



HAL
open science

Three essays on microeconomic models of demand and their applications in empirical industrial organisation

Ao Wang

► **To cite this version:**

Ao Wang. Three essays on microeconomic models of demand and their applications in empirical industrial organisation. Economics and Finance. Institut Polytechnique de Paris, 2020. English. NNT : 2020IPPAG003 . tel-02891977

HAL Id: tel-02891977

<https://theses.hal.science/tel-02891977>

Submitted on 7 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2020IPPAG003

Thèse de doctorat



Three essays on microeconomic models of demand and their applications in empirical industrial organisation

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à ENSAE

École doctorale n°626 Ecole Doctorale de l'Institut Polytechnique de Paris (ED IP Paris)

Spécialité de doctorat : Sciences économiques

Thèse présentée et soutenue à Palaiseau, le 12 juin 2020, par

Ao WANG

Composition du Jury :

Philippe Choné Professeur, CREST-ENSAE (UMR 9194)	Président
Laurent Lamy Ingénieur en Chef des Ponts, Eaux et Forêts, ENPC	Rapporteur
Matthew Shou-Chung Shum Professeur, California Institute of Technology	Rapporteur
Alessandro Iaria Lecturer, University of Bristol	Examineur
Xavier d'Haultfœuille Professeur, CREST-ENSAE (UMR 9194)	Directeur de thèse

Acknowledgments

I am extremely lucky to have Xavier d'Haultfœuille as my advisor. Smart, productive, constructive and humble, he defines well what is an excellent researcher. As a student and also a co-author, there are so many things that I learnt from him and will remain as life-long treasures. I owe him infinitely.

I also owe a lot to Alessandro Iaria, my great friend and co-author. I spend a lot of time with him: in his course of empirical IO at ENSAE, in the old building of CREST (MK2), and on Skype when we write papers. He is de facto my co-advisor and does a lot to make me what I am today. He taught me how to build up a paper like a logically perfect machine and a moving piece of art (sometimes we spend hours to improve one single sentence!). Many exciting ideas are born from our endless discussions. It is a great pleasure to work with him.

I would like to express my gratitude to Philippe Février as well, who has guided me to the world of structural econometrics and empirical industrial organisation at Ecole Polytechnique and Veltys. Businessman and researcher, he taught me the first lessons in economics and I learnt how economic intuition and structural models interact in research. Moreover, without Philippe, I may never meet Xavier and start my thesis with him.

CREST has provided an excellent research environment. I have participated in the *Formation Par la Recherche* at CREST-ENSAE and started my research with Xavier when I was still in master. I have also interacted a lot with colleagues at CREST, especially the IO and econometrics groups, and my research has greatly benefited from such interactions. Due to the generous financial support of CREST, I have been able to participate in many external academic activities. Without such support, I may have ended differently.

I want to thank all those who helped before and during the job market which was one of the most challenging moments in the past years. I really appreciate Xavier, Alessandro, Philippe Choné' and Laurent Lamy's effort for improving the paper and pushing me as far as possible on the job market. Without their continuous support and encouragement, the result would have been different. Thanks to Matt, I had a great time visiting Caltech and built up my job market paper. The paper has greatly benefited from the exchanges with Matt and other researchers at Caltech.

Finally, I deeply appreciate for my beloved wife's continuous and unconditional support. During the past years, we shared the most joyful moments as well as the most stressful ones. All will become our precious memories.

Contents

1	Identification and Estimation of Demand for Bundles	1
1.1	Introduction	1
1.2	Model and Notation	5
1.3	Local Identification	9
1.4	Global Identification	13
1.4.1	Exogenous Average Utilities	14
1.4.2	Endogenous Prices	15
1.5	Estimation	17
1.5.1	Invertibility of Product-Level Market Shares	18
1.5.2	A Maximum Likelihood Estimator	19
1.6	Empirical Illustration	20
1.6.1	Data and Definitions	21
1.6.2	Model Specification	23
1.6.3	Estimates of Demand for RTE Cereal Bundles	25
1.6.4	Hicksian Complementarity and Demand Synergies	29
1.6.5	Counterfactuals and Comparisons with Standard Model	35
1.7	Conclusions	36
2	A BLP Demand Model of Product-Level Market Shares with Complementarity	39
2.1	Introduction	39
2.2	Model	43
2.2.1	Model and Notation	43
2.2.2	Examples of Model (2.4) and Interpretation of Demand Synergy Parameters	45
2.2.3	Demand Synergies and Hicksian Substitutions	46
2.3	Identification	47
2.3.1	Economic Analyses and Sufficient Demand Primitives	49
2.3.2	Identification of Product-Level Market Share Functions	49
2.3.3	Identification of Bundle-Level Market Share Functions	52
2.3.4	Summary of Identification	54
2.4	Estimation and Implementation	54
2.4.1	Estimation Procedure	54
2.4.2	Implementation of Demand Inverse	55
2.5	Monte Carlo Simulations	56
2.6	Empirical Illustration: RTE Cereals and Milk	57
2.6.1	Data and Definitions	58
2.6.2	Model Specification	58
2.6.3	Demand Estimates	60
2.6.4	Price Elasticities	61
2.6.5	Counterfactual Simulations	66
2.7	Conclusion	68

3	Estimating the Gains (and Losses) of Revenue Management	69
3.1	Introduction	69
3.2	Institutional Background and Data	72
3.2.1	Revenue Management at iDTGV in 2007-2009	72
3.2.2	Data and descriptive statistics	74
3.3	Theoretical Model	74
3.3.1	Demand side	74
3.3.2	Supply side	75
3.4	Identification and Estimation	77
3.4.1	Counterfactuals and parameters of interests	77
3.4.2	Identification of (ε, ξ_b)	78
3.4.3	Partial identification of $\theta_0 = (\beta_0, \lambda_0, \mu_0)$	79
3.4.4	Inference on counterfactual revenues	81
3.5	Results	81
3.5.1	Demand estimation	81
3.5.2	Counterfactuals	85
3.6	Conclusion	89
	Appendices	91
A	Appendix to Chapter 1	93
A.1	Hendel (1999) and Dubé (2004) as Special Cases of Model (2.3)	93
A.2	Proof of Lemma 1	94
A.3	Proof of Rank Regularity Property	96
A.4	Preliminaries for Theorem 1	97
A.5	Proof of Theorem 1	97
A.6	Proof of Corollary 1	98
A.7	Testing Procedures for Assumption 2 and Assumption 3	99
A.8	Proof of Theorem 2	104
A.9	Proof of Theorem 3	108
A.10	Price-Setting Models Consistent with Assumption 4	109
A.11	Proof of Corollary 2	110
A.12	Proof of Theorem 4	110
A.13	Proof of Theorem 7	112
A.14	Proof of Theorem 6	113
A.15	Appendix Tables	120
B	Appendix to Chapter 2	121
B.1	Cross-Price Elasticities	121
B.2	Proof of Proposition 1	121
B.3	Proof of Theorem 8	122
B.3.1	Conditional Moment Restrictions and Convolution Equation	123
B.3.2	Unique Solution for the Convolution Equation (B.3) and Identification of Product-Level Market Share Functions	124

B.3.3	Sufficient Conditions for the Completeness of Location Families	126
B.4	Proof of Theorem 9	128
B.5	Proof of Corollary 3	133
B.6	Proof of Property 2	134
B.7	Construction of Product-Level Market Shares	134
B.8	Main Tables	135
B.9	Identification of Product-Level Market Share Functions Using Other IVs.	138
B.9.1	Exogenous product characteristics	138
B.10	Complementary Approach of Identification	145
C	Appendix to Chapter 3	147
C.1	A Micro-foundation of Assumption 1	147
C.2	Proof of Theorem 11	148
C.3	Expressions for the counterfactual revenues	149
C.3.1	Complete information	149
C.3.2	Incomplete Information	150
C.4	Demand Estimation with Aggregated Data	152
C.5	Robustness check: lognormal distribution on η_T	153
C.6	A key lemma for Theorem 11	154
C.7	Proof of Theorem 10	156
C.7.1	Complete Information	157
C.7.2	Incomplete Information	162
C.8	Intermediate- K stopping time pricing	168
	Bibliography	171

Identification and Estimation of Demand for Bundles*

Abstract. We present novel identification and estimation results for a mixed logit model of demand for bundles with endogenous prices given bundle-level market shares. Our approach hinges on an affine relationship between the utilities of single products and of bundles, on an essential real analytic property of the mixed logit model, and on the existence of exogenous cost shifters. We propose a new demand inverse in the presence of complementarity that enables to concentrate out of the likelihood function the (potentially numerous) market-product specific average utilities, substantially alleviating the challenge of dimensionality inherent in estimation. To illustrate the use of our methods, we estimate demand and supply in the US ready-to-eat cereal industry, where the proposed MLE reduces the numerical search from approximately 12000 to 130 parameters. Our estimates suggest that ignoring Hicksian complementarity among different products often purchased in bundles may result in misleading demand estimates and counterfactuals.

1.1 Introduction

In standard decision theory, consumer preferences are usually defined over bundles of products rather than over single products (Debreu (1959), Varian (1992), and Mas-Colell et al. (1995)), allowing for both substitutability and complementarity. Despite important exceptions (Manski and Sherman (1980), Hendel (1999), Dubé (2004), Gentzkow (2007), and Thomassen et al. (2017)), the models routinely used to estimate demand rely on the assumption that each of the products purchased in a bundle is chosen independently, precluding the possibility of complementarity and potentially leading to incorrect estimates and counterfactuals.

Models of demand for bundles face non-trivial identification challenges (Gentzkow, 2007), even in settings with a limited number of products (Fox and Lazzati (2017) and Allen and Rehbeck (2019a)). Moreover, the estimation of demand for bundles is subject to a challenge of dimensionality: the number of parameters can be too large to be handled numerically even with parsimonious specifications (Berry et al., 2014). These difficulties forced empirical researchers either to focus on applications with a limited number of products (typically two or three) or to make restrictive assumptions on the parameters capturing potential synergies among the products within bundles (typically a common parameter for all bundles and individuals).¹

We tackle these challenges and propose empirical methods that are practically useful in applications with more than a few products. In particular, we study the identification and estimation of a mixed

*This chapter is based on Iaria, A. and Wang, A. (2020). Identification and estimation of demand for bundles, CEPR Discussion Paper No. DP14363

¹Throughout the paper, we refer to the parameters capturing the potential synergies among products within bundles simply as *demand synergies* or *demand synergy parameters*.

logit model of demand for bundles with endogenous prices given observations on bundle-level market shares. Our arguments hinge on the affine relationship between the utilities of single products and of bundles typical of models along the lines of [Gentzkow \(2007\)](#)'s: the average utility of any bundle equals the sum of the average utilities of the single products plus an extra term capturing their potential demand synergies. This utility structure allows *(i)* for a novel identification approach based on the existence of exogenous but potentially unobserved cost shifters and *(ii)* to alleviate the challenge of dimensionality in estimation by means of a new demand inverse in the presence of complementarity.

Our approach is based on a symmetry assumption about the *average* demand synergies across markets: while the demand synergies for any specific bundle may be unobserved and heterogeneous across individuals, their average is required to be constant across markets with the same observable characteristics (e.g., demographics and prices). Importantly, we propose a specification test for this symmetry assumption on the basis of partial identification methods that can be performed prior to the estimation of the full model. Under this symmetry assumption and regularity conditions similar to [Rothenberg \(1971\)](#), we derive necessary and sufficient rank conditions for the local identification of the model with endogenous prices. This result formalizes [Gentzkow \(2007\)](#)'s insight that, when the average demand synergies are "similar" across markets, the availability of data on many markets will help identification. The "necessity" part of the result is informative about the limits of identification in models of demand for bundles: the separate identification of demand synergies and of the distribution of random coefficients is not immediate, and one needs observations on "enough" markets.

We provide novel sufficient conditions for the global identification of the model with endogenous prices to hold almost everywhere. Our argument combines three main ingredients: a finite number of elements in the identification set, an essential real analytic property of the mixed logit model, and the existence of exogenous cost shifters. We assume that the identification set does not have infinitely many elements and, building on [Chernozhukov et al. \(2007\)](#) and on [Romano and Shaikh \(2012\)](#), propose testable conditions to verify this in practice. We show that the mixed logit market share function is real analytic with respect to the market-product specific average utilities. This further shrinks the identification set in the presence of exogenous variation in the market-product specific average utilities. We then demonstrate that cost shifters can provide the required exogenous variation when the endogenous prices are generated by a large class of pure components and mixed bundling price-setting models.² We finally attain global identification almost everywhere by assuming the existence of exogenous cost shifters that are potentially unobserved but identifiable from observed market shares and prices. One can then interpret our identification strategy as based on the existence of "unobserved" but "identifiable" instruments, the exogenous cost shifters.

We propose a Maximum Likelihood Estimator (MLE) to be implemented with observed bundle-level market shares subject to sampling error and robust to price endogeneity. We account for sampling error to accommodate the typical necessity of computing bundle-level market shares from a sample of household-level purchases (as in [Gentzkow \(2007\)](#), [Kwak et al. \(2015\)](#), [Grzybowski and Verboven \(2016\)](#), [Ruiz et al. \(2017\)](#), and [Ershov et al. \(2018\)](#)). The estimation of demand for bundles is subject to a well known challenge of dimensionality: the number of market-product specific average utility parameters and of demand synergy parameters can be too large to be handled numerically ([Berry et al. \(2014\)](#)). We tackle this practical bottleneck by a novel demand inverse designed to handle complemen-

²For classic treatments of pure components and mixed bundling, see [Adams and Yellen \(1976\)](#), [Lewbel \(1985\)](#), [McAfee et al. \(1989\)](#). For more recent contributions, see [Armstrong and Vickers \(2010\)](#), [Chu et al. \(2011\)](#), [Armstrong \(2013\)](#), and [Zhou \(2017\)](#).

arity among products in models along the lines of [Gentzkow \(2007\)](#)'s. For any given value of the other parameters, we establish a one-to-one mapping between the observed product-level market shares and the market-product specific average utilities.³ This enables to concentrate out of the likelihood the potentially large number of market-product specific average utilities and to substantially simplify the MLE's numerical search: in our application, the numerical search is reduced from approximately 12000 to 130 parameters. We show that our assumptions for global identification guarantee consistency and asymptotic normality of this estimator.

We illustrate our methods in the context of the ready-to-eat (RTE) cereal industry in the USA. We revisit the classic studies by [Nevo \(2000, 2001\)](#), and allow for Hicksian complementarity among different RTE cereal brands in demand estimation.⁴ The households in our data are observed to purchase two or more different brands of RTE cereals in approximately 20% of their shopping trips. Our data record purchases rather than consumption: the purchases of different RTE cereal brands during the same shopping trip can clearly be motivated beyond synergies in consumption. For example, if households face shopping costs, one-stop shopping may be preferred to multi-stop shopping ([Pozzi \(2012\)](#) and [Thomassen et al. \(2017\)](#)). Moreover, if households delegate grocery shopping to one person, preference for variety may lead to the purchase of multiple brands on any shopping trip to accommodate the different needs of the household ([Hendel \(1999\)](#) and [Dubé \(2004\)](#)).

Our model encompasses these alternative mechanisms: the demand synergies are catch-all parameters that may reflect, for example, synergies in consumption, shopping costs, and preference for variety. We try to distinguish empirically the contribution of some of these possible mechanisms to the estimated demand synergies. Our results show that demand for RTE cereals exhibits substantial Hicksian complementarity and that around 75% of it does not seem to be explained by shopping costs or by preference for variety. We compare our estimation results from the full model to those from a model of demand for single brands (similar to [Nevo \(2000, 2001\)](#)) and show that ignoring Hicksian complementarity may result in misleading demand estimates and counterfactuals (see also [Fosgerau et al. \(2019\)](#)). Estimates from the full model support the classic [Cournot \(1838\)](#)'s insight that, in the presence of Hicksian complementarity, mergers can be welfare enhancing; while those from a standard model that does not allow for it predict that mergers are detrimental for consumer surplus.

Related Literature. There is a growing empirical literature leveraging the estimation of demand for bundles. [Manski and Sherman \(1980\)](#) study households' choices of motor vehicle holdings; [Hendel \(1999\)](#) studies preference for variety for personal computers, while [Dubé \(2004\)](#) and [Chan \(2006\)](#) for soft carbonated drinks; [Nevo et al. \(2005\)](#) study the decision of libraries to subscribe to economics and business journals; [Gentzkow \(2007\)](#) and [Gentzkow et al. \(2014\)](#) investigate competition and complementarity among newspapers; [Augereau et al. \(2006\)](#) the returns from adoption of technological standards; [Liu et al. \(2010a\)](#) and [Grzybowski and Verboven \(2016\)](#) the complementarity among telecommunication services; [Crawford and Yurukoglu \(2012\)](#) and [Crawford et al. \(2018\)](#) the problem

³Demand inverses at the *bundle*-level can simply rely on the classic results by [Berry \(1994\)](#) and [Berry et al. \(2013\)](#) as long as the *bundles* in the demand system are substitutes. However, if some of the *products* are complements, these classic results do not imply the invertibility of the demand system at the *product*-level. Our product-level demand inverse is instead based on the *P*-matrix property of [Gale and Nikaido \(1965\)](#), which does not require the products to be substitutes.

⁴Following [Samuelson \(1974\)](#) and [Gentzkow \(2007\)](#), we rely on the classic Hicksian notion of complementarity: we consider two brands as complements whenever their cross-price elasticity of (compensated) demand is negative. For recent discussions on complementarity in empirical models of demand, see [Manzini et al. \(2018\)](#), [Dubé \(2019\)](#), and [Iaria and Wang \(2019b\)](#).

of bundling and vertical restraints in cable television, while [Ho et al. \(2012\)](#) in the video rental industry; [Kretschmer et al. \(2012\)](#) study the adoption of complementary innovations; [Lee et al. \(2013\)](#) the complementarity between milk and RTE cereals; [Song et al. \(2017\)](#) the relationship between mergers and inter-firm bundling in the pharmaceutical industry; [Ruiz et al. \(2017\)](#) propose a machine learning model of demand for bundles, [Thomassen et al. \(2017\)](#) study the problem of transportation costs in grocery shopping; [Ershov et al. \(2018\)](#) the complementarity between potato chips and soft carbonated drinks; and [Fosgerau et al. \(2019\)](#) the complementarity between different brands of RTE cereals. We add to this empirical literature by providing novel identification and estimation methods for models along the lines of [Gentzkow \(2007\)](#)'s, specifically accounting for price endogeneity and alleviating the challenge of dimensionality inherent in estimation.

The global identification of non-linear models is notoriously complex to demonstrate ([Newey and McFadden \(1994\)](#) and [Lewbel \(2019\)](#)). Researchers typically resume to non-verifiable abstract conditions ([Rothenberg \(1971\)](#), [Bowden \(1973\)](#), and [Komunjer \(2012\)](#)) or focus on weaker identification concepts altogether, such as local identification ([Rothenberg \(1971\)](#), [Sargan \(1983\)](#), and [Lewbel \(2012\)](#)) or partial identification ([Manski \(1989\)](#), [Manski \(2003\)](#), and [Chesher and Rosen \(2017\)](#)). We contribute to this literature by providing sufficient conditions for global identification that are testable (on the basis of partial identification methods), rooted in economic theory (to address price endogeneity), and weaker than the classics ([Rothenberg \(1971\)](#), [Bowden \(1973\)](#), and [Komunjer \(2012\)](#)). The relative advantage of our conditions follows from a real analytic property we show to be satisfied by mixed logit models given any distribution of random coefficients (parametric or non-parametric), which allows us to relax the strict concavity of the likelihood function (or similar criterion functions). [Fox et al. \(2012\)](#) and [il Kim \(2014\)](#) also exploit the real analytic properties of logit models to achieve global identification, but in more restrictive frameworks. [il Kim \(2014\)](#) shows the real analytic property for multinomial logit and for nested logit models, while [Fox et al. \(2012\)](#) show it for mixed logit models with random coefficients defined over compact supports—ruling out, e.g., normal and log-normal distributions.

In the context of identification of models of demand for bundles, we add to the discussions by, for example, [Fox and Lazzati \(2017\)](#) and [Allen and Rehbeck \(2019a\)](#). [Fox and Lazzati \(2017\)](#) propose sufficient conditions for the non-parametric identification of demand for bundles (and binary games of complete information) on the basis of additively separable excluded regressors. [Allen and Rehbeck \(2019a\)](#) instead study the non-parametric identification of a large class of demand models, among which demand for bundles, by exploiting variation in the substitution and complementarity patterns among different products. While these papers make fewer distributional assumptions and can be preferred in situations with small choice sets and exogenous regressors, our arguments apply more readily to cases with larger choice sets, endogenous prices, and in general lead to practically convenient estimators.

Our mixed logit model of demand for bundles can be seen as a special case of the general non-parametric framework by [Berry and Haile \(2014\)](#). [Berry and Haile \(2014\)](#)' identification argument relies on the availability of *observed* instruments both to pin down the distribution of random coefficients and to address price endogeneity. In contrast, [Gentzkow \(2007\)](#)'s utility structure allows us to propose a complementary identification strategy based on *unobserved* instruments: we rely on the existence of “unobserved” but “identifiable” cost shifters and on conditional symmetry restrictions among the average demand synergies across markets. While less general in abstract terms, our arguments are more applicable to cases with limited observability of instruments and give rise to sizeable computational

advantages in estimation.⁵

Our estimator contributes to the modern literature on the estimation of demand systems started by [Berry et al. \(1995\)](#) (henceforth BLP). For example, [Berry et al. \(2004b\)](#), [Freyberger \(2015\)](#), and [Armstrong \(2016b\)](#) investigate the asymptotic properties of GMM estimators of demand systems with endogenous prices. While these GMM estimators are more widely applicable provided the availability of observable instruments, in the context of demand for bundles our MLE represents a numerically convenient alternative in which the instruments need to exist but do not need to be observed. More recently, [Compiani \(2019\)](#) proposes a non-parametric estimator of demand models that accommodates complementarity among products. There is a trade-off between our proposed estimator and [Compiani \(2019\)](#)'s. His non-parametric estimator is more flexible than ours, but it is subject to a curse of dimensionality that may constrain its applicability to settings with small choice sets. Our MLE is less affected by dimensionality and can be implemented with larger choice sets.

Since [Berry \(1994\)](#), the identification and the estimation of demand systems with endogenous prices has been relying on the ability to “invert” market share equations to uniquely determine the implied product-specific average utilities—the so called demand inverse. A standard requirement for the invertibility of demand systems is for the products to be substitutes, see [Berry et al. \(2013\)](#). This requirement can be problematic in contexts with complementary products: for example, in a model of demand for bundles of newspapers, [Fan \(2013\)](#) rules out by assumption any complementarity in order to rely on the classic demand inverse by [Berry \(1994\)](#) at the newspaper-level. Our novel demand inverse addresses this issue and allows to invert product-level market share equations in the presence of complementarity.

Organization. In the next section, we introduce model and notation. In sections 2.3 and 1.4, we present—respectively—our local and global identification results. In section 1.5, we propose our demand inverse and a related MLE. In section 1.6, we explore the practical relevance of our methods with an empirical illustration. In section 1.7, we conclude the paper with some final remarks. In (online) appendix section A, we report all the proofs and additional results.

1.2 Model and Notation

Imagine a cross-section of T independent markets denoted by \mathbf{T} , where each market $t \in \mathbf{T}$ is populated by $i = 1, \dots, I$ individuals. Individual i in market t makes purchases exclusively in market t and is a different person from individual i in any other market $t' \neq t$. For individuals in market t , let \mathbf{J}_t be the set of $j = 1, \dots, J_t$ market-specific products that can be purchased in isolation or in bundles. Let $\mathbf{C}_t = \mathbf{C}_{t1} \cup \{0\}$ be the choice set specific to market t , which includes: the collection of “inside” options \mathbf{C}_{t1} and the “outside” option $j = 0$ (i.e., the option not to purchase any product). In turn, the collection of inside options is defined as $\mathbf{C}_{t1} = \mathbf{J}_t \cup \mathbf{C}_{t2}$, where \mathbf{C}_{t2} denotes the set of market-specific bundles of products. The set of all available bundles across all markets is $\mathbf{C}_2 = \cup_{t=1}^T \mathbf{C}_{t2}$. We refer to the cardinality of these sets as: $C_t = |\mathbf{C}_t|$, $C_{t1} = |\mathbf{C}_{t1}|$, $C_{t2} = |\mathbf{C}_{t2}|$, and $C_2 = |\mathbf{C}_2|$. We denote by \mathbf{b} any element of the choice set \mathbf{C}_t , whereby some abuse of notation \mathbf{b} may refer to a bundle, a single product, or the outside option.

⁵The classic identification argument based on observed instruments requires the performance of high-dimensional demand inverses at the bundle-level, while our argument based on *unobserved* instruments allows for the performance of demand inverses only at the product-level. In practice, this implies the numerical inversion of a lower-dimensional demand system and leads to large computational advantages.

The indirect utility of individual i in market t from purchasing product j is:

$$\begin{aligned} U_{itj} &= u_{itj} + \varepsilon_{itj} \\ &= \delta_{tj} + \mu_{itj} + \varepsilon_{itj} \quad \text{and} \\ U_{it0} &= \varepsilon_{it0}, \end{aligned} \tag{1.1}$$

where $u_{itj} = \delta_{tj} + \mu_{itj}$, δ_{tj} is the market t -specific average utility of product $j \in \mathbf{J}_t$, μ_{itj} is an unobserved individual-specific utility deviation from δ_{tj} , while ε_{itj} and ε_{it0} are error terms. Throughout the paper, we treat the market t -specific average utilities as parameters to be identified and estimated. One can however reduce the number of parameters by using observable characteristics and making additional functional form assumptions.⁶

To ease exposition, when \mathbf{b} is a bundle, we refer to the products it contains as $j \in \mathbf{b}$. Following [Gentzkow \(2007\)](#), the indirect utility of individual i in market t from purchasing bundle $\mathbf{b} \in \mathbf{C}_{t2}$ is:

$$\begin{aligned} U_{it\mathbf{b}} &= \sum_{j \in \mathbf{b}} u_{itj} + \Gamma_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}} \\ &= \sum_{j \in \mathbf{b}} (\delta_{tj} + \mu_{itj}) + \Gamma_{t\mathbf{b}} + (\Gamma_{it\mathbf{b}} - \Gamma_{t\mathbf{b}}) + \varepsilon_{it\mathbf{b}} \\ &= \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}} + \left[\sum_{j \in \mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}} \right] + \varepsilon_{it\mathbf{b}} \\ &= \delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) + \mu_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}}, \end{aligned} \tag{1.2}$$

where $\Gamma_{it\mathbf{b}}$ is the individual-specific *demand synergy* among the products in bundle \mathbf{b} , which we specify as $\Gamma_{it\mathbf{b}} = \Gamma_{t\mathbf{b}} + \zeta_{it\mathbf{b}}$. $\Gamma_{t\mathbf{b}}$ is the average demand synergy for the products in bundle \mathbf{b} among the individuals in market t and $\zeta_{it\mathbf{b}}$ is an unobserved individual-specific deviation from this average. $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}}$ is the market t -specific average utility for bundle \mathbf{b} , $\mu_{it\mathbf{b}}$ is an unobserved individual-specific deviation from $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}})$, and $\varepsilon_{it\mathbf{b}}$ is an error term.

The demand synergy parameter $\Gamma_{it\mathbf{b}}$ captures the extra utility individual i in market t obtains from purchasing the products in bundle \mathbf{b} jointly rather than separately. When $\Gamma_{it\mathbf{b}} > 0$, the utility of the bundle is super-modular with respect to the utilities of the single products and, from i 's perspective, joint purchase brings more utility. Conversely, when $\Gamma_{it\mathbf{b}} < 0$, from i 's perspective the separate purchase of each $j \in \mathbf{b}$ brings more utility than their joint purchase. As we discuss below, in applications with observable bundle-specific characteristics (e.g., bundle-specific discounts), one can specify $\Gamma_{it\mathbf{b}}$ in terms of these characteristics.

We now turn to the distributional assumptions for the unobserved components of utility: $\mu_{it\mathbf{b}} = \sum_{j \in \mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}}$ and $\varepsilon_{it\mathbf{b}}$ for each $\mathbf{b} \in \mathbf{C}_t$. We assume that $\mu_{it\mathbf{b}}$ can be specified as a function of a vector of random coefficients β_{it} , so that $\mu_{it\mathbf{b}} = \mu_{it\mathbf{b}}(\beta_{it})$, and that β_{it} is distributed according to $F(\cdot; \Sigma_F)$, where Σ_F is a finite-dimensional parameter in a connected compact set $\Theta_{\Sigma_F} \subset \mathbb{R}^P$. As is typical, $\mu_{it\mathbf{b}}(\cdot)$ can also be a function of observable demographics (e.g., i 's income) and/or observable market-, product-, and bundle-specific characteristics (e.g., the price of bundle \mathbf{b} in market t). The error term $\varepsilon_{it\mathbf{b}}$ is assumed to be i.i.d. Gumbel.

Even though we make the assumption that $\varepsilon_{it\mathbf{b}}$ is i.i.d. Gumbel, as shown by [McFadden and Train \(2000\)](#), under mild regularity conditions any discrete choice model derived from random utility

⁶We provide more detail on this while discussing price endogeneity at the end of this section.

maximization can be approximated arbitrarily well by mixed logit models of the kind we consider. In addition, note that our mixed logit model is a generalization of [Gentzkow \(2007\)](#)'s, which restricts $F(\cdot; \Sigma_F)$ to be a normal distribution and $\Gamma_{it\mathbf{b}} = \Gamma_{\mathbf{b}}$ for all i 's and t 's. We add a layer of unobserved heterogeneity to the individual preferences specific to each bundle: for reasons unobserved to the econometrician, the products in any bundle can exhibit positive demand synergies for some individuals and negative for others.

Denote the market t -specific average utility vector by $\delta_t(\Gamma_t) = (\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}))_{\mathbf{b} \in \mathbf{C}_{t1}}$ and the vector collecting all the market t -specific average demand synergies by Γ_t . $\delta_t(\Gamma_t)$ does not only list the t -specific average utilities of bundles $\mathbf{b} \in \mathbf{C}_{t2}$, but also those of the single products $\mathbf{b} \in \mathbf{C}_{t1} \setminus \mathbf{C}_{t2}$ (where $\mathbf{C}_{t1} \setminus \mathbf{C}_{t2} = \mathbf{J}_t$): given that any single product has zero demand synergies, our notation for $\mathbf{b} = j \in \mathbf{C}_{t1} \setminus \mathbf{C}_{t2}$ is just $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \delta_{tj}$. Given our distributional assumptions, the market share function of $\mathbf{b} \in \mathbf{C}_t$ for individuals in market t takes the mixed logit form:

$$\begin{aligned} s_{it\mathbf{b}}(\delta_t(\Gamma_t); \Sigma_F) &= \int s_{it\mathbf{b}}(\delta_t(\Gamma_t), \beta_{it}) dF(\beta_{it}; \Sigma_F) \\ &= \int \frac{e^{\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) + \mu_{it\mathbf{b}}(\beta_{it})}}{\sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}(\Gamma_{t\mathbf{b}')} + \mu_{it\mathbf{b}'}(\beta_{it})}} dF(\beta_{it}; \Sigma_F), \end{aligned} \quad (1.3)$$

where $s_{it\mathbf{b}}(\delta_t(\Gamma_t), \beta_{it})$ is individual i 's purchase probability of \mathbf{b} in market t given β_{it} .

Complementarity and Substitutability. Following [Samuelson \(1974\)](#) and [Gentzkow \(2007\)](#), we rely on the classic notion of Hicksian complementarity: we consider two products as complements (substitutes) whenever their cross-price elasticity of demand is negative (positive).⁷ In a model similar to (2.3) with two products, j and k , and constant demand synergy parameters $\Gamma_{it(j,k)} = \Gamma_{(j,k)}$, [Gentzkow \(2007\)](#) shows that j and k are complements (substitutes) whenever $\Gamma_{(j,k)} > 0$ ($\Gamma_{(j,k)} < 0$). On the one hand, with more products and heterogeneous demand synergies, the relationship between Hicksian complementarity and $\Gamma_{it\mathbf{b}}$ is less clear-cut and the topic of ongoing research ([Iaria and Wang, 2019b](#)). On the other, though, standard models of demand for single products—obtained by constraining $\Gamma_{it\mathbf{b}} = -\infty$ for all i 's, t 's, and $\mathbf{b} \in \mathbf{C}_{t2}$ —rule out the possibility of Hicksian complementarity and force any two products to be substitutes. In this paper, we take a pragmatic approach and regard the complementarity or substitutability between products as an empirical question to be answered after the estimation of model (2.3).

Interpretation of Demand Synergies. Model (2.3) is agnostic about the exact meaning of $\Gamma_{it\mathbf{b}}$, which is a catch-all parameter that can reflect, for example, synergies in consumption, shopping costs, and preference for variety. In [Gentzkow \(2007\)](#)'s demand for on-line and printed newspapers, $\Gamma_{it\mathbf{b}}$ captures synergies in the consumption of the different news outlets. However, demand synergies—and consequently Hicksian complementarity—can also arise, for example, because of shopping costs ([Pozzi \(2012\)](#) and [Thomassen et al. \(2017\)](#)) or preference for variety within households ([Hendel \(1999\)](#) and [Dubé \(2004\)](#)). If individuals face shopping costs every time they visit a store, they may prefer to purchase all of their products at once rather than over several trips (one-stop shoppers). Moreover,

⁷In our application, we rule out income effects so that gross complementarity (in terms of elasticities of Marshallian demands) and Hicksian complementarity (in terms of elasticities of compensated demands) coincide. For discussions about complementarity in models of demand for bundles similar to those studied here, see [Manzini et al. \(2018\)](#), [Dubé \(2019\)](#), and [Iaria and Wang \(2019b\)](#).

if households delegate grocery shopping to one person, then preference for variety may lead to the purchase of multiple products on any shopping trip to accommodate the different needs within the household. Our model can rationalize shopping costs with positive demand synergies and, as we show in Appendix A.1, preference for variety with some additional structure on the demand synergy parameters.

