	0.070***	0.066***	0.069***	0.066***
	(0.005)	(0.007)	(0.006)	(0.007)	(0.004)	(0.005)	(0.005)	(0.005)
Bac professional and technical	0.177***	0.178***	0.139***	0.178***	0.149***	0.136***	0.148***	0.136***
	(0.008)	(0.010)	(0.009)	(0.010)	(0.008)	(0.007)	(0.008)	(0.007)
Bac general	0.189***	0.198***	0.164***	0.198***	0.148***	0.155***	0.148***	0.155***
	(0.010)	(0.011)	(0.011)	(0.011)	(0.009)	(0.007)	(0.009)	(0.007)
Bac +2	0.231***	0.267***	0.188***	0.267***	0.187***	0.203***	0.186***	0.203***
	(0.008)	(0.009)	(0.009)	(0.010)	(0.007)	(0.007)	(0.008)	(0.007)
Bac+3 and plus	0.359***	0.428***	0.281***	0.429***	0.304***	0.332***	0.303***	0.333***

	(0.010)	(0.015)	(0.013)	(0.016)	(0.009)	(0.011)	(0.011)	(0.012)
Management and High Intellectual professionals (base blue collar)	0.562***	0.592***	0.577***	0.591***	0.689***	0.719***	0.689***	0.718***
	(0.007)	(0.011)	(0.007)	(0.011)	(0.006)	(0.008)	(0.006)	(0.009)
High Skilled White Collar	0.195***	0.299***	0.194***	0.299***	0.299***	0.358***	0.299***	0.358***
	(0.004)	(0.007)	(0.005)	(0.007)	(0.004)	(0.006)	(0.004)	(0.006)
Low Skilled White Collar	-0.031***	0.125***	-0.053***	0.125***	0.029***	0.133***	0.029***	0.133***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)
Type of Contract	0.200***	0.076***	0.196***	0.075***	0.008	-0.006	0.008	-0.006
	(0.012)	(0.010)	(0.012)	(0.010)	(0.007)	(0.006)	(0.007)	(0.006)
Manufacturing (base trade)	0.054***	0.088***	-0.058***	0.090***	0.028***	0.060***	0.026***	0.063***
	(0.010)	(0.009)	(0.015)	(0.012)	(0.005)	(0.006)	(0.010)	(0.010)
Services	0.062***	0.054***	0.021**	0.055***	0.020***	0.047***	0.019***	0.048***
	(0.010)	(0.009)	(0.011)	(0.009)	(0.005)	(0.004)	(0.006)	(0.005)
λ			**-0.166***	**0.004**			**0.004**	**-0.006**
			(0.017)	(0.014)			(0.014)	(0.014)
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	3.273***	3.216***	3.553***	3.211***	3.384***	3.343***	3.384***	3.345***
	(0.025)	(0.028)	(0.039)	(0.033)	(0.020)	(0.020)	(0.020)	(0.020)
Observations	19,758	10,138	44,654	30,042	24,896	19,904	44,654	30,042
R-squared	0.63	0.64			0.65	0.61		
Adj. R-squared	0.63	0.64			0.65	0.61		

*Notes: Experience includes square and cubic term. Tenure includes squared term. Region includes 21 dummies (Region size indicates population size ranked from low to high, base category is region with lowest urban population). Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*

Table D-3 Detail threefold decomposition

	Large establishment			Small Establishment		
	Endowments	**Coefficients**	**Interaction**	**Endowments**	**Coefficients**	**Interaction**
Experience	0.036***	0.101	0.007	0.009***	0.148**	0.004**
	(0.005)	(0.084)	(0.005)	(0.002)	(0.065)	(0.002)
Exp. squared	-0.047***	-0.056	-0.006	-0.008***	-0.110	-0.004
	(0.009)	(0.101)	(0.010)	(0.003)	(0.076)	(0.003)
Exp. cube	0.016***	0.015	0.002	0.002	0.031	0.001
	(0.004)	(0.042)	(0.005)	(0.001)	(0.031)	(0.001)
Tenure	0.025***	0.068**	0.011**	0.005***	0.018	0.001
	(0.004)	(0.032)	(0.005)	(0.001)	(0.016)	(0.001)
Tenure. squared	-0.004	-0.132***	-0.035***	-0.006***	-0.035*	-0.003*
	(0.009)	(0.041)	(0.011)	(0.002)	(0.019)	(0.002)
Tenure cube	-0.000	0.055***	0.019***	0.003***	0.012	0.002
	(0.005)	(0.018)	(0.006)	(0.001)	(0.008)	(0.001)
Single	-0.000	-0.006***	0.001***	0.000	-0.005***	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Married	-0.001**	0.020***	0.003***	-0.000	0.018***	0.002***
	(0.000)	(0.003)	(0.000)	(0.000)	(0.002)	(0.000)
Other family status (divorced, widowed etc)	-0.000	-0.001	0.000	-0.000**	-0.001*	0.001*
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
No degree	0.000	0.002	-0.000	-0.004***	0.001	0.000
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Before Bac without degree	0.004***	0.008***	-0.001***	0.002***	0.002*	-0.000*
	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)
CAP/BEP	-0.009***	0.004**	0.002**	-0.005***	0.002	0.001
	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)	(0.000)
Bac professional and technical	-0.000	0.001	-0.000	-0.000	0.001**	-0.000**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)

Bac general	-0.000***	0.000	-0.000	-0.001***	-0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Bac +2	-0.008***	-0.004***	0.002***	-0.004***	-0.002*	0.001*
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
Bac+3 and plus	0.009***	-0.002***	-0.002***	0.004***	-0.001**	-0.000**
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Management and High Intellectual professionals	0.025***	0.003***	0.003***	0.030***	0.001***	0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
High Skilled White Collar	-0.000	-0.010***	0.000	-0.001**	-0.003**	0.000*
	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)
Low Skilled White Collar	0.040***	-0.032***	0.026***	0.068***	-0.028***	0.022***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)
Blue collar	-0.061***	0.016***	0.017***	-0.102***	0.007***	0.016***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Trade	0.001***	0.001	-0.000	0.001***	0.003***	-0.000***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Manufacturing	0.010***	-0.011***	-0.006***	0.004***	-0.002***	-0.002***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Services	-0.001*	0.009***	-0.004***	-0.002***	-0.005**	0.001**
	(0.001)	(0.003)	(0.001)	(0.000)	(0.002)	(0.001)
Base region Limousin	-0.000**	0.000	0.000	-0.000	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Franche-Comté	0.000	-0.000	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Auvergne	0.000	-0.001***	-0.000**	0.000**	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Champagne-Ardenne	-0.000***	0.001***	0.000**	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Basse-Normandie	-0.000	-0.001**	0.000**	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Bourgogne	-0.001***	0.001**	0.000**	0.000*	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Poitou-Charentes	-0.000**	-0.002***	0.000**	-0.000	-0.001**	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Alsace	-0.001***	0.002**	-0.000**	0.000	0.002***	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Haute-Normandie	0.000	0.002***	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Picardie	0.001***	-0.001*	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Languedoc-Roussillon	0.000***	0.001**	-0.000**	0.000*	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lorraine	0.001*	0.001	0.000	0.000	0.001***	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Centre	-0.000	-0.000	-0.000	-0.000	0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Midi-Pyrénées	0.000***	0.002***	-0.001**	0.000	-0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Bretagne	0.000	-0.004***	0.002***	-0.000**	-0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Aquitaine	0.000	0.004***	-0.000	0.000*	-0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Pays de la Loire	-0.000*	-0.001	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Nord-Pas-de-Calais	-0.000**	-0.001	-0.000	-0.000***	-0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Provence-Alpes-Côte d'Azur	-0.000	0.000	-0.000	-0.000	0.002***	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Rhône-Alpes	0.001***	0.001	0.000	0.000	0.002**	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Île-de-France	-0.007***	-0.017***	0.004***	-0.003***	-0.004***	0.000***
	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)	(0.000)
Type of Contract	0.002***	0.108***	0.003***	-0.000	0.010	0.000
	(0.000)	(0.016)	(0.001)	(0.000)	(0.009)	(0.000)
Constant		-0.016			0.050**	
		(0.028)			(0.022)	
Observations	29,896	29,896	29,896	44,800	44,800	44,800

*Notes: all categories of categorical variables are included so that the transformed coefficients do not depend on the choice of the omitted (base) category. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*

Table D-4 Detail twofold decomposition

	Large establishment		Small Establishment	
	Explained	**Unexplained**	**Explained**	**Unexplained**
Experience	0.038***	0.105	0.011***	0.150**
	(0.004)	(0.096)	(0.002)	(0.072)
Exp. squared	-0.047***	-0.062	-0.010***	-0.113
	(0.007)	(0.116)	(0.003)	(0.083)
Exp. cube	0.016***	0.017	0.002**	0.032
	(0.003)	(0.049)	(0.001)	(0.034)
Tenure	0.035***	0.068*	0.006***	0.018
	(0.003)	(0.038)	(0.002)	(0.017)
Tenure. squared	-0.032***	-0.139***	-0.008***	-0.036*
	(0.006)	(0.049)	(0.002)	(0.022)
Tenure cube	0.015***	0.059***	0.004***	0.012
	(0.003)	(0.021)	(0.001)	(0.009)
Single	0.001***	-0.005***	0.000	-0.005***
	(0.000)	(0.001)	(0.000)	(0.001)
Married	0.001***	0.022***	0.001***	0.019***
	(0.000)	(0.003)	(0.000)	(0.002)
Other family status (divorced, widowed etc)	-0.000	-0.000	-0.000	-0.001
	(0.000)	(0.001)	(0.000)	(0.000)
No degree	0.000	0.002	-0.005***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Before Bac without degree	0.004***	0.008***	0.001***	0.002*
	(0.000)	(0.002)	(0.000)	(0.001)
CAP/BEP	-0.007***	0.004**	-0.005***	0.002
	(0.000)	(0.002)	(0.000)	(0.002)
Bac professional and technical	-0.000	0.001	-0.000***	0.001**
	(0.000)	(0.001)	(0.000)	(0.001)
Bac general	-0.001***	0.000	-0.001***	-0.000
	(0.000)	(0.001)	(0.000)	(0.001)
Bac +2	-0.008***	-0.003**	-0.004***	-0.001
	(0.001)	(0.001)	(0.000)	(0.001)
Bac+3 and plus	0.007***	-0.003***	0.003***	-0.001
	(0.001)	(0.001)	(0.000)	(0.001)
Management and High Intellectual professionals	0.027***	0.005***	0.030***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
High Skilled White Collar	-0.000	-0.010***	-0.000**	-0.003**
	(0.000)	(0.002)	(0.000)	(0.001)
Low Skilled White Collar	0.053***	-0.020***	0.077***	-0.015***

	(0.001)	(0.002)	(0.001)	(0.001)
Blue collar	-0.048***	0.020***	-0.091***	0.012***
	(0.001)	(0.002)	(0.001)	(0.001)
Trade	0.001***	0.000	0.001***	0.003***
	(0.000)	(0.000)	(0.000)	(0.001)
Manufacturing	0.005***	-0.012***	0.003***	-0.003***
	(0.001)	(0.003)	(0.000)	(0.001)
Services	-0.004***	0.008***	-0.001***	-0.005**
	(0.001)	(0.002)	(0.000)	(0.002)
Base region Limousin	-0.000**	0.000	-0.000	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Franche-Comté	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Auvergne	-0.000***	-0.001***	0.000**	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Champagne-Ardenne	-0.000	0.001***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Basse-Normandie	0.000*	-0.001**	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Bourgogne	-0.000***	0.001**	0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Poitou-Charentes	-0.000**	-0.002***	-0.000**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Alsace	-0.001***	0.001**	0.000	0.002***
	(0.000)	(0.001)	(0.000)	(0.000)
Haute-Normandie	-0.000	0.002***	-0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)
Picardie	0.000***	-0.001*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Languedoc-Roussillon	0.000**	0.001**	0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Lorraine	0.001***	0.001	0.000***	0.001***
	(0.000)	(0.001)	(0.000)	(0.000)
Centre	-0.000*	0.000	0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Midi-Pyrénées	0.000***	0.001***	0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Bretagne	0.001***	-0.004***	-0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Aquitaine	-0.000	0.004***	0.000**	-0.000
	(0.000)	(0.001)	(0.000)	(0.001)
Pays de la Loire	-0.000*	-0.001	-0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)
Nord-Pas-de-Calais	-0.001***	-0.001*	-0.000***	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)
Provence-Alpes-Côte d'Azur	-0.000	0.000	-0.000	0.002***
	(0.000)	(0.001)	(0.000)	(0.001)
Rhône-Alpes	0.001***	0.001	0.000	0.002**
	(0.000)	(0.001)	(0.000)	(0.001)
Île-de-France	-0.005***	-0.015***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)

Type of Contract	0.004***	0.110***	-0.000	0.010
	(0.000)	(0.021)	(0.000)	(0.010)
Constant		-0.016		0.050**
		(0.033)		(0.024)
Observations	29,896	29,896	44,800	44,800

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.

Table D-5 Detail threefold decomposition after adjusting for selection

	Large establishment			Small Establishment		
	Endowments	**Coefficients**	**Endowments**	**Coefficients**	**Endowments**	**Coefficients**
Experience	0.035***	0.376***	0.024***	0.010***	0.183***	0.005**
	(0.005)	(0.085)	(0.006)	(0.002)	(0.065)	(0.002)
Exp. squared	-0.046***	-0.338***	-0.034***	-0.009***	-0.153**	-0.006*
	(0.009)	(0.098)	(0.011)	(0.003)	(0.075)	(0.003)
Exp. cube	0.016***	0.130***	0.016***	0.002**	0.049	0.002
	(0.004)	(0.040)	(0.005)	(0.001)	(0.030)	(0.001)
Tenure	0.026***	-0.170***	-0.027***	0.004***	-0.007	-0.000
	(0.003)	(0.023)	(0.004)	(0.001)	(0.010)	(0.000)
Tenure. squared	-0.005***	0.036***	0.010***	-0.001***	-0.006	-0.001
	(0.002)	(0.010)	(0.003)	(0.000)	(0.004)	(0.000)
Married (base single)	-0.001	0.034***	0.006***	0.000	0.030***	0.003***
	(0.001)	(0.005)	(0.001)	(0.000)	(0.004)	(0.000)
Other family status (divorced, widowed etc)	-0.000	0.003*	-0.001*	-0.001**	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Before Bac without degree (base no degree)	-0.002***	-0.002	0.000	-0.001***	0.001	-0.000
	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)
CAP/BEP	0.012***	-0.005**	-0.002**	0.005***	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Bac professional and technical	-0.000	-0.003***	0.000	-0.003***	0.001	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Bac general	-0.004***	-0.002**	0.001**	-0.005***	-0.001	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)

Bac +2	-0.025***	-0.015***	0.007***	-0.013***	-0.003*	0.001*
	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)
Bac+3 and plus	0.015***	-0.006***	-0.005***	0.006***	-0.001*	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Management and High Intellectual professionals (base blue collar)	0.044***	-0.001	-0.001	0.052***	-0.002***	-0.002***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
High Skilled White Collar	-0.001	-0.033***	0.001	-0.004***	-0.016***	0.001**
	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.000)
Low Skilled White Collar	-0.038***	-0.068***	0.055***	-0.053***	-0.053***	0.042***
	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)
Type of Contract	0.002***	0.115***	0.003***	-0.000	0.013	0.000
	(0.000)	(0.015)	(0.001)	(0.000)	(0.009)	(0.000)
Manufacturing (base trade)	0.022***	-0.061***	-0.036***	0.011***	-0.005***	-0.007***
	(0.003)	(0.008)	(0.005)	(0.002)	(0.002)	(0.003)
Services	-0.012***	-0.018**	0.007**	-0.008***	-0.020***	0.005***
	(0.002)	(0.007)	(0.003)	(0.001)	(0.005)	(0.001)
Franche-Comté (Base region Limousin)	-0.000	-0.001	0.000	-0.000	0.001**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Auvergne	0.001***	-0.002***	-0.001***	0.000	0.001	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Champagne-Ardenne	0.000*	-0.000	-0.000	0.000*	0.001***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Basse-Normandie	-0.001***	-0.001*	0.001*	0.000	0.001**	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Bourgogne	0.001**	-0.000	-0.000	-0.000	0.002**	-0.000**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Poitou-Charentes	-0.001**	-0.002***	0.000*	0.000	0.001	0.000

	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Alsace	-0.002***	-0.001	0.000	0.000	0.005***	0.000
	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)
Haute-Normandie	-0.000	0.001	-0.000	-0.000	0.002***	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Picardie	0.002***	-0.002**	-0.001**	0.000	0.002***	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Languedoc-Roussillon	-0.000	0.001	-0.000	0.000	0.002***	-0.000*
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Lorraine	0.003***	-0.002**	-0.001*	0.000**	0.003***	0.001***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
Centre	0.001***	-0.002**	-0.001**	0.000	0.003***	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Midi-Pyrénées	-0.001***	0.001	-0.000	-0.000	0.003***	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Bretagne	-0.002***	-0.005***	0.002***	0.000	0.002***	0.000**
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
Aquitaine	-0.000	0.003***	-0.000	-0.000	0.003***	-0.000**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Pays de la Loire	0.000*	-0.002	-0.000	0.000	0.002**	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Nord-Pas-de-Calais	0.001***	-0.004**	-0.001**	0.000	0.003**	0.001**
	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)
Provence-Alpes-Côte d'Azur	-0.000	-0.000	0.000	-0.000	0.006***	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.002)	(0.000)
Rhône-Alpes	0.003***	-0.001	-0.000	0.000	0.008***	0.000
	(0.001)	(0.002)	(0.001)	(0.000)	(0.002)	(0.000)
Île-de-France	-0.012***	-0.028***	0.006***	-0.004***	0.006*	-0.001*
	(0.001)	(0.005)	(0.001)	(0.001)	(0.003)	(0.000)
Constant		0.342***			0.041	
		(0.051)			(0.029)	
Observations	29,896	29,896	29,896	44,800	44,800	44,800

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

D-6 Detail threefold decompositionacross professions selection adjusted in large size establishments

	Prof1		Prof2		Prof3		Prof4	
	Endowments	Coefficients	Endowments	Coefficients	Endowments	Coefficients	Endowments	Coefficients
Experience	0.167***	0.215	0.038***	0.403***	0.014**	-0.425**	0.009	0.563**
	(0.040)	(0.206)	(0.014)	(0.125)	(0.007)	(0.215)	(0.011)	(0.240)
Exp. squared	-0.191***	-0.182	-0.003	-0.388***	-0.023*	0.358	-0.021	-0.623**
	(0.068)	(0.253)	(0.030)	(0.143)	(0.013)	(0.244)	(0.023)	(0.282)
Exp. cube	0.072**	0.068	-0.017	0.157***	0.009	-0.096	0.014	0.249**
	(0.033)	(0.111)	(0.016)	(0.057)	(0.006)	(0.097)	(0.012)	(0.113)
Tenure	-0.007	0.015	0.016**	-0.117***	0.020***	0.042	0.026***	-0.228***
	(0.012)	(0.058)	(0.007)	(0.041)	(0.006)	(0.052)	(0.005)	(0.037)
Tenure. squared	0.013	-0.037	0.003	0.017	-0.005**	-0.046**	-0.011***	0.091***
	(0.008)	(0.030)	(0.005)	(0.017)	(0.002)	(0.023)	(0.003)	(0.018)
Married (base single)	0.008*	0.010	-0.000	0.029***	-0.000	0.044***	-0.000	0.036***
	(0.004)	(0.014)	(0.001)	(0.008)	(0.000)	(0.011)	(0.000)	(0.010)
Other family status (divorced, widowed etc)	-0.001	-0.002	-0.001	-0.000	-0.000	0.004	-0.001	0.003
	(0.002)	(0.004)	(0.001)	(0.002)	(0.001)	(0.004)	(0.001)	(0.002)
Before Bac without degree (base no degree)	-0.000	0.001	0.000	0.001	0.002***	-0.008	-0.008***	-0.002
	(0.000)	(0.005)	(0.000)	(0.004)	(0.001)	(0.005)	(0.001)	(0.005)
CAP/BEP	-0.001	0.002	0.010**	-0.006	-0.003*	-0.034***	0.023***	0.004
	(0.001)	(0.006)	(0.004)	(0.004)	(0.002)	(0.009)	(0.003)	(0.003)
Bac professional and technical	0.000	0.001	0.004***	-0.003	-0.005***	-0.005**	0.002***	-0.001*
	(0.001)	(0.006)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Bac general	0.000	-0.002	-0.003***	-0.003	-0.001	-0.006***	-0.000	-0.001**