Random Intercepts and Demand Synergies. As argued by [Gentzkow \(2007\)](#), the random intercepts $(\mu_{itj})_{j=1}^{J_t}$ play an important conceptual role in the identification of demand synergies in mixed logit models of demand for bundles. Without random coefficients, the Independence from Irrelevant Alternatives (IIA) property would imply that the relative predicted market shares of any two bundles do not depend on the characteristics of any other bundle. Removing from the choice set a bundle almost identical to the preferred one (e.g., same products but one) or a bundle completely different from it (e.g., only different products) would equivalently have no impact on the remaining relative predicted market shares. The random intercepts mitigate this limitation in an intuitive way: the indirect utilities of all bundles including product j will share the random intercept μ_{itj} , so that bundles with a larger overlap of products will also have more correlated indirect utilities. Disentangling demand synergies from these random intercepts is the key identification challenge in models of demand for bundles: as shown by [Gentzkow \(2007\)](#), not accounting for possible correlations across the indirect utilities of bundles with overlapping products may lead to finding spurious demand synergies and Hicksian complementarities.

Average Utilities and Price Endogeneity. We treat the average utility δ_{tj} as a fixed effect to be identified and estimated, being unspecific about its exact dependence on price and other observed or unobserved market-product specific characteristics. For example, following [Berry \(1994\)](#) and BLP, a classical linear specification is $\delta_{tj} = x_{tj}\tau + \alpha p_{tj} + \xi_{tj}$, where x_{tj} is a vector of exogenous observed characteristics, p_{tj} is the observed price, $(\tau, \alpha)^T$ is a vector of preference parameters, and ξ_{tj} is a residual unobserved to the econometrician but observed to both individuals and price-setting firms. In this context, endogeneity arises whenever prices are chosen by firms on the basis of $(\xi_{tj})_{j=1}^{J_t}$.

Our local identification arguments are robust to cases of price endogeneity in which, for any bundle \mathbf{b} , the source of endogeneity is confined to $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}}$, with $\Gamma_{t\mathbf{b}}$ constrained to be constant across markets with the same market-bundle specific observables. In particular, as detailed in Assumption 2 below, we require $\Gamma_{t\mathbf{b}} = \Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)$, where $\Gamma_{\mathbf{b}}$ is a bundle-specific fixed effect and $g(\cdot, \cdot; \Sigma_g)$ is a function parametrized by Σ_g of the observed characteristics $x_{t\mathbf{b}}$ and of the observed price surcharge/discount $p_{t\mathbf{b}}$ (the difference between the price of bundle \mathbf{b} and $\sum_{j \in \mathbf{b}} p_{tj}$). For example, one can specify $g(\cdot, \cdot; \Sigma_g)$ as $g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \tau, \alpha) = x_{t\mathbf{b}}\tau + \alpha p_{t\mathbf{b}}$. While our assumptions on $\Gamma_{t\mathbf{b}}$ allow δ_{tj} to be any arbitrary function of $(x_{tj}, p_{tj}, \xi_{tj})$, they restrict the functional form of the market-bundle specific unobservables on the basis of which firms choose prices. For instance, with the above linear specifications for δ_{tj} and $\Gamma_{t\mathbf{b}}$, we have $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = (x_{t\mathbf{b}} + \sum_{j \in \mathbf{b}} x_{tj})\tau + \alpha(p_{t\mathbf{b}} + \sum_{j \in \mathbf{b}} p_{tj}) + \sum_{j \in \mathbf{b}} \xi_{tj} + \Gamma_{\mathbf{b}}$, with the market-bundle specific unobservable restricted to $\sum_{j \in \mathbf{b}} \xi_{tj} + \Gamma_{\mathbf{b}}$.

Our global identification arguments further require restrictions on δ_{tj} and on the price-setting model. As detailed in section 1.4.2, we require: (i) the average utility δ_{tj} to be additively separable in ξ_{tj} and an arbitrary function of (x_{tj}, p_{tj}) and (ii) the existence of exogenous cost shifters that are unobserved to the econometrician but identifiable from observed market shares and prices.

1.3 Local Identification

Suppose that the econometrician observes without error the market shares $s_{t\mathbf{b}}$ of each $\mathbf{b} \in \mathbf{C}_{t1}$ for each independent market $t = 1, \dots, T$.^{8,9} We focus on the case of a fixed number of products J_t and of a fixed number of independent markets T . We do not consider the case of panel data with repeated observations for each market. Similar to [Berry and Haile \(2014\)](#), our notion of identification concerns the conditions under which

$$\begin{aligned} s_{t\mathbf{b}}(\delta'_t(\Gamma'_t); \Sigma'_F) &= s_{t\mathbf{b}} \\ \text{subject to } \Gamma'_{t\mathbf{b}} &= \delta'_{t\mathbf{b}}(\Gamma'_{t\mathbf{b}}) - \sum_{j \in \mathbf{b}} \delta'_{tj}, \mathbf{b} \in \mathbf{C}_{t2} \end{aligned} \quad (1.4)$$

has a unique solution for $t \in \mathbf{T}$ and $\mathbf{b} \in \mathbf{C}_{t1}$, where $\delta'_t(\Gamma'_t) = (\delta'_{t\mathbf{b}}(\Gamma'_{t\mathbf{b}}))_{\mathbf{b} \in \mathbf{C}_{t1}}$ and $s_{t\mathbf{b}}(\delta'_t(\Gamma'_t); \Sigma'_F)$ is defined in (2.3). Define the $J_t \times 1$ market t -specific vector $\delta_{t\mathbf{J}_t} = (\delta_{tj})_{j \in \mathbf{J}_t}$, and the $C_{t1} \times 1$ market t -specific vectors $s_t(\cdot; \Sigma'_F) = (s_{t\mathbf{b}}(\cdot; \Sigma'_F))_{\mathbf{b} \in \mathbf{C}_{t1}}$ and $s_t = (s_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t1}}$.

Definition 1. *Model (2.3) is locally identified if and only if there exists a neighbourhood V of the true parameters $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma_1, \dots, \Gamma_T, \Sigma_F)$ such that $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma_1, \dots, \Gamma_T, \Sigma_F)$ is the unique solution to (1.4) in V .*

Definition 1 constrains our discussion of identification to the existence of a unique solution to system (1.4) in mixed logit model (2.3). We will refer to the existence of multiple solutions to this specific problem as to lack of identification. Because of the non-linear nature of model (2.3), we start by studying the problem of local identification. In section 1.4, we then investigate global identification, which requires stronger assumptions.

Building on [Berry et al. \(2013\)](#), our identification arguments rely on demand inverses derived from (1.4). Define the inverse market share for market $t \in \mathbf{T}$ as:

$$s_t^{-1}(\cdot; \Sigma_F) = (s_{t\mathbf{b}}^{-1}(\cdot; \Sigma_F))_{\mathbf{b} \in \mathbf{C}_{t1}} : \mathbf{S}_{t1} \Rightarrow \mathbb{R}^{C_{t1}}, \quad (1.5)$$

where $s_{t\mathbf{b}}^{-1}(\cdot; \Sigma_F)$ is the inverse market share for market $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_{t1}$, and

$$\mathbf{S}_{t1} = \{(s_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t1}} : s_{t\mathbf{b}} \in (0, 1), \sum_{\mathbf{b} \in \mathbf{C}_{t1}} s_{t\mathbf{b}} < 1\}$$

is the set of all feasible market share vectors for market $t \in \mathbf{T}$. The next Assumption imposes some regularity conditions on the parametric distribution of the random coefficients (first requirement) and that the products belonging to any bundle can also be purchased individually (second requirement).

Assumption 1.

1. The density of β_{it} , $\frac{dF(\beta_{it}; \Sigma'_F)}{d\beta_{it}}$, is continuously differentiable with respect to Σ'_F for any β_{it} .
2. If $\mathbf{b} \in \mathbf{C}_{t2}$, then $j \in \mathbf{J}_t$ for any $j \in \mathbf{b}$.

⁸This is only for the purpose of identification, in estimation we consider observed market shares subject to sampling error.

⁹[Sher and Kim \(2014\)](#), [Allen and Rehbeck \(2019a\)](#), and [Wang \(2019\)](#) study a different identification problem, where only the product-level market shares, rather than the bundle-level market shares, are observed.

The next Lemma verifies the sufficient conditions by [Berry et al. \(2013\)](#) for the bundle-level demand inverse (1.5) to be a continuously differentiable function.

Lemma 1.

- For any given $\Sigma'_F \in \Theta_{\Sigma_F}$, the inverse market share (1.5) is a function: for each $s_t \in \mathbf{S}_{t1}$, there exists a unique $\delta'_t \in \mathbb{R}^{C_{t1}}$ such that $s_t(\delta'_t; \Sigma'_F) = s_t$.
- Given Assumption 1, the inverse market share, $s_t^{-1}(s'_t; \Sigma'_F)$, is continuously differentiable with respect to (s'_t, Σ'_F) in a neighbourhood of (s_t, Σ_F) .

Proof. See Appendix A.2. □

It is not hard to see that without further restrictions on Γ_t or additional external information, model (2.3) can hardly be identified. In this paper, we build on [Gentzkow \(2007\)](#)'s insight that when $\Gamma_t = \Gamma$, the availability of purchase data for multiple markets will help identification, and propose the following restriction.

Assumption 2. $\Gamma_{t\mathbf{b}} = \Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)$ for $t \in \mathbf{T}$ and $\mathbf{b} \in \mathbf{C}_{t2}$, where $\Gamma_{\mathbf{b}}$ is a bundle-specific fixed effect, $x_{t\mathbf{b}}$ a vector of observed market-bundle specific characteristics, $p_{t\mathbf{b}}$ an observed price surcharge/discount for the joint purchase of the products in the bundle, and $g(\cdot, \cdot; \Sigma_g)$ a function of $(x_{t\mathbf{b}}, p_{t\mathbf{b}})$ known up to and continuously differentiable with respect to $\Sigma_g \in \Theta_{\Sigma_g} \subseteq \mathbb{R}^D$.

Assumption 2 restricts the variation in $\Gamma_{t\mathbf{b}}$ across markets to be fully captured by the variation in the observables $(x_{t\mathbf{b}}, p_{t\mathbf{b}})$ through the parametric function $g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)$. This allows to reduce the dimensionality of the collection of average demand synergies from $\sum_{t=1}^T C_{t2}$ to $D + C_2$ and in particular to treat $\Gamma_{\mathbf{b}}$ as a bundle-specific fixed effect to be identified and estimated. Note that, even though Assumption 2 requires all markets with given $(x_{t\mathbf{b}}, p_{t\mathbf{b}})$ to have the same average demand synergy $\Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)$, each individual in each market is allowed to have a specific demand synergy deviation $\zeta_{it\mathbf{b}}$, so that $\Gamma_{it\mathbf{b}}$ may potentially differ across individuals for any given market and bundle. Moreover, as we illustrate below, in applications with a large number of markets with overlapping choice sets, Assumption 2 potentially leads to many over-identifying moment restrictions and can be weakened, so to allow for more flexible specifications of $\Gamma_{t\mathbf{b}}$. In those cases, one could for example specify $(\Gamma_{t\mathbf{b}}, \Sigma_{tg}) = (\Gamma_{1\mathbf{b}}, \Sigma_{1g})$ for $t = 1, \dots, T_1$, $(\Gamma_{t\mathbf{b}}, \Sigma_{tg}) = (\Gamma_{2\mathbf{b}}, \Sigma_{2g})$ for $t = T_1 + 1, \dots, T_2$, and so on until each t belonged to one of Q groups of "similar" markets with $(\Gamma_{1\mathbf{b}}, \Sigma_{1g}) \neq (\Gamma_{2\mathbf{b}}, \Sigma_{2g}) \neq \dots \neq (\Gamma_{Q\mathbf{b}}, \Sigma_{Qg})$.

Remark 1. Assumption 2 gives rise to testable implications and can be verified in practice. In Appendix A.7, we present a specification test that builds on partial identification methods. Essentially, the proposed test checks whether there exists at least one profile of parameters $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma_1, \dots, \Gamma_T, \Sigma_F)$ satisfying Assumption 2 that solves demand system (1.4). A rejection of the test is evidence against Assumption 2 and highlights its incoherence with the data.

The presence of random coefficients, i.e. $\dim(\Sigma_F) > 0$, leads system (1.4) to have more unknowns than equations, introducing an identification problem not present in multinomial logit models. In general demand systems where the indirect utilities of different alternatives have no particular relationships, this dimensionality issue is typically addressed by including additional instruments beyond those necessary to address price endogeneity. However, in the case of [Gentzkow \(2007\)](#)'s demand for bundles, the specific structure that links the indirect utilities of bundles to those of single products allows to

reduce dimensionality from within the system. Assumption 2 embodies this strategy: by imposing a symmetry restriction among the average demand synergies across markets, the model can be identified without requiring additional instruments to those necessary to address price endogeneity.

Due to Lemma 1 and Assumption 2, at the true parameters Σ_F and market shares \mathcal{J}_t , one can re-express the first line of system (1.4) as:

$$\begin{aligned} \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g) &= s_{t\mathbf{b}}^{-1}(\mathcal{J}_t; \Sigma_F), \text{ for bundle } \mathbf{b} \in \mathbf{C}_{t2} \\ \delta_{tj} &= s_{tj}^{-1}(\mathcal{J}_t; \Sigma_F), \text{ for product } j \in \mathbf{b}. \end{aligned} \quad (1.6)$$

By substituting (1.6) into the second line of (1.4), one gets:

$$\Gamma_{\mathbf{b}} = s_{t\mathbf{b}}^{-1}(\mathcal{J}_t; \Sigma_F) - \sum_{j \in \mathbf{b}} s_{tj}^{-1}(\mathcal{J}_t; \Sigma_F) - g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g), \quad (1.7)$$

for $t \in \mathbf{T}$ and bundle $\mathbf{b} \in \mathbf{C}_{t2}$. Note that the left-hand side of system (1.7) does not depend on market t , while the right-hand side does. Consequently, at the true parameters $\Sigma = (\Sigma_F, \Sigma_g)$, true market shares of any two markets, \mathcal{J}_t and $\mathcal{J}_{t'}$, and any $\mathbf{b} \in \mathbf{C}_{t2} \cap \mathbf{C}_{t'2}$, one obtains:

$$s_{t\mathbf{b}}^{-1}(\mathcal{J}_t; \Sigma_F) - \sum_{j \in \mathbf{b}} s_{tj}^{-1}(\mathcal{J}_t; \Sigma_F) - g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g) = s_{t'\mathbf{b}}^{-1}(\mathcal{J}_{t'}; \Sigma_F) - \sum_{j \in \mathbf{b}} s_{t'j}^{-1}(\mathcal{J}_{t'}; \Sigma_F) - g(x_{t'\mathbf{b}}, p_{t'\mathbf{b}}; \Sigma_g). \quad (1.8)$$

Our identification strategy exploits all such moment conditions for any pair of markets $t \neq t'$ and any $\mathbf{b} \in \mathbf{C}_{t2} \cap \mathbf{C}_{t'2}$. As we will see below, under certain conditions, these moment restrictions can uniquely determine the true parameters $\Sigma = (\Sigma_F, \Sigma_g)$. Then, due to (1.7), the true parameters $\Sigma = (\Sigma_F, \Sigma_g)$ can uniquely determine the remaining portion $\Gamma_{\mathbf{b}}$ of the true demand synergies, for any $\mathbf{b} \in \mathbf{C}_{t2}$. Denote $g_t(\Sigma_g) = (g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g))_{\mathbf{b} \in \mathbf{C}_{t2}}$. Finally, because of Lemma 1, one can uniquely recover $\delta_t(\Gamma + g_t(\Sigma_g)) = (\delta_{t1}, \dots, \delta_{tJ_t}, (\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)))_{\mathbf{b} \in \mathbf{C}_{t2}})$.

Note that for any $\mathbf{b} \in \mathbf{C}_2$, there exists t such that $\mathbf{b} \in \mathbf{C}_{t2}$. Then, for any $\mathbf{b} \in \mathbf{C}_2$, define $\mathbf{T}_{\mathbf{b}} = \{t : \mathbf{b} \in \mathbf{C}_{t2}, t \in \mathbf{T}\}$. If $\mathbf{T}_{\mathbf{b}}$ has more than one element, we order them from t_1 to $t_{|\mathbf{T}_{\mathbf{b}}|}$. By applying the right-hand side of (1.7) to t_a and to t_{a+1} and by taking the difference, for $a = 1, \dots, |\mathbf{T}_{\mathbf{b}}| - 1$, we then obtain $|\mathbf{T}_{\mathbf{b}}| - 1$ moment conditions:¹⁰

$$\begin{aligned} m_{\mathbf{b}}(\Sigma'_F, \Sigma'_g; \mathcal{J}) &= \left[s_{t_a \mathbf{b}}^{-1}(\mathcal{J}_{t_a}; \Sigma'_F) - \sum_{j \in \mathbf{b}} s_{t_a j}^{-1}(\mathcal{J}_{t_a}; \Sigma'_F) - s_{t_{a+1} \mathbf{b}}^{-1}(\mathcal{J}_{t_{a+1}}; \Sigma'_F) \right. \\ &\quad \left. + \sum_{j \in \mathbf{b}} s_{t_{a+1} j}^{-1}(\mathcal{J}_{t_{a+1}}; \Sigma'_F) + g(x_{t_{a+1} \mathbf{b}}, p_{t_{a+1} \mathbf{b}}; \Sigma'_g) - g(x_{t_a \mathbf{b}}, p_{t_a \mathbf{b}}; \Sigma'_g) \right]_{a=1}^{|\mathbf{T}_{\mathbf{b}}|-1}, \quad (1.9) \\ m_{\mathbf{b}}(\Sigma'; \mathcal{J})|_{\Sigma'=\Sigma} &= 0. \end{aligned}$$

Moment conditions (1.9) rely on relationship (1.7) and the fact that markets t_a and t_{a+1} have the same bundle-specific fixed effect $\Gamma_{\mathbf{b}}$. As a consequence, at the true parameter values $\Sigma' = \Sigma$, $m_{\mathbf{b}}(\Sigma'; \mathcal{J})|_{\Sigma'=\Sigma} = (\Gamma_{\mathbf{b}} - \Gamma_{\mathbf{b}})_{a=1}^{|\mathbf{T}_{\mathbf{b}}|-1} = 0$. Define $m(\Sigma') = m(\Sigma'; \mathcal{J})$ as a function of $\Sigma' = (\Sigma'_F, \Sigma'_g) \in \Theta_{\Sigma} =$

¹⁰For notational simplicity, we suppress the dependence of the moment conditions from the market-bundle specific observables $(x_{t\mathbf{b}}, p_{t\mathbf{b}})_{t=1}^T$.

$\Theta_{\Sigma_F} \times \Theta_{\Sigma_g}$ that stacks together the above moment conditions for all the bundles with $|\mathbf{T}_b| \geq 2$: $m(\Sigma') = (m_b(\Sigma'; \mathcal{J}))_{b \in \mathcal{C}_2, |\mathbf{T}_b| \geq 2}$. We then have $m(\Sigma')|_{\Sigma'=\Sigma} = 0$, which consists of $\sum_{b \in \mathcal{C}_2, |\mathbf{T}_b| \geq 2} (|\mathbf{T}_b| - 1)$ moment conditions with $P + D = \dim(\Sigma')$ unknowns.

In what follows, inspired by [Rothenberg \(1971\)](#), we show that full column rank condition $\text{rank}\left(\frac{\partial m(\Sigma')}{\partial \Sigma'}\right)|_{\Sigma'=\Sigma} = \dim(\Sigma) = P + D$ is necessary and sufficient for identification among the *rank regular* $\Sigma \in \Theta_\Sigma$.^{11,12} Rank regularity is a broader concept than full column rank: if $\frac{\partial m(\Sigma')}{\partial \Sigma'}|_{\Sigma'=\Sigma}$ is of full column rank, then Σ is rank regular.¹³

Theorem 1. Local Identification: *Suppose Assumptions 1 and 2 hold, and $\Sigma \in \Theta_\Sigma$ is rank regular for $\frac{\partial m(\Sigma')}{\partial \Sigma'}$. Then, model (2.3) is locally identified if and only if $\frac{\partial m(\Sigma')}{\partial \Sigma'}|_{\Sigma'=\Sigma}$ is of full column rank.*

Proof. See Appendix A.5 □

Theorem 1 establishes the link between the number of markets with overlapping choice sets and the local identification of model (2.3). Note that, if the number of markets with bundle \mathbf{b} available in the choice set increases, so that $|\mathbf{T}_b|$ becomes larger, then the number of moment conditions in (1.9) increases. In this sense, Theorem 1 formalizes the intuition that having data on additional markets with overlapping choice sets, or analogously on larger overlapping choice sets for certain markets, will help identification. Specifically, suppose that Σ is rank regular and that its dimension, $P + D$, is greater than the number of moment conditions, $\sum_{b \in \mathcal{C}_2, |\mathbf{T}_b| \geq 2} (|\mathbf{T}_b| - 1)$. Then, the rank of $\frac{\partial m(\Sigma')}{\partial \Sigma'}|_{\Sigma'=\Sigma}$ cannot exceed the number of its rows, $\sum_{b \in \mathcal{C}_2, |\mathbf{T}_b| \geq 2} (|\mathbf{T}_b| - 1)$, which in turn is smaller than the number of its columns, $P + D$. As a consequence, $\frac{\partial m(\Sigma')}{\partial \Sigma'}|_{\Sigma'=\Sigma}$ is not of full column rank and model (2.3) is not identified.

While theoretically useful, the concept of rank regularity is abstract and not easily verifiable. The next Corollary shows that whenever the dimension of Σ is larger than the number of moment conditions and the Jacobian matrix $\frac{\partial m(\Sigma')}{\partial \Sigma'}|_{\Sigma'=\Sigma}$ is of full *row* rank, then Σ is rank regular and model (2.3) is *not* identified.¹⁴

Corollary 1. Lack of Local Identification: *Suppose Assumptions 1 and 2 hold, and the number of moment conditions, $\sum_{b \in \mathcal{C}_2, |\mathbf{T}_b| \geq 2} (|\mathbf{T}_b| - 1)$ is strictly smaller than the dimension of Σ , $P + D$. Then, if the Jacobian matrix $\frac{\partial m(\Sigma')}{\partial \Sigma'}|_{\Sigma'=\Sigma}$ is of full row rank, model (2.3) is not locally identified.*

Proof. See Appendix A.6. □

¹¹ $\Sigma \in \Theta_\Sigma$ is rank regular for the continuously differentiable function $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ if there exists a neighbourhood U of Σ such that $\text{rank}\left(\frac{\partial m(\Sigma')}{\partial \Sigma'}\right) = \text{rank}\left(\frac{\partial m(\Sigma')}{\partial \Sigma'}\right)|_{\Sigma'=\Sigma}$ for each $\Sigma' \in U$.

¹²[Rothenberg \(1971\)](#) shows the usefulness of the concept of rank regularity for local identification in non-linear models. Our Theorem 1 adapts [Rothenberg \(1971\)](#)'s Theorem 1 (p. 579) to our environment. Note that the concept of rank regularity is not vacuous in our context and there is plenty of such points: the set of rank regular points of $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is open and dense in Θ_Σ . For a proof of this property, see Appendix A.3.

¹³In fact, $\left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]^T \left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]$ has positive determinant at $\Sigma' = \Sigma$. Moreover, $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is continuously differentiable with respect to Σ' . Then, the determinant of $\left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]^T \left[\frac{\partial m(\Sigma')}{\partial \Sigma'}\right]$ is also continuous with respect to Σ' and therefore positive in a neighbourhood of $\Sigma' = \Sigma$. As a consequence, $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is of full column rank in a neighbourhood of $\Sigma' = \Sigma$ and has constant rank $P + D$ in the same neighbourhood of $\Sigma' = \Sigma$.

¹⁴Note that lack of local identification is the strongest negative result one can get: if the model is not locally identified, then for sure it will not be globally identified.

1.4 Global Identification

Up to this point, we have focused on the local uniqueness of solutions to system (1.9). Without any further restriction, the set of solutions to system (1.9) over the entire domain of parameters may not be singleton. There are at least two approaches to dealing with this global multiplicity. Partial identification, which entails the characterization of the set of global solutions to system (1.9), i.e. the identified set, and global identification, which consists in strengthening the conditions for local identification until the identified set is singleton over the entire domain of parameters. We opt for the second approach and, in what follows, discuss sufficient conditions for global identification. Our choice is motivated by estimation convenience: as detailed in section 1.5, our global identification conditions imply a convenient MLE.

Denote by $m(\Sigma'; \mathbf{T}_0)$ moment conditions (1.9) constructed from the subset of markets $\mathbf{T}_0 \subsetneq \mathbf{T}$ and evaluated at $\Sigma' = (\Sigma'_F, \Sigma'_g)$. The starting point of our global identification argument is to restrict the number of solutions to system (1.9) with the following testable Assumption:

Assumption 3. *There exists $\mathbf{T}_0 \subsetneq \mathbf{T}$ such that $m(\Sigma'; \mathbf{T}_0) = 0$ has a finite number of solutions in Θ_Σ .*

Assumption 3 alleviates the severity of the identification problem to a finite collection of observationally equivalent candidates. All of our global identification results go through also in the more general case of a countable collection of solutions. However, for notational simplicity, we limit our exposition to the finite case. Assumption 3 implies local identification, but is weaker than the typical rank conditions used to achieve global identification. For example, the classic results by [Rothenberg \(1971\)](#) require the Jacobian of the gradient of the log-likelihood function to be non-singular everywhere, so that the log-likelihood function is strictly concave ([Bowden, 1973](#)). While strict concavity is guaranteed by logit and probit models ([Amemiya \(1985\)](#) pp. 273-274), it is not by mixed logit models. Coherently with mixed logit model (2.3), Assumption 3 does not impose strict concavity of the log-likelihood function.

Remark 2. *While Assumption 3 is high-level, in Appendix A.7 we present a verifiable sufficient condition that implies it (Proposition 4): $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}$ is of full column rank when evaluated at any of the solutions to $m(\Sigma'; \mathbf{T}_0) = 0$. Building on the partial identification methods by [Chernozhukov et al. \(2007\)](#) and by [Romano and Shaikh \(2012\)](#), in Appendix A.7 we also propose a testing procedure to verify this sufficient condition in practice. A rejection of the test is evidence in support of Assumption 3.*

Assumption 3 is not new to the global identification literature and is also used, for example, by [Komunjer \(2012\)](#). To obtain global identification, [Komunjer \(2012\)](#) additionally requires the moment function to have non-negative Jacobian and to be proper.¹⁵ We avoid these further restrictions by relying on the following real analytic property of the mixed logit model.¹⁶

Theorem 2. Real Analytic Property: *For any F , $s_t(\delta_t; F)$ is real analytic with respect to δ_t in $\mathbb{R}^{C_{t1}}$, for $t = 1, \dots, T$.*

Proof. See Appendix A.8. □

¹⁵A function $f : X \rightarrow Y$ between two topological spaces is *proper* if the preimage of every compact set in Y is compact in X .

¹⁶A function $f : \mathcal{X} \rightarrow \mathbb{R}$ is *real analytic* in \mathcal{X} if for each $x_0 \in \mathcal{X}$, there exists a neighbourhood U of x_0 such that $f(x)$ is equal to its Taylor expansion $\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n$ for any $x \in U$.

Theorem 2 shows the market share function of the mixed logit model to be real analytic with respect to the average utilities given any distribution of random coefficients (parametric or non-parametric). Fox et al. (2012) and il Kim (2014) also exploit the real analytic properties of logit models to achieve global identification, but in more restrictive frameworks. il Kim (2014) shows the real analyticity of multinomial logit and nested logit models (section IV), while Fox et al. (2012) show it for mixed logit models with random coefficients defined over compact supports (Lemma 5 and section 6)—ruling out, e.g., normal and log-normal distributions.

While our local identification results do not rely on the nature of the variation in $\delta_{t\mathbf{J}_t}$, our global identification depends on whether the variation in $\delta_{t\mathbf{J}_t}$ is exogenous across markets: price endogeneity restricts this variation and leads to additional difficulties. To overcome these difficulties, we propose the use of mild restrictions on the price-setting model. In what follows, we treat separately the simpler case of exogenous variation in $\delta_{t\mathbf{J}_t}$, and that of price endogeneity.

1.4.1 Exogenous Average Utilities

Here we consider the case of exogenous variation in $\delta_{t\mathbf{J}_t}$ across markets. Given Assumption 3, denote the finite set of solutions to $m(\Sigma'; \mathbf{T}_0) = 0$ in Θ_Σ by $S = \{\Sigma^r : r = 0, \dots, R\}$, where $\Sigma^0 = (\Sigma_F^0, \Sigma_g^0)$ represents the true value $\Sigma = (\Sigma_F, \Sigma_g)$. On the basis of Lemma 1, define the corresponding Γ^r for $r = 0, 1, \dots, R$. The real analytic property of $s_t(\delta'_t; \Sigma'_F)$ allows to eliminate the extra solutions Σ^r , $r = 1, \dots, R$, by exploiting the additional variation provided by $\delta_{t\mathbf{J}_t}$ for $t \in \mathbf{T} \setminus \mathbf{T}_0$. Intuitively, the real analytic property guarantees that S is non-singleton, i.e. lack of identification, only on a union of R zero measure sets of $\delta_{t\mathbf{J}_t}$, $t \in \mathbf{T} \setminus \mathbf{T}_0$. Because the union of any finite number of zero measure sets has still zero measure, the real analytic property—combined with Assumption 3—ensures global identification almost everywhere given the additional variation provided by $\delta_{t\mathbf{J}_t}$, $t \in \mathbf{T} \setminus \mathbf{T}_0$.

Define the set of matrices $\mathbf{M} = \{M_t : t = 1, \dots, T\}$, where each M_t is a matrix of dimension $C_{t2} \times C_{t1}$. Remember that C_{t2} is the number of bundles and C_{t1} the number of inside options (i.e., bundles plus single products). M_t is made of two sub-matrices: $M_t = [M_t^1, M_t^2]$. M_t^1 is a matrix of -1 's and 0 's of dimension $C_{t2} \times J_t$, where the columns represent single products and the rows bundles. Each row of M_t^1 identifies with -1 's the product composition of the corresponding bundle. M_t^2 is instead an identity matrix \mathbf{I} of dimension $C_{t2} \times C_{t2}$, with the rows corresponding to bundles. For example, suppose the choice set (without outside option) in market t to be $\{1, 2, 3, (1, 2), (1, 3), (2, 3)\}$ and the corresponding average utility vector to be $\delta_t = (\delta_{t1}, \delta_{t2}, \delta_{t3}, \delta_{t(1,2)}, \delta_{t(1,3)}, \delta_{t(2,3)})^\top$, with $C_{t1} = 6$ and $C_{t2} = 3$. Then,

$$M_t = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 1 \end{bmatrix}.$$

Remember that $g_t(\Sigma_g) = (g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g))_{\mathbf{b} \in \mathbf{C}_{t2}}$ and that $\delta_t(\Gamma + g_t(\Sigma_g)) = (\delta_{t\mathbf{J}_t}, (\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)))_{\mathbf{b} \in \mathbf{C}_{t2}})$, where $\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)) = \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)$. For $r = 1, \dots, R$, define:

$$\Delta_r^{\text{ID}} = \{(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} : \exists t \in \mathbf{T} \setminus \mathbf{T}_0 \text{ such that } M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) \neq \Gamma^r + g_t(\Sigma_g^r)\}.$$

Denote by $\delta_{t\mathbf{J}_t}^0$ the true value of $\delta_{t\mathbf{J}_t}$ for which $s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0) = \mathcal{J}_t$ and define $\Delta^{\text{ID}} = \bigcap_{r=1, \dots, R} \Delta_r^{\text{ID}}$.

Theorem 3. Global Identification with Exogenous Prices: Suppose Assumptions 1–3 hold and Θ_Σ is compact. Then, it follows that:

- System (1.9) has a unique solution in Θ_Σ and model (2.3) is globally identified if and only if $(\delta_{t\mathbf{J}_t}^0)_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{ID}$.
- If $\Delta_r^{ID} \neq \emptyset$ for $r = 1, \dots, R$, then the Lebesgue measure of $\times_{t \in \mathbf{T} \setminus \mathbf{T}_0} \mathbb{R}^{J_t} \setminus \Delta^{ID}$ is zero.

Proof. See Appendix A.9. □

While the first result of Theorem 3 provides necessary and sufficient conditions for global identification, the second underlines their practical usefulness. The set Δ^{ID} is “very large” and will include the true $(\delta_{t\mathbf{J}_t}^0)_{t \in \mathbf{T} \setminus \mathbf{T}_0}$ in “almost all” cases: global identification will be achieved almost everywhere.

1.4.2 Endogenous Prices

We now extend the global identification results from Theorem 3 to the case of endogenous prices, where the variation in $\delta_{t\mathbf{J}_t}$ across markets is restricted by the price-setting behaviour of firms. We add mild restrictions to the price-setting model and assume the existence of exogenous cost shifters that are unobserved to the econometrician but identifiable from observed market shares and prices. [Berry and Haile \(2014\)](#) rely on a similar restriction (Assumption 7b, p. 1769) for the global identification of a simultaneous system of demand and supply by instrumental variables. However, because of the specific utility structure of model (2.3) under Assumption 2, our argument is different and does not require the instrumental variables (i.e., the cost shifters) to be observed to the econometrician but only to be identifiable.