	(0.001)	(0.005)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Bac +2	-0.019***	-0.035*	-0.050***	-0.022*	-0.003*	-0.006***	0.001*	0.000
	(0.006)	(0.021)	(0.006)	(0.012)	(0.002)	(0.002)	(0.000)	(0.000)
Bac+3 and plus	0.017*	-0.059*	0.001	-0.002***	0.001	-0.000	0.000	0.000
	(0.009)	(0.031)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Type of Contract	0.006**	0.163***	0.001**	0.079**	0.001	0.084***	-0.000	0.063**
	(0.002)	(0.054)	(0.000)	(0.034)	(0.000)	(0.031)	(0.000)	(0.028)
Manufacturing (base trade)	0.008	-0.039	0.009	-0.025**	0.003*	-0.002	-0.005***	-0.069***
	(0.015)	(0.050)	(0.010)	(0.012)	(0.002)	(0.010)	(0.002)	(0.025)
Services	0.004	0.003	-0.022***	-0.049***	0.001	0.033**	-0.002	0.019***
	(0.012)	(0.041)	(0.008)	(0.018)	(0.001)	(0.015)	(0.002)	(0.006)
Franche-Comté (Base region Limousin)	-0.000	0.001	0.000	-0.000	-0.000	-0.001	-0.001*	-0.004***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)
Auvergne	0.000	-0.001	0.001**	-0.001*	0.000	-0.001	0.002*	-0.007***
	(0.001)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)
Champagne-Ardenne	0.001	-0.001	-0.000	-0.001	0.000	0.001	0.004***	-0.004**
	(0.002)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)
Basse-Normandie	0.001	-0.003	-0.001**	-0.003**	-0.000	-0.002	-0.005***	-0.007***
	(0.002)	(0.003)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.002)
Bourgogne	0.000	-0.001	0.001	-0.000	0.000	0.001	0.001	-0.009***
	(0.001)	(0.004)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.003)
Poitou-Charentes	0.000	-0.002	-0.002***	-0.001	-0.000	-0.003**	0.000	-0.007***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Alsace	-0.000	-0.004	-0.005***	-0.002	-0.001*	-0.005*	0.003*	-0.006**
	(0.001)	(0.012)	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.003)
Haute-Normandie	0.001	-0.001	0.002***	0.001	0.000	-0.003**	-0.009***	-0.008*
	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)
Picardie	0.002	-0.001	0.002***	-0.001	0.000	-0.001	-0.002	-0.016***
	(0.003)	(0.003)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.004)
Languedoc-	0.001	0.002	-0.000	0.001	0.000	-0.001	0.002**	-0.001**

Roussillon								
	(0.002)	(0.006)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Lorraine	0.004	-0.003	0.003**	0.001	0.001*	-0.003	0.010***	-0.011***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Centre	0.003	-0.002	0.003**	-0.001	0.001	-0.003*	-0.004**	-0.017***
	(0.005)	(0.004)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.005)
Midi-Pyrénées	0.001	-0.002	-0.000	0.000	-0.000	-0.000	-0.007***	-0.005
	(0.003)	(0.007)	(0.000)	(0.002)	(0.000)	(0.002)	(0.002)	(0.003)
Bretagne	0.001	-0.002	-0.003**	-0.005**	-0.001*	-0.008***	-0.006***	-0.012***
	(0.002)	(0.005)	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.003)
Aquitaine	0.000	0.000	0.000	0.002	-0.001*	-0.004	0.008***	-0.001
	(0.003)	(0.007)	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)
Pays de la Loire	0.001	-0.000	0.003***	-0.003**	-0.000	0.001	-0.003*	-0.017***
	(0.004)	(0.005)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.004)
Nord-Pas-de-Calais	0.002	-0.004	0.001*	-0.004	0.001*	-0.008***	0.002	-0.018***
	(0.004)	(0.008)	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.005)
Provence-Alpes-Côte d'Azur	0.001	-0.002	-0.003***	-0.003	0.001*	-0.001	0.010***	-0.002**
	(0.002)	(0.009)	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)
Rhône-Alpes	0.006	-0.004	0.003***	-0.003	0.000	-0.002	0.004*	-0.010**
	(0.012)	(0.017)	(0.001)	(0.004)	(0.001)	(0.003)	(0.002)	(0.005)
Île-de-France	-0.051	-0.101	-0.015***	-0.031***	-0.003	-0.026***	-0.006**	-0.020***
	(0.050)	(0.164)	(0.003)	(0.011)	(0.003)	(0.007)	(0.003)	(0.006)
Constant		0.176		0.124		0.336***		0.517***
		(0.336)		(0.094)		(0.088)		(0.107)
Observations	3,904	3,904	9,221	9,221	5,433	5,433	11,338	11,338

*Prof1= Management and High Intellectual professionals, prof2=High Skilled White Collar, prof3=Low Skilled White Collar, prof4=Blue Collar*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

D-7 Detail threefold decomposition using basic hourly wage

	Large establishment			Small Establishment		
	Endowments	Coefficients	Endowments	Coefficients	Endowments	Coefficients
Experience	0.033***	0.107	0.007	0.010***	0.123**	0.003*
	(0.005)	(0.077)	(0.005)	(0.002)	(0.062)	(0.002)
Exp. squared	-0.039***	-0.065	-0.007	-0.010***	-0.099	-0.004
	(0.008)	(0.091)	(0.009)	(0.003)	(0.072)	(0.003)
Exp. cube	0.012***	0.012	0.002	0.002**	0.030	0.002
	(0.004)	(0.037)	(0.005)	(0.001)	(0.029)	(0.001)
Tenure	0.008***	-0.025*	-0.004*	0.002***	-0.014*	-0.000
	(0.002)	(0.015)	(0.002)	(0.001)	(0.008)	(0.000)
Tenure. squared	0.005***	0.009	0.003	-0.000	0.004	0.000
	(0.002)	(0.008)	(0.002)	(0.000)	(0.004)	(0.000)
Tenure cube	0.000	-0.005***	0.001***	-0.000	-0.005***	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Single	-0.001**	0.014***	0.002***	-0.000	0.015***	0.001***
	(0.000)	(0.003)	(0.000)	(0.000)	(0.002)	(0.000)
Married	-0.000*	0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Other family status (divorced, widowed etc)	0.000	-0.003***	0.000	-0.004***	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
No degree	0.005***	0.005***	-0.001***	0.002***	0.002	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Before Bac without degree	-0.009***	0.003*	0.001*	-0.006***	0.006***	0.002***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
CAP/BEP	-0.000	-0.000	0.000	0.000	0.001	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Bac professional and technical	-0.000	0.001*	-0.000*	-0.000**	-0.001	0.000

	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Bac general	-0.007***	0.001	-0.000	-0.005***	-0.002**	0.001**
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
Bac +2	0.009***	-0.001**	-0.001**	0.004***	-0.001*	-0.000*
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Bac+3 and plus	0.027***	0.005***	0.004***	0.030***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Management and High Intellectual professionals	-0.000	-0.010***	0.000	-0.001***	-0.001	0.000
	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)
High Skilled White Collar	0.042***	-0.033***	0.027***	0.071***	-0.023***	0.018***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Low Skilled White Collar	-0.065***	0.013***	0.014***	-0.097***	0.003***	0.006***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Blue collar	0.002***	0.002***	-0.001***	0.001***	0.003***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Trade	0.013***	-0.023***	-0.014***	0.005***	-0.001*	-0.001*
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Manufacturing	-0.000	0.013***	-0.005***	-0.001***	-0.006***	0.002***
	(0.001)	(0.003)	(0.001)	(0.000)	(0.002)	(0.001)
Services	-0.000**	0.001***	0.000*	-0.000	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Base region Limousin	0.000	0.000	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Franche-Comté	0.000	-0.001*	-0.000*	0.000**	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Auvergne	-0.000***	0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Champagne-Ardenne	-0.000*	-0.002***	0.001***	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Basse-Normandie	-0.001***	0.001***	0.001***	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Bourgogne	-0.000**	-0.002***	0.000**	-0.000	-0.001***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Poitou-Charentes	-0.000***	0.002***	-0.001***	0.000	0.003***	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Alsace	-0.000	0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Haute-Normandie	0.000*	-0.000	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Picardie	0.001***	0.001	-0.000	0.000	-0.001*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Languedoc-Roussillon	0.001***	-0.002***	-0.001***	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lorraine	-0.000	0.001**	0.000**	-0.000	0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Centre	-0.000	0.001**	-0.000**	0.000	0.001*	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Midi-Pyrénées	0.000**	-0.004***	0.002***	-0.000**	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Bretagne	0.000	0.003***	-0.000	0.000***	0.001	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Aquitaine	-0.000*	0.001**	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Pays de la Loire	-0.000	-0.002***	-0.000***	-0.000**	-0.001**	-0.000**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Nord-Pas-de-Calais	-0.000	-0.000	0.000	-0.000	0.002**	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Provence-Alpes-Côte d'Azur	0.001***	0.002***	0.001**	0.000	0.001*	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Rhône-Alpes	-0.007***	-0.010***	0.002***	-0.003***	-0.004***	0.000***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Île-de-France	0.003***	0.133***	0.004***	0.000***	0.011	0.000

	(0.000)	(0.015)	(0.001)	(0.000)	(0.008)	(0.000)
Constant		-0.054**			0.036*	
		(0.026)			(0.021)	
Observations	29,851	29,851	29,851	44,536	44,536	44,536

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Appendix - E Chapter6

Table-6.1 Results by Estimation method (2005-2006)

	GEG	GRG	RG	GEG	GRG	RG	GEG	GRG	RG	GEG	GRG	RG
Control variable		FE			RE			BE			FD	
50-199 employees (base 1-49)	1.292***	0.500***	0.085	1.925***	0.916***	0.627***	1.314**	0.460*	-0.105	1.203**	0.351*	-0.186
	(0.28)	(0.19)	(0.20)	(0.30)	(0.15)	(0.18)	(0.60)	(0.24)	(0.40)	(0.49)	(0.21)	(0.29)
200 & more employees	0.981***	0.445***	0.707***	0.243	0.857***	1.013***	-0.203	0.671***	0.497	0.87	0.321*	0.531**
	(0.34)	(0.15)	(0.17)	(0.18)	(0.12)	(0.14)	(0.25)	(0.17)	(0.30)	(0.54)	(0.16)	(0.23)
Gender			0.841***			0.835***			0.993***			0.756***
			(0.20)			(0.16)			(0.29)			(0.23)
Tenure	0.022**	0.025***	0.031***	0.009***	0.024***	0.019***	0.008***	0.015***	0.008	0.023**	0.028***	0.033***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Status	0.935***	1.097***	1.031***	0.813***	1.319***	1.159***	1.345***	2.575***	2.472***	0.979**	1.362***	1.282***
	(0.28)	(0.22)	(0.18)	(0.23)	(0.16)	(0.18)	(0.43)	(0.32)	(0.51)	(0.40)	(0.29)	(0.23)
Upper Secondary general (base Primary education/lower secondary)		0.837***	0.34		1.207***	0.759***		0.968***	0.838**		0.701**	0.361
		(0.23)	(0.27)		(0.18)	(0.22)		(0.24)	(0.33)		(0.27)	(0.28)
Upper Secondary technical		1.086***	0.577		1.245***	1.038***		0.795***	1.364**		0.864**	0.488
		(0.29)	(0.35)		(0.23)	(0.30)		(0.29)	(0.55)		(0.36)	(0.36)
University level 1st and 2nd stage		1.081***	1.260***		1.085***	1.227***		0.418	0.191		0.964**	1.260***
		(0.33)	(0.28)		(0.23)	(0.35)		(0.29)	(0.60)		(0.37)	(0.32)
University level 3rd stage /Doctoral studies		1.708***	0.860***		1.218***	0.560**		0.171	-0.096		1.630***	0.842***
		(0.24)	(0.23)		(0.20)	(0.23)		(0.27)	(0.39)		(0.28)	(0.25)

Bassin Parisien (base ile de France)	1.894***			1.877***			0.98			1.863**		
	(0.61)			(0.45)			(0.68)			(0.74)		
NORD – PAS-DE-CALAIS	0.8			2.742***			3.962***			0.825		
	(0.80)			(0.71)			(0.93)			(0.90)		
EST	0.726			1.766***			1.929**			0.795		
	(0.93)			(0.62)			(0.93)			(1.07)		
OUEST	1.892***			1.906***			1.504*			1.901**		
	(0.66)			(0.52)			(0.85)			(0.73)		
SUD-OUEST	0.929*			2.920***			4.204***			0.857		
	(0.55)			(0.57)			(0.94)			(0.70)		
CENTRE-EST	0.755			3.169***			4.719***			0.813		
	(0.80)			(0.62)			(1.18)			(0.94)		
MÉDITERRANÉE	1.399*			1.879***			1.598			1.429		
	(0.82)			(0.68)			(0.98)			(0.91)		
Trade (base Manufacturing)	1.618**	0.483	0.326	2.311***	0.856***	0.464	0.857	1.725***	0.899	1.669**	0.742**	0.531
	(0.70)	(0.31)	(0.29)	(0.63)	(0.29)	(0.31)	(0.85)	(0.44)	(0.70)	(0.83)	(0.32)	(0.35)
Services	0.862**	0.232	0.133	1.164***	0.147	0.345***	0.963***	-0.068	0.157	0.885**	0.285*	0.198
	(0.37)	(0.17)	(0.19)	(0.18)	(0.10)	(0.12)	(0.26)	(0.11)	(0.15)	(0.41)	(0.16)	(0.23)
(Base Category Management and High Intellectual professionals) High Skilled White Collar	-1.089***	-1.035***	-1.009***	-2.175***	-1.712***	-1.510***	-2.776***	-3.305***	-3.502***	-1.129***	-1.344***	-1.359***
	(0.36)	(0.25)	(0.27)	(0.19)	(0.22)	(0.27)	(0.38)	(0.38)	(0.58)	(0.42)	(0.30)	(0.34)
Low Skilled White Collar	-2.824***	-1.385***	-1.102**	-2.875***	-2.407***	-2.124***	-3.405***	-4.313***	-3.757***	-2.888***	-1.923***	-1.647***

	(0.39)	(0.31)	(0.41)	(0.19)	(0.23)	(0.29)	(0.39)	(0.42)	(0.85)	(0.51)	(0.47)	(0.49)
Blue collar	-1.549***	-0.650*	-1.130***	-2.158***	-1.814***	-1.960***	-2.827***	-4.071***	-4.318***	-1.602***	-1.132**	-1.494***
	(0.43)	(0.38)	(0.36)	(0.23)	(0.28)	(0.30)	(0.46)	(0.48)	(0.72)	(0.56)	(0.48)	(0.43)
CDD(base CDI)	-0.335	-0.837***	-0.424	-1.201***	-0.398	-0.840**	-0.62	0.056	-0.441	-0.334	-0.590**	-0.091
	(0.90)	(0.27)	(0.27)	(0.37)	(0.26)	(0.33)	(0.44)	(0.41)	(0.81)	(1.02)	(0.29)	(0.35)
Other contracts	-3.137**	-0.985*	0.637	0.165	-0.78	-0.284	-1.042	-1.572**	-1.672	-3.021*	-1.200**	0.272
	(1.34)	(0.51)	(0.91)	(0.86)	(0.49)	(0.57)	(0.97)	(0.75)	(1.21)	(1.63)	(0.56)	(1.00)
Constant	-1.272	-0.177	-0.522	-0.706	-0.423	-1.315	-3.446	2.629*	0.405	-0.195	-0.605*	-0.764**
	(2.42)	(0.66)	(0.74)	(1.89)	(0.73)	(0.94)	(2.72)	(1.39)	(2.67)	(0.80)	(0.31)	(0.33)
Observations	100	160	80	100	160	80	100	160	80	50	80	40
R-squared	0.99	0.99	1				1	1	1	0.99	0.99	1
Number of id	50	80	40	50	80	40	50	80	40			
Adj. R-squared	0.99	0.99	1				0.99	0.99	1	0.98	0.99	1

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.The dependent variable is log of gross hourly wage.

Table-6.2 Results by choice of cohort (2005-2006)

	GEG				GRG				RG			
Control variable	FE	RE	BE	FD	FE	RE	BE	FD	FE	RE	BE	FD
50-199 employees (base 1-49)	1.292***	1.925***	1.314**	1.203**	0.500***	0.916***	0.460*	0.351*	0.085	0.627***	-0.105	-0.186
	(0.28)	(0.30)	(0.60)	(0.49)	(0.19)	(0.15)	(0.24)	(0.21)	(0.20)	(0.18)	(0.40)	(0.29)
200 & more employees	0.981***	0.243	-0.203	0.87	0.445***	0.857***	0.671***	0.321*	0.707***	1.013***	0.497	0.531**
	(0.34)	(0.18)	(0.25)	(0.54)	(0.15)	(0.12)	(0.17)	(0.16)	(0.17)	(0.14)	(0.30)	(0.23)
Gender									0.841***	0.835***	0.993***	0.756***
									(0.20)	(0.16)	(0.29)	(0.23)
Tenure	0.022**	0.009***	0.008***	0.023**	0.025***	0.024***	0.015***	0.028***	0.031***	0.019***	0.008	0.033***
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)
Status	0.935***	0.813***	1.345***	0.979**	1.097***	1.319***	2.575***	1.362***	1.031***	1.159***	2.472***	1.282***
	(0.28)	(0.23)	(0.43)	(0.40)	(0.22)	(0.16)	(0.32)	(0.29)	(0.18)	(0.18)	(0.51)	(0.23)
Upper Secondary general (base Primary education/lower secondary)					0.837***	1.207***	0.968***	0.701**	0.34	0.759***	0.838**	0.361
					(0.23)	(0.18)	(0.24)	(0.27)	(0.27)	(0.22)	(0.33)	(0.28)
Upper Secondary technical					1.086***	1.245***	0.795***	0.864**	0.577	1.038***	1.364**	0.488
					(0.29)	(0.23)	(0.29)	(0.36)	(0.35)	(0.30)	(0.55)	(0.36)
University level 1st and 2nd stage					1.081***	1.085***	0.418	0.964**	1.260***	1.227***	0.191	1.260***
					(0.33)	(0.23)	(0.29)	(0.37)	(0.28)	(0.35)	(0.60)	(0.32)
University level 3rd stage /Doctoral studies					1.708***	1.218***	0.171	1.630***	0.860***	0.560**	-0.096	0.842***
					(0.24)	(0.20)	(0.27)	(0.28)	(0.23)	(0.23)	(0.39)	(0.25)
Bassin Parisien (base ile de	1.894***	1.877***	0.98	1.863**								

France)												
	(0.61)	(0.45)	(0.68)	(0.74)								
NORD – PAS-DE-CALAIS	0.8	2.742***	3.962***	0.825								
	(0.80)	(0.71)	(0.93)	(0.90)								
EST	0.726	1.766***	1.929**	0.795								
	(0.93)	(0.62)	(0.93)	(1.07)								
OUEST	1.892***	1.906***	1.504*	1.901**								
	(0.66)	(0.52)	(0.85)	(0.73)								
SUD-OUEST	0.929*	2.920***	4.204***	0.857								
	(0.55)	(0.57)	(0.94)	(0.70)								
CENTRE-EST	0.755	3.169***	4.719***	0.813								
	(0.80)	(0.62)	(1.18)	(0.94)								
MÉDITERRANÉE	1.399*	1.879***	1.598	1.429								
	(0.82)	(0.68)	(0.98)	(0.91)								
Trade (base Manufacturing)	1.618**	2.311***	0.857	1.669**	0.483	0.856***	1.725***	0.742**	0.326	0.464	0.899	0.531
	(0.70)	(0.63)	(0.85)	(0.83)	(0.31)	(0.29)	(0.44)	(0.32)	(0.29)	(0.31)	(0.70)	(0.35)
Services	0.862**	1.164***	0.963***	0.885**	0.232	0.147	-0.068	0.285*	0.133	0.345***	0.157	0.198
	(0.37)	(0.18)	(0.26)	(0.41)	(0.17)	(0.10)	(0.11)	(0.16)	(0.19)	(0.12)	(0.15)	(0.23)
(Base Category Management and High Intellectual professionals) High Skilled White Collar	- 1.089***	- 2.175***	- 2.776***	- 1.129***	- 1.035***	- 1.712***	- 3.305***	- 1.344***	- 1.009***	- 1.510***	- 3.502***	- 1.359***
	(0.36)	(0.19)	(0.38)	(0.42)	(0.25)	(0.22)	(0.38)	(0.30)	(0.27)	(0.27)	(0.58)	(0.34)
Low Skilled White Collar	- 2.824***	- 2.875***	- 3.405***	- 2.888***	- 1.385***	- 2.407***	- 4.313***	- 1.923***	-1.102**	- 2.124***	- 3.757***	- 1.647***
	(0.39)	(0.19)	(0.39)	(0.51)	(0.31)	(0.23)	(0.42)	(0.47)	(0.41)	(0.29)	(0.85)	(0.49)
Blue collar	-	-	-	-	-0.650*	-	-	-1.132**	-	-	-	-