Similar to BLP, we specify the average utility δ_{tj} as additively separable in a systematic component and an unobserved residual: $\delta_{tj} = \Delta_{tj}(x_{tj}, p_{tj}) + \xi_{tj}$, with x_{tj} a vector of observed exogenous characteristics, p_{tj} the observed endogenous price, $\Delta_{tj}(\cdot, \cdot)$ any arbitrary function of (x_{tj}, p_{tj}) (potentially different across markets and products), and ξ_{tj} a residual unobserved to the econometrician. Even though we rely on the additive separability of $\Delta_{tj}(x_{tj}, p_{tj})$ and ξ_{tj} , the target of our identification is still their sum δ_{tj} . Endogeneity arises whenever firms choose prices (also) on the basis of the market-specific residuals, which we denote by $\xi_{t\mathbf{J}_t} = (\xi_{tj})_{j \in \mathbf{J}_t} \in \mathbb{R}^{J_t}$. Because we essentially treat each δ_{tj} as a fixed effect, price endogeneity complicates global identification to the extent that it constrains the variation of $\delta_{t\mathbf{J}_t}$ across markets (the key identifying variation used in Theorem 3). As an extreme example, suppose that prices are chosen so that $\Delta_{tj}(x_{tj}, p_{tj}) = -\xi_{tj}$, then $\delta_{tj} = 0$ for every t and j . This rules out *any* variability in $\delta_{t\mathbf{J}_t}$, introducing the need for alternative sources of identification. To simplify exposition, in what follows we sometimes drop the dependence on $\Delta_{tj}(x_{tj}, p_{tj})$ from our notation.

Here we discuss the case of pure components pricing, where each firm chooses the prices of the individual products it owns and the price of any bundle is given by the sum of the prices of its components. With pure components pricing, the econometrician observes the prices of the individual products $p_{t\mathbf{J}_t} = (p_{tj})_{j \in \mathbf{J}_t}$, while the price surcharges/discounts for the joint purchase of products in bundles are all constrained to zero, so that $p_{t\mathbf{b}} = 0$ and $g_{t\mathbf{b}}(\Sigma'_g) = g_{t\mathbf{b}}(x_{t\mathbf{b}}, 0; \Sigma'_g)$ for $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_{t2}$. Our arguments can be readily modified to accommodate alternative pricing strategies such

as mixed bundling (see [Armstrong \(2016a\)](#) for a survey on non-linear pricing).^{17,18}

Denote by $c_{t\mathbf{J}_t} = (c_{tj})_{j \in \mathbf{J}_t} \in \mathbb{R}_+^{J_t}$ a vector of cost shifters, one for each of the products in market t . These cost shifters could for example be the marginal costs of the products sold in market t . Similar to $\xi_{t\mathbf{J}_t}$, also the cost shifters $c_{t\mathbf{J}_t}$ are assumed to be unobserved to the econometrician. In this sense, cost shifters can be seen as “unobserved” instruments: their existence provides exogenous identifying variation, but they do not need to be observed to the econometrician. As for the case of exogenous average utilities, we propose a characterization of the set of unobservables $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ that suffices for the global identification of $(\delta_{t\mathbf{J}_t}, \Gamma, \Sigma)$.

Let $D_{t\xi} \times D_{tc}$ denote the domain of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ for $t \in \mathbf{T}$. Suppose that the firms in market t choose prices according to pure components given the true (Γ^0, Σ^0) and $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \in D_{t\xi} \times D_{tc}$. Denote the set of equilibrium prices given $\xi_{t\mathbf{J}_t}$ and $c_{t\mathbf{J}_t}$ by $p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \subseteq \mathbb{R}_+^{J_t}$, given $\xi_{t\mathbf{J}_t}$ by $\mathbf{P}_t(\xi_{t\mathbf{J}_t}) = \cup_{c_{t\mathbf{J}_t} \in D_{tc}} p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$, and the grand collection of all possible equilibrium prices by $\mathbf{P}_t = \cup_{\xi_{t\mathbf{J}_t} \in D_{t\xi}} \mathbf{P}_t(\xi_{t\mathbf{J}_t})$. The vector of observed prices is an equilibrium of the price-setting model, so that $p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$.

Assumption 4.

- (*Cost Shifters at the Product-Level*) D_{tc} is open in \mathbb{R}^{J_t} for $t \in \mathbf{T}$.
- (*Identifiability of Cost Shifters*) $c_{t\mathbf{J}_t}$ is a C^1 function of $(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \in \{(\xi'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t}) : \xi'_{t\mathbf{J}_t} \in D_{t\xi}, p'_{t\mathbf{J}_t} \in \mathbf{P}_t(\xi'_{t\mathbf{J}_t})\}$: $c_{t\mathbf{J}_t} = \phi_t(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t})$.

The second part of Assumption 4 resembles Assumption 7b by [Berry and Haile \(2014\)](#) and is the key to our global identification with price endogeneity. [Berry and Haile \(2014\)](#) show that their Assumption 7b is implied by a variety of common price-setting models of oligopoly with differentiated products (Remark 1, p. 1766). Their result follows from the assumption of “connected substitutes” on the demand system (Definition 1, p. 1759): loosely speaking, this rules out any negative cross-price elasticity between any two products. In the case of pure components pricing, the relevant demand system has only J_t product-level equations (the system of product-level market shares) rather than C_{t1} bundle-level equations. While model (2.3) satisfies the connected substitutes property at the bundle-level, it may not at the product-level (i.e., products may be complements) and hence Remark 1 by [Berry and Haile \(2014\)](#) does not apply to our case.

By combining the bundle-level connected substitutes property with the specific utility structure of model (2.3) under Assumption 2, in Appendix A.10 we show that Assumption 4 is satisfied by common pure components pricing models. We show that it is consistent with any number of firms (monopoly, duopoly, or oligopoly) playing a complete information simultaneous Bertrand-Nash game with any profile of demand synergies (substitutability and/or complementarity). Importantly, Assumption 4 leaves the cardinality of $p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ unrestricted: the price-setting model is allowed to have a unique, several, or infinitely many equilibria.

Denote by $s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); p'_{t\mathbf{J}_t}, \Sigma'_F)$ the market share function in market t evaluated at prices $p'_{t\mathbf{J}_t} = (p'_{tj})_{j \in \mathbf{J}_t}$ and structural parameters $(\delta'_{t\mathbf{J}_t}, \Gamma', \Sigma')$, and remember that $\delta'_{tj} = \Delta'_{tj}(x_{tj}, p_{tj}) + \xi'_{tj}$ for

¹⁷With mixed bundling pricing, every firm chooses one price for each bundle it sells and the price of any bundle of products owned by different firms is the sum of the prices of its components. In this case, the price surcharge/discount for the joint purchase of products in bundles, $p_{t\mathbf{b}}$, may differ from zero for any t and $\mathbf{b} \in \mathbf{C}_{t2}$.

¹⁸Identification results for the case of mixed-bundling are available on request.

each $j \in \mathbf{J}_t$. Given Assumption 3, define for each $r = 1, \dots, R$:

$$\Xi_r^{\text{ID}} = \{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} : \exists t \in \mathbf{T} \setminus \mathbf{T}_0 \text{ such that } M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) \neq \Gamma^r + g_t(\Sigma_g^r) \\ \text{for any } p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})\}.$$

and $\Xi^{\text{ID}} = \bigcap_{r=1}^R \Xi_r^{\text{ID}}$. We make the following technical Assumption:

Assumption 5. *For every $r = 1, \dots, R$, there exists $t \in \mathbf{T} \setminus \mathbf{T}_0$, so that for almost every $p_{t\mathbf{J}_t} \in \mathbf{P}_t$, there exists $\xi'_{t\mathbf{J}_t}$, such that $\Gamma^r + g_t(\Sigma_g^r) \neq M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r)$.*

Even though Assumption 5 is abstract, it is implied by more concrete conditions. The following Corollary shows that, for example, by strengthening the real analytic property of mixed logit models from Theorem 2, Assumption 5 is satisfied:

Corollary 2. *Suppose that the following conditions hold:*

- For $r = 1, \dots, R$, $\Xi_r^{\text{ID}} \neq \emptyset$.
- For any $t \in \mathbf{T} \setminus \mathbf{T}_0$ and (Γ, Σ) , $s_t(\delta'_t(\Gamma + g_t(\Sigma)); p'_{t\mathbf{J}_t}, \Sigma_F)$ is real analytic with respect to $(\delta'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t})$.

Then, Assumption 5 holds.

Proof. See Appendix A.11. □

Corollary 2 tightens the real analyticity of the market share function to hold also with respect to the prices $p'_{t\mathbf{J}_t}$ (in addition to the average utilities $\delta'_{t\mathbf{J}_t}$). If price enters the indirect utility linearly (as is typical in applied work), then Corollary 2 will hold when the price coefficient is for example constant, or bounded, or when its moments increase at most exponentially.

Denote by $(\xi_{t\mathbf{J}_t}^0, c_{t\mathbf{J}_t}^0, p_{t\mathbf{J}_t}^0)$ the true value of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}, p_{t\mathbf{J}_t})$ for which $s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}^0, \Sigma_F^0) = s_t$. We now present the main identification result of the paper.

Theorem 4. Global Identification with Endogenous Prices: *Suppose Assumptions 1–4 hold and Θ_Σ is compact. Then, it follows that:*

- If $(\xi_{t\mathbf{J}_t}^0, c_{t\mathbf{J}_t}^0)_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Xi^{\text{ID}}$, system (1.9) has a unique solution in Θ_Σ and model (2.3) is globally identified.
- If Assumption 5 holds, the Lebesgue measure of $\times_{t \in \mathbf{T} \setminus \mathbf{T}_0} [D_{t\xi} \times D_{tc}] \setminus \Xi^{\text{ID}}$ is zero.

Proof. See Appendix A.12. □

As for Theorem 3, the first part of Theorem 4 provides sufficient conditions for global identification, while the second highlights that global identification will be achieved almost everywhere.

1.5 Estimation

We propose a Maximum Likelihood Estimator (MLE) to be implemented with observed bundle-level market shares subject to sampling error and robust to price endogeneity. We account for sampling error to accommodate the typical necessity of computing bundle-level market shares from a sample of household-level purchases (as in [Gentzkow \(2007\)](#), [Kwak et al. \(2015\)](#), [Grzybowski and Verboven](#)

(2016), Ruiz et al. (2017), and Ershov et al. (2018)). We consider asymptotics over the number of individuals I within each market, keeping fixed the number of markets and bundles, and demonstrate that our identification conditions imply the proposed MLE to be consistent and asymptotically normal.

Even though theoretically attractive, the standard MLE of model (2.3) is subject to a challenge of dimensionality even under Assumption 2: the number of demand parameters can still be too large to be handled numerically (Berry et al., 2014). As an example, suppose that in every market there are J products and individuals purchase bundles of size K . Without further restrictions, model (2.3) under Assumption 2 would imply J^K demand synergy parameters Γ , P parameters Σ_F for the distribution of random coefficients, D parameters Σ_g for the function g_t , and $J \times T$ average utility parameters $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T})$. The estimation of $J^K + P + D + J \times T$ parameters may be hard, especially because identification requires a large T .

We propose to reduce the dimensionality of the MLE's numerical search by means of a novel demand inversion specific to Gentzkow (2007)'s model that concentrates $(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T})$ out of the likelihood function.¹⁹ As a consequence, our proposed MLE effectively reduces the numerical search from $(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$, i.e. $J^K + P + D + J \times T$ parameters, to (Γ', Σ') , i.e. $J^K + P + D$ parameters.

Remark 3. *Our estimation discussion focuses on the case of exogenous average utilities. However, when the assumptions from the previous section hold and the model is globally identified, the estimation results presented below will also hold for the case of price endogeneity with no modification. The exogenous cost shifters that play the role of instruments in our identification arguments need to exist but do not need to be observed. The estimation of $(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ will not require the “explicit” use of instruments also in the presence of price endogeneity.*

1.5.1 Invertibility of Product-Level Market Shares

Here we propose a novel demand inverse designed to handle complementarity among products in models along the lines of Gentzkow (2007)'s. For any given value of the other parameters, we establish a one-to-one mapping between the observed product-level market shares and the market-product specific average utilities. We then illustrate how this demand inverse can be used to greatly simplify the practical implementation of the MLE of demand for bundles.

Define the observed product-level market share of product $j \in \mathbf{J}_t$ as $\mathcal{J}_{tj} = \sum_{\mathbf{b} \in \mathbf{C}_{t1}: j \in \mathbf{b}} \mathcal{J}_{t\mathbf{b}}$ and denote the vector stacking \mathcal{J}_{tj} for all products in market t by $\mathcal{J}_{t\mathbf{J}_t} = (\mathcal{J}_{tj})_{j \in \mathbf{J}_t}$. Similarly, define the product-level market share function of each product $j \in \mathbf{J}_t$ as $s_{tj}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \sum_{\mathbf{b} \in \mathbf{C}_{t1}: j \in \mathbf{b}} s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$ and denote the vector stacking $s_{tj}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma')$ for all products in market t by $s_{t\mathbf{J}_t}(\cdot; \Gamma', \Sigma') = (s_{tj}(\cdot; \Gamma', \Sigma'))_{j \in \mathbf{J}_t}$.

Theorem 5. Demand Inverse: *Suppose that Assumptions 1 and 2 hold. Then, for any $(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma$, there exists at most one $\delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathcal{J}_{t\mathbf{J}_t}$.*

Proof. See Appendix A.13. □

When (Γ', Σ') is equal to the true value (Γ, Σ) , Theorem 7 implies that the only $\delta'_{t\mathbf{J}_t}$ that satisfies $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma, \Sigma) = \mathcal{J}_{t\mathbf{J}_t}$ is the true $\delta_{t\mathbf{J}_t}$. As a result, the function $s_{t\mathbf{J}_t}(\cdot; \Gamma, \Sigma)$ is globally invertible.

¹⁹As we clarify below, our demand inverse differs from the classic one by Berry (1994) and Berry et al. (2013), which in our context corresponds to the demand inverse presented in Lemma 1.

When $(\Gamma', \Sigma') \neq (\Gamma, \Sigma)$, it is possible that there is no $\delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = s_{t\mathbf{J}_t}$.²⁰ Because the existence of some $\delta'_{t\mathbf{J}_t}$ that rationalizes the observed product-level market shares can always be verified numerically (following the procedure outlined below), in what follows we assume it and denote the bijection mapping $s_{t\mathbf{J}_t}$ to $\delta'_{t\mathbf{J}_t}$ by:

$$\begin{aligned}\delta'_{t\mathbf{J}_t} &= s_{t\mathbf{J}_t}^{-1}(s_{t\mathbf{J}_t}; \Gamma', \Sigma') \\ &= \delta_{t\mathbf{J}_t}(s_{t\mathbf{J}_t}; \Gamma', \Sigma').\end{aligned}$$

Theorem 7 differs from the classic demand inverse by [Berry \(1994\)](#) (then generalized by [Berry et al. \(2013\)](#)). In our context, [Berry \(1994\)](#) implies a bijection between the observed *bundle*-level market shares and the *full* vector of market-bundle specific average utilities. We rely on this classic demand inverse throughout the paper and, for completeness, adapt it to our framework in Lemma 1. Differently, Theorem 7 establishes a bijection between a *transformation* of the observed bundle-level market shares—the *product*-level market shares—and a *sub*-vector of the market-bundle specific average utilities—the market-*product* specific average utilities. While the invertibility of the product-level market shares on the basis of [Berry \(1994\)](#) would require the products to be substitutes, Theorem 7 applies also to the case of complementary products.

1.5.2 A Maximum Likelihood Estimator

We now allow for the possibility that observed market shares are subject to sampling error, due for example to the necessity of measuring them from household-level purchase data. Denote by $I_{t\mathbf{b}}$ the number of individuals in market t observed to choose \mathbf{b} and by $\hat{s}_{t\mathbf{b}} = \frac{I_{t\mathbf{b}}}{I_t}$ the corresponding observed market share. To simplify exposition, in what follows we drop any notational dependence from the observables and denote $g_t(\Sigma'_g) = (g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma'_g))_{\mathbf{b} \in \mathbf{C}_{t2}}$. The log-likelihood function evaluated at $(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ can be written as:

$$\ell_I(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma'; \hat{s}_1, \dots, \hat{s}_T) = \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_t} \hat{s}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F), \quad (1.10)$$

where $\hat{s}_t = (\hat{s}_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_t}$ for $t = 1, \dots, T$. Denote the domain of the parameters by $\Theta = \Theta_\delta \times \Theta_\Gamma \times \Theta_\Sigma$, where Θ_δ , Θ_Γ , and Θ_Σ are compact. Given Theorem 7, we propose the following MLE that concentrates $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T})$ out of the log-likelihood function:

$$\begin{aligned}(\hat{\Gamma}, \hat{\Sigma}) &\equiv \arg \max_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} \ell_I((\delta_{t\mathbf{J}_t}(\hat{s}_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma'; \hat{s}_1, \dots, \hat{s}_T), \\ &= \arg \max_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} \ell_I^c(\Gamma', \Sigma'; \hat{s}_1, \dots, \hat{s}_T) \\ \hat{\delta}_{t\mathbf{J}_t} &\equiv \delta_{t\mathbf{J}_t}(\hat{s}_{t\mathbf{J}_t}; \hat{\Gamma}, \hat{\Sigma}), \quad t = 1, \dots, T.\end{aligned} \quad (1.11)$$

To simplify notation, denote the true parameters $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ by $\theta = (\theta_\delta, \Gamma, \Sigma)$ and the MLE $(\hat{\theta}_\delta, \hat{\Gamma}, \hat{\Sigma})$ by $\hat{\theta}$. The next Theorem establishes the asymptotic properties of $\hat{\theta}$.

²⁰For example, a model of demand for single products (i.e., $\Gamma' = -\infty$) cannot rationalize situations in which the sum of the observed product-level market shares is larger than one. (This can happen because the same $s_{t\mathbf{b}}$ contributes to the product-level market share of any $j \in \mathbf{b}$, giving rise to “multiple counting” of $s_{t\mathbf{b}}$ when summing s_{tj} over j .) In such cases, the demand inverse is therefore not feasible when evaluated at $\Gamma' = -\infty$.

Theorem 6. MLE estimator: Suppose Assumptions 1–3 hold, the true $(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{ID}$, $\hat{s}_{t\mathbf{b}} \xrightarrow{p} s_{t\mathbf{b}}$ for $t = 1, \dots, T$, $\mathbf{b} \in \mathbf{C}_t$, and the standard regularity conditions detailed in Appendix A.14 hold. Then:

- (Consistency) $\hat{\theta} \xrightarrow{p} \theta$.
- (Asymptotic Normality) There exist matrices $W_1, W_2 > 0$ such that $\sqrt{I}(\hat{\theta}_\delta - \theta_\delta) \xrightarrow{d} \mathcal{N}(0, W_1)$ and $\sqrt{I}[(\hat{\Gamma}, \hat{\Sigma}) - (\Gamma, \Sigma)] \xrightarrow{d} \mathcal{N}(0, W_2)$.

Proof. See Appendix A.14. □

Estimator (1.11) is neither a standard MLE nor a concentrated MLE. A standard MLE would maximize (1.10) with respect to (δ, Γ, Σ) , while (1.11) only maximizes it with respect to (Γ, Σ) . Differently from a concentrated MLE, which also would maximize (1.10) only with respect to (Γ, Σ) , estimator (1.11) is however not as efficient as the standard MLE. The demand inverse from Theorem 7 only uses observed product-level market shares (rather than bundle-level), and this causes a loss of information in the process of concentrating out θ_δ from the log-likelihood function. MLE (1.11) trades-off computational ease against estimation efficiency.

Implementation. In the spirit of BLP, the demand inverse from Theorem 7 enables to break down the numerical search for $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ into two steps that can be solved sequentially while implementing (1.11):

- Step 1. For any given guess of (Γ', Σ') and market $t = 1, \dots, T$, compute $\delta'_{t\mathbf{J}_t} = \delta_{t\mathbf{J}_t}(\hat{s}_{t\mathbf{J}_t}; \Gamma', \Sigma')$ by the Newton-Raphson method as the unique solution to system $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \hat{s}_{t\mathbf{J}_t}$. To implement the Newton-Raphson method, note that the derivative $\frac{\partial s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma')}{\partial \delta'_{t\mathbf{J}_t}} = [\mathbf{I} - M_t^{1\mathbf{T}}] \frac{\partial s_{t\mathbf{C}_{t1}}}{\partial \delta'_{t\mathbf{C}_{t1}}} [\mathbf{I} - M_t^{1\mathbf{T}}]^{\mathbf{T}}$ is everywhere symmetric and positive-definite, where M_t^1 is defined prior to Theorem 3. Because the solution to the system is guaranteed to be at most unique, whenever the algorithm finds one, the numerical search can end.²¹ Given this solution, compute the derivative $\frac{\partial \delta_{t\mathbf{J}_t}}{\partial (\Gamma', \Sigma')} = - \left[\frac{\partial s_{t\mathbf{J}_t}}{\partial \delta'_{t\mathbf{J}_t}} \right]^{-1} \left[\frac{\partial s_{t\mathbf{J}_t}}{\partial (\Gamma', \Sigma')} \right]$ and move on to Step 2. In case the algorithm cannot find a solution, then Theorem 7 implies that $(\Gamma', \Sigma') \neq (\Gamma, \Sigma)$: try a new guess of (Γ', Σ') and go back to the beginning of Step 1.
- Step 2. Plug $\delta'_{t\mathbf{J}_t}$ for $t = 1, \dots, T$ from Step 1 into $\ell_I((\delta'_{t\mathbf{J}_t})_{t=1, \dots, T}, \Gamma', \Sigma'; \hat{s}_1, \dots, \hat{s}_T)$ and obtain $\ell_I^c(\Gamma', \Sigma'; \hat{s}_1, \dots, \hat{s}_T)$. Compute its derivative with respect to (Γ', Σ') , $\frac{\partial \ell_I^c}{\partial (\Gamma', \Sigma')} = \sum_{t=1}^T \frac{\partial \ell_I}{\partial \delta'_{t\mathbf{J}_t}} \frac{\partial \delta_{t\mathbf{J}_t}}{\partial (\Gamma', \Sigma')} + \frac{\partial \ell_I}{\partial (\Gamma', \Sigma')}$. Check whether the current guess of (Γ', Σ') numerically maximizes $\ell_I^c(\Gamma', \Sigma'; \hat{s}_1, \dots, \hat{s}_T)$. If yes, the current value of the parameters is $\hat{\theta}$. If not, use $\frac{\partial \ell_I^c}{\partial (\Gamma', \Sigma')}$ to search for a new guess of (Γ', Σ') and go back to Step 1.

1.6 Empirical Illustration

We illustrate our methods in the context of the ready-to-eat (RTE) cereal industry in the USA. We revisit the classic studies by Nevo (2000, 2001), and allow for Hicksian complementarity among different

²¹For a useful discussion about the Newton-Raphson method in the context of demand estimation, see Conlon and Gortmaker (2019).

brands in demand estimation. The households in our data are observed to purchase two or more different RTE cereal brands in approximately 20% of their shopping trips. In the data, we observe purchases rather than consumption. In terms of purchases, demand for bundles can arise for reasons different from synergies in consumption (as in [Gentzkow \(2007\)](#)): shopping costs (as in [Pozzi \(2012\)](#) and [Thomassen et al. \(2017\)](#)) and preference for variety (as in [Hendel \(1999\)](#) and [Dubé \(2004\)](#)) represent two likely alternatives.

Our model can rationalize shopping costs with positive demand synergies and, as we show in [Appendix A.1](#), preference for variety with some additional structure on the demand synergy parameters. We try to distinguish empirically the relative contribution of these mechanisms to the estimated demand synergies. Our results show that demand for RTE cereals exhibits substantial Hicksian complementarity and that around 75% of it does not seem to be explained by shopping costs or by preference for variety. We compare our estimation results from the full model to those from a model of demand for single brands (similar to [Nevo \(2000, 2001\)](#)) and show that ignoring Hicksian complementarity may result in misleading demand estimates and counterfactuals. Despite the different econometric approach and data, our results are in line with those by [Fosgerau et al. \(2019\)](#), who also document Hicksian complementarity among different RTE cereal brands in the USA.

1.6.1 Data and Definitions

We use household-level and store-level IRI data on ready-to-eat (RTE) cereals for the period 2008-2011 for the city of Pittsfield in the USA. We report a succinct description of the data used and refer the reader to [Bronnenberg et al. \(2008\)](#) for a more thorough discussion.

We focus on the $I = 2897$ households who are observed to purchase RTE cereals at least once from 2008 until 2011. For these households, we observe some demographics (e.g., income group and family size) and a panel of shopping trips $r = 1, \dots, 756663$ to 7 different grocery stores over a period of 208 weeks. A shopping trip is defined as a purchase occasion of a household to a grocery store in a given day. Each shopping trip records all the Universal Product Codes (UPCs) purchased by a household across all product categories sold by the store: during 83256 of these, RTE cereals are observed to be purchased. We define a market as a store-week combination $t = 1, \dots, 1431$.

Over the sample period, the households are observed to purchase 1173 different UPCs of RTE cereals. For feasibility, we reduce the number of different RTE cereal products by collecting UPCs into what we call brands. We define $J = 16$ different brands on the basis of producers and ingredients. We classify producers into six groups: General Mills, Kellogg’s, Quaker, Post, Small Producers, and Private Labels. The first four correspond to the four largest RTE cereal producers, “Small Producers” correspond to the remaining producers, and “Private Labels” correspond to the UPCs directly branded by the retailers (i.e., the stores). We collect the UPCs of each of the producers into three types on the basis of their ingredients: cereal type R refers to “Regular,” F/W to “Fiber/Whole Grain,” and S to “Added Sugar.” [Appendix Table B.1](#) lists these RTE cereal brands and their average market shares across the shopping trips with some RTE cereal purchase.²² We use the store-level data to compute brand-level prices for each brand j and store-week combination t , p_{tj} . Each p_{tj} is computed as the average price per 16oz across the UPCs belonging to brand j in store-week t .

We make the standard assumption that RTE cereal purchases do not determine store choice and take store choice as exogenous in our econometric model. We consider household i to choose the outside

²²Market shares are computed over the shopping trips observed in each store-week combination.

option, which we denote by $j = 0$, whenever no RTE cereal brand is purchased during shopping trip r (in general, something must be purchased for a shopping trip to be in the data). Around 89% of all shopping trips do not involve any purchase of RTE cereals.

During each shopping trip r , a household i is considered to purchase RTE cereal brand j whenever they are observed to purchase at least a UPC of brand j . Households are considered to purchase bundles only when purchasing at least two *different* brands of RTE cereals during the *same* shopping trip. In our view, this is a conservative measure of households' demand for bundles.²³ For computational convenience, we focus our analysis on the shopping trips with observed purchases of at most two different RTE cereal brands, thus discarding 3.27% of the shopping trips with some RTE cereal purchase. In 17.69% of the shopping trips with some RTE cereal purchase, households are observed to purchase two different brands of RTE cereals.

Table 1.1 describes how the average bundle size purchased changes among households with different observable characteristics. The top panel of Table 1.1 shows that larger families are more likely to purchase larger bundles. This accords to the idea of preference for variety by [Hendel \(1999\)](#) and [Dubé \(2004\)](#): in order to satisfy more heterogeneous preferences (e.g., different genders and ages), larger households are more likely to purchase a wider variety of RTE cereal brands on each shopping trip. The central panel of Table 1.1 highlights the potential relevance of shopping costs, as suggested by [Pozzi \(2012\)](#) and [Thomassen et al. \(2017\)](#): households observed to shop with a higher frequency (facing lower shopping costs) are less likely to purchase bundles of different RTE cereal brands on any shopping trip.²⁴ The bottom panel of Table 1.1 divides the households into three income groups and does not suggest apparent relationships between income and average purchased bundle size.²⁵

We construct choice sets at the level of the store-week t : any household during any shopping trip in t is assumed to face choice set \mathbf{C}_t . This is made of three components: single brands, bundles of size 2, and the outside option. From the store-level data, we observe which of the 16 brands of RTE cereals are available in each store-week t . Denote this set of available brands by \mathbf{J}_t . Households can also purchase bundles $(j_1, j_2) \in (\mathbf{J}_t \times \mathbf{J}_t) \setminus \{(k_1, k_2) | k_1 = k_2\}$ made of pairs of different RTE cereal brands. Finally, households may decide not to purchase any RTE cereal brand, $j = 0$. By combining these purchase possibilities, the choice set faced during all shopping trips in t is $\mathbf{C}_t = \{0\} \cup \mathbf{J}_t \cup (\mathbf{J}_t \times \mathbf{J}_t) \setminus \{(k_1, k_2) | k_1 = k_2\}$.²⁶

²³For instance, the purchases of different RTE cereal brands across different shopping trips within the same t are considered as independent purchases of single brands rather than bundles. To keep the dimensionality of the problem manageable, we do not count as bundles the purchases of multiple units of the same brand within the same shopping trip. Accommodating either less conservative definitions of bundles or purchases of multiple units of the same brand would not represent any conceptual challenge for the proposed methods.

²⁴We compute the “weekly shopping frequency” as the average number of shopping trips per week for each household over the entire four-year period of our sample. The median among the 2897 households is 1.80 shopping trips per week.

²⁵We create the three income groups on the basis of 12 income classes originally provided in the IRI data, which are ordered in increasing level of income from 1 to 12. We code as “low income” the classes 1-4, “medium income” the classes 5-8, and we group in “high income” the remaining classes 9-12.

²⁶The choice set \mathbf{C}_t also excludes those bundles that are never purchased during any of the shopping trips in t . Even though all brands in \mathbf{J}_t have positive market shares by construction, some combination of brands (j_1, j_2) from $(\mathbf{J}_t \times \mathbf{J}_t) \setminus \{(k_1, k_2) | k_1 = k_2\}$ may not be observed to be jointly purchased.

Table 1.1: Average Number of Different Brands per Shopping Trip

	#Households	Ave. Bundle Size
Family Size		
1	732	1.12
2	1184	1.16
≥ 3	981	1.22
Weekly Shopping Frequency		
(0, 2]	1779	1.19
(2, 3]	810	1.17
> 3	308	1.14
Income Group		
<i>low</i>	679	1.18
<i>medium</i>	1169	1.16
<i>high</i>	1049	1.19

Notes: The Table shows the distribution of family size, weekly shopping frequency, and income group among the 2897 households in our data. See text for the definitions of these variables. For each value of these variables, we report the average number of different RTE cereal brands observed to be purchased per shopping trip by the corresponding households.

1.6.2 Model Specification

Any household i is observed going on several shopping trips, each taking place in a specific store-week combination t (our definition of market). The indirect utility of household i by purchasing brand $j \in \mathbf{J}_t$ during shopping trip r in market t is:

$$\begin{aligned} U_{irtj} &= u_{itj} + \varepsilon_{irtj} \\ &= \delta_{tj} + \mu_{itj} + \varepsilon_{irtj}, \end{aligned} \tag{1.12}$$

$$\mu_{itj} = -p_{tj} \exp(d_i^\alpha \alpha + v_i) + \eta_{ij}$$

where $u_{itj} = \delta_{tj} + \mu_{itj}$, δ_{tj} is market t -specific average utility for RTE cereal brand $j \in \mathbf{J}_t$, μ_{itj} is a household i -specific utility deviation from δ_{tj} , and ε_{irtj} is an idiosyncratic error term. p_{tj} is the price of brand j in store-week combination t , and $d_i^\alpha \alpha + v_i$ is a vector of household i -specific price coefficients made of two components: an observable part that is a function of the household characteristics d_i^α (to be detailed in the next section) and an unobserved random component v_i . η_{ij} is an unobserved household i -specific preference for brand j , which is constant across i 's shopping trips and potentially correlated across brands.

Specification (1.12) encapsulates the entire effect of price p_{tj} in the household i -specific μ_{itj} . In terms of the notation used in section 1.4.2, this implies $\Delta_{tj}(p_{tj}, x_{tj}) = 0$ and $\delta_{tj} = \xi_{tj}$. Even though we use household-level data, we face price endogeneity if, for instance, the producer of RTE cereal brand j sets price p_{tj} taking the average utility δ_{tj} into consideration. Our proposed estimator essentially addresses this endogeneity problem by treating the average utility δ_{tj} for each brand j in each market t as a fixed effect.

The indirect utility of i by purchasing bundle \mathbf{b} during shopping trip r in market t is:

$$\begin{aligned}
U_{irt\mathbf{b}} &= \sum_{j \in \mathbf{b}} u_{itj} + \Gamma_{i\mathbf{b}} + \varepsilon_{irt\mathbf{b}} \\
&= \sum_{j \in \mathbf{b}} (\delta_{tj} + \mu_{itj}) + \Gamma_{\mathbf{b}} + \zeta_{i\mathbf{b}} + \varepsilon_{irt\mathbf{b}} \\
&= \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}} + \left[\sum_{j \in \mathbf{b}} \mu_{itj} + (d_i^\gamma \gamma + \tilde{\zeta}_{i\mathbf{b}}) \right] + \varepsilon_{irt\mathbf{b}} \\
&= \delta_{t\mathbf{b}} + \mu_{it\mathbf{b}} + \varepsilon_{irt\mathbf{b}},
\end{aligned} \tag{1.13}$$

where $\delta_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{\mathbf{b}}$ is market t -specific average utility for bundle \mathbf{b} , $\mu_{it\mathbf{b}}$ is household i -specific utility deviation from $\delta_{t\mathbf{b}}$, $\Gamma_{i\mathbf{b}}$ is household i -specific demand synergy among the brands in bundle \mathbf{b} , and $\varepsilon_{irt\mathbf{b}}$ is an idiosyncratic error term. The demand synergy parameter $\Gamma_{i\mathbf{b}} = \Gamma_{\mathbf{b}} + \zeta_{i\mathbf{b}}$ captures the extra utility household i obtains from buying the RTE cereal brands in bundle \mathbf{b} jointly rather than separately. It is the sum of $\Gamma_{\mathbf{b}}$, common to all households, and of $\zeta_{i\mathbf{b}} = d_i^\gamma \gamma + \tilde{\zeta}_{i\mathbf{b}}$, where $d_i^\gamma \gamma$ is a function of observed household characteristics d_i^γ (to be detailed in the next section) and $\tilde{\zeta}_{i\mathbf{b}}$ is an unobserved random component. Because of pure components pricing, i.e. $p_{t\mathbf{b}} = 0$, and the absence of other bundle-specific observed product characteristics, i.e. $x_{t\mathbf{b}} = 0$, we constrain function $g(\cdot, \cdot | \Sigma_g) = 0$ (see Assumption 2).²⁷ We attempt to empirically distinguish the relative contribution to $\Gamma_{i\mathbf{b}}$ of two alternative mechanisms. In particular, we specify $d_i^\gamma \gamma$ to include measures of family size (to proxy for preference for variety) and of average weekly shopping frequency (to proxy for shopping costs).

Finally, the indirect utility of household i by choosing the outside option during shopping trip r in market t is assumed to be:

$$U_{irt0} = \varepsilon_{irt0}. \tag{1.14}$$

Suppose that ε_{irt0} and the $\varepsilon_{irt\mathbf{b}}$'s are i.i.d. Gumbel. Express $\mu_{it\mathbf{b}} = \mu_{it\mathbf{b}}(\beta_i)$ as a function of the unobservable $\beta_i = (v_i, \eta_i, \tilde{\zeta}_i) = (v_i, (\eta_{ij})_{j \in \mathbf{J}}, (\tilde{\zeta}_{i\mathbf{b}})_{\mathbf{b} \in \mathbf{B}})$.²⁸ Then, given β_i and $(\delta_{t\mathbf{J}_t}, \alpha, \gamma, \Gamma) = ((\delta_{tj})_{j \in \mathbf{J}_t}, \alpha, \gamma, (\Gamma_{\mathbf{b}})_{\mathbf{b} \in \mathbf{B}})$, household i 's purchase probability of $\mathbf{b} \in \mathbf{C}_t$ during shopping trip r in market t is:

$$s_{irt\mathbf{b}}(\delta_{t\mathbf{J}_t}, \alpha, \gamma, \Gamma; \beta_i) = \frac{e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_i)}}{\sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_i)}}. \tag{1.15}$$

We assume $\beta_i = (v_i, \eta_i, \tilde{\zeta}_i)$ to be normally distributed and denote its c.d.f. by $\Phi(\cdot; \Sigma_F)$. Let $y_{itr\mathbf{b}} \in \{0, 1\}$ be an indicator for whether household i purchased \mathbf{b} during shopping trip r in market t , with $\sum_{\mathbf{b} \in \mathbf{C}_t} y_{itr\mathbf{b}} = 1$. Let T_i denote the set of markets for which we observe shopping trips by household i . For each $t \in T_i$, define R_{it} as the set of shopping trips by household i that took place in market t . By integrating over the distribution of β_i , we obtain the likelihood of i 's observed purchases $y_i = (y_{itr\mathbf{b}})_{t \in T_i, r \in R_{it}, \mathbf{b} \in \mathbf{C}_t}$:

$$L_i(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma_F; y_i) = \int \prod_{t \in T_i} \prod_{r \in R_{it}} \prod_{\mathbf{b} \in \mathbf{C}_t} (s_{irt\mathbf{b}}(\delta_{t\mathbf{J}_t}, \alpha, \gamma, \Gamma; \beta_i))^{y_{itr\mathbf{b}}} d\Phi(\beta_i; \Sigma_F). \tag{1.16}$$

²⁷We follow [Nevo \(2000, 2001\)](#) in assuming that RTE cereal producers set prices at the brand-level rather than at the bundle-level (i.e., pure components pricing): households purchasing multiple RTE cereal brands during the same shopping trip are assumed to pay the sum of the prices of the single brands.

²⁸ \mathbf{J} and \mathbf{B} are defined as, respectively, the union of all \mathbf{J}_t and of all $(\mathbf{J}_t \times \mathbf{J}_t) \setminus \{(k_1, k_2) | k_1 = k_2\}$ for $t = 1, \dots, T$.

By aggregating over the $I = 2897$ households, the likelihood function for the entire set of observed purchases is:

$$L_I(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma_F; y_1, \dots, y_I) = \prod_{i=1}^{2897} L_i(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma_F, y_i). \quad (1.17)$$

We estimate the demand parameters $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \alpha, \gamma, \Gamma, \Sigma_F)$ on the basis of MLE (1.11) derived from likelihood function (1.17).²⁹ To get a sense of the practical advantages implied by our novel demand inverse, in the current application the proposed MLE reduces the numerical search (with respect to a standard MLE) from 12351 to 133 parameters, i.e. $(\alpha, \gamma, \Gamma, \Sigma_F)$.

Restricted Models. In what follows, we refer to the model specified in (1.15)-(1.17) as to the *full model* of demand for bundles. To better evaluate the practical relevance of allowing for complementarity, we also estimate two restricted versions of the full model.

In the first restricted model, we constrain $\Gamma_{i\mathbf{b}} = 0$ in estimation for all i 's and \mathbf{b} 's. A comparison of the full model with this first restricted model highlights the importance of controlling for the demand synergies $\Gamma_{i\mathbf{b}}$ while keeping everything else unchanged. Despite the absence of demand synergies, this restricted model can still give rise Hicksian complementarity.³⁰

In the second restricted model, we rule out Hicksian complementarity by constraining $\Gamma_{i\mathbf{b}} = -\infty$ in estimation for all i 's and \mathbf{b} 's. In other words, choice sets are restricted not to include any bundle. This amounts to specifying a standard model of demand for single RTE cereal brands (along the lines of Nevo (2000, 2001)) with choice set $\mathbf{C}_t = \{0\} \cup \mathbf{J}_t$ in each t . This second restricted model is estimated on the basis of the same purchase observations as the other two models. However, the observations are used differently: the second restricted model does not differentiate between simultaneous (during the same shopping trip) and sequential (during different shopping trips) purchases of different brands.

1.6.3 Estimates of Demand for RTE Cereal Bundles

In this section, we present our estimation results for the full model. We postpone a comparison of the estimation results from the three different models to the next two sections, where we discuss price elasticities and counterfactual simulations.

We capture observed heterogeneity in price sensitivity $-\exp(d_i^\alpha \alpha + v_i)$ by d_i^α , a vector of nine mutually exclusive dummies indicating household i 's income group (low, medium, and high) and family size (one, two, and larger than two). Unobserved heterogeneity in price sensitivity is instead captured by the random coefficient v_i , which we assume to be i.i.d. normal with standard deviation σ_v . For each of the three estimated models, the top panel of Table 2.2 reports estimates of the average price sensitivity $\mathbb{E}[-\exp(d_i^\alpha \alpha + v_i) | d_i^\alpha]$ for each value of d_i^α . These results do not seem to suggest any systematic heterogeneity in price sensitivity among households with different incomes and family sizes.

We specify the RTE cereal brand-specific random intercepts as $\eta_{ij} = \eta_{i,\text{type}} + \tilde{\eta}_{ij}$, where $\eta_{i,\text{type}}$ captures household i 's unobserved and correlated preferences across cereal types $\{R, F/W, S\}$ and $\tilde{\eta}_{ij}$

²⁹Even though (1.17) is expressed in terms of individual purchases (y_1, \dots, y_I) rather than of sampled market shares $(\hat{y}_1, \dots, \hat{y}_T)$, it can be easily shown that the corresponding MLE satisfies the conditions of Theorem 6.

³⁰In this first restricted model, the cross-price elasticities can still be negative because the choice set $\mathbf{C}_t = \{0\} \cup \mathbf{J}_t \cup (\mathbf{J}_t \times \mathbf{J}_t) \setminus \{(k_1, k_2) | k_1 = k_2\}$ is not *complete*. \mathbf{C}_t would be complete if it included also the bundles made of two units of the same brand. Gentzkow (2007)'s Proposition 1 at page 719, which states that a positive demand synergy is necessary and sufficient for Hicksian complementarity, only applies to models with complete choice sets.

captures i.i.d. unobserved preference for brand j . Remember that cereal type R refers to “Regular,” F/W to “Fiber/Whole Grain,” and S to “Added Sugar.” Note that any two brands with similar ingredients will share the same $\eta_{i,\text{type}}$. We assume $\tilde{\eta}_{ij}$ to be distributed normal with standard deviation $\sigma_{\tilde{\eta}}$. Following [Gentzkow \(2007\)](#), we also assume $(\eta_{i,R}, \eta_{i,F/W}, \eta_{i,S})$ to be distributed jointly normal with standard deviations and pairwise correlations denoted by, respectively, $\sigma_{\eta_{\text{type}}}$ and $\text{corr}_{\text{type}, \text{type}'}$, $\text{type}, \text{type}' \in \{R, F/W, S\}$. We allow single-person households and multi-person households to have different joint normal distributions of $(\eta_{i,R}, \eta_{i,F/W}, \eta_{i,S})$. We assume $\tilde{\eta}_{ij}$ and $\eta_{i,\text{type}}$ to be mutually independent.

The estimates of the distribution of η_{ij} and of the other random coefficients are reported in the central panel of [Table 2.2](#). Overall, the estimates are highly significant and underline the importance of controlling for unobserved heterogeneity, not only in terms of price sensitivity, but also of brand-specific random intercepts and of demand synergies (we return to these in more detail below). Households’ unobserved preferences for healthier F/W and children S cereal brands are positively correlated, while unobserved preferences for regular R cereal brands seem to correlate negatively with both F/W and S cereal brands. Households of different family sizes do not seem to have systematically different distributions of $(\eta_{i,R}, \eta_{i,F/W}, \eta_{i,S})$.

We specify the demand synergy of household i for bundle \mathbf{b} as:

$$\begin{aligned} \Gamma_{i\mathbf{b}} &= \Gamma_{\mathbf{b}} + d_i^{\gamma} \gamma + \tilde{\zeta}_{i\mathbf{b}} \\ &= \Gamma_{\mathbf{b}} + \gamma_2 \mathbf{1}\{\text{family size}_i = 2\} + \gamma_{\geq 3} \mathbf{1}\{\text{family size}_i \geq 3\} \\ &\quad + \gamma_s \mathbf{1}\{\text{normal shopping frequency}_i\} + \tilde{\zeta}_{i\mathbf{b}}, \end{aligned} \tag{1.18}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and “normal shopping frequency $_i$ ” denotes whether the average weekly shopping frequency of household i lies below the 95th percentile.³¹ Parameter γ_k captures systematic differences between the average demand synergies of households of family size k and single-person households. We include family size in the specification of $\Gamma_{i\mathbf{b}}$ as a proxy for preference for variety. γ_s instead measures differences in the average demand synergies between households observed to shop at a normal frequency, i.e. in the bottom 95% of the distribution, and households who shop very often, i.e. in the top 5% of the distribution. We control for *normal* shopping frequency in (1.18) as a proxy for *larger* shopping costs. These are meant to rationalize the purchase patterns documented in [Table 1.1](#): larger families may have to satisfy more heterogeneous preferences within the household, while more frequent shoppers may be less likely to purchase multiple brands on any shopping trip. $\tilde{\zeta}_{i\mathbf{b}}$ represents a i -specific unobserved component of demand synergy for bundle \mathbf{b} , which we assume to be i.i.d. normal with standard deviation $\sigma_{\tilde{\zeta}}$. $\tilde{\zeta}_{i\mathbf{b}}$ allows for the possibility that the brands in bundle \mathbf{b} have positive demand synergies for some households and negative for others. Estimates of the $\Gamma_{\mathbf{b}}$ ’s are reported in [Table 1.3](#), while estimates of the remaining demand synergy parameters are reported at the bottom of [Table 2.2](#).

The estimates from [Table 1.3](#) suggest that several pairs of RTE cereal brands have positive $\Gamma_{\mathbf{b}}$ ’s.³² Importantly for competition policy, as we will explore in the next section, there appear to be positive $\Gamma_{\mathbf{b}}$ ’s not only among brands within the same producer, but also among brands sold by different producers. For example, the first column of [Table 1.3](#) shows that single-person households exhibit

³¹In our sample of households, the 95th percentile of the average weekly shopping frequency (i.e., the average number of shopping trips in a week) is 3.67.

³²Note that these estimates come from the full model, neither of the restricted models allows for demand synergies. See the Table notes for an interpretation of the missing values.

positive demand synergies between General Mills and most of the 15 remaining brands. Moreover, in line with the evidence from Table 1.1, the estimated demand synergy shifters γ_2 and $\gamma_{\geq 3}$ from the bottom panel of Table 2.2 are positive and increasing in family size, i.e. $\gamma_{\geq 3} \geq \gamma_2$. We interpret this as evidence of preference for variety: larger families exhibit more positive demand synergies among different RTE cereal brands than smaller families. Intuitively, larger families may be more likely to purchase different brands in order to satisfy more heterogeneous RTE cereal tastes within the household (e.g., adults and children of different ages). Differently, γ_s is positive but not significantly different from zero, highlighting that—after controlling for everything else—households with different shopping frequencies are similarly likely to purchase bundles on any shopping trip. The standard deviation $\sigma_{\tilde{\zeta}}$ of the random coefficient $\tilde{\zeta}_{i\mathbf{b}}$ is estimated to be small but significant, suggesting the presence of household-specific heterogeneity in demand synergies beyond differences in family size and weekly shopping frequency.

Table 1.2: Demand Estimates for Full and Restricted Models

	Full Model	Restricted Model 1 $\Gamma_{i\mathbf{b}} = 0$	Restricted Model 2 $\Gamma_{i\mathbf{b}} = -\infty$
<i>Average Price Sensitivities</i>			
low income, family size = 1	-0.44 (0.164)	-0.49 (0.196)	-1.27 (0.180)
family size = 2	-0.47 (0.175)	-0.51 (0.202)	-1.30 (0.185)
family size ≥ 3	-0.39 (0.147)	-0.38 (0.151)	-1.17 (0.167)
medium income, family size = 1	-0.47 (0.174)	-0.52 (0.210)	-1.29 (0.183)
family size = 2	-0.42 (0.157)	-0.46 (0.183)	-1.27 (0.180)
family size ≥ 3	-0.49 (0.183)	-0.49 (0.195)	-1.31 (0.185)
high income, family size = 1	-0.39 (0.146)	-0.43 (0.173)	-1.20 (0.170)
family size = 2	-0.42 (0.157)	-0.45 (0.180)	-1.26 (0.179)
family size ≥ 3	-0.40 (0.150)	-0.41 (0.165)	-1.25 (0.177)
<i>Random Coefficients</i>			
price, σ_v	0.36 (0.004)	0.39 (0.004)	0.14 (0.002)
brand intercepts, $\sigma_{\bar{\eta}}$	0.87 (0.005)	0.87 (0.005)	0.81 (0.005)
demand synergies, σ_{ζ}	0.06 (0.010)		
Single-Person Households			
σ_{η_R}	0.50 (0.037)	0.47 (0.046)	0.20 (0.037)
$\sigma_{\eta_{F/W}}$	0.54 (0.014)	0.52 (0.024)	0.65 (0.017)
σ_{η_S}	0.96 (0.030)	0.90 (0.034)	0.97 (0.030)
$\text{corr}_{R,F/W}$	-0.86 (0.014)	-0.84 (0.016)	-0.89 (0.012)
$\text{corr}_{R,S}$	-0.52 (0.042)	-0.60 (0.044)	-0.61 (0.036)
$\text{corr}_{F/W,S}$	0.29 (0.042)	0.41 (0.056)	0.47 (0.044)
Multi-Person Households			
σ_{η_R}	0.10 (0.016)	0.37 (0.016)	0.45 (0.015)
$\sigma_{\eta_{F/W}}$	0.54 (0.008)	0.70 (0.008)	0.86 (0.009)
σ_{η_S}	0.96 (0.017)	1.10 (0.017)	1.24 (0.015)
$\text{corr}_{R,F/W}$	-0.93 (0.006)	-0.91 (0.006)	-0.95 (0.004)
$\text{corr}_{R,S}$	-0.79 (0.016)	-0.73 (0.015)	-0.78 (0.010)
$\text{corr}_{F/W,S}$	0.73 (0.019)	0.81 (0.011)	0.85 (0.008)
<i>Demand Synergies, γ</i>			
family size = 2, γ_2	0.14 (0.017)		
family size ≥ 3 , $\gamma_{\geq 3}$	0.36 (0.015)		
normal shop. freq., γ_s	0.001 (0.012)		

Notes: Each column of the Table reports estimates from one of three model specifications: the full model, restricted model 1 (which constrains $\Gamma_{i\mathbf{b}} = 0$ in estimation), and restricted model 2 (which constrains $\Gamma_{i\mathbf{b}} = -\infty$ in estimation, i.e. standard demand model for single brands). The top panel reports the estimated average price sensitivity $\mathbb{E}[-\exp(d_i^\alpha \alpha + v_i)]|d_i^\alpha|$ for each value of d_i^α and the corresponding standard deviation (in brackets), computed as $\sqrt{\text{Var}[\exp(d_i^\alpha \alpha + v_i)]|d_i^\alpha|}$. The central panel reports estimates of the parameters characterizing the distribution of the random coefficients, while the bottom panel those of the demand synergy parameters associated to different family sizes and weekly shopping frequencies. For the estimates in the central and bottom panel, standard errors are reported in brackets. Cereal type R refers to “Regular,” F/W to “Fiber/Whole Grain,” and S to “Added Sugar.”

Evidence in Support of Assumption 2. As discussed in Remark 1, Assumption 2 can be verified in practice. In Appendix A.7, we present a specification test for Assumption 2 that builds on partial identification methods. A rejection of the test is evidence against Assumption 2. In the context of our empirical illustration, the test statistic evaluated at the estimates from Tables 2.2 and 1.3 is 9910, which is smaller than the critical value for rejection at the 10% level, 21081 (a chi-square with 20819 degrees of freedom). This strongly suggests that $\Theta_I(\mathbf{T})$ in (A.13) is not empty, providing reassuring evidence in support of Assumption 2.

1.6.4 Hicksian Complementarity and Demand Synergies

Table 1.4 reports the average (across markets) estimated own- and cross-price elasticities of demand from the full model. Each entry reports the percent change in the brand-level market share of the column RTE cereal brand with respect to a 1% increase in the price of the row RTE cereal brand. Given the estimated market share function $\hat{s}_{t\mathbf{b}}$ for each $\mathbf{b} \in \mathbf{C}_{t1}$ in market t , the estimated brand-level market share function of brand $j \in \mathbf{J}_t$ is defined as $\hat{s}_{tj} = \sum_{\mathbf{b} \in \mathbf{C}_{t1}: j \in \mathbf{b}} \hat{s}_{t\mathbf{b}}$. Table 1.4 provides pervasive evidence of Hicksian complementarity. For example, the first column shows that households exhibit statistically significant complementarity between General Mills and several of the 15 remaining brands. According to intuition, Hicksian complementarity seems to be more pronounced among those brands with larger positive $\Gamma_{\mathbf{b}}$ (see Table 1.3).

In our specification, Hicksian complementarity among different RTE cereal brands can be explained by alternative mechanisms: correlation in the unobserved preferences for single brands (η_{ij}), preference for variety (γ_2 and $\gamma_{\geq 3}$), shopping costs (γ_s), bundle-specific fixed effects ($\Gamma_{\mathbf{b}}$'s)—which, among other things, may account for synergies in consumption—, and residual unobserved heterogeneity ($\tilde{\zeta}_{t\mathbf{b}}$). To shed light on the relative contributions of these mechanisms, we sequentially “switch them off” from the estimated full model and re-compute the cross-price elasticities. Table 1.7 summarizes the results.

Differently from Gentzkow (2007), the unobserved preferences for single brands (η_{ij}) contribute to the substitutability among RTE cereal brands (−2.00%), possibly because of the negative correlation between $\eta_{i,R}$ and both $\eta_{i,F/W}$ and $\eta_{i,S}$ (see Table 2.2). The average of the cross-price elasticities instead increases (becoming less negative) as we progressively switch off the various components of $\Gamma_{i\mathbf{b}}$: residual unobserved heterogeneity (+0.09%), shopping costs (+0.05%), and especially preference for variety (+15.22%).

However, the most dramatic changes occur when we further set the bundle-specific fixed effects $\Gamma_{\mathbf{b}}$'s either to zero (+54.94%) or to $-\infty$ (+31.69%). While this is expected in the case of $\Gamma_{i\mathbf{b}} = -\infty$ (standard demand model for single brands), the average of the cross-price elasticities already changes from negative (complementarity) to positive (substitutability) when setting each $\Gamma_{i\mathbf{b}}$ to zero. Collectively, these results suggest that most of the estimated complementarity is explained by preference for variety and by the bundle-specific fixed effects.

Standard models of demand for single brands rule out Hicksian complementarity among different RTE cereal brands and restrict the cross-price elasticities to be positive. Ignoring the presence of complementarity among different brands may lead to incorrect demand estimates and misleading price elasticities. To quantify the extent of this problem, we compare the price elasticities computed on the basis of the estimates from the full model (Table 1.4) to those computed on the basis of the

estimates from the restricted models (Tables 1.5 and 1.6).³³ Several of the estimated cross-price elasticities have opposite signs, mistakenly suggesting substitutability rather than complementarity among different pairs of RTE cereal brands. To further explore the economic consequences of accounting for complementarity in demand estimation, we next compare counterfactual simulations implied by the estimates from the full model to those implied by the estimates from the restricted models.

³³Demand estimates from the full model can be found in the first column of Table 2.2 and in Table 1.3, while those from the restricted models can be found in the second and third columns of Table 2.2 (the restricted models do not include demand synergies).

Table 1.3: Estimated Average Demand Synergy Parameters, Γ_b

	G. Mills			Kellogg's			Quaker			Post			Private			Small Producers			
	<i>F/W</i>	<i>R</i>	<i>S</i>	<i>F/W</i>	<i>R</i>	<i>S</i>	<i>F/W</i>	<i>R</i>	<i>S</i>	<i>F/W</i>	<i>R</i>	<i>S</i>	<i>F/W</i>	<i>R</i>	<i>S</i>	<i>F/W</i>	<i>R</i>	<i>S</i>	
G. Mills
Kellogg's	<i>R</i>	0.93 (0.036)																	
	<i>F/W</i>	1.83 (0.027)																	
	<i>S</i>	2.12 (0.059)																	
Quaker	<i>R</i>	1.82 (0.851)				1.50 (0.935)													
	<i>F/W</i>	1.17 (0.050)			0.35 (0.992)	0.84 (0.295)													
	<i>S</i>	1.05 (0.081)			2.72 (0.918)	1.41 (0.344)				2.52 (0.090)									
Post	<i>R</i>	1.66 (0.985)			2.22 (0.983)	2.04 (0.986)				3.62 (0.988)									
	<i>F/W</i>	0.97 (0.036)			0.95 (0.072)	0.89 (0.050)				1.30 (0.273)	1.76 (0.925)	1.05 (0.273)		4.01 (0.964)					
	<i>S</i>	0.96 (0.145)			0.56 (0.908)	1.19 (0.298)				0.86 (0.931)	0.78 (0.855)	1.49 (0.912)		1.97 (0.208)					
Private	<i>R</i>	0.96 (0.059)			0.53 (0.252)	0.62 (0.097)				1.41 (0.971)	1.41 (0.674)	0.49 (0.833)		0.90 (0.115)					
	<i>F/W</i>	0.49 (0.067)			0.56 (0.318)	0.40 (0.098)				1.79 (0.104)	0.27 (0.731)	0.12 (0.949)		0.67 (0.245)		2.31 (0.082)			
	<i>S</i>	0.57 (0.084)			0.79 (0.370)	0.73 (0.115)				0.84 (0.311)	-0.01 (0.446)	0.88 (0.527)		0.72 (0.395)		2.83 (0.081)			
Small P.	<i>R</i>	0.21 (0.143)			-0.01 (0.723)	0.54 (0.357)				1.66 (0.990)	1.66 (0.877)	0.68 (0.912)		0.19 (0.799)		0.71 (0.725)		0.01 (0.893)	
	<i>F/W</i>	0.96 (0.033)			0.50 (0.078)	0.39 (0.059)				0.32 (0.301)	0.21 (0.171)	0.34 (0.659)		0.52 (0.074)		0.54 (0.227)		0.99 (0.389)	
	<i>S</i>	0.23 (0.544)			-0.36 (0.908)	0.54 (0.532)				1.39 (0.720)	-0.20 (0.890)	0.17 (0.930)		0.64 (0.707)		1.56 (0.819)		1.50 (0.567)	
																2.90 (0.096)		1.80 (0.893)	

Notes: The Table reports the estimated average demand synergies for all bundles of RTE cereal brands for single-person households (standard errors in brackets). Cereal type *R* refers to "Regular," *F/W* to "Fiber/Whole Grain," and *S* to "Added Sugar." An off-diagonal missing value "." refers to an observed zero market share for the corresponding bundle (we do not observe the purchase of the bundle in any of the shopping trips in our data). Differently, missing values "." along the diagonal refer to our definition of bundle: we only consider as bundles the joint purchases of *different* brands of RTE cereals.

Table 1.4: Average Estimated Own- and Cross-Price Elasticities (Full Model)

	G. Mills			Kellogg's			Quaker			Post			Private			Small Producers			
	F/W	R	S																
G. Mills	-1.226 (0.154)	-0.016 (0.002)	-0.022 (0.003)	-0.007 (0.001)	-0.002 (0.007)	-0.006 (0.001)	-0.002 (0.001)	-0.001 (0.001)	-0.022 (0.004)	-0.001 (0.001)	-0.006 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.006 (0.001)	-0.001 (0.001)	-0.024 (0.004)	0.001 (0.003)	-0.024 (0.002)	0.001 (0.002)
Kellogg's	-0.047 (0.007)	-1.387 (0.184)	-0.073 (0.011)	-0.033 (0.004)	-0.010 (0.017)	-0.006 (0.002)	-0.001 (0.006)	-0.002 (0.003)	-0.012 (0.003)	-0.002 (0.002)	-0.001 (0.010)	-0.001 (0.007)	-0.001 (0.001)	-0.001 (0.007)	0.000 (0.011)	0.001 (0.002)	0.003 (0.012)	0.001 (0.002)	0.001 (0.003)
	-0.053 (0.007)	-0.060 (0.009)	-1.152 (0.147)	-0.048 (0.006)	0.003 (0.001)	-0.004 (0.001)	-0.002 (0.001)	-0.012 (0.003)	-0.012 (0.003)	-0.001 (0.004)	-0.001 (0.007)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.000 (0.002)	0.005 (0.002)	0.000 (0.007)	0.005 (0.002)	0.001 (0.003)
	-0.047 (0.009)	-0.074 (0.010)	-0.133 (0.018)	-1.134 (0.146)	0.003 (0.010)	0.000 (0.011)	-0.007 (0.012)	-0.004 (0.000)	-0.004 (0.009)	0.001 (0.003)	0.000 (0.003)	0.001 (0.006)	0.001 (0.006)	0.003 (0.010)	0.007 (0.013)	0.013 (0.004)	0.002 (0.007)	0.013 (0.013)	0.004 (0.003)
Quaker	-0.036 (0.135)	-0.067 (0.114)	0.023 (0.010)	-0.012 (0.036)	-1.538 (0.230)	0.002 (0.024)	-0.022 (0.033)	-0.022 (0.033)	-0.054 (0.081)	0.001 (0.000)	0.002 (0.004)	0.001 (0.001)	0.005 (0.022)	-0.013 (0.019)	-0.014 (0.032)	0.027 (0.004)	-0.014 (0.032)	0.027 (0.004)	0.003 (0.001)
	-0.050 (0.010)	-0.017 (0.007)	-0.014 (0.005)	0.000 (0.014)	0.001 (0.011)	-1.126 (0.144)	-0.024 (0.004)	-0.010 (0.009)	-0.030 (0.006)	0.000 (0.004)	-0.004 (0.014)	-0.019 (0.006)	-0.009 (0.006)	-0.004 (0.014)	0.002 (0.018)	0.009 (0.033)	0.002 (0.018)	0.009 (0.033)	-0.002 (0.019)
	-0.048 (0.014)	-0.009 (0.038)	-0.017 (0.012)	-0.021 (0.036)	-0.029 (0.042)	-0.059 (0.101)	-1.196 (0.155)	0.001 (0.000)	-0.016 (0.042)	-0.002 (0.009)	0.003 (0.015)	0.004 (0.009)	0.004 (0.028)	0.003 (0.015)	0.000 (0.009)	0.011 (0.007)	0.000 (0.009)	0.011 (0.007)	0.003 (0.004)
Post	-0.054 (0.102)	-0.024 (0.028)	-0.038 (0.098)	0.009 (0.002)	-0.029 (0.002)	-0.149 (0.131)	0.001 (0.000)	-0.861 (0.110)	-0.223 (0.190)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.058 (0.113)	0.003 (0.001)	0.019 (0.001)	0.006 (0.001)	0.003 (0.000)	0.019 (0.003)	0.006 (0.001)
	-0.072 (0.012)	-0.013 (0.003)	-0.017 (0.004)	-0.002 (0.004)	-0.009 (0.014)	-0.011 (0.002)	-0.002 (0.006)	-0.017 (0.015)	-1.160 (0.146)	-0.005 (0.005)	-0.004 (0.003)	-0.002 (0.011)	-0.001 (0.009)	-0.004 (0.003)	-0.001 (0.009)	-0.002 (0.002)	0.002 (0.007)	-0.002 (0.002)	0.000 (0.004)
	-0.044 (0.047)	0.007 (0.035)	-0.017 (0.070)	0.005 (0.031)	0.003 (0.000)	0.000 (0.023)	-0.006 (0.020)	-0.006 (0.000)	-0.118 (0.106)	-1.111 (0.137)	0.005 (0.018)	0.001 (0.006)	0.002 (0.013)	0.005 (0.018)	0.007 (0.001)	0.019 (0.044)	0.007 (0.001)	0.019 (0.044)	0.000 (0.007)
Private	-0.059 (0.010)	-0.003 (0.034)	-0.002 (0.005)	0.006 (0.007)	-0.006 (0.008)	-0.004 (0.016)	0.001 (0.007)	0.001 (0.000)	-0.012 (0.010)	-0.001 (0.003)	-0.001 (0.012)	-0.029 (0.005)	-0.043 (0.006)	-0.968 (0.120)	-0.001 (0.006)	-0.001 (0.021)	-0.001 (0.006)	-0.001 (0.021)	-0.003 (0.008)
	-0.009 (0.009)	0.001 (0.029)	0.006 (0.006)	0.006 (0.020)	0.003 (0.001)	-0.027 (0.009)	-0.003 (0.006)	-0.028 (0.054)	-0.006 (0.042)	0.001 (0.002)	-0.037 (0.007)	-0.900 (0.109)	-0.042 (0.006)	-0.037 (0.007)	0.002 (0.012)	-0.016 (0.008)	0.002 (0.012)	-0.016 (0.008)	-0.003 (0.012)
	-0.012 (0.009)	-0.003 (0.054)	-0.002 (0.011)	0.015 (0.013)	-0.005 (0.015)	-0.001 (0.012)	-0.001 (0.018)	-0.001 (0.000)	-0.004 (0.042)	0.001 (0.004)	-0.067 (0.010)	-0.046 (0.007)	-0.893 (0.113)	-0.067 (0.010)	0.003 (0.005)	-0.014 (0.043)	0.003 (0.005)	-0.014 (0.043)	-0.005 (0.009)
Small P.	0.011 (0.037)	0.011 (0.045)	0.000 (0.032)	0.003 (0.012)	-0.007 (0.015)	0.002 (0.024)	0.000 (0.005)	0.001 (0.000)	0.006 (0.026)	0.002 (0.006)	-0.002 (0.007)	0.001 (0.011)	0.002 (0.004)	-0.002 (0.007)	0.002 (0.004)	-0.209 (0.034)	-1.572 (0.206)	-0.209 (0.034)	-0.004 (0.016)
	-0.072 (0.011)	0.001 (0.002)	0.006 (0.003)	0.006 (0.006)	0.004 (0.001)	0.003 (0.011)	0.002 (0.004)	0.001 (0.000)	-0.002 (0.002)	0.001 (0.002)	0.000 (0.006)	-0.004 (0.002)	-0.003 (0.009)	0.000 (0.008)	-0.052 (0.173)	-1.333 (0.005)	-0.052 (0.008)	-1.333 (0.173)	-0.022 (0.005)
	0.030 (0.079)	0.012 (0.033)	0.012 (0.049)	0.020 (0.014)	0.003 (0.001)	-0.008 (0.073)	0.005 (0.007)	0.001 (0.000)	0.002 (0.035)	0.000 (0.004)	-0.010 (0.031)	-0.006 (0.028)	-0.012 (0.023)	-0.010 (0.040)	-0.011 (0.023)	-0.249 (0.055)	-0.011 (0.040)	-0.249 (0.055)	-0.917 (0.113)

Notes: The Table reports the average estimated own- (diagonal) and cross-price (off-diagonal) elasticities from the full model, where averages are computed across markets. Each entry reports the percent change in the brand-level market share of the column RTE cereal brand with respect to a 1% increase in the price of the row RTE cereal brand. Given the estimated market share function \hat{s}_{tB} for each bundle $\mathbf{b} \in \mathbf{C}_{tB}$ in market t , the estimated brand-level market share function of any brand $j \in \mathbf{J}_t$ is defined as $\hat{s}_{tj} = \sum_{\mathbf{b} \in \mathbf{C}_{tj}} \hat{s}_{tB}$. Cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar." The missing values "." refer to situations where the corresponding column and row brands are never simultaneously sold in the same market. The standard errors of the estimated elasticities are in brackets obtained from a parametric bootstrap as in Nevo (2000, 2001) with 50 draws.