	1.549***	2.158***	2.827***	1.602***		1.814***	4.071***		1.130***	1.960***	4.318***	1.494***
	(0.43)	(0.23)	(0.46)	(0.56)	(0.38)	(0.28)	(0.48)	(0.48)	(0.36)	(0.30)	(0.72)	(0.43)
CDD(base CDI)	-0.335	-1.201***	-0.62	-0.334	-0.837***	-0.398	0.056	-0.590**	-0.424	-0.840**	-0.441	-0.091
	(0.90)	(0.37)	(0.44)	(1.02)	(0.27)	(0.26)	(0.41)	(0.29)	(0.27)	(0.33)	(0.81)	(0.35)
Other contracts	-3.137**	0.165	-1.042	-3.021*	-0.985*	-0.78	-1.572**	-1.200**	0.637	-0.284	-1.672	0.272
	(1.34)	(0.86)	(0.97)	(1.63)	(0.51)	(0.49)	(0.75)	(0.56)	(0.91)	(0.57)	(1.21)	(1.00)
Constant	-1.272	-0.706	-3.446	-0.195	-0.177	-0.423	2.629*	-0.605*	-0.522	-1.315	0.405	-0.764**
	(2.42)	(1.89)	(2.72)	(0.80)	(0.66)	(0.73)	(1.39)	(0.31)	(0.74)	(0.94)	(2.67)	(0.33)
Observations	100	100	100	50	160	160	160	80	80	80	80	40
R-squared	0.99		1	0.99	0.99		1	0.99	1		1	1
Number of id	50	50	50		80	80	80		40	40	40	
Adj. R-squared	0.99		0.99	0.98	0.99		0.99	0.99	1		1	1

Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.The dependent variable is log of gross hourly wage.

Table-6.3 Results by Estimation method for three years (2002, 2005-2006)

	GEG	GRG	RG	GEG	GRG	RG	GEG	GRG	RG	GEG	GRG	RG
control variable	FE			RE			BE			FD		
50-199 employees (base 1-49)	3.531***	0.579***	0.366*	3.803***	0.813***	0.615***	2.522***	1.108***	0.636	1.544***	0.25	-0.194
	(0.21)	(0.14)	(0.20)	(0.27)	(0.13)	(0.17)	(0.47)	(0.40)	(0.58)	(0.54)	(0.19)	(0.32)
200 & more employees	1.041***	0.322***	0.318***	0.844***	0.514***	0.453***	0.438	1.642***	1.686***	1.342***	0.229	0.328
	(0.26)	(0.10)	(0.10)	(0.21)	(0.09)	(0.11)	(0.34)	(0.27)	(0.41)	(0.49)	(0.16)	(0.22)
Gender			0.357**			0.512***			1.097**			0.634***
			(0.15)			(0.17)			(0.51)			(0.21)
Tenure	0.026***	0.041***	0.043***	0.007***	0.039***	0.034***	0.010***	0.022***	0.015*	0.016	0.033***	0.032***
	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Status	0.069	0.287***	0.276***	0.057	0.270***	0.255***	0.251**	0.376**	0.269	0.931***	1.232***	1.361***
	(0.04)	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)	(0.10)	(0.17)	(0.26)	(0.33)	(0.35)	(0.29)
Upper Secondary general (base Primary education/lower secondary)		1.318***	0.868***		1.330***	1.090***		1.946***	1.702**		0.865**	0.09
		(0.18)	(0.20)		(0.13)	(0.20)		(0.47)	(0.73)		(0.34)	(0.33)

Upper Secondary technical		1.846***	1.678***		1.792***	1.723***		2.391***	3.446***		0.831**	0.25
		(0.13)	(0.13)		(0.11)	(0.15)		(0.47)	(0.71)		(0.41)	(0.41)
University level 1st and 2nd stage		2.271***	1.989***		2.163***	1.877***		1.068**	1.641**		1.353***	1.065***
		(0.16)	(0.24)		(0.14)	(0.25)		(0.45)	(0.79)		(0.50)	(0.30)
University level 3rd stage /Doctoral studies		2.408***	2.132***		2.231***	1.906***		1.660***	1.199***		1.633***	1.056***
		(0.10)	(0.13)		(0.11)	(0.18)		(0.25)	(0.38)		(0.35)	(0.26)
Bassin Parisien (base ile de France)	0.989**			0.374			0.957			1.494**		
	(0.40)			(0.44)			(1.12)			(0.65)		
NORD – PAS-DE-CALAIS	2.353**			3.933***			5.917***			0.806		
	(0.96)			(0.79)			(1.06)			(0.83)		
EST	0.748			1.605**			2.780**			0.208		
	(0.62)			(0.67)			(1.18)			(1.14)		
OUEST	0.247			0.287			0.611			1.558*		
	(0.53)			(0.62)			(1.36)			(0.80)		
SUD-OUEST	0.575			3.519***			2.184*			1.438*		
	(0.82)			(0.71)			(1.28)			(0.75)		

CENTRE-EST	2.997***			4.086***			3.809***			0.503		
	(0.71)			(0.69)			(1.23)			(0.82)		
MÉDITERRANÉE	-0.274			1.507*			2.135			0.415		
	(0.82)			(0.89)			(1.33)			(0.85)		
Trade (base Manufacturing)	0.905**	0.359*	0.038	1.710***	0.664***	0.428	2.587***	1.436**	0.173	1.166	0.729**	0.592
	(0.35)	(0.18)	(0.23)	(0.32)	(0.19)	(0.27)	(0.66)	(0.59)	(1.13)	(0.82)	(0.32)	(0.46)
Services	1.148***	0.414***	0.506***	1.248***	0.490***	0.590***	1.638***	0.205	0.14	0.566	0.187	0.174
	(0.16)	(0.10)	(0.11)	(0.15)	(0.07)	(0.10)	(0.23)	(0.14)	(0.23)	(0.38)	(0.14)	(0.28)
(Base Category Management and High Intellectual professionals) High Skilled White Collar	-0.750***	-0.662***	-0.535**	-1.255***	-0.698***	-0.588***	-1.627***	-0.842***	-1.437**	-0.801**	-1.149***	-1.178***
	(0.23)	(0.15)	(0.21)	(0.10)	(0.12)	(0.19)	(0.12)	(0.30)	(0.63)	(0.33)	(0.33)	(0.37)
Low Skilled White Collar	-0.889***	-0.791***	-0.473***	-1.831***	-0.950***	-0.659***	-2.450***	-1.290***	-0.253	-1.971***	-1.608***	-1.176**
	(0.18)	(0.13)	(0.14)	(0.11)	(0.08)	(0.18)	(0.16)	(0.19)	(0.71)	(0.55)	(0.58)	(0.56)
Blue collar	-0.642***	0.288***	0.200*	-1.203***	0.133	-0.079	-1.498***	-0.822***	-1.866***	-1.297***	-1.097**	-1.363***
	(0.17)	(0.10)	(0.11)	(0.07)	(0.08)	(0.14)	(0.13)	(0.23)	(0.38)	(0.40)	(0.52)	(0.42)

CDD(base CDI)	-1.787**	-0.934***	-0.732**	-1.715***	-0.819***	-0.888***	-1.782***	-0.041	-1.315	-0.649	-0.736**	-0.555
	(0.78)	(0.25)	(0.28)	(0.42)	(0.21)	(0.28)	(0.51)	(0.66)	(1.12)	(1.18)	(0.32)	(0.49)
Other contracts	-1.399**	-0.727***	-0.918**	-0.279	-0.831***	-1.073***	-0.064	-0.926	-0.529	-2.812**	-1.166**	0.218
	(0.64)	(0.26)	(0.41)	(0.47)	(0.23)	(0.31)	(0.59)	(0.72)	(1.05)	(1.31)	(0.54)	(0.92)
Constant	3.593*	0.674	0.786	2.021	0.472	0.612	2.781	0.44	-0.99	0.614	-0.683***	-0.642
	(2.09)	(0.55)	(0.88)	(2.16)	(0.62)	(1.06)	(3.59)	(1.65)	(3.25)	(0.78)	(0.24)	(0.43)
Observations	150	240	120	150	240	120	150	240	120	50	80	40
R-squared	0.97	0.99	1				0.99	0.99	0.99	0.99	1	1
Number of id	50	80	40	50	80	40	50	80	40			

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.The dependent variable is log of gross hourly wage.

Table-6.4 Results by choice of cohorts for three years (2002, 2005-2006)

	GEG				GRG				RG			
control variable	FE	RE	BE	FD	FE	RE	BE	FD	FE	RE	BE	FD
50-199 employees (base 1-49)	3.531***	3.803***	2.522***	1.544***	0.579***	0.813***	1.108***	0.25	0.366*	0.615***	0.636	-0.194
	(0.21)	(0.27)	(0.47)	(0.54)	(0.14)	(0.13)	(0.40)	(0.19)	(0.20)	(0.17)	(0.58)	(0.32)
200 & more employees	1.041***	0.844***	0.438	1.342***	0.322***	0.514***	1.642***	0.229	0.318***	0.453***	1.686***	0.328
	(0.26)	(0.21)	(0.34)	(0.49)	(0.10)	(0.09)	(0.27)	(0.16)	(0.10)	(0.11)	(0.41)	(0.22)
Gender									0.357**	0.512***	1.097**	0.634***
									(0.15)	(0.17)	(0.51)	(0.21)
Tenure	0.026***	0.007***	0.010***	0.016	0.041***	0.039***	0.022***	0.033***	0.043***	0.034***	0.015*	0.032***
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Status	0.069	0.057	0.251**	0.931***	0.287***	0.270***	0.376**	1.232***	0.276***	0.255***	0.269	1.361***
	(0.04)	(0.04)	(0.10)	(0.33)	(0.02)	(0.02)	(0.17)	(0.35)	(0.03)	(0.03)	(0.26)	(0.29)
Upper Secondary general (base Primary education/lower secondary)					1.318***	1.330***	1.946***	0.865**	0.868***	1.090***	1.702**	0.09
					(0.18)	(0.13)	(0.47)	(0.34)	(0.20)	(0.20)	(0.73)	(0.33)
Upper Secondary technical					1.846***	1.792***	2.391***	0.831**	1.678***	1.723***	3.446***	0.25
					(0.13)	(0.11)	(0.47)	(0.41)	(0.13)	(0.15)	(0.71)	(0.41)
University level 1st and 2nd stage					2.271***	2.163***	1.068**	1.353***	1.989***	1.877***	1.641**	1.065***
					(0.16)	(0.14)	(0.45)	(0.50)	(0.24)	(0.25)	(0.79)	(0.30)
University level 3rd stage /Doctoral studies					2.408***	2.231***	1.660***	1.633***	2.132***	1.906***	1.199***	1.056***

					(0.10)	(0.11)	(0.25)	(0.35)	(0.13)	(0.18)	(0.38)	(0.26)
Bassin Parisien (base ile de France)	0.989**	0.374	0.957	1.494**								
	(0.40)	(0.44)	(1.12)	(0.65)								
NORD – PAS-DE-CALAIS	2.353**	3.933***	5.917***	0.806								
	(0.96)	(0.79)	(1.06)	(0.83)								
EST	0.748	1.605**	2.780**	0.208								
	(0.62)	(0.67)	(1.18)	(1.14)								
OUEST	0.247	0.287	0.611	1.558*								
	(0.53)	(0.62)	(1.36)	(0.80)								
SUD-OUEST	0.575	3.519***	2.184*	1.438*								
	(0.82)	(0.71)	(1.28)	(0.75)								
CENTRE-EST	2.997***	4.086***	3.809***	0.503								
	(0.71)	(0.69)	(1.23)	(0.82)								
MÉDITERRANÉE	-0.274	1.507*	2.135	0.415								
	(0.82)	(0.89)	(1.33)	(0.85)								
Trade (base Manufacturing)	0.905**	1.710***	2.587***	1.166	0.359*	0.664***	1.436**	0.729**	0.038	0.428	0.173	0.592
	(0.35)	(0.32)	(0.66)	(0.82)	(0.18)	(0.19)	(0.59)	(0.32)	(0.23)	(0.27)	(1.13)	(0.46)
Services	1.148***	1.248***	1.638***	0.566	0.414***	0.490***	0.205	0.187	0.506***	0.590***	0.14	0.174
	(0.16)	(0.15)	(0.23)	(0.38)	(0.10)	(0.07)	(0.14)	(0.14)	(0.11)	(0.10)	(0.23)	(0.28)
(Base Category Management and High Intellectual professionals) High Skilled White Collar	-0.750***	-1.255***	-1.627***	-0.801**	-0.662***	-0.698***	-0.842***	-1.149***	-0.535**	-0.588***	-1.437**	-1.178***

	(0.23)	(0.10)	(0.12)	(0.33)	(0.15)	(0.12)	(0.30)	(0.33)	(0.21)	(0.19)	(0.63)	(0.37)
Low Skilled White Collar	-0.889***	-1.831***	-2.450***	-1.971***	-0.791***	-0.950***	-1.290***	-1.608***	-0.473***	-0.659***	-0.253	-1.176**
	(0.18)	(0.11)	(0.16)	(0.55)	(0.13)	(0.08)	(0.19)	(0.58)	(0.14)	(0.18)	(0.71)	(0.56)
Blue collar	-0.642***	-1.203***	-1.498***	-1.297***	0.288***	0.133	-0.822***	-1.097**	0.200*	-0.079	-1.866***	-1.363***
	(0.17)	(0.07)	(0.13)	(0.40)	(0.10)	(0.08)	(0.23)	(0.52)	(0.11)	(0.14)	(0.38)	(0.42)
CDD(base CDI)	-1.787**	-1.715***	-1.782***	-0.649	-0.934***	-0.819***	-0.041	-0.736**	-0.732**	-0.888***	-1.315	-0.555
	(0.78)	(0.42)	(0.51)	(1.18)	(0.25)	(0.21)	(0.66)	(0.32)	(0.28)	(0.28)	(1.12)	(0.49)
Other contracts	-1.399**	-0.279	-0.064	-2.812**	-0.727***	-0.831***	-0.926	-1.166**	-0.918**	-1.073***	-0.529	0.218
	(0.64)	(0.47)	(0.59)	(1.31)	(0.26)	(0.23)	(0.72)	(0.54)	(0.41)	(0.31)	(1.05)	(0.92)
Constant	3.593*	2.021	2.781	0.614	0.674	0.472	0.44	-0.683***	0.786	0.612	-0.99	-0.642
	(2.09)	(2.16)	(3.59)	(0.78)	(0.55)	(0.62)	(1.65)	(0.24)	(0.88)	(1.06)	(3.25)	(0.43)
Observations	150	150	150	50	240	240	240	80	120	120	120	40
R-squared	0.97		0.99	0.99	0.99		0.99	1	1		0.99	1
Number of id	50	50	50		80	80	80		40	40	40	
Adj. R-squared	0.96		0.99	0.98	0.99		0.99	0.99	1		0.99	1

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Size as continuous variable- case-1

Table-6.5 Results by Estimation method (2005-2006) (proportion for size category)

	GEG	GRG	RG	GEG	GRG	RG	GEG	GRG	RG	GEG	GRG	RG
control variable	FE			RE			BE			FD		
Log of Size	0.470***	0.370***	0.311***	0.487***	0.553***	0.470***	0.467	0.947***	0.978***	0.894***	0.671***	0.483***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.05)	(0.32)	(0.08)	(0.14)	(0.05)	(0.07)	(0.09)
Gender			0.484***			0.225			-0.157			0.403**
			(0.13)			(0.15)			(0.25)			(0.16)
Tenure	0.026***	0.028***	0.034***	0.013***	0.013***	0.016***	0.009***	0.005**	0.002	0.016***	0.017***	0.027***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
Status	1.072***	1.078***	0.976***	0.814***	1.352***	1.423***	0.853	0.339	0.475	0.293**	0.608***	0.860***
	(0.14)	(0.16)	(0.15)	(0.15)	(0.12)	(0.16)	(0.88)	(0.31)	(0.47)	(0.12)	(0.17)	(0.16)
Upper Secondary general (base Primary education/lower secondary)		0.298	-0.068		0.436***	0.420**		0.392**	0.356		-0.003	-0.211
		(0.20)	(0.26)		(0.14)	(0.21)		(0.16)	(0.22)		(0.15)	(0.28)
Upper Secondary technical		0.07	0.289		0.256	0.414		0.261	0.417		-0.103	0.106
		(0.26)	(0.27)		(0.18)	(0.28)		(0.19)	(0.39)		(0.20)	(0.31)
University level 1st and 2nd stage		0.412	0.955***		0.254	0.763**		-0.018	-0.349		-0.101	0.666*
		(0.27)	(0.30)		(0.19)	(0.35)		(0.20)	(0.43)		(0.20)	(0.35)

University level 3rd stage /Doctoral studies		0.808***	0.887***		0.137	0.135		-0.369**	-0.564*		0.330**	0.518*
		(0.19)	(0.19)		(0.16)	(0.23)		(0.18)	(0.29)		(0.16)	(0.26)
Bassin Parisien (base ile de France)	0.2			0.567*			-0.418			-0.09		
	(0.37)			(0.33)			(0.80)			(0.23)		
NORD – PAS-DE-CALAIS	0.227			1.219**			2.701**			-0.112		
	(0.55)			(0.53)			(1.16)			(0.31)		
EST	0.46			0.965**			1.114			-0.268		
	(0.44)			(0.42)			(1.06)			(0.32)		
OUEST	0.217			1.126***			0.924			-0.193		
	(0.48)			(0.39)			(0.92)			(0.29)		
SUD-OUEST	-0.335			1.084***			2.451**			-0.413*		
	(0.38)			(0.40)			(1.14)			(0.22)		
CENTRE-EST	0.575			1.346***			3.218**			0.047		
	(0.44)			(0.41)			(1.41)			(0.24)		
MÉDITERRANÉE	-0.22			0.625			1.678			-0.205		
	(0.52)			(0.51)			(1.05)			(0.39)		
Trade (base Manufacturing)	1.189**	0.692***	0.456**	1.784***	0.669***	0.352	0.544	0.760***	0.987*	0.19	0.157	0.108
	(0.49)	(0.23)	(0.20)	(0.44)	(0.21)	(0.26)	(1.00)	(0.28)	(0.49)	(0.35)	(0.23)	(0.26)
Services	-0.041	0.254**	0.215	0.314**	-0.017	0.067	0.271	-0.042	-0.08	-0.434***	0.153	0.104

	(0.21)	(0.12)	(0.13)	(0.14)	(0.07)	(0.10)	(0.26)	(0.07)	(0.10)	(0.13)	(0.11)	(0.16)
(Base Category Management and High Intellectual professionals) High Skilled White Collar	-1.312***	-1.273***	-1.466***	-1.854***	-1.955***	-2.271***	-2.094**	-1.092***	-1.091*	-0.667***	-0.972***	-1.472***
	(0.18)	(0.17)	(0.25)	(0.13)	(0.16)	(0.24)	(0.92)	(0.34)	(0.56)	(0.17)	(0.18)	(0.27)
Low Skilled White Collar	-2.086***	-1.923***	-1.416***	-2.422***	-2.936***	-2.813***	-2.510**	-2.105***	-2.417***	-1.228***	-1.491***	-1.345***
	(0.24)	(0.25)	(0.31)	(0.13)	(0.17)	(0.32)	(1.02)	(0.36)	(0.64)	(0.18)	(0.22)	(0.37)
Blue collar	-1.695***	-1.449***	-1.575***	-1.999***	-2.638***	-2.757***	-2.119**	-1.942***	-2.190***	-1.025***	-1.082***	-1.579***
	(0.30)	(0.24)	(0.29)	(0.15)	(0.21)	(0.29)	(0.99)	(0.39)	(0.63)	(0.16)	(0.20)	(0.28)
CDD(base CDI)	0.634	-0.34	-0.077	-0.197	-0.148	-0.338	-0.002	-0.238	-0.136	0.52	-0.530***	-0.275
	(0.63)	(0.23)	(0.22)	(0.31)	(0.18)	(0.30)	(0.43)	(0.24)	(0.56)	(0.36)	(0.17)	(0.26)
Other contracts	-0.995	-0.886*	0.219	-1.370*	-1.312***	-1.112**	-1.798*	-1.123**	-1.581*	-0.23	-0.499	0.425
	(0.99)	(0.48)	(0.60)	(0.72)	(0.35)	(0.47)	(0.93)	(0.46)	(0.83)	(0.56)	(0.40)	(0.64)
Constant	1.825	-0.651	-2.288**	3.338**	-0.421	-1.920**	-0.703	0.332	0.633	2.805***	1.483***	1.007**
	(1.42)	(0.66)	(0.85)	(1.43)	(0.51)	(0.92)	(2.60)	(0.73)	(1.49)	(0.33)	(0.28)	(0.45)
Observations	100	160	80	100	160	80	100	160	80	50	80	40
R-squared	0.99	1	1				0.99	1	1	1	1	1
Number of id	50	80	40	50	80	40	50	80	40			

Adj. R-squared	0.99	1	1				0.99	1	1	1	1	1

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.The dependent variable is log of gross hourly wage.