Table 1.5: Average Estimated Own- and Cross-Price Elasticities (Restricted Model 1, $\Gamma_{ib} = 0$ for each i and b)

	G. Mills			Kellogg's			Quaker			Post			Private			Small Producers				
	F/W	R	F/W	R	S	F/W	R	S	F/W	R	S	F/W	R	S	F/W	R	S			
G.Mills	F/W	-1.247 (0.068)	0.005 (0.000)	0.007 (0.000)	0.003 (0.000)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.000 (0.000)	0.005 (0.000)	0.001 (0.000)	0.002 (0.000)	0.001 (0.000)	0.002 (0.000)	0.001 (0.000)	0.006 (0.000)	0.001 (0.000)	0.001 (0.000)		
Kellogg's	R	0.014 (0.001)	-1.435 (0.081)	0.008 (0.001)	0.004 (0.000)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.000 (0.000)	0.006 (0.000)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.007 (0.000)	0.007 (0.000)	0.001 (0.000)	0.001 (0.000)		
	F/W	0.016 (0.001)	0.007 (0.000)	-1.197 (0.065)	0.004 (0.000)	0.001 (0.000)	0.004 (0.000)	0.004 (0.000)	0.002 (0.000)	0.000 (0.000)	0.008 (0.001)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.009 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	
Quaker	S	0.020 (0.001)	0.008 (0.001)	0.012 (0.001)	-1.208 (0.065)	0.001 (0.000)	0.004 (0.000)	0.003 (0.000)	0.001 (0.000)	0.010 (0.005)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)	0.004 (0.000)	0.011 (0.000)	0.011 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	
	R	0.013 (0.001)	0.006 (0.000)	0.007 (0.001)	0.003 (0.000)	-1.597 (0.092)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.000 (0.000)	0.006 (0.000)	0.002 (0.000)	0.003 (0.000)	0.007 (0.000)	0.002 (0.000)	0.036 (0.002)	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)	
Post	F/W	0.021 (0.001)	0.008 (0.001)	0.013 (0.001)	0.005 (0.000)	0.001 (0.000)	-1.173 (0.064)	0.002 (0.000)	0.000 (0.000)	0.009 (0.001)	0.001 (0.000)	0.002 (0.000)	0.003 (0.000)	0.002 (0.000)	0.011 (0.000)	0.011 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	
	S	0.023 (0.002)	0.008 (0.001)	0.014 (0.001)	0.009 (0.001)	0.001 (0.000)	0.005 (0.000)	-1.259 (0.069)	0.001 (0.000)	0.001 (0.000)	0.011 (0.001)	0.001 (0.000)	0.003 (0.000)	0.002 (0.000)	0.013 (0.000)	0.013 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)
Private	R	0.006 (0.000)	0.002 (0.000)	0.004 (0.000)	0.011 (0.001)	0.000 (0.000)	0.003 (0.000)	0.002 (0.000)	-0.963 (0.051)	0.002 (0.000)	0.002 (0.000)	0.004 (0.000)	0.004 (0.000)	0.022 (0.001)	0.022 (0.000)	0.007 (0.000)	0.007 (0.000)	0.007 (0.000)	0.007 (0.000)	
	F/W	0.017 (0.001)	0.007 (0.000)	0.011 (0.001)	0.005 (0.000)	0.001 (0.000)	0.003 (0.000)	0.002 (0.000)	0.000 (0.000)	-1.201 (0.065)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.009 (0.001)	0.009 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
SmallP.	S	0.031 (0.002)	0.010 (0.001)	0.015 (0.001)	0.013 (0.001)	0.003 (0.000)	0.006 (0.000)	0.003 (0.000)	0.000 (0.000)	0.017 (0.001)	-1.158 (0.063)	0.004 (0.000)	0.002 (0.000)	0.010 (0.001)	0.017 (0.000)	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)
	R	0.014 (0.001)	0.006 (0.000)	0.008 (0.001)	0.004 (0.000)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.001 (0.000)	0.007 (0.000)	0.001 (0.000)	-1.015 (0.054)	0.002 (0.000)	0.002 (0.000)	0.007 (0.000)	0.007 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
SmallP.	F/W	0.018 (0.001)	0.008 (0.001)	0.012 (0.001)	0.005 (0.000)	0.005 (0.000)	0.004 (0.000)	0.002 (0.000)	0.000 (0.000)	0.008 (0.001)	0.001 (0.000)	0.002 (0.000)	-0.944 (0.049)	0.009 (0.000)	0.009 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
	S	0.020 (0.001)	0.009 (0.001)	0.013 (0.001)	0.008 (0.001)	0.001 (0.000)	0.004 (0.000)	0.003 (0.000)	0.003 (0.000)	0.009 (0.001)	0.001 (0.000)	0.003 (0.000)	0.002 (0.000)	0.002 (0.000)	0.011 (0.000)	0.011 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)
SmallP.	R	0.016 (0.001)	0.007 (0.000)	0.009 (0.001)	0.004 (0.000)	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	0.001 (0.000)	0.007 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	0.008 (0.001)	0.008 (0.000)	-1.636 (0.091)	0.008 (0.001)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
	F/W	0.017 (0.001)	0.007 (0.001)	0.011 (0.001)	0.005 (0.000)	0.005 (0.000)	0.004 (0.000)	0.002 (0.000)	0.001 (0.000)	0.008 (0.001)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)	-1.370 (0.076)	0.002 (0.000)	0.002 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
SmallP.	S	0.023 (0.002)	0.009 (0.001)	0.014 (0.001)	0.010 (0.001)	0.004 (0.000)	0.005 (0.000)	0.003 (0.000)	0.001 (0.000)	0.011 (0.001)	0.002 (0.000)	0.003 (0.000)	0.001 (0.000)	0.002 (0.000)	0.013 (0.001)	0.013 (0.001)	-0.994 (0.053)	-0.994 (0.053)	-0.994 (0.053)	-0.994 (0.053)

Notes: The Table reports the average estimated own- (diagonal) and cross-price (off-diagonal) elasticities from restricted model 1 (which constrains $\Gamma_{ib} = 0$ in estimation), where averages are computed across markets. Each entry reports the percent change in the brand-level market share of the column RTE cereal brand with respect to a 1% increase in the price of the row RTE cereal brand. Given the estimated market share function \hat{s}_{ib} for each bundle $b \in C_{t,1}$ in market t , the estimated brand-level market share function of any brand $j \in J_t$ is defined as $\hat{s}_{tj} = \sum_{b \in C_{t,1}: j \in b} \hat{s}_{ib}$. Cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar." The missing values "." refer to situations where the corresponding column and row brands are never simultaneously sold in the same market. The standard errors of the estimated elasticities are in brackets obtained from a parametric bootstrap as in [Nevo \(2000, 2001\)](#) with 50 draws.

Table 1.6: Average Estimated Own- and Cross-Price Elasticities (Restricted Model 2, $\Gamma_{ib} = -\infty$ for each i and b)

	G. Mills			Kellogg's			Quaker			Post			Private			Small Producers		
	F/W	R	F/W	S	R	F/W	S	R	F/W	S	R	F/W	S	R	F/W	S		
G. Mills	F/W	-4.027 (0.699)	0.094 (0.018)	0.135 (0.025)	0.049 (0.009)	0.014 (0.003)	0.039 (0.007)	0.015 (0.003)	0.003 (0.001)	0.095 (0.018)	0.007 (0.001)	0.026 (0.005)	0.022 (0.004)	0.025 (0.005)	0.109 (0.020)	0.010 (0.002)		
Kellogg's	R	0.276 (0.052)	-4.716 (0.838)	0.117 (0.022)	0.042 (0.008)	0.014 (0.003)	0.030 (0.006)	0.014 (0.003)	0.003 (0.001)	0.085 (0.016)	0.006 (0.001)	0.027 (0.005)	0.019 (0.004)	0.026 (0.005)	0.097 (0.018)	0.008 (0.002)		
	F/W	0.323 (0.060)	0.093 (0.017)	-3.869 (0.670)	0.054 (0.010)	0.014 (0.003)	0.040 (0.007)	0.017 (0.003)	0.004 (0.001)	0.109 (0.020)	0.007 (0.001)	0.026 (0.005)	0.023 (0.004)	0.024 (0.004)	0.115 (0.021)	0.009 (0.002)		
Quaker	S	0.329 (0.060)	0.098 (0.018)	0.154 (0.028)	-3.868 (0.667)	0.010 (0.002)	0.038 (0.007)	0.027 (0.005)	0.004 (0.001)	0.120 (0.022)	0.011 (0.002)	0.026 (0.005)	0.040 (0.007)	0.023 (0.004)	0.121 (0.023)	0.016 (0.003)		
	R	0.276 (0.051)	0.100 (0.019)	0.102 (0.019)	0.040 (0.007)	-5.253 (0.945)	0.033 (0.006)	0.013 (0.002)	0.037 (0.007)	0.084 (0.016)	0.005 (0.001)	0.037 (0.007)	0.020 (0.004)	0.026 (0.005)	0.096 (0.018)	0.009 (0.002)		
Post	F/W	0.349 (0.065)	0.092 (0.017)	0.152 (0.028)	0.050 (0.009)	0.014 (0.003)	-3.781 (0.655)	0.018 (0.003)	0.004 (0.001)	0.101 (0.019)	0.007 (0.001)	0.028 (0.005)	0.026 (0.005)	0.026 (0.005)	0.122 (0.023)	0.010 (0.002)		
	S	0.336 (0.062)	0.093 (0.017)	0.151 (0.028)	0.085 (0.016)	0.016 (0.003)	0.046 (0.008)	-4.053 (0.704)	0.004 (0.001)	0.115 (0.021)	0.010 (0.002)	0.028 (0.005)	0.041 (0.008)	0.026 (0.005)	0.130 (0.025)	0.016 (0.003)		
Private	R	0.212 (0.038)	0.046 (0.009)	0.101 (0.018)	0.030 (0.005)	0.052 (0.010)	0.032 (0.006)	0.005 (0.001)	-2.961 (0.506)	0.048 (0.009)	0.012 (0.002)	0.008 (0.001)	0.019 (0.003)	0.011 (0.002)	0.063 (0.011)	0.019 (0.003)		
	F/W	0.314 (0.058)	0.095 (0.018)	0.151 (0.028)	0.057 (0.010)	0.014 (0.003)	0.037 (0.007)	0.018 (0.003)	0.004 (0.001)	-3.849 (0.665)	0.007 (0.001)	0.029 (0.005)	0.022 (0.004)	0.023 (0.004)	0.112 (0.021)	0.011 (0.002)		
Small P.	S	0.364 (0.067)	0.093 (0.017)	0.143 (0.026)	0.105 (0.019)	0.009 (0.002)	0.046 (0.008)	0.024 (0.004)	0.164 (0.029)	-3.678 (0.633)	0.032 (0.006)	0.016 (0.003)	0.040 (0.007)	0.024 (0.004)	0.149 (0.027)	0.021 (0.004)		
	R	0.264 (0.048)	0.092 (0.017)	0.110 (0.020)	0.039 (0.007)	0.016 (0.003)	0.032 (0.006)	0.013 (0.002)	0.004 (0.001)	0.088 (0.016)	0.005 (0.001)	-3.179 (0.544)	0.019 (0.003)	0.025 (0.005)	0.090 (0.017)	0.007 (0.001)		
Small P.	F/W	0.306 (0.056)	0.089 (0.016)	0.142 (0.026)	0.050 (0.009)	0.013 (0.003)	0.043 (0.008)	0.016 (0.003)	0.004 (0.001)	0.103 (0.018)	0.006 (0.001)	0.028 (0.005)	0.023 (0.004)	0.023 (0.004)	0.102 (0.019)	0.010 (0.002)		
	S	0.325 (0.059)	0.095 (0.017)	0.144 (0.026)	0.088 (0.016)	0.013 (0.002)	0.043 (0.008)	0.027 (0.005)	0.099 (0.018)	0.099 (0.018)	0.010 (0.002)	0.028 (0.005)	-2.925 (0.491)	0.022 (0.004)	0.113 (0.021)	0.015 (0.003)		
Small P.	R	0.291 (0.055)	0.101 (0.019)	0.115 (0.021)	0.040 (0.007)	0.012 (0.002)	0.034 (0.006)	0.014 (0.003)	0.003 (0.001)	0.083 (0.015)	0.005 (0.001)	0.029 (0.005)	0.019 (0.003)	-5.300 (0.942)	0.099 (0.019)	0.009 (0.002)		
	F/W	0.325 (0.060)	0.097 (0.018)	0.144 (0.027)	0.053 (0.010)	0.014 (0.003)	0.041 (0.008)	0.018 (0.003)	0.003 (0.001)	0.101 (0.019)	0.008 (0.001)	0.027 (0.005)	0.023 (0.004)	0.025 (0.005)	-4.479 (0.785)	0.011 (0.002)		
Small P.	S	0.324 (0.059)	0.098 (0.018)	0.147 (0.027)	0.089 (0.016)	0.011 (0.002)	0.041 (0.008)	0.027 (0.005)	0.004 (0.001)	0.109 (0.020)	0.012 (0.002)	0.029 (0.005)	0.041 (0.008)	0.022 (0.004)	0.126 (0.023)	-3.111 (0.527)		

Notes: The Table reports the average estimated own- (diagonal) and cross-price (off-diagonal) elasticities from restricted model 2 (which constrains $\Gamma_{ib} = -\infty$ in estimation), where averages are computed across markets. Each entry reports the percent change in the market share of the column RTE cereal brand with respect to a 1% increase in the price of the row RTE cereal brand. Cereal type R refers to "Regular," F/W to "Fiber/Whole Grain," and S to "Added Sugar." The missing values "." refer to situations where the corresponding column and row brands are never simultaneously sold in the same market. The standard errors of the estimated elasticities are in brackets obtained from a parametric bootstrap as in Nevo (2000, 2001) with 50 draws.

Table 1.7: Cross-Price Elasticities and Demand Synergies

	Average	Relative Change
Estimated Full Model	-0.0126	—
Estimated Full Model, then switch off η_{ij}	-0.0130	-2.00%
Estimated Full Model, then switch off η_{ij} and $\tilde{\zeta}_{i\mathbf{b}}$	-0.0130	0.09%
Estimated Full Model, then switch off η_{ij} , $\tilde{\zeta}_{i\mathbf{b}}$, and γ_s	-0.0130	0.05%
Estimated Full Model, then switch off η_{ij} , $\tilde{\zeta}_{i\mathbf{b}}$, γ_s , γ_2 , and $\gamma_{\geq 3}$	-0.0100	15.22%
Estimated Full Model, switch off η_{ij} and set $\Gamma_{i\mathbf{b}} = 0$	0.0007	54.94%
Estimated Full Model, switch off η_{ij} and set $\Gamma_{i\mathbf{b}} = -\infty$	0.0068	31.69%

Note: The Table reports the average (across markets) of the cross-price elasticities of all pairs of different RTE cereal brands. All the cross-price elasticities are obtained from the full model, where the estimated parameters in the specification of $\Gamma_{i\mathbf{b}}$ from (1.18) are “switched off” sequentially. For example, the third row is obtained from the estimated full model by setting the standard deviations of η_{ij} and $\tilde{\zeta}_{i\mathbf{b}}$ to zero, while the fourth is obtained by further setting the estimated γ_s to zero. The column “Relative Change” reports the percent change in the average of the cross-price elasticities from any two consecutive rows: for instance, the relative change from the fourth row to the fifth is 15.22%.

1.6.5 Counterfactuals and Comparisons with Standard Model

Here we evaluate the economic relevance of allowing for Hicksian complementarity by comparing some counterfactuals from the full model, with $\Gamma_{i\mathbf{b}}$ as in (1.18), to those from two restricted models: restricted model 1, which constrains $\Gamma_{i\mathbf{b}} = 0$ in estimation, and restricted model 2—the standard demand model for single brands—which constrains $\Gamma_{i\mathbf{b}} = -\infty$ in estimation thereby forcing substitutability. For the counterfactuals, we take the observed scenario of pure components pricing and oligopolistic competition among RTE cereal producers as a reference (see [Nevo \(2000, 2001\)](#) for the institutional details), and simulate the changes in prices, profits, and consumer surplus implied by different market structures.³⁴

The results of these counterfactuals are reported in Table 1.8. The Table reports relative changes in prices (top panel), profits (central panel), and consumer surplus (bottom panel) associated with each of three counterfactual market structures (columns) as simulated by each of the three estimated models (rows). We consider four alternative market structures: “competition,” where we suppose that each single brand is owned and sold by a different (fictional) producer (for a total of 16 producers); “oligopoly,” which corresponds to the observed oligopolistic competition among six producers; “duopoly,” where we suppose that five of the producers (General Mills, Kellogg’s, Quaker, Post, and the Small Producers) perfectly collude and compete as one against the private labels (whose prices are chosen by the retailer); and “monopoly,” where we suppose that the six producers perfectly collude as a monopolist.

The simulation results from the full model confirm the classic insight by [Cournot \(1838\)](#): mergers between producers selling complementary brands can be socially desirable. In pure components pricing, the prices of all single brands—and consequently of all bundles—decrease as the level of competition weakens: while industry-level profit remains basically unchanged, consumer surplus increases with market concentration. As market structure becomes more concentrated, producers internalize more of the externalities due to complementarity and consequently choose lower prices, as can be seen from the relative increase in consumer surplus from -5.27% for competition to +7.62% for monopoly.

³⁴Given our estimates of demand and marginal costs, we simulate each profile of counterfactual prices—independently for each market—using the necessary first order conditions for a Nash equilibrium of the corresponding pure components pricing game. For example, in a monopoly, the same agent chooses a specific price for each single brand so to maximize industry profits.

Strikingly, the restricted models produce opposite predictions: as market structure becomes more concentrated, prices increase and consumer surplus decreases. As shown in Tables 1.5 and 1.6, both restricted models predict positive cross-price elasticities and therefore substitutability among any pair of RTE cereal brands. Not surprisingly then, any merger between producers selling substitutable brands will lead to higher prices and ultimately hurt consumers. These results underline the economic importance of allowing for both substitutability and complementarity in demand estimation: while estimates from the full model provide supportive evidence for the classic [Cournot \(1838\)](#)'s insight that mergers can be welfare enhancing, those from a standard model that does not allow for Hicksian complementarity can only predict that mergers will be detrimental for consumer surplus.

Table 1.8: Counterfactual Simulations

	Competition	Oligopoly	Duopoly	Monopoly
<i>Price change</i>				
Full Model	+8.08% (1.35%)	0%	-5.03% (1.16%)	-5.34% (1.65%)
Restricted Model 1, $\Gamma_{i\mathbf{b}} = 0$	-0.54% (0.03%)	0%	+3.69% (0.21%)	+5.26% (0.30%)
Restricted Model 2, $\Gamma_{i\mathbf{b}} = -\infty$	-0.56% (0.10%)	0%	+4.17% (0.72%)	+6.06% (1.07%)
<i>Profit change</i>				
Full Model	-0.47% (0.04%)	0%	+0.27% (0.03%)	+0.30% (0.05%)
Restricted Model 1, $\Gamma_{i\mathbf{b}} = 0$	-0.03% (0.002%)	0%	+0.10% (0.003%)	+0.12% (0.004%)
Restricted Model 2, $\Gamma_{i\mathbf{b}} = -\infty$	-0.36% (0.01%)	0%	+1.64% (0.03%)	+2.00% (0.03%)
<i>Consumer Surplus change</i>				
Full Model	-5.27% (0.22%)	0%	+6.29% (0.37%)	+7.62% (0.62%)
Restricted Model 1, $\Gamma_{i\mathbf{b}} = 0$	+0.49% (0.01%)	0%	-3.23% (0.04%)	-4.54% (0.06%)
Restricted Model 2, $\Gamma_{i\mathbf{b}} = -\infty$	+1.53% (0.01%)	0%	-11.94% (0.07%)	-16.41% (0.09%)

Notes: The Table reports average counterfactual changes in prices (top panel), profits (central panel), and consumer surplus (bottom panel) of pure components pricing under alternative simulated market structures with respect to the observed oligopoly. Each column refers to a specific market structure: the second column refers to the observed oligopoly in the data while the others refer to simulated counterfactuals (see text for details). Each row refers to one of three model specifications: the full model, restricted model 1 (which constrains $\Gamma_{i\mathbf{b}} = 0$ in estimation), and restricted model 2 (which constrains $\Gamma_{i\mathbf{b}} = -\infty$ in estimation, i.e. standard demand model for single brands). The standard errors associated to the estimated relative changes are in brackets and obtained from a parametric bootstrap as in [Nevo \(2000, 2001\)](#) with 50 draws.

1.7 Conclusions

We present a novel identification and estimation strategy of a mixed logit model of demand for bundles with endogenous prices given observations on bundle-level market shares. We propose a novel demand inverse in the presence of complementarity that allows to concentrate out of the likelihood function the (potentially numerous) market-product specific average utilities and to substantially alleviate the challenge of dimensionality inherent in estimation. Finally, we estimate demand and supply in the US ready-to-eat cereal industry, where our estimator reduces the numerical search from approximately

12000 to 130 parameters. Our results suggest that ignoring Hicksian complementarity among products often purchased in bundles may result in misleading demand estimates and counterfactuals.

Our identification and estimation arguments are developed for mixed logit models with parametric distributions of random coefficients. In light of the well known challenge of dimensionality that affects the estimation of demand for bundles (Berry et al. (2014)), our priority is to propose estimation methods that can be practically useful in applications that involve more than a few products. While our parametric assumptions clearly help in alleviating the challenge of dimensionality in estimation, they require the econometrician to take a stand on the functional form of the distribution of random coefficients. An avenue for future research is the study of the semi-parametric identification of mixed logit models of demand for bundles, where the distribution of random coefficients is allowed to be non-parametric and more robust against misspecification.

The implementation of our methods requires the observation of bundle-level market shares rather than of the more readily available aggregate market shares of single products. Even though direct measures of bundle-level market shares are widely available only for a few industries, such as media and telecommunication (Crawford and Yurukoglu (2012) and Crawford et al. (2018)), it is usually possible to construct indirect measures of bundle-level market shares from samples of household-level purchases (Gentzkow (2007), Kwak et al. (2015), Grzybowski and Verboven (2016), Ruiz et al. (2017), and Ershov et al. (2018)). In some important industries, however, only measures of aggregate market shares of single products are widely available (e.g., the car industry, see Berry et al. (1995, 2004a)) even though households are known to purchase bundles of products (Manski and Sherman (1980)). When only aggregate market shares of single products are available, our proposed methods do not apply. An important direction for future research is thus the identification and estimation of models of demand for bundles on the basis of aggregate market shares of single products (see Sher and Kim (2014), Allen and Rehbeck (2019a), and Wang (2019)).

A BLP Demand Model of Product-Level Market Shares with Complementarity

Abstract. Applied researchers most often estimate the demand for differentiated products assuming that at most one item can be purchased. Yet multiple purchases are pervasive. Ignoring this feature can lead to erroneous counterfactuals, in particular, because complementarities are ruled out. I consider the identification and estimation of a random coefficient discrete choice model of bundles, namely sets of products, when only product-level market shares are available. This last feature arises when only aggregate market shares, as opposed to individual purchases, are available, a very common phenomenon in practice. Following the classical approach with aggregate data, I consider a two-step method. First, using a novel inversion result where demand can exhibit Hicksian complementarity, I recover the mean utilities of products from the product-level market shares. Second, to infer the structural parameters from the mean utilities while dealing with price endogeneity, I use instrumental variables. I provide low-level conditions under which the model is globally identified through moment conditions based on such instruments. Finally, I illustrate the practical implementation of the methods and estimate the demand for Ready-To-Eat (RTE) cereals and milk in the US. The demand estimates suggest that RTE cereals and milk are overall Hicksian complementary and these complementarities are heterogeneous across bundles. Ignoring such complementarities results in misleading counterfactuals.

2.1 Introduction

Since the seminal work of [Berry \(1994\)](#) and [Berry et al. \(1995\)](#) (henceforth BLP), BLP-type models have been widely used in empirical demand literature and also gain popularity beyond empirical industrial organisation.¹ Applied researchers most often estimate BLP models of single products. An assumption behind these models is that individuals can only choose at most one item of a single product. Yet, multiple purchases are pervasive. In particular, this assumption rules out Hicksian complementarities among products.² As a result, estimating demand models of single products may lead to biased estimates and misleading counterfactuals. To relax this assumption, researchers typically use models of demand for bundles.³ Estimating such models usually requires individual choice data at *bundle level* (e.g. scanner data, survey data). However, in some important industries, only aggregate purchase data at *product level* is widely available.⁴

¹For example, analysis of voting data ([Rekkas \(2007\)](#), [Milligan and Rekkas \(2008\)](#), [Gordon and Hartmann \(2013\)](#), [Merlo and Paula \(2017\)](#), [Gillen et al. \(2019\)](#)), asset pricing ([Kojien and Yogo, 2019](#)).

²Hicksian complementarity is defined as negative (compensated) cross-price elasticity between two products. For a survey of different concepts of complementarity, see [Samuelson \(1974\)](#).

³In the empirical literature, the terminology “bundle” is often defined as a set of products/services/decisions purchased by individuals. In this paper, I use this definition and formalise it in Assumption 6.

⁴[Manski and Sherman \(1980\)](#) estimated a model that explains the composition of holdings in two-vehicle households. While the purchase data in the automotive industry is often at product level.

This paper proposes a random coefficient discrete-choice model of demand for bundles using aggregate demand at product level. Similar to classic BLP models of single products, aggregate demand at product level is defined in form of “market share”, i.e. aggregate choice probability (or total purchase) of a product, and is largely available in most industries. Differently, the proposed model does not restrict products to be substitutes and, notably, allows for Hicksian complementarities among products that can be jointly chosen in a bundle. Moreover, it enables to encompass, in a structural way, different mechanisms that may drive the Hicksian complementarity, and still allows for endogenous prices.⁵ The identification arguments of the model are constructive and lead to a practically useful Generalized Method of Moments (GMM) estimator. In particular, it can handle potentially large choice sets of bundles and its implementation is straightforward, essentially as a standard BLP estimator.

Different from the current literature of demand for bundles, the proposed model only requires the availability of market shares at product level, rather than at bundle level. Consequently, the market shares of bundles are not observed to the researcher and hence the identification and estimation of the model become more difficult. Depending on the research questions, the task of identification is more or less challenging. Specifically, I consider economic analyses (e.g. price elasticities, marginal costs, mergers) under linear and nonlinear pricing.⁶ I show that the economic analyses under linear pricing require to identify the product-level market share functions, while those under nonlinear pricing require to further identify the bundle-level market share functions. The latter is more challenging because only product-level market shares are available.

I then study the identification in two sequential parts. In the first part, I follow the classical approach in demand models with aggregate data and employ a two-step strategy to identify the product-level market share functions. In the first step, I invert the product-level market shares to the mean utilities of products. Because of possible Hicksian complementarities among products, the typical conditions that guarantee the invertibility of product-level market shares (connected substitutes conditions, see [Berry et al. \(2013\)](#)) may not hold. To solve this challenge, I use a novel demand inverse argument that hinges on two elements. First, the affine relationship between the utilities of bundles and its single products: the average utility of any bundle equals the sum of those of its single products plus an extra term capturing their potential *demand synergy*. Second, the P-matrix property by [Gale and Nikaido \(1965\)](#) which crucially does not restrict the products to be Hicksian substitutes. In the second step, I use instrumental variables (IVs) to deal with endogenous prices and propose low-level sufficient conditions to achieve the identification.⁷ Concretely, I assume the large support of the IVs and their independence with respect to the (unobserved) demand and supply shocks. I then leverage [Mattner \(1992\)](#) and [D’Haultfoeuille \(2011\)](#) and show that the product-level market share functions are identified when demand and supply shocks are normally distributed, or the random coefficients have compact support in the models of multiple choice of products across categories.

In the second part, assuming the identification of the product-level market share functions, I study that of the bundle-level market share functions. This requires to disentangle the demand synergies among products from the unobserved correlations in the utilities of products ([Gentzkow, 2007](#)). I show

⁵Examples of such mechanisms include shopping cost ([Pozzi \(2012\)](#), [Thomassen et al. \(2017\)](#)), preference for variety ([Hendel \(1999\)](#), [Dubé \(2004\)](#)), and synergies in consumption ([Gentzkow, 2007](#)).

⁶Under linear pricing, firms set prices for single products and the price of a bundle is the sum of the prices of its single products. Under nonlinear pricing, firms further set discounts or surcharges on the bundles of their own products.

⁷In general, one can leverage completeness conditions along the lines of [Berry and Haile \(2014\)](#) to achieve the identification. Despite its generality, these conditions are often high-level. For weaker forms of completeness conditions, the testability, and the sufficient conditions of completeness, see [Mattner \(1993\)](#), [D’Haultfoeuille \(2011\)](#), [Canay et al. \(2013\)](#), [Andrews \(2017\)](#), [Freyberger \(2017\)](#), [Hu and Shiu \(2018\)](#).

that the identification of the bundle-level market share functions is achieved in often used models where the bundle size is at most two.⁸ Moreover, I also prove that using only product-level market shares may have limited power in identifying bundle-level market share functions in other types of models. I provide an example of non-identification in a model of multi-unit choice.

I propose a GMM estimation procedure, similar to the two-step procedure in BLP models of single products. In the first step, given a guess of the demand synergy parameters and the distribution of the random coefficients, I invert the observed product-level market shares to the mean utilities of products. In the second step, I instrument out the unobserved demand shocks in the mean utilities of products and construct the GMM objective function. However, there are nontrivial challenges that BLP models of single products do not have. In particular, the implementation of the demand inverse is complicated due to possible Hicksian complementarities among products: the fixed-point algorithm proposed by [Berry et al. \(1995\)](#) may not have the contraction-mapping property. As a result, it may not converge when applied to implement the demand inverse in the first step above. To solve this challenge, I propose to use Jacobian-based algorithms. To enhance their numerical performance, I suggest using an initial point directly constructed from the observed product-level market shares and show that using this initial point can significantly improve the numerical performance of Jacobian-based methods. In Monte Carlo simulations, this reduces the convergence time by 70% relative to using standard initial point even in large applications (the number of bundles being about 10,000).

Finally, I illustrate the practical implementation of the proposed methods and estimate the demand for Ready-To-Eat (RTE) cereals and milk in the US. First, the demand estimates suggest that RTE cereals and milk are overall Hicksian complementary. I simulate a merger between a national RTE cereal producer and a national milk producer. The results are aligned with [Cournot \(1838\)](#)'s insight: in the presence of Hicksian complementarity, mergers can be welfare enhancing. In contrast, a BLP model of RTE cereals (or milk) assumes Hicksian independence between RTE cereals and milk and predicts no welfare effect in the merger simulation. Second, I find that some types of RTE cereals are more (or less) complementary to certain types of milk and a model of demand for bundles with restricted demand synergies may not capture this feature. In a nonlinear pricing counterfactual, I illustrate that using such restricted models may lead to important bias in welfare prediction.

Related Literature Empirical literature dealing with multiple purchases typically employs demand models of bundles with individual choice data at bundle level.⁹ See [Berry et al. \(2014\)](#) for a survey of complementary choices and sections 4.2-4.3 of [Dubé \(2018\)](#) for a survey of econometric modeling of complementary goods. Differently, the methods in this paper rely on aggregate demand data at product level and can be applied when bundle-level demand data is not accessible.¹⁰ In particular, the proposed methods are different from those in [Iaria and Wang \(2019a\)](#) in three aspects. First, the methods in two papers work under different data availabilities. Those in this paper work when only aggregate data at product level is available, while those of the other paper apply when bundle-level demand data is accessible. Second, the source of identification is different. I exploit the exogenous

⁸See [Gentzkow \(2007\)](#), [Fan \(2013\)](#), [Grzybowski and Verboven \(2016\)](#).

⁹Examples include consumer choice in supermarket ([Hendel \(1999\)](#), [Dubé \(2004\)](#), [Lee et al. \(2013\)](#), [Kwak et al. \(2015\)](#), [Thomassen et al. \(2017\)](#), [Ershov et al. \(2018\)](#)), household choice among motor vehicles ([Manski and Sherman, 1980](#)), choice of telecommunication services ([Liu et al. \(2010b\)](#), [Crawford and Yurukoglu \(2012\)](#), [Grzybowski and Verboven \(2016\)](#), [Crawford et al. \(2018\)](#)), subscription decision ([Nevo et al. \(2005\)](#), [Gentzkow \(2007\)](#)), firms' decision on technology adoptions ([Augereau et al. \(2006\)](#), [Kretschmer et al. \(2012\)](#)).

¹⁰Aggregate demand data at product level, e.g. aggregate purchase data, vote shares, is standard and widely available. While, demand data at bundle level is typically obtained via survey or scanners in the supermarket.

variation in IVs to achieve identification, while that paper fully exploits the affine relationship between the utilities of the bundle and its single products due to the availability of bundle-level demand data. Third, the estimation methods are different. This paper uses a GMM estimation procedure, while that paper proposes a likelihood-type estimator that resolves the dimensionality challenge of market-product fixed effects.