Table-6.6 Results by choice of cohorts (2005-2006) (proportion for size category)

	GEG				GRG				RG			
control variable	FE	RE	BE	FD	FE	RE	BE	FD	FE	RE	BE	FD
Log of Size	0.470***	0.487***	0.467	0.894***	0.370***	0.553***	0.947***	0.671***	0.311***	0.470***	0.978***	0.483***
	(0.04)	(0.04)	(0.32)	(0.05)	(0.04)	(0.03)	(0.08)	(0.07)	(0.05)	(0.05)	(0.14)	(0.09)
Gender									0.484***	0.225	-0.157	0.403**
									(0.13)	(0.15)	(0.25)	(0.16)
Tenure	0.026***	0.013***	0.009***	0.016***	0.028***	0.013***	0.005**	0.017***	0.034***	0.016***	0.002	0.027***
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Status	1.072***	0.814***	0.853	0.293**	1.078***	1.352***	0.339	0.608***	0.976***	1.423***	0.475	0.860***
	(0.14)	(0.15)	(0.88)	(0.12)	(0.16)	(0.12)	(0.31)	(0.17)	(0.15)	(0.16)	(0.47)	(0.16)
Upper Secondary general (base Primary education/lower secondary)					0.298	0.436***	0.392**	-0.003	-0.068	0.420**	0.356	-0.211
					(0.20)	(0.14)	(0.16)	(0.15)	(0.26)	(0.21)	(0.22)	(0.28)
Upper Secondary technical					0.07	0.256	0.261	-0.103	0.289	0.414	0.417	0.106
					(0.26)	(0.18)	(0.19)	(0.20)	(0.27)	(0.28)	(0.39)	(0.31)
University level 1st and 2nd stage					0.412	0.254	-0.018	-0.101	0.955***	0.763**	-0.349	0.666*

					(0.27)	(0.19)	(0.20)	(0.20)	(0.30)	(0.35)	(0.43)	(0.35)
University level 3rd stage /Doctoral studies					0.808***	0.137	-0.369**	0.330**	0.887***	0.135	-0.564*	0.518*
					(0.19)	(0.16)	(0.18)	(0.16)	(0.19)	(0.23)	(0.29)	(0.26)
Bassin Parisien (base ile de France)	0.2	0.567*	-0.418	-0.09								
	(0.37)	(0.33)	(0.80)	(0.23)								
NORD – PAS-DE-CALAIS	0.227	1.219**	2.701**	-0.112								
	(0.55)	(0.53)	(1.16)	(0.31)								
EST	0.46	0.965**	1.114	-0.268								
	(0.44)	(0.42)	(1.06)	(0.32)								
OUEST	0.217	1.126***	0.924	-0.193								
	(0.48)	(0.39)	(0.92)	(0.29)								
SUD-OUEST	-0.335	1.084***	2.451**	-0.413*								
	(0.38)	(0.40)	(1.14)	(0.22)								
CENTRE-EST	0.575	1.346***	3.218**	0.047								
	(0.44)	(0.41)	(1.41)	(0.24)								
MÉDITERRANÉE	-0.22	0.625	1.678	-0.205								
	(0.52)	(0.51)	(1.05)	(0.39)								
Trade (base	1.189**	1.784***	0.544	0.19	0.692***	0.669***	0.760***	0.157	0.456**	0.352	0.987*	0.108

Manufacturing)												
	(0.49)	(0.44)	(1.00)	(0.35)	(0.23)	(0.21)	(0.28)	(0.23)	(0.20)	(0.26)	(0.49)	(0.26)
Services	-0.041	0.314**	0.271	-0.434***	0.254**	-0.017	-0.042	0.153	0.215	0.067	-0.08	0.104
	(0.21)	(0.14)	(0.26)	(0.13)	(0.12)	(0.07)	(0.07)	(0.11)	(0.13)	(0.10)	(0.10)	(0.16)
(Base Category Management and High Intellectual professionals) High Skilled White Collar	-1.312***	-1.854***	-2.094**	-0.667***	-1.273***	-1.955***	-1.092***	-0.972***	-1.466***	-2.271***	-1.091*	-1.472***
	(0.18)	(0.13)	(0.92)	(0.17)	(0.17)	(0.16)	(0.34)	(0.18)	(0.25)	(0.24)	(0.56)	(0.27)
Low Skilled White Collar	-2.086***	-2.422***	-2.510**	-1.228***	-1.923***	-2.936***	-2.105***	-1.491***	-1.416***	-2.813***	-2.417***	-1.345***
	(0.24)	(0.13)	(1.02)	(0.18)	(0.25)	(0.17)	(0.36)	(0.22)	(0.31)	(0.32)	(0.64)	(0.37)
Blue collar	-1.695***	-1.999***	-2.119**	-1.025***	-1.449***	-2.638***	-1.942***	-1.082***	-1.575***	-2.757***	-2.190***	-1.579***
	(0.30)	(0.15)	(0.99)	(0.16)	(0.24)	(0.21)	(0.39)	(0.20)	(0.29)	(0.29)	(0.63)	(0.28)
CDD(base CDI)	0.634	-0.197	-0.002	0.52	-0.34	-0.148	-0.238	-0.530***	-0.077	-0.338	-0.136	-0.275
	(0.63)	(0.31)	(0.43)	(0.36)	(0.23)	(0.18)	(0.24)	(0.17)	(0.22)	(0.30)	(0.56)	(0.26)
Other contracts	-0.995	-1.370*	-1.798*	-0.23	-0.886*	-1.312***	-1.123**	-0.499	0.219	-1.112**	-1.581*	0.425
	(0.99)	(0.72)	(0.93)	(0.56)	(0.48)	(0.35)	(0.46)	(0.40)	(0.60)	(0.47)	(0.83)	(0.64)

Constant	1.825	3.338**	-0.703	2.805***	-0.651	-0.421	0.332	1.483***	-2.288**	-1.920**	0.633	1.007**
	(1.42)	(1.43)	(2.60)	(0.33)	(0.66)	(0.51)	(0.73)	(0.28)	(0.85)	(0.92)	(1.49)	(0.45)
Observations	100	100	100	50	160	160	160	80	80	80	80	40
R-squared	0.99		0.99	1	1		1	1	1		1	1
Number of id	50	50	50		80	80	80		40	40	40	
Adj. R-squared	0.99		0.99	1	1		1	1	1		1	1

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Size as continuous variable- case-2 (Average of size)

Table -6.7 Results by Estimation method (2005-2006) (Average of size category)

	GEG	GRG	RG	GEG	GRG	RG	GEG	GRG	RG	GEG	GRG	RG
control variable	FE			RE			BE			FD		
Log of Size	0.202***	0.148***	0.126***	0.181***	0.238***	0.242***	-0.112**	0.201***	0.211***	0.444***	0.127***	0.144**
	(0.05)	(0.03)	(0.04)	(0.04)	(0.02)	(0.03)	(0.05)	(0.03)	(0.06)	(0.07)	(0.05)	(0.06)
Gender			0.562**			0.587***			0.696**			0.580**
			(0.24)			(0.15)			(0.29)			(0.27)
Tenure	0.018*	0.031***	0.018***	0.012***	0.018***	0.011***	0.011***	0.011***	0.009*	0.007	0.032***	0.017**
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Status	1.169***	1.105***	1.113***	1.182***	1.278***	1.317***	2.374***	2.233***	1.571***	0.318	1.185***	1.042***
	(0.25)	(0.22)	(0.18)	(0.23)	(0.14)	(0.17)	(0.32)	(0.29)	(0.52)	(0.36)	(0.26)	(0.25)
Upper Secondary general (base Primary education/lower secondary)		0.245	0.384		0.828***	0.628***		0.860***	0.856**		0.255	0.381
		(0.20)	(0.34)		(0.17)	(0.23)		(0.21)	(0.32)		(0.21)	(0.38)
Upper Secondary technical		0.409	0.143		0.846***	0.676**		0.753***	1.270**		0.373	0.158
		(0.29)	(0.35)		(0.21)	(0.28)		(0.26)	(0.53)		(0.32)	(0.41)
University level 1st and 2nd stage		0.551	0.654		0.673***	0.666*		0.136	0.483		0.608	0.648

		(0.35)	(0.39)		(0.23)	(0.34)		(0.27)	(0.56)		(0.37)	(0.44)
University level 3rd stage /Doctoral studies		1.181***	0.970***		0.692***	0.142		-0.093	-0.003		1.237***	0.921**
		(0.25)	(0.27)		(0.19)	(0.25)		(0.24)	(0.39)		(0.26)	(0.35)
Bassin Parisien (base ile de France)	1.080*			1.312**			1.054			0.998*		
	(0.54)			(0.53)			(0.77)			(0.56)		
NORD – PAS-DE-CALAIS	0.635			1.731**			3.499***			0.167		
	(0.84)			(0.76)			(1.01)			(0.90)		
EST	-0.253			0.847			1.885*			-1.430**		
	(0.80)			(0.64)			(1.02)			(0.66)		
OUEST	0.958*			1.532***			1.433			1.177**		
	(0.56)			(0.58)			(0.85)			(0.58)		
SUD-OUEST	0.065			1.575***			3.619***			0.457		
	(0.56)			(0.59)			(0.99)			(0.48)		
CENTRE-EST	0.518			2.140***			4.435***			-0.438		
	(0.59)			(0.65)			(1.24)			(0.56)		
MÉDITERRANÉE	0.635			1.807**			1.553			-0.096		
	(0.71)			(0.79)			(0.98)			(0.66)		
Trade (base Manufacturing)	1.14	0.448*	0.654**	2.618***	0.873***	0.295	1.107	1.840***	0.893	0.405	0.561**	0.584
	(0.76)	(0.25)	(0.29)	(0.65)	(0.25)	(0.29)	(0.82)	(0.39)	(0.69)	(0.73)	(0.28)	(0.38)