Identifying and estimating models of demand for bundles from aggregate demand at product level is a challenging task. Moreover, prices are often endogenous, which introduces additional difficulty in identification. To the best of my knowledge, this is the first paper that provides a systematic treatment of both issues in BLP-type models of demand for bundles that may exhibit Hicksian complementarity.¹¹ This paper differs from Fosgerau et al. (2019) who model Hicksian complementarity via overlapping nests. Sher and Kim (2014)'s identification arguments crucially rely on substitutes assumptions in consumers' utility,¹² while this paper does not restrict utility functions to be submodular or supermodular. Allen and Rehbeck (2019a)'s main results imply the identification of the product-level market share functions in discrete choice models with additively separable heterogeneity. Instead, the current paper achieves the identification using IVs and further provides identification results of the bundle-level market share functions. Their following paper, Allen and Rehbeck (2019b), gives identification results of the distributional features of the random coefficients in the case of two products (and therefore one bundle). Notably, except for Fosgerau et al. (2019), all other papers mentioned above assume away endogenous prices.

This paper also contributes to the research agenda of the identification and estimation of discrete choice demand models. Berry et al. (2013) propose the connected substitutes conditions that guarantee the invertibility of the market share functions. In model of demand for bundles with only product-level market shares being available, these conditions rely on the products to be substitutes.¹³ This paper uses a novel demand inverse argument to deal with possible Hicksian complementarities among products. Some papers have employed similar concepts of demand inverse. In a model of multiple choice of products across categories, Song and Chintagunta (2006) implement the demand inversion of market shares at brand level. However, they do not have theoretical results on the invertibility of the brand-level market share functions. Iaria and Wang (2019a) formally prove the invertibility of the product-level market share functions in model of demand for bundles and use it to reduce the dimensionality of fixed effects in estimation. In contrast, I use this demand inverse as a key identification argument when only product-level market shares are available.

Organisation In the next section, I introduce the model and necessary notations. I provide several examples in the literature that can be formulated via the model and, in particular, illustrate how this model can allow for Hicksian complementarity. In section 2.3, I present identification results. In section 2.4, I describe the GMM estimation procedure and its implementation. In section 2.5, I use Monte Carlo simulations to test numerical performances of the demand inverse implemented by

¹¹Dunker et al. (2017) also deal with price endogeneity in identification. However, instead of using the product-level market shares, they assume the availability of a vector of bundle-level market shares that has the same dimension as the number of products.

¹²When each consumer is assumed to consume at most one unit of each good, they impose submodularity restriction in consumers' utility (see their Assumption 2); when multi-unit demand is allowed, they use a stronger "M-natural concavity" restriction (see their Assumption 3).

¹³Fan (2013) studies newspaper market in the US and assumed households subscribe to at most two newspapers. She gives sufficient conditions for the connected substitutes conditions proposed by Berry et al. (2013) and rules out Hicksian complementarities among different newspapers.

Jacobian-based algorithms. In section 2.6, I examine the practical performance of the methods with an empirical illustration. Section 2.7 concludes. All proofs are in Appendices B.1-B.6. Figures and tables can be found in Appendix B.8.

2.2 Model

2.2.1 Model and Notation

Denote market by $t = 1, \dots, T$. The definition of market depends on the concrete application. For example, one could have different geographic areas in the case of cross sectional data, or different periods in the case of panel data, or a combination of these. For individuals in market t , let \mathbf{J}_t be the set of J_t market-specific products that can be purchased in isolation or in bundles. A bundle \mathbf{b} is defined as a collection of single products in \mathbf{J}_t and denote the set of available bundles in market t by \mathbf{C}_{t2} . Denote the outside option by 0. Individuals in market t can either choose a product $j \in \mathbf{J}_t$, a bundle $\mathbf{b} \in \mathbf{C}_{t2}$, or the outside option 0. Denote by $\mathbf{C}_{t1} = \mathbf{J}_t \cup \mathbf{C}_{t2}$ the set of available products and bundles, and by $\mathbf{C}_t = \mathbf{C}_{t1} \cup \{0\}$ the choice set of individuals in market t . Let p_{tj} denote the price of product j in market t , and $x_{tj} \in \mathbb{R}^K$ the market-product specific vector of other characteristics of j in market t . I follow the linear index assumption in BLP models of demand (see [Berry and Haile \(2014\)](#)). For individual i in market t , the indirect utility from choosing product j is:

$$\begin{aligned}
 U_{itj} &= u_{itj} + \varepsilon_{itj} \\
 &= x_{tj}\beta_i - \alpha_i p_{tj} + \eta_{ij} + \xi_{tj} + \varepsilon_{itj} \\
 &= x_{tj}^{(1)}\beta^{(1)} + x_{tj}^{(2)}\beta_i^{(2)} - \alpha_i p_{tj} + \eta_{ij} + \xi_{tj} + \varepsilon_{itj} \\
 &= [x_{tj}\beta - \alpha p_{tj} + \eta_j + \xi_{tj}] + [x_{tj}^{(2)}\Delta\beta_i^{(2)} - \Delta\alpha_i p_{tj} + \Delta\eta_{ij}] + \varepsilon_{itj} \\
 &= \delta_{tj} + \mu_{itj} + \varepsilon_{itj},
 \end{aligned} \tag{2.1}$$

where $u_{itj} = \delta_{tj} + \mu_{itj}$ with $\delta_{tj} = x_{tj}\beta - \alpha p_{tj} + \eta_j + \xi_{tj}$ being market t -specific mean utility of product $j \in \mathbf{J}_t$ and $\mu_{itj} = x_{tj}^{(2)}\Delta\beta_i^{(2)} - \Delta\alpha_i p_{tj} + \Delta\eta_{ij}$ being an individual i -specific utility deviation from δ_{tj} , while ε_{itj} is an idiosyncratic error term. $x_{tj}^{(1)} \in \mathbb{R}^{K_1}$ is the vector of product characteristics that enter U_{itj} with deterministic coefficient(s), $\beta^{(1)}$, i.e. consumers have homogeneous taste on $x_{tj}^{(1)}$, while $x_{tj}^{(2)} \in \mathbb{R}^{K_2}$ and p_{tj} enter U_{itj} with potentially individual i -specific coefficients, $\beta_i^{(2)}$ and α_i . They capture consumers' heterogeneous tastes on $x_{tj}^{(2)}$ and sensitivities to price change. The term η_{ij} captures individual i 's perception of the quality of product j , with η_j capturing average quality product j and $\Delta\eta_{ij}$ individual deviation from η_j . Any characteristics of product j that does not vary across markets is encapsulated in η_j . ξ_{tj} is a market-product specific demand shock of product j , observed to both firms and individuals but not observed to the econometrician.

Throughout the paper, denote product j being in bundle \mathbf{b} by $j \in \mathbf{b}$. The indirect utility for individual

i in market t from choosing bundle $\mathbf{b} \in \mathbf{C}_{t2}$ is:

$$\begin{aligned}
 U_{it\mathbf{b}} &= \sum_{j \in \mathbf{b}} u_{itj} + \Gamma_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}} \\
 &= \sum_{j \in \mathbf{b}} (\delta_{tj} + \mu_{itj}) + \Gamma_{t\mathbf{b}} + (\Gamma_{it\mathbf{b}} - \Gamma_{t\mathbf{b}}) + \varepsilon_{it\mathbf{b}} \\
 &= \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}} + \left[\sum_{j \in \mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}} \right] + \varepsilon_{it\mathbf{b}} \\
 &= \delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) + \mu_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}},
 \end{aligned} \tag{2.2}$$

where $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}}$ is market t -specific mean utility of bundle \mathbf{b} , $\mu_{it\mathbf{b}}$ is an individual i -specific utility deviation from $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}})$, $\Gamma_{it\mathbf{b}}$ and $\Gamma_{t\mathbf{b}}$ are the individual-market it - and market t -specific *demand synergies* among the products of bundle \mathbf{b} , $\zeta_{it\mathbf{b}}$ is (observed or unobserved) individual deviation from average demand synergies $\Gamma_{t\mathbf{b}}$, and $\varepsilon_{it\mathbf{b}}$ is an idiosyncratic error term. Demand synergies $\Gamma_{it\mathbf{b}}$'s are new parameters that classical demand models of single products do not have. They capture the extra utility individual i obtains from choosing the products in bundle \mathbf{b} 's in market t jointly rather than separately. As one will see in section 2.2.3, these parameters enable to capture rich substitution patterns among products, and in particular, Hicksian complementarity. Moreover, $\Gamma_{it\mathbf{b}}$'s are catch-all parameters and can encompass a wide range of mechanisms that make it more (or less) appealing to choose products jointly than separately. Examples of such mechanisms include discount (or surcharge) on joint purchase of products, shopping cost (one-stop vs multi-stop shoppers), preference for variety of products (single-person vs multi-person households), and synergies in consumption. In section 2.2.2, I will show this point via several examples of demand models in the literature.

Finally, the indirect utility of choosing the outside option 0 is normalized as $U_{it0} = \varepsilon_{it0}$, where ε_{it0} is an idiosyncratic error term. To complete the model, I assume that $\mu_{itj} = x_{tj}^{(2)} \Delta \beta_i^{(2)} - \Delta \alpha_i p_{tj} + \Delta \eta_{ij} = \mu_{tj}(\theta_{it}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t})$ and $\mu_{it\mathbf{b}} = \sum_{j \in \mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}} = \mu_{t\mathbf{b}}(\theta_{it}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t})$ are functions of random coefficients $\theta_{it} = (\Delta \beta_i^{(2)}, \Delta \alpha_i, (\Delta \eta_{ij})_{j \in \mathbf{J}}, (\zeta_{it\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2})$, where $\mathbf{J} = \cup_{t=1}^T \mathbf{J}_t$, $\mathbf{C}_2 = \cup_{t=1}^T \mathbf{C}_{t2}$, and θ_{it} is distributed according to $F \in \Theta_F$.¹⁴ Moreover, ε_{it0} , ε_{itj} , and $\varepsilon_{it\mathbf{b}}$ are assumed to be i.i.d. Gumbel for all $j \in \mathbf{J}_t$ and $\mathbf{b} \in \mathbf{C}_{t2}$.

Denote the vector of market t -specific mean utilities for products in \mathbf{J}_t by $\delta_{t\mathbf{J}_t} = (\delta_{tj})_{j \in \mathbf{J}_t}$, and the vector collecting all average demand synergies by $\Gamma_t = (\Gamma_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$. Define $\delta_t(\Gamma_t) = (\delta_{t\mathbf{J}_t}, (\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}))_{\mathbf{b} \in \mathbf{C}_{t2}})$. The market share function of $\mathbf{b} \in \mathbf{C}_{t1}$ in market t is:¹⁵

$$\begin{aligned}
 s_{t\mathbf{b}}(\delta_t(\Gamma_t); x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, F) &= \int s_{t\mathbf{b}}(\delta_t(\Gamma_t); x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \theta_{it}) dF(\theta_{it}) \\
 &= \int \frac{e^{\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) + \mu_{t\mathbf{b}}(\theta_{it}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t})}}{\sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}(\Gamma_{t\mathbf{b}'}) + \mu_{t\mathbf{b}'}(\theta_{it}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t})}} dF(\theta_{it}),
 \end{aligned} \tag{2.3}$$

where $s_{t\mathbf{b}}(\delta_t(\Gamma_t); x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \theta_{it})$ is individual i 's choice probability of \mathbf{b} in market t given θ_{it} . I then define *product-level* market share function of $j \in \mathbf{J}_t$ as the weighted sum of the market share functions

¹⁴Typically, the distribution of θ_{it} depends on individual i 's demographic characteristics $d_i \in \mathbf{D}$. In this case, F is a mixture of distributions of $\theta_{it}|d_i$: $F = \sum_{d_i \in \mathbf{D}} \pi_t(d_i) F(\cdot|d_i)$, where $\pi_t(\cdot)$ is the distribution of demographics in market t .

¹⁵I abuse the expression $\mathbf{b} \in \mathbf{C}_{t1}$ for both product j and bundle \mathbf{b} .

of \mathbf{b} 's that contain j in market t :

$$\begin{aligned} s_{tj}(\delta_{t\mathbf{J}_t}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma_t, F) &= \sum_{\mathbf{b} \in \mathbf{C}_{t1}} w_{j\mathbf{b}} s_{t\mathbf{b}}(\delta_t(\Gamma_t); x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, F) \\ &= \int \frac{\sum_{\mathbf{b} \in \mathbf{C}_{t1}} w_{j\mathbf{b}} e^{\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) + \mu_{t\mathbf{b}}(\theta_{it}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t})}}{\sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'}(\Gamma_{t\mathbf{b}'}) + \mu_{t\mathbf{b}'}(\theta_{it}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t})}} dF(\theta_{it}), \end{aligned} \quad (2.4)$$

where $w_{j\mathbf{b}}$ is the number of times j appears in \mathbf{b} and known for any $j \in \mathbf{J}_t$ and $\mathbf{b} \in \mathbf{C}_{t1}$. When $j \notin \mathbf{b}$, we have $w_{j\mathbf{b}} = 0$ and the market share of bundle \mathbf{b} does not contribute to the product-level market share of j . When $j \in \mathbf{b}$, then $w_{j\mathbf{b}}$ is a positive integer. In the setting of qualitative choices (i.e. there is no bundle that contains multiple units of the same product) $w_{j\mathbf{b}} = 1$ for $j \in \mathbf{b}$. Then, (2.4) represents the population-level marginal choice probability of j . In the setting of quantity choices (i.e. a bundle may contain multiple units of the same product), $w_{j\mathbf{b}}$ is equal to the units of product j purchased in the form of bundle \mathbf{b} and may be larger than 1. Then, (2.4) represents the population-level total purchases of product j . In both settings, the aggregation in (2.4) is consistent with the aggregate demand data typically available to the researcher.

2.2.2 Examples of Model (2.4) and Interpretation of Demand Synergy Parameters

Model (2.4) is quite general and a wide range of economic models can be formulated via specific restrictions on \mathbf{C}_{t2} and/or on demand synergies $\Gamma_{it\mathbf{b}}$. In this section, I provide some examples in the literature. In particular, I will explain the economic interpretation of $\Gamma_{it\mathbf{b}}$'s in each setting.

Demand for Single Products. This model can be seen as a particular case of (2.4) with $\mathbf{C}_{t2} = \emptyset$, or equivalently, $\Gamma_{it\mathbf{b}} = -\infty$, for all $\mathbf{b} \in \mathbf{C}_{t2}$. This restriction on $\Gamma_{it\mathbf{b}}$ rules out simultaneous choices of more than one single product and restricts products to be Hicksian substitutes.

Multiple Choice of Products, within Category. [Gentzkow \(2007\)](#) considers household's choice over bundles of at most 2 different newspapers: $\mathbf{C}_{t2} = \{(j_1, j_2) : j_1 < j_2, j_1, j_2 \in \mathbf{J}_t\}$ and $\mathbf{C}_t = \{0\} \cup \mathbf{J}_t \cup \mathbf{C}_{t2}$. In general, one can allow for choice over bundles of up to K different products: $\mathbf{C}_{t2} = \{(j_1, \dots, j_k) : j_1 < \dots < j_k, j_1, \dots, j_k \in \mathbf{J}_t\}$. As shown in [Iaria and Wang \(2019a\)](#), demand synergy $\Gamma_{it\mathbf{b}}$ can proxy various mechanisms, including preference for variety, shopping cost, synergies in consumption.

Multiple Choice of Products, across Categories. [Grzybowski and Verboven \(2016\)](#) and [Ershov et al. \(2018\)](#) consider choice over bundles of products across different categories. In the simplest case where a bundle is defined as a collection of 2 different products (chips and soda) with each belonging to a different category (salty snacks and carbonated drinks), we have $\mathbf{C}_{t2} = \mathbf{J}_1 \times \mathbf{J}_2 = \{(j_1, j_2) : j_1 \in \mathbf{J}_1, j_2 \in \mathbf{J}_2\}$. In the example of potato chips and carbonated soda ([Ershov et al., 2018](#)), $\Gamma_{it\mathbf{b}}$'s are interpreted as synergies in consumption.

Quantity Choice: Multiple Units. As a deviation from demand models for single products, individuals purchase not only one out of J products but also a discrete quantity of the chosen product. This can be captured by $\mathbf{C}_{t2} = \{(j, \dots, j) : j \in \mathbf{J}, \text{the length of } (j, \dots, j) \leq L\}$, where L is the maximal

units individuals can purchase. In the simplest case, individuals can purchase the outside option 0, a unit of product $j \in \mathbf{J}$ (single product), or a bundle of two identical units (j, j) , $j \in \mathbf{J}$. Demand synergy $\Gamma_{it(j,j)}$ is then interpreted as extra utility individual i 's obtains from the purchased additional unit of product j relative to the first unit: $\Gamma_{it(j,j)} < 0 (> 0)$ implies a decreasing (increasing) marginal utility of purchasing product j . If $\Gamma_{it(j,j)} = 0$, then individual i 's utility from purchasing the second unit of product j remains the same as that from the first unit.

Quantity Choice: Multiple Discreteness. Demand model of multiple discreteness (see [Hendel \(1999\)](#) and [Dubé \(2004\)](#)) can be seen as a further extension of demand for multiple units that includes bundles defined as a collection of multiple units of different products: $\mathbf{b} = (\underbrace{(j, \dots, j)}_{n_j})_{j \in \mathbf{J}}$, where

n_j is the number of units of product j . As shown in [Iaria and Wang \(2019a\)](#) (Appendix 8.1), [Dubé \(2004\)](#)'s model of multiple discreteness can be formulated by specifying $\Gamma_{it\mathbf{b}} = \sum_{j \in \mathbf{J}} \Gamma_{it(j, \dots, j)}$, where $\Gamma_{it(\underbrace{j, \dots, j}_{n_j})} \leq 0$ for any $n_j > 1$ and $j \in \mathbf{J}$. The non-positive $\Gamma_{it(j, \dots, j)}$ represents non-increasing marginal utility of consuming additional units of product j during one consumption moment and the additivity in $\Gamma_{it\mathbf{b}}$ across $j \in \mathbf{J}$ represents the independence between consumption moments.¹⁶

Multi-Category Multi-Store Demand. [Thomassen et al. \(2017\)](#) studies a multi-category multi-store demand model, where individual purchases multiple units in each of K product categories and purchase all the units of the same category in the same store. Consider the simplest case where individual purchases at most one unit in each of 2 product categories (k_1 and k_2) from 2 stores (S_1 and S_2). This can be captured by $\mathbf{J} = \{j = (j^1, j^2) : j^1 = k_1, k_2, j^2 = S_1, S_2\}$ and $\mathbf{C}_2 = \{(j, r) : j, r \in \mathbf{J}, j^1 \neq r^1\}$. A product is defined as a Cartesian product of categories and stores with first coordinate being category and the second being store (category 1 in store 2) and a bundle is defined as a Cartesian product of two products that differ in their first coordinate (category 1 in store 2 and category 2 in store 2). Demand synergy $\Gamma_{it(j,r)}$ is interpreted as shopping cost due to store choice: $\Gamma_{it(j,r)} = 0$ if $j^2 = r^2$ (purchase products of both categories in the same store), and negative otherwise (purchase products of one category in store 1 and those of the other in store 2).

2.2.3 Demand Synergies and Hicksian Substitutions

One prominent feature of model (2.4) is that it allows for flexible substitution patterns in demand via demand synergy parameters. In particular, the cross-price elasticities in model (2.4) can be negative, i.e. Hicksian complementarity, which is ruled out by demand models of single products.¹⁷ In this section, I will elaborate this point in a model of multiple choice of products within category: $\mathbf{C}_2 = \{(j_1, j_2) : j_1 < j_2, j_1, j_2 \in \mathbf{J}\}$. Similar analysis can be conducted with other types of models (2.4).

To ease exposition, I drop the notation of market t and product characteristics in the product-level market share functions. We then compute the derivative of the product-level market share prices with respect to prices:

$$\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} = - \int \alpha_i \frac{\partial s_{i\mathbf{J}}}{\partial \delta_{\mathbf{J}}} dF(\theta_i),$$

¹⁶Due to [Dubé \(2004\)](#)'s perfect substitute specification, individual will consume up to one product during one consumption moment.

¹⁷Due to linear index structure, there is no income effect and therefore negative (positive) cross-price elasticities are interpreted as Hicksian complementarity (substitutability).

where $\frac{\partial s_{i\mathbf{J}}}{\partial \delta_{\mathbf{J}}} = \left(\frac{\partial s_{ij}}{\partial \delta_r} \right)_{j,r \in \mathbf{J}}$ and

$$\frac{\partial s_{ij}}{\partial \delta_r} = -[s_{ij} \cdot s_{ir} - s_{ijr}] = -[s_j(\delta(\Gamma); \theta_i) s_r(\delta(\Gamma); \theta_i) - s_{jr}(\delta(\Gamma); \theta_i)].$$

The cross-price elasticity between j and r is:

$$\varepsilon_{jr} = \frac{p_r}{s_j} \int \alpha_i [s_j(\delta(\Gamma); \theta_i) s_r(\delta(\Gamma); \theta_i) - s_{jr}(\delta(\Gamma); \theta_i)] dF(\theta_i).$$

See Appendix B.1 for details of computation.

Different from models of demand for single products, the cross-price elasticity ε_{jr} has an additional term $-s_{jr}(\delta(\Gamma); \theta_i)$. When this term is relatively large, i.e. the joint choice probability for products j and r is relatively large, we may have a negative ε_{jr} , i.e. Hicksian complementarity between j and r . In the case of two products and one bundle, i.e. $\mathbf{J} = \{1, 2\}$ and $\mathbf{C}_2 = \{(1, 2)\}$, [Gentzkow \(2007\)](#) shows that $\Gamma_{jr} = 0$ is the cut-off value for Hicksian substitutability and complementarity: $\varepsilon_{12} < 0$ if and only if $\Gamma_{(1,2)} > 0$. When there are more than 2 products, even though $\Gamma_{(j,r)} = 0$ may not be the cut-off value for Hicksian substitute or complementary between j and r , similar intuition remains valid. To see this, note that whether j and r are substitute, complementary or independent, i.e. $\varepsilon_{jr} > 0$, $\varepsilon_{jr} < 0$ or $\varepsilon_{jr} = 0$, is determined by the weighted average of $s_{ij} \cdot s_{ir} - s_{ijr}$. The latter is further determined by the magnitude of synergy parameter between j and r , Γ_{jr} , relative to other demand synergy parameters. If Γ_{jr} is sufficiently negative, then s_{ijr} is close to zero and thus $\varepsilon_{jr} > 0$. As an extreme case, when $\Gamma_{jr} = -\infty$, i.e. bundle (j, r) is not in the choice set, j and r are always substitute and therefore ε_{jr} is positive. If Γ_{jr} is positive and large enough relative to $\Gamma_{j'r'}$ for all $(j', r') \neq (j, r)$, then $s_{ij} - s_{ijr}$ and $s_{ir} - s_{ijr}$ are negligible relative to s_{ijr} . Then, the sign of ε_{jr} is determined by the population average of $s_{ijr}^2 - s_{ijr}$. Since s_{ijr} is strictly between 0 and 1, $s_{ijr}^2 - s_{ijr}$ is always negative and therefore $\varepsilon_{jr} < 0$. If Γ_{jr} takes some medium value in $(-\infty, \infty)$, we can expect $\varepsilon_{jr} = 0$ and therefore j and r are independent.

2.3 Identification

I first give the assumptions the identification and estimation will rely on.

Assumption 6. For any $t \in \mathbf{T}$,

- (i). (Data availability) The product-level market shares, $s_{tj} = \sum_{\mathbf{b} \in \mathbf{C}_{t1}} w_{j\mathbf{b}} s_{t\mathbf{b}}$, are observed to the econometrician for $j \in \mathbf{J}_t$.
- (ii). (Mix and match) If bundle $\mathbf{b} \in \mathbf{C}_{t2}$, then $j \in \mathbf{J}_t$, for any $j \in \mathbf{b}$.
- (iii). (Many-market) The total number of products, $|\cup_{t=1}^T \mathbf{J}_t|$, and bundles $|\cup_{t=1}^T \mathbf{C}_{t2}|$, are fixed while the number of markets, $T = |\mathbf{T}|$, is large.

Assumption 6(i) specifies situations where only product-level (rather than bundle-level) market shares are available to the econometrician. To simplify the exposition, I assume that product-level market shares are observed without error, i.e. the number of individuals in each market is infinite. In estimation, one can allow for the number of individuals to increase fast enough relative to the number

of markets and the main results of the paper still hold.¹⁸ Assumption 6(ii) clarifies that bundles are formed because individuals *mix and match* products, i.e. a bundle is defined as a set of products chosen by individuals. The definition of product may vary from application to application. If some single products are only sold together (e.g. business-class flight is only available via bundle of business-class seat and large allowance of luggage), as long as purchase data on such combination is available, i.e. Assumption 6(i) holds, then one can define such combination as a product and Assumption 6(ii) is not violated. Finally, Assumption 6(iii) focuses on the *many-market* setting where the numbers of products and bundles are fixed while the number of markets increases.

As clarified in Assumption 6(iii), I consider the many-market setting in this paper. As a result, without further restrictions, as T increases, the number of demand synergy parameters to be identified (i.e. Γ_t for all $t \in \mathbf{T}$) increases simultaneously. This challenge of dimensionality introduces substantial difficulty in identification and incidental parameters in estimation. To overcome this challenge, I propose the following assumption along the lines of [Gentzkow \(2007\)](#)'s model of demand for bundles (and also its generalized version in [Iaria and Wang \(2019a\)](#)):

Assumption 7. For any bundle $\mathbf{b} \in \mathbf{C}_{t2}$ and $t \in \mathbf{T}$,

$$\Gamma_{t\mathbf{b}} = g(x_{t\mathbf{b}}; \Sigma_g) + \Gamma_{\mathbf{b}},$$

where $x_{t\mathbf{b}}$ a vector of observed market-bundle specific non-price characteristics, $g(\cdot; \Sigma_g)$ a function of $x_{t\mathbf{b}}$ parametrized by and continuously differentiable with respect to $\Sigma_g \in \Theta_{\Sigma_g}$, and $\Gamma_{\mathbf{b}}$ is a bundle-specific fixed effect.

Assumption 7 reduces the dimension of the demand synergy parameters to the sum of $\dim(\Sigma_g)$ and $\dim(\Gamma) = \dim((\Gamma_{\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}) = |\mathbf{C}_2|$, which remains fixed as T increases. The main motivation for this assumption is that the bundle-level market shares are not observed to the econometrician. If they were all available, then one could directly identify and estimate model (2.3), rather than model (2.4), à la BLP with bundle-level instruments and then Assumption 7 is not required.

Different from the model used in [Iaria and Wang \(2019a\)](#), Assumption 7 assumes linear pricing in the factual, i.e. the observed price of a bundle is the sum of the prices of its single products. This excludes nonlinear pricing in the factual, i.e. bundle-specific discounts or surcharges. While it is possible to extend the main results in this paper to allow for nonlinear pricing in the factual, I focus on the situations covered by Assumption 7 and will explore this extension in future research.

Note that even with Assumption 7, one can still simulate counterfactuals under nonlinear pricing. In such counterfactuals, this assumption implies that the source of unobserved variations across markets is limited to the market-product specific demand shocks $\xi_{t\mathbf{J}_t}$. Then, conditional on the observed characteristics of products and bundles, prices vary across markets only due to variations in $\xi_{t\mathbf{J}_t}$.

Assumption 7 summarizes various economic situations with or without exogenous characteristics of bundles. Both situations can be similarly treated in the following identification and estimation discussion. To simplify the exposition, I will focus on the leading case $g \equiv 0$, i.e. $\Gamma_{t\mathbf{b}} = \Gamma_{\mathbf{b}}$.

¹⁸In models of demand for single products, [Freyberger \(2015\)](#) allows for sampling errors in the observed market shares. He shows the consistency and asymptotic normality of the GMM estimator by requiring the number of individuals to increase fast enough relative to the number of markets.

2.3.1 Economic Analyses and Sufficient Demand Primitives

The demand primitives to be identified are motivated by research questions. In the context of demand for bundles, researchers are usually interested in economic analyses under linear and nonlinear pricing strategies. Under linear pricing strategy, firms only set prices of their single products; the price of a bundle is defined as the sum of the prices of its single products. Under nonlinear pricing strategy, firms can not only set prices of its single products but also on the bundles of their own products. Then, the price of a bundle can be different from the sum of the prices of its products when there is a discount or surcharge. Identifying the full structural parameters does allow to conduct all these analyses. However, this may be overly sufficient. The next proposition shows precisely which demand primitives are needed to conduct the two kinds of analyses in merger simulations.

Proposition 1. *Suppose that Assumptions 6-7 hold and $\alpha_i = \alpha$.¹⁹ Also suppose that the observed prices $p_{t\mathbf{J}_t}$ and those after mergers are generated from a simultaneous Bertrand price-setting game under complete information with constant marginal cost c_{tj} for $j \in \mathbf{J}_t$, and are unique.*

- *If (α, β, η) and $s_{t\mathbf{J}_t}(\cdot; x_{t\mathbf{J}_t}^{(2)}, \Gamma, F)$ are identified, then,*
 - *Price elasticities at $p_{t\mathbf{J}_t}$ are identified.*
 - *c_{tj} 's are identified.*
 - *The changes of prices, profits, consumer surplus, social welfare before and after the merger under linear pricing are identified.*
- *If Γ and $s_{t\mathbf{b}}(\cdot; x_{t\mathbf{J}_t}^{(2)}, F)$, for any $\mathbf{b} \in \mathbf{C}_{t1}$ are further identified, and moreover, $c_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} c_{tj}$ for $\mathbf{b} \in \mathbf{C}_{t2}$, then, the changes of prices, profits, consumer surplus, social welfare before and after the merger under nonlinear pricing are identified.*

Proof. See Appendix B.2. □

Remark 4. *The condition $c_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} c_{tj}$ implies that there is no additional cost for firms to set bundle-specific prices. The second statement of Proposition 1 still holds if there is such additional cost and it is known to the researcher.*

Proposition 1 specifies the set of sufficient demand primitives with which the researcher can simulate mergers under linear and nonlinear pricing, respectively. Identifying (α, β, η) and $s_{t\mathbf{J}_t}(\cdot; x_{t\mathbf{J}_t}^{(2)}, \Gamma, F)$ is enough for merger simulations under linear pricing. For those under nonlinear pricing, it suffices to further identify Γ and $s_{t\mathbf{b}}(\cdot; x_{t\mathbf{J}_t}^{(2)}, F)$ for all $\mathbf{b} \in \mathbf{C}_{t2}$. In what follows, I will organise the identification discussion in two parts. In the first part, I discuss the identification of (α, β, η) and $s_{t\mathbf{J}_t}(\cdot; x_{t\mathbf{J}_t}^{(2)}, \Gamma, F)$; in the second part, I assume the identification of (α, β, η) and $s_{t\mathbf{J}_t}(\cdot; x_{t\mathbf{J}_t}^{(2)}, \Gamma, F)$ and continue to identify Γ and $s_{t\mathbf{b}}(\cdot; x_{t\mathbf{J}_t}^{(2)}, F)$ for $\mathbf{b} \in \mathbf{C}_{t2}$.

2.3.2 Identification of Product-Level Market Share Functions

I follow the classical approach in demand models of aggregate market shares and use a two-step identification strategy. In the first step, I recover the mean utilities of products using a novel demand

¹⁹Because $\alpha_i = \alpha$, prices enter the indirect utilities only via the mean utilities. In this proposition and the proof, I drop the notation $p_{t\mathbf{J}_t}$ from $s_t(\cdot; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma, F)$.

inverse to deal with possible Hicksian complementarities among products; in the second step, I construct moment conditions using IVs to deal with endogenous prices and identify product-level demand primitives.

Demand Inverse in Model (2.4) with Complementarity. Let $\mathbf{J} = \cup_{t=1}^T \mathbf{J}_t$ denote the set of all available single products, and $\mathbf{C}_2 = \cup_{t=1}^T \mathbf{C}_{t2}$ the set of all available bundles. The first step hinges on the invertibility of product-level market share functions:

Theorem 7. (*Theorem Demand Inverse of Iaria and Wang (2019a)*)

Suppose that Assumption 6-7 holds. Then, for any (Γ', F') , there exists at most one $\delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma', F') = \mathcal{J}_{t\mathbf{J}_t}$.

When (Γ', F') are the true parameters (Γ, F) , the vector of the true mean utilities of products, $\delta_{t\mathbf{J}_t}$, is the unique solution of $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma, F) = \mathcal{J}_{t\mathbf{J}_t}$. As a result, the function $s_{t\mathbf{J}_t}(\cdot; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma, F)$ is globally invertible. Denote its inverse by:

$$\delta_{t\mathbf{J}_t} = s_{t\mathbf{J}_t}^{-1}(\mathcal{J}_{t\mathbf{J}_t}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma, F). \quad (2.5)$$

Iaria and Wang (2019a) first proved this result and use it to reduce the dimensionality of fixed effects in a likelihood estimation procedure.²⁰ In this paper, it is used as a fundamental identification argument and an essential step to form moment conditions when only product-level market shares are available. There are two key differences relative to the classical demand inverse in demand models of single products. First, the invertibility of market share functions in demand models of single products follows from the connected substitutes conditions (Berry et al., 2013) which require the products to be Hicksian substitutes. These conditions may not apply to model (2.4) because the products can be Hicksian complementary. The invertibility of product-level market share functions in Theorem 7 is built on the affine relationship between the utilities of bundles and single products (see equation (2.2)) and on the P-matrix property by Gale and Nikaido (1965), which-crucially-does not require the products to be Hicksian substitutes. Second, the demand inverse in Theorem 7 may not be implemented by the fixed-point contraction mapping algorithm proposed by Berry et al. (1995). This is because the contraction mapping property of the algorithm may not hold when (some) products are Hicksian complementary in model (2.4). I propose to use Jacobian-based algorithms to implement this demand inverse.²¹ See section 2.4.2 for details of the implementation.