Services	0.262	0.358**	0.365*	0.559***	0.102	0.181*	0.639***	-0.139	0.093	-0.142	0.387**	0.32
	(0.22)	(0.16)	(0.19)	(0.20)	(0.09)	(0.10)	(0.22)	(0.09)	(0.14)	(0.23)	(0.18)	(0.24)
(Base Category Management and High Intellectual professionals) High Skilled White Collar	-0.758**	-1.009***	-0.929***	-2.295***	-1.680***	-1.809***	-3.730***	-2.970***	-2.424***	0.194	-1.104***	-0.862**
	(0.34)	(0.22)	(0.29)	(0.19)	(0.20)	(0.27)	(0.29)	(0.36)	(0.62)	(0.30)	(0.23)	(0.35)
Low Skilled White Collar	-1.827***	-1.483***	-1.542***	-2.962***	-2.649***	-2.526***	-4.415***	-4.065***	-2.560***	-0.272	-1.624***	-1.419*
	(0.40)	(0.37)	(0.55)	(0.20)	(0.21)	(0.30)	(0.32)	(0.37)	(0.83)	(0.47)	(0.40)	(0.71)
Blue collar	-1.670***	-0.821**	-1.288***	-2.355***	-2.035***	-2.515***	-3.923***	-3.890***	-3.184***	-0.586	-0.945**	-1.214***
	(0.34)	(0.36)	(0.35)	(0.23)	(0.26)	(0.30)	(0.36)	(0.44)	(0.75)	(0.39)	(0.40)	(0.42)
CDD(base CDI)	-0.382	-0.775***	-0.618	-0.092	-0.151	-0.648**	0.091	0.248	-0.515	-1.173**	-0.703**	-0.659
	(0.55)	(0.25)	(0.52)	(0.40)	(0.23)	(0.32)	(0.43)	(0.34)	(0.78)	(0.54)	(0.29)	(0.58)
Other contracts	-1.029	-1.112***	-0.444	-1.393	-1.233***	-0.908	-2.123**	-2.069***	-1.979	-1.452	-1.208**	-0.28
	(1.34)	(0.40)	(0.98)	(0.93)	(0.45)	(0.56)	(0.88)	(0.64)	(1.27)	(1.50)	(0.47)	(1.16)
Constant	-1.136	-1.325**	-0.281	-1.657	-0.091	-0.084	-1.169	2.677**	-1.577	2.797***	-0.242	0.252
	(2.60)	(0.65)	(1.27)	(2.06)	(0.69)	(0.94)	(2.55)	(1.13)	(2.54)	(0.61)	(0.30)	(0.49)
Observations	100	160	80	100	160	80	100	160	80	50	80	40
R-squared	0.99	0.99	1				0.99	1	1	0.99	0.99	1

Number of id	50	80	40	50	80	40	50	80	40			
Adj. R-squared	0.99	0.99	1				0.99	1	1	0.99	0.99	1

Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table-6.8 Results by choice of cohorts (2005-2006) (Average of size category)

	GEG				GRG				RG			
control variable	FE	RE	BE	FD	FE	RE	BE	FD	FE	RE	BE	FD
Log of Size	0.202***	0.181***	-0.112**	0.444***	0.148***	0.238***	0.201***	0.127***	0.126***	0.242***	0.211***	0.144**
	(0.05)	(0.04)	(0.05)	(0.07)	(0.03)	(0.02)	(0.03)	(0.05)	(0.04)	(0.03)	(0.06)	(0.06)
Gender									0.562**	0.587***	0.696**	0.580**
									(0.24)	(0.15)	(0.29)	(0.27)
Tenure	0.018*	0.012***	0.011***	0.007	0.031***	0.018***	0.011***	0.032***	0.018***	0.011***	0.009*	0.017**
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Status	1.169***	1.182***	2.374***	0.318	1.105***	1.278***	2.233***	1.185***	1.113***	1.317***	1.571***	1.042***
	(0.25)	(0.23)	(0.32)	(0.36)	(0.22)	(0.14)	(0.29)	(0.26)	(0.18)	(0.17)	(0.52)	(0.25)
Upper Secondary general (base Primary education/lower secondary)					0.245	0.828***	0.860***	0.255	0.384	0.628***	0.856**	0.381
					(0.20)	(0.17)	(0.21)	(0.21)	(0.34)	(0.23)	(0.32)	(0.38)
Upper Secondary technical					0.409	0.846***	0.753***	0.373	0.143	0.676**	1.270**	0.158
					(0.29)	(0.21)	(0.26)	(0.32)	(0.35)	(0.28)	(0.53)	(0.41)
University level 1st and 2nd stage					0.551	0.673***	0.136	0.608	0.654	0.666*	0.483	0.648
					(0.35)	(0.23)	(0.27)	(0.37)	(0.39)	(0.34)	(0.56)	(0.44)

University level 3rd stage /Doctoral studies					1.181***	0.692***	-0.093	1.237***	0.970***	0.142	-0.003	0.921**
					(0.25)	(0.19)	(0.24)	(0.26)	(0.27)	(0.25)	(0.39)	(0.35)
Bassin Parisien (base ile de France)	1.080*	1.312**	1.054	0.998*								
	(0.54)	(0.53)	(0.77)	(0.56)								
NORD – PAS-DE-CALAIS	0.635	1.731**	3.499***	0.167								
	(0.84)	(0.76)	(1.01)	(0.90)								
EST	-0.253	0.847	1.885*	-1.430**								
	(0.80)	(0.64)	(1.02)	(0.66)								
OUEST	0.958*	1.532***	1.433	1.177**								
	(0.56)	(0.58)	(0.85)	(0.58)								
SUD-OUEST	0.065	1.575***	3.619***	0.457								
	(0.56)	(0.59)	(0.99)	(0.48)								
CENTRE-EST	0.518	2.140***	4.435***	-0.438								
	(0.59)	(0.65)	(1.24)	(0.56)								
MÉDITERRANÉE	0.635	1.807**	1.553	-0.096								
	(0.71)	(0.79)	(0.98)	(0.66)								
Trade (base Manufacturing)	1.14	2.618***	1.107	0.405	0.448*	0.873***	1.840***	0.561**	0.654**	0.295	0.893	0.584
	(0.76)	(0.65)	(0.82)	(0.73)	(0.25)	(0.25)	(0.39)	(0.28)	(0.29)	(0.29)	(0.69)	(0.38)

Services	0.262	0.559***	0.639***	-0.142	0.358**	0.102	-0.139	0.387**	0.365*	0.181*	0.093	0.32
	(0.22)	(0.20)	(0.22)	(0.23)	(0.16)	(0.09)	(0.09)	(0.18)	(0.19)	(0.10)	(0.14)	(0.24)
(Base Category Management and High Intellectual professionals) High Skilled White Collar	-0.758**	-2.295***	-3.730***	0.194	-1.009***	-1.680***	-2.970***	-1.104***	-0.929***	-1.809***	-2.424***	-0.862**
	(0.34)	(0.19)	(0.29)	(0.30)	(0.22)	(0.20)	(0.36)	(0.23)	(0.29)	(0.27)	(0.62)	(0.35)
Low Skilled White Collar	-1.827***	-2.962***	-4.415***	-0.272	-1.483***	-2.649***	-4.065***	-1.624***	-1.542***	-2.526***	-2.560***	-1.419*
	(0.40)	(0.20)	(0.32)	(0.47)	(0.37)	(0.21)	(0.37)	(0.40)	(0.55)	(0.30)	(0.83)	(0.71)
Blue collar	-1.670***	-2.355***	-3.923***	-0.586	-0.821**	-2.035***	-3.890***	-0.945**	-1.288***	-2.515***	-3.184***	-1.214***
	(0.34)	(0.23)	(0.36)	(0.39)	(0.36)	(0.26)	(0.44)	(0.40)	(0.35)	(0.30)	(0.75)	(0.42)
CDD(base CDI)	-0.382	-0.092	0.091	-1.173**	-0.775***	-0.151	0.248	-0.703**	-0.618	-0.648**	-0.515	-0.659
	(0.55)	(0.40)	(0.43)	(0.54)	(0.25)	(0.23)	(0.34)	(0.29)	(0.52)	(0.32)	(0.78)	(0.58)
Other contracts	-1.029	-1.393	-2.123**	-1.452	-1.112***	-1.233***	-2.069***	-1.208**	-0.444	-0.908	-1.979	-0.28
	(1.34)	(0.93)	(0.88)	(1.50)	(0.40)	(0.45)	(0.64)	(0.47)	(0.98)	(0.56)	(1.27)	(1.16)
Constant	-1.136	-1.657	-1.169	2.797***	-1.325**	-0.091	2.677**	-0.242	-0.281	-0.084	-1.577	0.252
	(2.60)	(2.06)	(2.55)	(0.61)	(0.65)	(0.69)	(1.13)	(0.30)	(1.27)	(0.94)	(2.54)	(0.49)
Observations	100	100	100	50	160	160	160	80	80	80	80	40

R-squared	0.99		0.99	0.99	0.99		1	0.99	1		1	1
Number of id	50	50	50		80	80	80		40	40	40	
Adj. R-squared	0.99		0.99	0.99	0.99		1	0.99	1		1	1

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.The dependent variable is log of gross hourly wage.

Table 6.9 Hausman Test

		Pseudo panel data-1 Gender_education_Generation		**Pseudo panel data-2 Gender_Region_Generation**		**Pseudo panel data-3 Region_Generation**	
Different variations with variable size	**Hausman Test**	*Model-1 (only size)*	*Model-2 (all covariates)*	*Model-1 (only size)*	*Model-2 (all covariates)*	*Model-1 (only size)*	*Model-2 (all covariates)*
Size as categorical variable (2005-2006)	chi2(*)	2,86	160,02	10,91	28,48	2,07	3197,9
	Prob>chi2	0,239	0,000	0,004	0,019	0,356	0,000
size as continuous variable (proportion of each category) (2005-2006)	chi2(*)	3,38	16,51	NA	106,83	31,83	101,74
	Prob>chi2	0,066	0,488		0,000	0,000	0,000
size as continuous variable (average of each category) (2005-2006)	chi2(*)	1,24	NA	22,37	91,68	25,63	21,72
	Prob>chi2	0,265		0,000	0,000	0,000	0,115
Size as categorical variable (2002-2006)	chi2(*)	10,15	105,82	10,98	75,97	11,4	NA
	Prob>chi2	0,006	0,000	0,004	0,000	0,003	

For each type of pseudo panel data two models are estimated, firstly, only log of hourly wage as dependent and employer size and independent variable and both FE and RE is estimated and after that Hausman test with stata command 'hausman fixed random' is tested. Secondly, all covariates along with size are used and Hausman test is computed. The procedure is repeated with each modification of employer size variable.

Notes: 'NA'means that the fitted model fails to meet the asymptotic assumptions of the Hausman test;