When $(\Gamma', F') \neq (\Gamma, F)$, it is possible that there is no $\delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma', F') = \mathcal{J}_{t\mathbf{J}_t}$.²² In this case, such (Γ', F') are directly ruled out of the identification set of (Γ, F) . In what follows, the identification discussion will restrict to (Γ', F') such that $\delta'_{t\mathbf{J}_t}$ exists.

Instrumental Variable Approach. To clarify the source of identification, I suppose that $\mathbf{J}_t = \mathbf{J}$ and $\mathbf{C}_{t2} = \mathbf{C}_2$ for any $t \in \mathbf{T}$, i.e. there is no variation in the set of products and bundles across

²⁰See their Theorem MLE and implementation section.

²¹Conlon and Gortmaker (2019) provide a review of numerical methods for implementation of demand inverse in demand models of single products.

²²For example, if the data generating process is such that the sum of the observed product-level market shares is larger than one, then any demand models of single products ($\Gamma_{\mathbf{b}} = -\infty$ for any \mathbf{b}) cannot rationalize the observed product-level market shares and hence the demand inverse is not feasible with $\Gamma' = -\infty$.

markets. Combining equation (2.5) and $\delta_{t\mathbf{J}} = x_{t\mathbf{J}}\beta - \alpha p_{t\mathbf{J}} + \eta_{\mathbf{J}} + \xi_{t\mathbf{J}}$, I obtain:

$$x_{t\mathbf{J}}\beta - \alpha p_{t\mathbf{J}} + \eta_{\mathbf{J}} + \xi_{t\mathbf{J}} = s_{\mathbf{J}}^{-1}(s_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F). \quad (2.6)$$

The source of price endogeneity is $\xi_{t\mathbf{J}}$: $\xi_{t\mathbf{J}}$ are observed to firms and therefore $p_{t\mathbf{J}}$ are set based on $\xi_{t\mathbf{J}}$. Consequently, $p_{t\mathbf{J}}$ and $\xi_{t\mathbf{J}}$ are correlated, while $\xi_{t\mathbf{J}}$ are not observed to the econometrician. Beyond the price endogeneity, Γ and F constitute parameters that cannot be pinned down without further assumption. I use IVs to solve these challenges:

Assumption 8. *There are random variables $z_{t\mathbf{J}} = (z_{tj})_{j \in \mathbf{J}}$, such that $\mathbb{E}[\xi_{t\mathbf{J}} | z_{t\mathbf{J}}, x_{t\mathbf{J}}] = 0$ almost everywhere.*

Assumption 8 gives rise to conditional moment restrictions:

$$\mathbb{E}[\xi_j(\beta, \alpha, \eta, \Gamma, F; s_{t\mathbf{J}}, x_{t\mathbf{J}}, p_{t\mathbf{J}}) | z_{t\mathbf{J}}, x_{t\mathbf{J}}] = 0 \quad a.e., \quad (2.7)$$

for $j \in \mathbf{J}$, where $\xi_j(\beta, \alpha, \eta, \Gamma, F; s_{t\mathbf{J}}, x_{t\mathbf{J}}, p_{t\mathbf{J}}) = s_j^{-1}(s_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) - x_{tj}\beta + \alpha p_{tj} - \eta_j$. The identification of $s_{\mathbf{J}}(\cdot; \Gamma, F)$ (or equivalently its inverse $s_{\mathbf{J}}^{-1}(\cdot; \Gamma, F)$) by moment conditions (2.7) can follow from general arguments in nonlinear models using IVs. In demand models of single products, one can leverage completeness conditions of joint distribution of $(z_{t\mathbf{J}}, x_{t\mathbf{J}}, s_{t\mathbf{J}}, p_{t\mathbf{J}})$ with respect to $(s_{t\mathbf{J}}, p_{t\mathbf{J}})$ (Berry and Haile, 2014). Intuitively, this requires sufficiently rich variation in $(z_{t\mathbf{J}}, x_{t\mathbf{J}})$ that can distinguish any function of endogenous variables $(s_{t\mathbf{J}}, p_{t\mathbf{J}})$ from others. In the context of (2.7), the same general arguments also apply. I need variation in $(z_{t\mathbf{J}}, x_{t\mathbf{J}})$ to distinguish $\xi_{\mathbf{J}}(\beta, \alpha, \eta, \Gamma, F; \dots)$ from $\xi_{\mathbf{J}}(\beta', \alpha', \eta', \Gamma', F'; \dots)$ for any $(\beta', \alpha', \eta', \Gamma', F') \neq (\beta, \alpha, \eta, \Gamma, F)$. As long as such variation is available, having demand synergy parameters Γ does not conceptually introduce additional difficulty for identification.

Despite the generality, these arguments and required conditions are often high-level. In what follows, I leverage usual economic settings and propose *low-level* sufficient conditions for the identification of market-level market share functions. To simplify the exposition, I will focus on cost-type variables and take product characteristics $x_{t\mathbf{J}}$ as fixed.²³ In Appendix B.9, I propose similar sufficient conditions for other commonly used instruments: BLP-type instruments, exogenous product characteristics.

Suppose that the ownership of each product is the same across markets and that prices are generated from a simultaneous Bertrand pricing game under complete information with constant marginal cost c_{tj} for $j \in \mathbf{J}$. Without loss of generality, I specify $c_{tj} = z_{tj} + w_{tj}$, where z_{tj} is cost shifter for product j and w_{tj} is exogenous supply shock that is observed to firms but not observed to the econometrician. The main identification result of the product-level market share functions is the following:

Theorem 8. *Suppose that Assumptions 6-8 and regularity condition 2 of Appendix B.3 holds. Moreover, the following conditions hold:*

1. $z_{t\mathbf{J}}$ is independent of $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$ and the support of $z_{t\mathbf{J}}$ is \mathbb{R}^J .
2. $\alpha_i = \alpha \neq 0$
3. Given $x^{(2)}$, $p_{t\mathbf{J}} = p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}; x^{(2)})$ is a function of $(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$.

²³Common examples of cost-type variables and its proxies are input prices, variables correlated with marginal costs, prices of the same products in other markets (e.g., Hausman-type instruments).

4. For any $(\Gamma', F') \neq (\Gamma, F)$, there exists $\delta'_J \neq \delta''_J$ such that $s_{\mathbf{J}}(\delta'_J; x^{(2)}, \Gamma, F) = s_{\mathbf{J}}(\delta'_J; x^{(2)}, \Gamma', F')$ and $s_{\mathbf{J}}(\delta''_J; x^{(2)}, \Gamma, F) = s_{\mathbf{J}}(\delta''_J; x^{(2)}, \Gamma', F')$.

Then,

- If (ξ, w) is Gaussian distributed, then α and $s_{t\mathbf{J}}(\cdot; x^{(2)}, \Gamma, F)$ are identified.
- Suppose that the data generating process is a model of multiple choice of products across categories (see section 2.2.2). If the distribution F has compact support, then under regularity condition 3 of Appendix B.3, α and $s_{t\mathbf{J}}(\cdot; x^{(2)}, \Gamma, F)$ are identified.

Remark 5. The two statements of Theorem 8 are complementary: the first statement achieves the identification by restricting the distribution of demand and supply shocks and remains agnostic on the distribution of random coefficients.²⁴ While, the second statement restricts the distribution of the random coefficients to have compact support and does not posit on the distribution of (ξ, w) .

Remark 6. Once α and $s_{t\mathbf{J}}(\cdot; x^{(2)}, \Gamma, F)$ are identified for any $x^{(2)}$, $\delta_{t\mathbf{J}}$ can be recovered from the demand inverse in Theorem 7. Then, the identification of β and η follows from standard linear IV arguments.

Proof. See Appendix B.3. □

The first condition reinforces Assumption 8 to strong exogeneity of cost shifters $z_{t\mathbf{J}}$ and assumes large support of $z_{t\mathbf{J}}$. The second condition simplifies the price coefficient to be homogeneous for all individuals but still allows for random coefficients on other product characteristics. The third condition imposes the uniqueness of the Bertrand price competition and is typically necessary to simulate counterfactuals in empirical research. The fourth condition requires that the image sets of two different product-level market share functions intersect at two different vectors of the mean utilities of products. This is a mild requirement. In fact, in the case of demand models of single products, these image sets are identical and equal to $\{(\delta'_j)_{j \in \mathbf{J}} : \sum_{j \in \mathbf{J}} \delta'_j < 1, \delta'_j > 0\}$.

When model (2.4) degenerates to a demand model of single products, i.e. $\Gamma = -\infty$, Theorem 8 implies the identification of α and the market share functions of products in classic models of demand. Then, one can invert the market share functions and recover the mean utilities of products. The identification of β and η follows by applying standard linear IV arguments.

Also relying on product-level choice data, the main result of [Allen and Rehbeck \(2019a\)](#) implies the identification of product-level market share functions in the context of model (2.4) with additive separable unobservable heterogeneity. While their identification strategy crucially relies on the assumption of additively separable unobservable heterogeneity and does not allow for endogenous prices, I exploit exogenous variation in cost shifters and product characteristics to deal with price endogeneity and achieve the identification of product-level market share functions.

2.3.3 Identification of Bundle-Level Market Share Functions

In this section, I assume that the product-level market share functions are identified and aim to identify Γ and $s_{t\mathbf{b}}(\cdot; x_{\mathbf{J}}^{(2)}, F)$.²⁵ It suffices to separably identify demand synergies Γ and the distribution

²⁴The identification in Theorem 8 can also be achieved when the distribution of (ξ, w) has “fat tail”. See [Mattner \(1992\)](#) and [D’Haultfoeuille \(2011\)](#) for details.

²⁵Note that the results in this section do not require $\alpha_i = \alpha$ as in Proposition 1. Without loss of generality, $x_{i_j}^{(2)}$ will also include the price variable.

of random coefficients, F . The key challenge of this task is that only product-level market shares are observed. I first provide identification results for a class of models widely used in the empirical literature.²⁶

Theorem 9. *Suppose that $\mathbf{C}_2 = \{(j, j') : j < j', j, j' \in \mathbf{J}\}$, or $\mathbf{C}_2 = \{(j_1, j_2) : j_1 \in \mathbf{J}_1, j_2 \in \mathbf{J}_2\}$, $\Gamma_{it\mathbf{b}} = \Gamma_{\mathbf{b}}$ for $\mathbf{b} \in \mathbf{C}_2$, and $s_{\mathbf{J}}(\cdot; x_{t\mathbf{J}}^{(2)}, \Gamma, F)$ is identified in \mathbb{R}^J . Then, Γ and $s_{\mathbf{b}}(\cdot; x_{t\mathbf{J}}^{(2)}, F)$ are identified in \mathbb{R}^{C_1} for any $\mathbf{b} \in \mathbf{C}_1$.*

Proof. See Appendix B.4. □

Remark 7. *If for some bundle \mathbf{b} , the true $\Gamma_{\mathbf{b}}$ is equal to $-\infty$, i.e. bundle \mathbf{b} is not in the choice set, then Theorem 9 implies that $\Gamma_{\mathbf{b}} = -\infty$ is identified.*

Theorem 9 shows that product-level demand data already suffices to identify bundle-level demand primitives (synergy parameters and bundle-level market share functions) in models of multiple choices of products within/across categories. Consequently, researchers are able to conduct the nonlinear pricing counterfactuals in Proposition 1 using these models. It is worthy pointing out that Theorem 9 does not automatically imply that F is identified. In BLP-type models of demand, this further step of identification of F can be achieved under mild conditions. For example, Wang (2020) (Theorems 2 and 3) show that it suffices to have a *single* variation in $x_t^{(2)}$ across markets. I refer to that paper for more details.

However, the separable identification in Theorem 9 may not be achieved in some other kinds of model (2.4). The following corollary gives an example.

Corollary 3 (Non-separable identification of Γ and $s_{\mathbf{b}}(\cdot; F)$). *Suppose that the data generating process is a model of multi-unit choice: $\mathbf{J} = \{1\}$ and $\mathbf{C}_2 = \{(1, 1)\}$, $\Gamma_{i(1,1)} = \Gamma > -\infty$. Moreover, the product-level market share function:*

$$s_{1.}(\delta; \Gamma, F) = \int \frac{e^{\delta+\mu} + 2e^{2\delta+2\mu+\Gamma}}{1 + e^{\delta+\mu} + e^{2\delta+2\mu+\Gamma}} dF(\mu). \quad (2.8)$$

is identified. Then, there exists (Γ, F) such that the Γ and $s_{(1,1)}(\cdot; F)$ are not separably identified.

Proof. See Appendix B.5. □

Corollary 3 illustrates the limited power of product-level market shares in models of multi-unit choice to separably identify Γ and F . Intuitively, one cannot distinguish Γ and $s_{(1,1)}(\cdot; F)$ because it is impossible to shift the mean utility of the first unit without shifting that of the second unit. When bundle-level demand data is available, Iaria and Wang (2019a) shows how to identify and estimate model of demand for bundles by exploring the same bundle-specific fixed effects $\Gamma_{\mathbf{b}}$ across markets. This gives rise to additional moment restrictions that separately identify Γ and $s_{\mathbf{b}}(\cdot; F)$. With only product-level demand data, this source of identification is no longer available in model (2.8). As a consequence, unless imposing further assumptions on synergy parameters or the distribution of the random coefficients, the availability of bundle-level demand data may be necessary to disentangle Γ and F and to conduct nonlinear pricing counterfactuals in models of multi-unit choice.²⁷

²⁶See Gentzkow (2007), Fan (2013), Kwak et al. (2015), Grzybowski and Verboven (2016) for example.

²⁷For the identification with bundle-level demand data, see Fox and Lazzati (2017), Allen and Rehbeck (2019b), and Iaria and Wang (2019a).

2.3.4 Summary of Identification

In section 2.3.2, I provide sufficient conditions for the identification of product-level market share functions using IVs. In section 2.3.3, I study the identification of bundle-level market share functions assuming the identification of the product-level market share functions. The proposed sufficient conditions may be overly sufficient. To complete the discussion, I provide a complementary but non-constructive approach built on moment conditions (B.2). See Appendix B.10 for details.

2.4 Estimation and Implementation

In this section, I propose a GMM estimation procedure for model (2.4) and discuss its implementation. The proposed estimation procedure is conceptually similar to that used in BLP models of single products. However, due to the nature of bundle choice, the implementation has non-trivial challenges. I consider parametric estimation of model (2.4). Concretely, F is characterized by $\Sigma \in \Theta_\Sigma \subset \mathbb{R}^P$. Define the true value of parameter vector as $\theta_0 = (\alpha_0, \beta_0, \eta_0, \Sigma_0, \Gamma_0)$. I also assume that $(x_{t\mathbf{J}_t}, z_{t\mathbf{J}_t})$ are valid instruments and θ_0 is identified.

2.4.1 Estimation Procedure

I construct unconditional moment conditions from (B.2) using a finite set of functions of $(x_{t\mathbf{J}_t}, z_{t\mathbf{J}_t})$, $\Phi = \{\phi_g(x_{t\mathbf{J}_t}, z_{t\mathbf{J}_t})\}_{g=1}^G$:

$$m(\theta'; \{s_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}, x_{t\mathbf{J}_t}, z_{t\mathbf{J}_t}\}_{t=1}^T, \Phi) = (\mathbb{E}[\xi_{tj}(\beta', \alpha', \eta', \Gamma', F'; s_{t\mathbf{J}_t}, x_{t\mathbf{J}_t}, p_{t\mathbf{J}_t})\phi_g(z_{t\mathbf{J}_t}, x_{t\mathbf{J}_t})])_{g=1}^G,$$

The finite-sample counterparts are:

$$m_T(\theta'; \{s_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}, x_{t\mathbf{J}_t}, z_{t\mathbf{J}_t}\}_{t=1}^T, \Phi) = \left(\frac{1}{T} \sum_{t=1}^T \frac{1}{J_t} \sum_{j=1}^{J_t} [s_{tj}^{-1}(s_{t\mathbf{J}_t}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma', \Sigma') - x_{tj}\beta' + \alpha'p_{tj} - \eta'_j] \phi_g(x_{tj}, z_{tj}) \right)_{g=1}^G. \quad (2.9)$$

Then, the GMM estimator of θ_0 , $\hat{\theta}_T^{GMM}$, is defined as:

$$\hat{\theta}_T^{GMM} = \arg \min_{\theta' \in \Theta} m_T(\theta'; \{s_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}, x_{t\mathbf{J}_t}, z_{t\mathbf{J}_t}\}_{t=1}^T, \Phi)^T W_T m_T(\theta'; \{s_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}, x_{t\mathbf{J}_t}, z_{t\mathbf{J}_t}\}_{t=1}^T, \Phi), \quad (2.10)$$

where Θ is a compact set and $W_T \in \mathbb{R}^{G \times G}$ is a weighting matrix that converges to a positive-definite matrix W in probability. If θ_0 lies in the interior of Θ , then under standard regularity conditions (see Newey and McFadden (1994)), $\hat{\theta}_T^{GMM}$ is consistent and asymptotically normal.²⁸

A basic requirement for the good finite-sample performance of (2.10) is that we have at least as many moment conditions as the dimension of $(\alpha_0, \beta_0, \eta_0, \Gamma_0, \Sigma_0)$. In particular, we have $\dim(\Gamma_0)$ demand synergy parameters in (2.10) that BLP models of single products do not have. Therefore, we need at least $\dim(\Gamma_0)$ more moment conditions. If the number of valid instruments or variability of these instruments is limited, then one can also specify Σ_0 and reduce its dimensionality according to the economic setting. For example, two products of the same producer, or of similar nutrition

²⁸If some parameters (e.g., distributional parameters Σ) are on the boundary, the GMM estimator may not be asymptotically normal. See Ketz (2019) for an inference procedure that is valid when distributional parameters are on the boundary and Andrews (2002) for a general treatment.

ingredients, may have greater or smaller synergies. Then, one can specify the demand synergy among the two products as a function of the distance between the characteristics of the products.

In BLP models of single products, a suggested practice is to approximate the optimal instruments in the form of Amemiya (1977) and Chamberlain (1987) that achieve the semi-parametric efficiency bound. Reynaert and Verboven (2014) and Conlon and Gortmaker (2019) report significant gain by using Berry et al. (1995)'s GMM estimator with optimal instruments. However, the difficulty of approximating optimal instruments still remains in the estimation procedure (2.10). A good approximation of optimal instruments relies on the knowledge of the true parameters. Moreover, when the number of products is large, even low order of such approximation may be subject to a curse of dimensionality and the number of needed basis functions is exponentially proportional to the number of products. Gandhi and Houde (2016) provide a solution that breaks the dependence of basis functions on the identity of products under symmetry conditions among products. The number of basis functions is then invariant with respect to the number of products. However, due to potentially heterogeneous synergy parameters across bundles, the identity of products does matter in (2.10). In the empirical application of section 2.6, I do not employ the approach of optimal instruments and will explore it in future research.

2.4.2 Implementation of Demand Inverse

A key step of the estimation procedure is the implementation of the demand inverse in Theorem 7. It seeks for the solution of the following equation:

$$s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma', F') - \mathcal{J}_{t\mathbf{J}_t} = 0. \quad (2.11)$$

In practice, $s_{t\mathbf{J}_t}(\cdot; \Gamma', F')$ are usually computed using Monte Carlos simulations. The researcher first draws independently a finite number of sets of random numbers with which one can approximate the distribution F' .²⁹ Then, the market share functions are computed on the basis of the approximated version of F' which is numerically a discrete distribution with finite support and therefore is defined in a compact support. In the analysis of the numerical performance of solving (2.11), I will assume that the distribution F' has compact support.

In the context of demand models of single products, Berry et al. (1995) propose a fixed-point iterative algorithm to implement the demand inverse in estimation. An essential property of this algorithm is contraction mapping, which guarantees the convergence of the iteration. However, the contraction-mapping property may not hold if one uses the same iterative algorithm to solve (2.11) because products can be Hicksian complementary. To solve this challenge, I propose to use Jacobian-based approach to solve (2.11). This approach is not new in the literature. Conlon and Gortmaker (2019) tests performances of different Jacobian-based algorithms to solve the demand inverse in demand models of single products and find supportive evidences for the efficiency of Jacobian-based numerical methods. A leading example is Newton-Raphson method:

$$\begin{aligned} \delta^{(0)} &= \delta^{(0)}, \\ \delta^{(n+1)} &= \delta^{(n)} - J_s^{-1}(\delta^{(n)})[s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma', F') - \mathcal{J}_{t\mathbf{J}_t}], \end{aligned} \quad (2.12)$$

where $J_s(\delta'_{t\mathbf{J}_t}) = \frac{\partial s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; x_{t\mathbf{J}_t}^{(2)}, p_{t\mathbf{J}_t}, \Gamma', F')}{\partial \delta'_{t\mathbf{J}_t}}$. In the context of (2.11), Algorithm (2.12) is well-define because

²⁹A typical method is to simulate a fixed set of random numbers from uniform distribution in $[0, 1]$ and use $(F')^{-1}$ to transform these random numbers to that under the distribution F' .

the Jacobian matrix $J_s(\delta'_{t\mathbf{J}})$ is everywhere symmetric and positive-definite. Moreover, the uniqueness of solution is guaranteed by Theorem 7. If (2.12) converges, then it converges to the unique solution of (2.11).

It is well-known that the numerical performance of Jacobian-based algorithms such as (2.12) depends on the quality of the initial point $\delta^{(0)}$: the closer $\delta^{(0)}$ is to the solution $\delta'_{t\mathbf{J}}$, the faster Algorithm (2.12) converges. In general, there is no theoretical guidance for choosing a good initial point.³⁰ Motivated by the practical implementation of the demand inverse, i.e. F' in (2.11) is approximated by a discrete distribution with finite support, I propose an initial point that is directly constructed from the observed product-level market shares. In the next proposition, I prove that in models of multiple choice of products across K categories, the proposed initial point is “close” to the solution of (2.11):

Proposition 2. *Suppose that $s_{t\mathbf{J}}$ in (2.11) are generated from a model of multiple choice of products across K categories, for $K \geq 1$ (see Section 2.2.2) and the distribution F' has compact support \mathbf{D}_F . Denote the solution to (2.11) by $\delta'_{\mathbf{J}}$. For products of category k , define $\delta_{k*}^{(0)} = \left(\delta_{jk*}^{(0)} \right)_{j \in \mathbf{J}_k}$, where $\delta_{jk*}^{(0)} = \ln \frac{s_j}{1 - \sum_{j \in \mathbf{J}_k} s_j}$. Define $\delta_*^{(0)} = (\delta_{k*}^{(0)})_{k=1, \dots, K}$. Then, there exists $A(\mathbf{D}_F, \Gamma') > 0$ such that*

$$|\delta'_{\mathbf{J}} - \delta_*^{(0)}| \leq A(\mathbf{D}_F, \Gamma').$$

Proof. See Appendix B.6. □

Even though it is hard to derive similar results in a general model (2.4), Proposition 2 sheds light on how to find a good initial point for Jacobian-based algorithms: it suggests to use initial points as if the data generate process is a multinomial logit model. In a model where the bundle size is up to size K , such a point along the lines of Proposition 2 can be defined as: for $j \in \mathbf{J}_t$,

$$\delta_{j*}^{(0)} = \ln \frac{s_j}{K - \sum_{j \in \mathbf{J}_t} s_j}. \tag{2.13}$$

Here $K - \sum_{j \in \mathbf{J}_t} s_j$ serves as the “market share” of the outside option.³¹ In the next section, I will explore potential efficiency gain of using $\delta_*^{(0)}$ in Monte Carlos simulations.

2.5 Monte Carlos Simulations

In this section, I explore the numeric performance of Jacobian-based algorithms in the implementation of the demand inverse of product-level market shares. I compare convergence time across different algorithms and initial points—particularly, the proposed initial point in section 2.4.2—for different sizes of product set.

Table 2.1 summarizes the main results of the simulations. The data generating process is a discrete choice model of bundles up to size two and the prices are generated from a Bertrand pricing game under complete information with constant marginal costs. I simulate 50 markets with the same structural parameters. The unobserved demand shocks $\xi_{t\mathbf{J}}$ are Gaussian and i.i.d. across markets. Then, I implement the demand inverse using the true model and also using a demand model of single products in

³⁰One of few theoretical results on global convergence of Newton-Raphson method is Newton-Kantorovich Theorem.

³¹When the bundle size is up to K , since product-level market shares of two different products overlap on up to $K - 1$ bundle-level market shares. Consequently, the sum of all product-level market shares is strictly smaller than K .

each market and report the median convergence time (in seconds).³² Moreover, I test the performances of such demand inverses with the initial point $\delta_*^{(0)}$ defined in (2.13) and also $\delta^{(0)} = 0$. I replicate this setting for different sizes of product sets ($J = 10, 50, \text{ and } 100$). For example, in the case of $J = 100$, the true model has 5051 alternatives (100 single products, 4950 bundles of two different products, and an outside option). The demand inverse $s_{\mathbf{J}}^{-1}$ will then treat the observed market shares as those generated from the true model, while the demand inverse $s_{\bar{\mathbf{J}}}^{-1}$ will treat the same observed market shares as if they are generated from a demand model of 100 single products. The three algorithms used in the simulations are built-in algorithms of *fsolve* in Matlab.

Table 2.1: Demand Inverse of Product-Level Market Shares: Convergence Time in Seconds

Algorithm	Trust-Region-Dogleg				Trust-Region-Reflective				Levenberg-Marquardt			
	$s_{\mathbf{J}}^{-1}$		$s_{\bar{\mathbf{J}}}^{-1}$		$s_{\mathbf{J}}^{-1}$		$s_{\bar{\mathbf{J}}}^{-1}$		$s_{\mathbf{J}}^{-1}$		$s_{\bar{\mathbf{J}}}^{-1}$	
Init. Point	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0
# Products												
$J = 10$	0.04	0.09	0.03	0.08	0.05	0.09	0.04	0.08	0.08	0.09	0.07	0.09
50	0.49	1.45	0.10	2.64	0.41	1.31	0.13	0.34	1.31	1.92	0.19	0.12
100	4.50	12.22	0.12	3.21	3.32	12.15	0.27	0.60	12.25	20.21	0.33	0.18

Notes: Trust-region-dogleg, trust-region, and Levenberg-Marquardt algorithms are built-in algorithms of the function *fsolve* in Matlab. All of them are large-scale and minimize the sum of squares of the components of (2.11). Median convergence time (in seconds) of 50 independently simulated markets is reported. Tolerance level in the stop criterion of all algorithms is set to 10^{-16} .

There are two main findings. First, for $s_{\mathbf{J}}^{-1}$, using the recommended initial point $\delta_*^{(0)}$ remarkably reduces convergence time in all cases. The gain is larger when the number of products is larger. When $J = 100$, trust-region-dogleg and trust-region-reflective algorithms reduce around 70% convergence time by using $\delta_*^{(0)}$ than using $\delta^{(0)} = 0$. The efficiency gain for $s_{\bar{\mathbf{J}}}^{-1}$ using $\delta_*^{(0)}$ is similar.³³ Second, using $\delta_*^{(0)}$ does not seem to increase the number of iterations for the algorithm to converge as the problem size increases. For example, the convergence time for $J = 100$ by using the three Jacobian-based algorithms with $\delta_*^{(0)}$ is roughly 100 times of that for $J = 10$. Because the bundle size is at most two, then the size of choice set increases quadratically with respect to the number of products and therefore the number of required computations for one evaluation of market share functions also increases quadratically. While, the total convergence time when using $\delta_*^{(0)}$ seems to increase only quadratically for $s_{\bar{\mathbf{J}}}^{-1}$ with respect to the number of products. This may imply that the number of iterations does not increase as J increases.

2.6 Empirical Illustration: RTE Cereals and Milk

In this section, I illustrate the practical implementation of the proposed methods and estimate the demand for Ready-To-Eat (RTE) cereals and milk in the US. I use models of demand for bundles with different specifications of $\Gamma_{\mathbf{b}}$'s that are more or less flexible. Depending on the research question, some of the specifications may be too restrictive and consequently lead to biased results. I illustrate the potential biases due to these restrictions in several counterfactual simulations.

³²The structural parameters are chosen so that the sum of the simulated product-level market shares is always smaller than one. This allows to implement the demand inverse of single products on these simulated product-level market shares.

³³In this demand inverse, K in (2.13) is 1.

2.6.1 Data and Definitions

I use the store-week level datasets of the RTE cereal and milk categories from the IRI data. The IRI data has been used in the empirical literature of demand (see [Nevo \(2000, 2001\)](#)). I will give a succinct description and refer to these papers and also [Bronnenberg et al. \(2008\)](#) for a thorough discussion.

In this illustrative application, I focus on the period 2008-2011 and the city of Pittsfield in the US. I define a market t as a combination of store and week and obtain 1387 markets. In each market, the sales (in lbs and dollars) of RTE cereals and fluid milk are observed at Universal Product Code (UPC) level. For the RTE cereal category, similarly to [Nevo \(2001\)](#), I define a product as a combination of brand, flavour, fortification, and type of grain. For fluid milk category, I define a product as a combination of brand, flavour, fortification, fat content, and type of milk. Then, the sales of product j of category $k \in \mathbf{K} = \{RTE\ cereal, fluid\ milk\}$ in market t is the sum of the sales in lbs of all the UPC's that this product collects. The price of j of category k in market t , p_{tj}^k , is defined as the ratio between its sales in dollars and in lbs. To simplify the implementation, within each category, I keep the products with the largest sales in lbs. I then obtain 25 RTE cereal products and 20 fluid milk products, and denote the set of these products by \mathbf{J}_k for $k \in \mathbf{K}$, respectively.³⁴

For each market, I consider the weekly consumption of breakfast cereals as the market size for RTE cereal category and weekly consumption of fluid milk for milk category. To calibrate the market size for each category, I assume that households go shopping once per week for breakfast cereals and fluid milk. Then, the market size for RTE cereal category (or milk) is the product of the weekly per capita consumption of breakfast cereals (or fluid milk) and the sampled population size. I obtain the former information from external sources and the latter from the IRI data. Finally, for each market, the product-level market share of $j \in \mathbf{J}_k$ is then the ratio between its total sales and the market size for category k . Appendix B.7 provides computational details of the construction of the product-level market shares and Tables B.1-B.2 in Appendix B.8 summarize the characteristics of the products.

2.6.2 Model Specification

For each store-week combination t , denote the set of available products in category $k \in \mathbf{K}$ by $\mathbf{J}_{tk} \subset \mathbf{J}_k$. Denote by 1 the RTE cereal category, 2 the milk category, and then $\mathbf{J}_t = \mathbf{J}_{t1} \cup \mathbf{J}_{t2}$. The set of bundles \mathbf{C}_{t2} is defined as $\mathbf{J}_{t1} \times \mathbf{J}_{t2}$, where each bundle contains a RTE cereal product and a milk product.³⁵ Household's choice set is then defined as $\mathbf{C}_t = \mathbf{J}_t \cup \mathbf{C}_{t2} \cup \{0\}$, where 0 represents the outside option.³⁶ The size of \mathbf{C}_t is 546 (45 products, 500 bundles, 1 outside option) if all products in \mathbf{J} are available in market t .

For household i in market t , the indirect utility from choosing product $j \in \mathbf{J}_{tk}$ is:

$$\begin{aligned} U_{itj}^k &= -p_{tj}^k \alpha_i + \eta_{ij}^k + \xi_{tj}^k + \varepsilon_{itj}^k \\ &= [-p_{tj}^k \alpha + \eta_j^k + \xi_{tj}^k] + [\Delta \eta_{ij}^k - \Delta \alpha_i p_{tj}^k] + \varepsilon_{itj}^k \\ &= \delta_{tj}^k + \mu_{itj}^k + \varepsilon_{itj}^k, \\ \mu_{itj}^k &= \Delta \eta_{ij}^k - (d_i \Delta \alpha + v_i) p_{tj}^k, \end{aligned}$$

³⁴The purchase of the 25 RTE cereal products represents 38% of the total purchase of RTE cereals in the IRI data, and that of the 20 fluid milk represents around 88%.

³⁵I do not include bundles of products of the same category.

³⁶According to the definition of products and the market sizes, the outside option collects RTE cereals or milk products not included in \mathbf{J}_{tk} , relevant products not present in the categories (e.g. cereal biscuits), and the bundles of these products (e.g. cereal biscuits and milk).

and

$$\Delta\eta_{ij}^k = \begin{cases} \Delta\eta_{i,\text{flavor}(j)} + \Delta\eta_{i,\text{fortification}(j)} + \Delta\eta_{i,\text{brand}(j)}^1, & \text{if } k = 1, \\ \Delta\eta_{i,\text{flavor}(j)} + \Delta\eta_{i,\text{fortification}(j)} + \Delta\eta_{i,\text{brand}(j)}^2 + \Delta\eta_{i,\text{fat content}(j)}^2, & \text{if } k = 2, \end{cases}$$

where δ_{tj}^k is market t -specific mean utility for $j \in \mathbf{J}_{tk}$, μ_{itj}^k is a household i -specific utility deviation from δ_{tj}^k , and ε_{itj}^k is an idiosyncratic error term. The coefficient α is population-average price coefficient, and $\Delta\alpha_i = d_i\Delta\alpha + v_i$ is household i -specific price coefficient deviation from α and is the sum of an observed part that is a function of the household characteristics d_i (income groups) and an unobserved component v_i . $\Delta\eta_{ij}^k$ is an unobserved household i -specific preference for product j of category k , where $\Delta\eta_{i,\text{flavor}(j)}$ captures household i 's unobserved preference for the flavour of j of category k (unflavoured, flavoured), $\Delta\eta_{i,\text{fortification}(j)}$ captures i 's unobserved preference for the nutrition in product j of category k (unfortified, fortified), $\Delta\eta_{i,\text{brand}(j)}^k$ captures i 's unobserved preference for the brand of j of category k , and $\Delta\eta_{i,\text{fat content}(j)}^2$ captures i 's unobserved preference for the fat content in milk j (whole fat, low fat, skimmed). Note that because RTE cereals and milk have both the flavour and fortification characteristics, then for products $j \in \mathbf{J}_{t1}$ and $r \in \mathbf{J}_{t2}$, if they have the same flavour (or fortification type), then $\Delta\eta_{i,\text{flavor}(j)} = \Delta\eta_{i,\text{flavor}(r)}$ (or $\Delta\eta_{i,\text{fortification}(j)} = \Delta\eta_{i,\text{fortification}(r)}$).

The indirect utility of household i in market t from choosing bundle $\mathbf{b} = (j, r)$ is:

$$\begin{aligned} U_{it\mathbf{b}} &= [-p_{tj}^1\alpha_i + \eta_{ij}^1 + \xi_{tj}^1] + [-p_{tr}^2\alpha_i + \eta_{ir}^2 + \xi_{tr}^2] + \Gamma_{\mathbf{b}} + \varepsilon_{it\mathbf{b}} \\ &= [\delta_{tj}^1 + \delta_{tr}^2 + \Gamma_{\mathbf{b}}] + [\mu_{itj}^1 + \mu_{itr}^2] + \varepsilon_{it\mathbf{b}} \\ &= \delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}}) + \mu_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}}, \end{aligned}$$

where $\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}}) = \delta_{tj}^1 + \delta_{tr}^2 + \Gamma_{\mathbf{b}}$ is market t -specific mean utility for bundle \mathbf{b} , $\mu_{it\mathbf{b}}$ is household i -specific utility deviation from $\delta_{t\mathbf{b}}$, $\Gamma_{\mathbf{b}} = \Gamma_{(j,r)}$ is demand synergy between RTE cereal j and milk r , and $\varepsilon_{it\mathbf{b}}$ is an idiosyncratic error term. The demand synergy parameter $\Gamma_{(j,r)}$ captures the extra utility household obtains from buying RTE cereal j and milk r jointly rather than separately. One prominent reason for the joint purchase is synergy in consumption, i.e. members in the household consume together RTE cereals and milk for their breakfasts.³⁷ The matching between the characteristics of RTE cereal j and milk r may determines the extra utility $\Gamma_{(j,r)}$. Consequently, I specify $\Gamma_{(j,r)}$ as a function of the characteristics of j and r :

$$\begin{aligned} \Gamma_{(j,r)}(\gamma) &= \gamma_0 + \mathbf{1}\{j \text{ is multi-grain}\}\gamma_1 + \mathbf{1}\{j \text{ is granola}\}\gamma_2 \\ &\quad + \mathbf{1}\{r \text{ is skimmed}\}\gamma_3 + \mathbf{1}\{r \text{ is low fat}\}\gamma_4 \\ &\quad + \mathbf{1}\{j \text{ is flavoured}\}\gamma_5 + \mathbf{1}\{r \text{ is chocolate milk}\}\gamma_6 \\ &\quad + \mathbf{1}\{j \text{ is flavored and } r \text{ is chocolate milk}\}\gamma_7 \\ &\quad + \mathbf{1}\{j \text{ is fortified, } r \text{ is chocolate milk}\}\gamma_8 + \mathbf{1}\{j \text{ is fortified, } r \text{ is fortified}\}\gamma_9. \end{aligned} \tag{2.14}$$

The parameter γ_0 represents the synergy in consumption of the reference bundle (unflavoured unfortified uni-grain RTE cereal and unflavoured whole-fat milk). γ_1 and γ_2 quantify additional synergies due to other types of grains (multi-grain, granola). γ_3 and γ_4 measures additional synergies due to lower fat content (skimmed, low fat). γ_5 , γ_6 and γ_7 proxy additional synergies due to flavour combinations.

³⁷Another reason can be shopping cost, i.e. household may not want to go shopping twice to buy RTE cereals and milk. However, this is assumed away by the assumption that household goes shopping once per week for breakfast cereals and milk.

γ_8 and γ_9 quantify additional synergies due to the combinations of fortified nutrition in RTE cereals and milk characteristics.

Finally, the indirect utility of household i in market t from choosing the outside option is normalised to be $U_{it0} = \varepsilon_{it0}$. Denote the random coefficients by

$$\theta_{it} = (v_i, \Delta\eta_{i,\text{unflavoured}}, \Delta\eta_{i,\text{flavoured}}, \Delta\eta_{i,\text{unfortified}}, \Delta\eta_{i,\text{fortified}}, \{\Delta\eta_{i,\text{br.}}^1\}_{\text{br.} \in \mathbf{B}_1}, \{\Delta\eta_{i,\text{br.}}^2\}_{\text{br.} \in \mathbf{B}_2}, \Delta\eta_{i,\text{whole fat}}^2, \Delta\eta_{i,\text{low fat}}^2, \Delta\eta_{i,\text{skimmed}}^2),$$

where \mathbf{B}_k denotes the set of brands in category k . I assume that θ_{it} follows a Gaussian distribution F and the components are uncorrelated. This specification already allows for unobserved correlation among products of the same characteristics within and across categories. Define $\delta_{t\mathbf{J}_t} = (\delta_{t\mathbf{J}_{t1}}^1, \delta_{t\mathbf{J}_{t2}}^2)$ and $p_{t\mathbf{J}_t} = (p_{t\mathbf{J}_{t1}}^1, p_{t\mathbf{J}_{t2}}^2)$. Write $\mu_{itj}^k = \mu_{itj}^k(d_i, \theta_{it}, p_{tj}^k)$. Finally, assume that ε_{it0} , ε_{itj}^1 's, ε_{itj}^2 's, and $\varepsilon_{it\mathbf{b}}$'s are i.i.d. Gumbel. Then, the product-level market share function of $j \in \mathbf{J}_{t1}$ is:

$$s_{tj}(\delta_{t\mathbf{J}_t}; p_{t\mathbf{J}_t}, \gamma, F) = \int \frac{e^{\delta_{tj}^1 + \mu_{itj}^1(d_i, \theta_{it}, p_{tj}^1)} \left[1 + \sum_{r \in \mathbf{J}_{t2}} e^{\delta_{tr}^2 + \mu_{tr}^2(d_i, \theta_{it}, p_{tr}^2) + \Gamma_{(j,r)}(\gamma)} \right]}{1 + \sum_{k=1,2} \sum_{j' \in \mathbf{J}_{tk}} e^{\delta_{tj'}^k + \mu_{tj'}^k(d_i, \theta_{it}, p_{tj'}^k)} + \sum_{(j',r) \in \mathbf{J}_{t1} \times \mathbf{J}_{t2}} e^{\delta_{tj'}^1 + \delta_{tr}^2 + \mu_{tj'}^1(d_i, \theta_{it}, p_{tj'}^1) + \mu_{tr}^2(d_i, \theta_{it}, p_{tr}^2) + \Gamma_{(j',r)}(\gamma)}} dF(\theta_{it}) d\Pi_t(d_i), \quad (2.15)$$

where $\Pi_t(\cdot)$ is the distribution function of demographics d_i in market t . The formula for $r \in \mathbf{J}_{t2}$ is similar.

In what follows, I will estimate three models (2.15) with different specifications of $\Gamma_{\mathbf{b}}$'s: Model I with $\Gamma_{\mathbf{b}} = 0$ for all $\mathbf{b} \in \mathbf{C}_2$, Model II with $\Gamma_{\mathbf{b}} = \gamma_0$, i.e. any bundle of RTE cereal and milk has the same synergy in consumption, and a full model with $\Gamma_{\mathbf{b}}$ specified in (2.14). Note that Model I is equivalent to two separate BLP models of demand for single products, respectively for RTE cereals and milk, with the same price coefficients.

2.6.3 Demand Estimates

Demand estimates are summarized in Table 2.2. In column "IV regression", I estimate a multinomial logit with $\Gamma_{\mathbf{b}} = 0$ for all $\mathbf{b} \in \mathbf{C}_2$. Columns "Model I", "Model II", and "Full Model" show the estimates by using Model I, Model II, and the full model, respectively. In all the models, I control for product-specific intercepts and use the same Hausman-type instruments. These instruments include the prices of the same products in the same store and week but in other cities (Boston for RTE cereals and Hartford for milk), the prices of other products of the same category with the same product characteristics.

The price coefficient (α) is estimated -0.59 in the multinomial logit model. The other three models with random coefficients show important heterogeneity in price sensitivities across income groups. Without surprise, households with higher income are estimated to have a lower (in absolute value) price coefficient and therefore less sensitive to price change. The standard variance of the unobserved heterogeneity in the price coefficient (σ_v) is estimated small. Moreover, after controlling for the product-specific intercepts, households' preference seem to be almost homogeneous for products within some types (e.g. unflavoured, fortification, fat content). One potential reason is that products are little differentiated within each of these types. In contrast, households' preference for flavoured products seems to be more heterogeneous. This is also intuitive because the flavours of RTE cereals and milk

are much more horizontally differentiated and different households may have their favourite flavours. I also find that households' preference for RTE cereal brands is much more heterogeneous than that for milk brands.

In model II, the synergy parameter γ_0 is estimated to be 0.902 and significant. This specification constraints all bundles of RTE cereal and milk to have the same synergy, regardless of their characteristics. In the full model, the demand synergies are allowed to vary across bundles. In column "Full Model", $\gamma_0 = -1.540$ represents the synergy in consumption between the unflavoured unfortified uni-grain RTE cereal and unflavoured whole-fat milk. Regarding the characteristics of RTE cereals, γ_1 is estimated positive, meaning that multi-grain cereals are preferred (over uni-grain ones) when consumed with milk. Moreover, granola, which contains oats and other whole grains as well ingredients such as dried fruit and nuts, is estimated to be even preferred over multi-grain cereals ($\gamma_2 > \gamma_1$). Concerning the characteristics of milk, products with lower fat are estimated to be preferred when consumed with cereals ($\gamma_3 > \gamma_4 > 0$). Another interesting finding is about households' preference for flavour combinations of cereals and milk. Flavoured cereals are estimated to be preferred over unflavoured ones ($\gamma_5 > 0$). For chocolate milk, households' preference seems to be more complicated. When consumed with unflavoured cereals, chocolate milk is preferred over unflavoured milk ($\gamma_6 > 0$). While, I find that it is seldom consumed with flavoured or fortified cereals, i.e. γ_7 and γ_8 are estimated very negative.³⁸ In fact, flavoured (or fortified) RTE cereals are usually frosted. Very negative γ_7 and γ_8 may reflect households' disutility for bundles with too much sugar. Finally, I also find that bundles of fortified cereals and milk are less attractive than the reference one ($\gamma_9 < 0$). In the data, the types of added nutrition in RTE cereals and milk are the same, e.g. vitamins, calcium. $\gamma_9 < 0$ may reflect that the same types of added nutrition in cereals and milk are substitute.

2.6.4 Price Elasticities

I compute the average (across markets) estimated self- and cross-price elasticities obtained from the full model. Because there is no income effect in the empirical specification, negative (positive) cross-price elasticities are then interpreted as Hicksian complementarity (substitutability). To facilitate the exposition, I report the price elasticities at the level of product characteristics and producers. This will illustrate how RTE cereals and milk are complementary along each of these dimensions. The results are illustrated in Tables 2.3-2.6. Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row producer (or characteristics) with respect to a 1% increase in the prices of the products collected by the column producer (or characteristics).³⁹

Overall, RTE cereals are estimated to have larger self-price elasticities than milk. This may reflect that households view milk more necessary than RTE cereals and therefore are less sensitive to changes in the prices of milk. Given the specification of model (2.15), RTE cereals are always substitutes to each other and the cross-price elasticities among them are positive. Similarly, the cross-price elasticities

³⁸In Table 2.2, the estimates of γ_7 and γ_8 are $-\infty$. This means that the model with $\gamma_7, \gamma_8 = -\infty$, i.e. the corresponding bundles are not in the choice set, performs statistically as well as the one without these restrictions in terms of the value of the GMM objective function. Concretely, I first estimate a model with γ_7 and γ_8 being finite. I find that the components of γ_7 and γ_8 in the minimiser of the GMM objective function are very negative. Then, I estimate the model with $\gamma_7, \gamma_8 = -\infty$. The difference in the value of the GMM objective function is less than 10^{-9} , or equivalently, one cannot reject the "hypothesis" that $\gamma_7, \gamma_8 = -\infty$. In the future, I will consider a formal testing procedure.

³⁹Concretely, denote by \mathbf{C}_m the set of products that row (column) producer m represents. Then, the price elasticity between brands m and n , E_{mn} , is defined as $E_{mn} = \frac{\sum_{j \in \mathbf{J}_m} \delta_j \cdot \sum_{r \in \mathbf{J}_n} \varepsilon_{jr}}{\sum_{j \in \mathbf{J}_m} \delta_j}$, where ε_{jr} is the price elasticity between products j and r .

among milk products are also positive. These are shown by the positive off-diagonal elements in the diagonal blocks (RTE cereals-RTE cereals, Milk-Milk) of Tables 2.3-2.6. Differently, the cross-price elasticities between RTE cereal and milk products, i.e. the elements in the off-diagonal blocks (RTE cereals-Milk, Milk-RTE cereals), can be either positive or negative.

Table 2.3 show some interesting findings about the substitution patterns along the dimensions of grain type and fat content. Granola is estimated to be complementary to milk with any level of fat and skimmed milk is complementary to cereals with any kind of grain. Moreover, milk with lower fat is uniformly more complementary to any kind of grain than milk with higher fat. This reveals that households do not seek for fat in milk when drinking it with RTE cereals. As to flavours (Tables 2.4-2.5), unflavoured cereals and flavoured milk (and the reverse) are shown to be complementary. While, flavoured cereals and chocolate milk are estimated to be (strong) substitutes. Coherent with the estimates of γ 's in Table 2.2, the relationship between chocolate milk and RTE cereals is more complicated. Chocolate milk is estimated to be complementary to unflavoured or unfortified cereals. However, it is estimated to be substitute to flavoured or fortified cereals. Finally, I find that most RTE cereals and milk are complementary at producer level.

As a comparison, the demand synergies in model I are constrained to be zero. Then, the cross-price elasticities between RTE cereals and milk are mechanically zero. In model II, all the bundles are restricted to have the same demand synergy which is estimated to be positive (see the column "Model II" of Table 2.2). I re-do the exercises in Tables 2.3-2.6 using the demand estimates from model II (see Tables B.3-B.6 of Appendix B.8). In contrast to those obtained from the full model, the results show that RTE cereals and milk are complementary along every dimension.

Table 2.2: Demand Estimates

	IV Regression $\Gamma_{\mathbf{b}} = 0$	Model I $\Gamma_{\mathbf{b}} = 0$	Model II $\Gamma_{\mathbf{b}} = \gamma_0$	Full Model
Price Coef.				
uniform, α	-0.59 (0.011)			
(baseline) low income, α_1		-1.369 (0.042)	-1.128 (0.060)	-1.062 (0.075)
medium income, $\Delta\alpha_2$		0.218 (0.164)	0.148 (0.096)	0.164 (0.0764)
high income, $\Delta\alpha_3$		0.947 (0.045)	0.718 (0.029)	0.712 (0.0269)
Random Coef.				
σ_v		0.115 (0.070)	0.086 (0.072)	0.046 (0.1244)
$\sigma_{\text{unflavoured}}$		0.015 (2.754)	0.018 (1.490)	0.023 (2.6772)
$\sigma_{\text{flavoured}}$		2.353 (0.259)	1.684 (0.038)	1.010 (0.1615)
$\sigma_{\text{unfortified}}$		0.048 (3.047)	0.004 (3.870)	0.017 (2.2994)
$\sigma_{\text{fortified}}$		0.010 (5.393)	0.010 (2.388)	0.015 (6.8129)
σ_{fat}		0.077 (1.099)	0.062 (0.660)	0.034 (1.6161)
$\sigma_{\text{cereal brand}}$		0.780 (0.049)	0.660 (0.058)	0.847 (0.0705)
$\sigma_{\text{milk brand}}$		0.005 (6.156)	0.003 (5.368)	0.004 (6.9505)
Demand Synergies				
γ_0			0.902 (0.155)	-1.540 (0.3437)
multi-grain, γ_1				0.533 (0.0359)
granola, γ_2				4.363 (0.0891)
skimmed, γ_3				2.880 (0.2111)
low fat, γ_4				0.514 (0.1282)
flavoured cereal, γ_5				1.816 (0.2324)
chocolate milk, γ_6				13.625 (0.2621)
flavoured cereal and chocolate milk, γ_7				$-\infty$
fortified cereal, chocolate milk, γ_8				$-\infty$
fortified cereal and milk, γ_9				-1.538 (0.3512)
GMM Objective Function		0.1636	0.1599	0.1434

Notes: Standard errors are reported in brackets. For all the models, instruments are the same and product-specific intercepts are included. In the “IV Regression”, week dummies and store dummies are also included.

Table 2.3: Average Estimated Own- and Cross-Price Elasticities (Full Model): Grain Type and Fat Content

	RTE cereals			Milk		
	uni-grain	multi-grain	granola	skimmed	low fat	whole fat
RTE cereals, uni-grain	-1.407	0.194	0.009	-0.032	0.007	0.009
multi-grain	0.266	-1.492	0.009	-0.034	0.001	0.009
granola	0.220	0.168	-1.335	-0.084	-0.071	-0.005
Milk, skimmed	-0.350	-0.243	-0.053	-0.252	0.047	0.023
low fat	0.010	-0.005	-0.020	0.018	-0.262	0.028
whole fat	0.056	0.045	-0.005	0.018	0.054	-0.307

Table 2.4: Average Estimated Own- and Cross-Price Elasticities (Full Model): Flavours

	RTE cereals		Milk	
	unflavoured	flavoured	unflavoured	chocolate
RTE cereals, unflavoured	-1.397	0.190	0.016	-0.014
flavoured	0.145	-1.381	-0.051	0.003
Milk, unflavoured	0.031	-0.130	-0.214	0.001
chocolate	-1.319	0.378	0.071	-0.264

Table 2.5: Average Estimated Own- and Cross-Price Elasticities (Full Model): Fortification and Flavours

	RTE cereals		Milk	
	unfortified	fortified	unflavoured	chocolate
RTE cereals, unfortified	-1.263	0.060	-0.029	-0.006
fortified	0.393	-1.668	0.043	0.002
Milk, unflavoured	-0.124	0.025	-0.214	0.001
chocolate	-1.004	0.063	0.071	-0.264

Table 2.6: Average Estimated Own- and Cross-Price Elasticities (Full Model): Brands

	RTE cereals					Milk				
	General Mills	Kashi	Kellogg's	Post	Private Label	Garelick Farms	Guidas	High Lawn Farm	Hood	Private Label
RTE cereals, General Mills	-1.444	0.009	0.146	0.037	0.009	-0.007	-0.034	-0.005	-0.004	-0.022
Kashi	0.154	-1.678	0.152	0.035	0.011	-0.005	-0.005	-0.0003	-0.0002	-0.001
Kellogg's	0.152	0.010	-1.424	0.032	0.009	-0.002	0.006	-0.0005	-0.0003	0.002
Post	0.173	0.010	0.146	-1.541	0.010	-0.005	-0.013	-0.002	-0.001	-0.009
Private Label	0.175	0.013	0.163	0.038	-1.335	-0.044	.	-0.012	-0.011	-0.094
Milk, Garelick Farms	-0.092	-0.002	-0.018	-0.014	-0.032	-0.323	.	0.005	0.005	0.054
Guidas	-0.044	-0.0003	0.0004	-0.004	.	.	-0.173	0.012	0.002	0.014
High Lawn Farm	-0.109	0.001	-0.006	-0.011	-0.020	0.015	0.119	-0.479	0.004	0.046
Hood	-0.154	-0.001	-0.032	-0.018	-0.030	0.015	0.114	0.007	-0.497	0.045
Private Label	-0.076	0.001	0.011	-0.007	-0.022	0.016	0.123	0.007	0.005	-0.251

Notes: The . refers to the situation where the products of the row and column brands are not available at the same time in any market in the data.

2.6.5 Counterfactual Simulations

I simulate two counterfactuals using the demand estimates obtained from the full model, model I and model II. In the factual and counterfactual scenarios, I assume that producers play a simultaneous Bertrand price-setting game with complete information. Moreover, for each model, the market-product specific marginal costs are assumed constant and remain unchanged in the counterfactuals. Note that Private Label is present in both categories. I assume that it maximises the total profit generated by its RTE cereal and milk products and their bundles in all scenarios.

Merger across category. The first scenario is a situation where General Mills and Garelick Farms merge to a new producer and every producer after the merger still implements linear pricing strategy as in the factual scenario. In this exercise, I compare the merger outcomes (price change, consumer surplus change, etc.) predicted by model I and the full model. The results are summarised in Table 2.7.

In model I, because all the demand synergies between RTE cereals and milk are restricted to be zero, their cross-price elasticities are always zero and therefore the merger between General Mills and Garelick Farms will not lead to any change in prices and consumer surplus relative to those in the factual scenario (row “Model I” in Table 2.7). The full model estimates that RTE cereals and milk products exhibit substantial complementarity at the producer level (Table 2.6). The merger outcomes are coherent with Cournot (1838)’s intuition that mergers between producers selling complementary products can be socially desirable. Intuitively, the merged producer internalises the complementarity in the pricing and consequently reduces the prices of General Mills RTE cereals and Garelick Farms milk (row “Full Model” in Table 2.7). This finally leads to an increase of consumer surplus by 2.02%. As a check, I also simulate the same merger using the demand estimates of model II. The results are similar and illustrated in the last row of Table 2.7.

Nonlinear pricing of Private Label. To study the potential bias due to the restriction $\Gamma_{\mathbf{b}} = \gamma_0$ for all $\mathbf{b} \in \mathbf{C}_2$, I simulate a scenario where Private Label implements nonlinear pricing strategy on its own products and bundles of RTE cereal and milk, while the market structures in both markets remain the same as in the factual scenario. In this exercise, I compare the predictions (prices, consumer surplus, etc.) on the basis of the demand estimates from model II, which restricts $\Gamma_{\mathbf{b}} = \gamma_0$ for all $\mathbf{b} \in \mathbf{C}_2$, and the full model. The estimated cross-price elasticities by both models show the existence of substantial complementarities between RTE cereals and milk. However, model II estimates that the RTE cereals and milk are complementary along all the dimensions (see Tables B.3-B.6), while the full model reveals that some types of RTE cereals are more (or less) complementary to certain types of milk and vice versa. Intuitively, the consequences of the nonlinear pricing of private label are sensitive to the magnitude of the complementarities between private label RTE cereals and milk relative to those between RTE cereals and milk of other producers. The main goal of this exercise is to illustrate the extent of bias in the nonlinear pricing analysis, if the specification of the demand synergy parameters (and therefore the substitution patterns between RTE cereals and milk) is potentially restrictive. Table 2.8 summarises the results.

Both models predict that the prices of private label products increase. At the same time, the bundles of private label products will have price discounts. However, the full model predicts much larger increases in the average prices of private label RTE cereals and milk and also a larger discount.

Moreover, two models predict opposite change in consumer surplus. The full model predicts a decrease by 0.11%, while model II predicts an increase by 1.83%.

Summary. Both model II and the full model capture the complementarity between RTE cereals and milk, while model I restricts them to be independent. Consequently, model II and the full model generate similar predictions for the first counterfactual simulation where General Mills and Garelick Farms merge to one single producer. In the second counterfactual where Private Label implements a nonlinear pricing strategy, the magnitude of the complementarities between private label RTE cereals and milk relative to those between products of other producers become important. Restricting all $\Gamma_{\mathbf{b}}$'s to be the same in model II seems too coarse and may not be appropriate in such counterfactuals. While, specifying $\Gamma_{\mathbf{b}}$ as a function of the characteristics of the products it includes provides more flexibility and is more suitable for the purpose.

Table 2.7: Merger Simulation: General Mills and Garelick Farms

	Price change		Consumer Surplus change
	RTE Cereals	Milk	
Full Model	-0.67%	-3.49%	2.02%
Model I, $\Gamma_{\mathbf{b}} = 0$	0%	0%	0%
Model II, $\Gamma_{\mathbf{b}} = \gamma_0$	-0.49%	-4.49%	3.44%

Notes: The Table reports average changes in prices (first two columns) and consumer surplus (last column) under alternative simulated market structure with respect to the observed oligopoly. The first row refers to the full model (column "Full Model" in Table 2.2). The second row refers to model I which restricts all $\Gamma_{\mathbf{b}}$'s to be 0 in estimation (column "Model I" in Table 2.2). The third row refers to the model II which constrains all $\Gamma_{\mathbf{b}} = \gamma_0$ in estimation (column "Model II" in Table 2.2). The counterfactual is simulated for markets where all RTE cereal products and private label products are available.

Table 2.8: Merger Simulation: Nonlinear Pricing of Private Label

	Price change				Discount (\$)	Consumer Surplus change
	Private Label		Other Brands			
	RTE Cereals	Milk	RTE Cereals	Milk		
Full Model	50.78%	18.92%	0.05%	0.34%	3.147	-0.11%
Model II, $\Gamma_{\mathbf{b}} = \gamma_0$	23.36%	1.73%	-0.001%	-0.27%	2.381	1.83%

Notes: The Table reports average changes in prices (first four columns), discount in dollars on the bundles of private label RTE cereals and milk (the fifth column), consumer surplus (last column) under alternative pricing strategy of Private Label with respect to the observed one. The first row refers to the full model (column "Full Model" in Table 2.2). The second row refers to the model II which constrains all $\Gamma_{\mathbf{b}} = \gamma_0$ in estimation (column "Model II" in Table 2.2). The counterfactual simulation is conducted for markets where all RTE cereal products and private label products are available.

2.7 Conclusion

This paper considers the identification and estimation of a random coefficient discrete choice model of bundles, namely sets of products, when only product-level market shares are available. This last feature arises when only aggregate market shares, as opposed to individual purchases, are available, a very common phenomenon in practice. Following the classical approach with aggregate data, I consider a two-step method. First, using a novel inversion result where demand can exhibit complementarity, I recover the mean utilities of products from the product-level market shares. Second, to infer the structural parameters from the mean utilities while dealing with price endogeneity, I use IVs. I provide low-level conditions under which the model is globally identified through moment conditions based on such instruments. Finally, I illustrate the practical implementation of the methods and estimate the demand for RTE cereals and milk in the US. The demand estimates suggest that RTE cereals and milk are overall Hicksian complementary and these complementarities are heterogeneous across bundles. Ignoring such complementarities results in misleading counterfactuals.

As shown in section 2.3, merger simulations under linear pricing only require the identification of the product-level market share functions. This implies that one may not need to point estimate the demand synergy parameters and the distribution of the random coefficients in the GMM procedure to conduct such analyses. However, the estimation procedure in the current paper still assumes that the full model is identified. An interesting avenue for future research is to develop an adapted inference procedure for these counterfactuals that do not require the identification of the full model.

In practice, even though bundle-level market shares may not be available, other bundle-level information may still be accessible. For example, a household with a membership card may receive a discount if she purchases a specific bundle of products. An extension of the current paper is to explore identification under endogenous and observed bundle-level prices.

Similar to [Allen and Rehbeck \(2019a\)](#), the identification of the product-level market share functions remains agnostic about whether a bundle is in the choice set, i.e. $\Gamma_{\mathbf{b}} \neq -\infty$. As shown in [Theorem 9](#), in some models, one can identify whether $\Gamma_{\mathbf{b}} \neq -\infty$. However, allowing for some $\Gamma_{\mathbf{b}}$ being $-\infty$ may introduce boundary problems in estimation and therefore complicates inference (see [Andrews \(2002\)](#)). In practice, an important question is how to select out those $\Gamma_{\mathbf{b}}$ that are $-\infty$.

Finally, in the context of models of demand for single products, [Reynaert and Verboven \(2014\)](#) report remarkable efficiency gain by using optimal instruments. As mentioned in section 2.4, one may have non-trivial difficulties to construct the optimal instruments in the context of demand for bundles. An important question is whether and to what extent a similar approach can be used to further improve the practical performance of the proposed methods.

Estimating the Gains (and Losses) of Revenue Management*

Abstract. If the rise of revenue management has increased flexibility in the way firms set their prices, firms usually still impose constraints on their pricing strategy. There is yet scarce evidence on the gains or losses of such strategies compared to uniform pricing or fully flexible strategies. In this paper, we quantify these gains and losses and identify their underlying sources in the context of French railway transportation. This is complicated by the censoring on demand and the absence of exogenous price variations. We develop an original identification strategy on the demand that combines temporal variations in relative prices and moment inequalities stemming from basic rationality on consumers' side and weak optimality conditions on the firm's pricing strategy. Our results suggest significant gains of the actual revenue management compared to uniform pricing, but also substantial losses compared to the optimal pricing strategy. Finally, we highlight the key role of revenue management for acquiring information when demand is uncertain.

3.1 Introduction

The rise of e-commerce has led to a growing flexibility of prices. Firms seek to adjust quickly to demand shocks and exploit consumers' heterogeneity. This comes nevertheless at a cost. First, adjusting prices continuously requires specialized teams or good algorithms, usually both. Also, simple rules are usually set to simplify the pricing strategy. These rules may nonetheless be suboptimal. In this paper, we seek to identify how much gains can be expected by adopting flexible strategies compared to uniform pricing. We also seek to quantify the importance of losses of current strategies compared to the optimal ones, under various constraints imposed on such strategies. Finally, by varying these constraints and the assumptions behind the counterfactuals, we aim at identifying the main sources of these gains or losses.

We address these questions by studying revenue management at iDTGV, a subsidiary of the French railway monopoly, SNCF. Between 2004 and 2017, this firm provided low-cost trains from Paris to several towns in France, and the corresponding returns. Its revenue management was based on quantities, as is often the case in companies selling perishable goods (e.g. flight tickets, hotel rooms, rented cars for given periods etc.), namely, for the economy class on which we focus hereafter, 12 classes of prices sorted in ascending order and referred to as fare classes hereafter were defined. The price within each fare class for a given trip such as Paris-Bordeaux was set almost constant during the period we studied. Then, for each train, revenue managers could decide, at any moment before its departure, to close the current fare class and open the next one, therefore increasing the prices of the seats. We

*This paper is a joint work with Xavier D'Haultfœuille (CREST-ENSAE), Philippe Février (Veltys), and Lionel Wilner (INSEE)

investigate hereafter the relative benefits of this common pricing strategy compared to uniform pricing or strategies with a higher number of fare classes.¹

In order to compute such counterfactuals, we first show that in our context, recovering the price elasticity coefficient, relative demand parameters (of, e.g. Bordeaux versus Toulouse in Paris-Toulouse trains) and the total demand at a given price are sufficient to recover a rich set of counterfactual revenues. In particular, these counterfactual revenues do not depend on the timing of consumers' arrival. This is convenient here, as we do not observe such information. We can compute not only revenues under uniform pricing, but also revenues under optimal dynamic pricing, with any number of fare classes. Importantly also, we can compute such counterfactuals assuming either that iDTGV has complete or incomplete information on the demand for a given train.

The identification of price elasticity, relative demand parameters and the total demand at a given price are however complicated by two issues that are likely to arise in many markets of perishable goods. First, and as already observed by [Swan \(1990\)](#), [Lee \(1990\)](#) and [Stefanescu \(2012\)](#), we face a severe censoring problem here: demand at a given price is generally larger than the number of seats sold at that price. Second, prices vary only within the grids of 12 prices corresponding to each of the 12 fare classes. Hence, we cannot rely on usual instruments such as cost shifters.

To identify price elasticity, we rely on a new argument tailored to our application but that may apply to other contexts as well. Specifically, we exploit the fact that revenue management is done at a line level (e.g. Paris-Toulouse), while the train serves several cities (e.g. Bordeaux and Toulouse). This means that fare classes close at the same time for all destinations within the same line. Relative prices between, e.g. Bordeaux and Toulouse, then vary simultaneously whenever a fare class closes. We prove that by relating these variations between relative prices and the proportion of consumers buying tickets for one destination versus another, price elasticity can be identified. This can be achieved provided that price elasticities and the proportion of consumers seeking to buy a ticket for one destination versus another remain constant over time. We can test both conditions empirically and our results suggest that they are reasonable in our context.

Identification of the distribution over trains of total demand at a given price is also involved, in particular because of the censoring problem mentioned above. We first show that basic rationality conditions on the consumers lead to inequalities relating this total demand with the number of seats. We complement these inequalities by weak optimality conditions on the actual revenue management. Specifically, we assume that this revenue management was better on average than a uniform pricing practice based on incomplete information. Given our very purpose, it is important here not to impose too strong optimality conditions, such as optimality vis-à-vis all dynamic strategies, as these conditions would very much drive our final results. Also, our conditions have the advantage of being relatively simple to exploit for identification and estimation. At the end, our conditions based on consumers' rationality and weak optimality of the actual revenue management can be combined to form a set of moment inequalities.

Using this methodology, we first estimate a price elasticity at around -2, which is below but still in the range of most estimates in the transportation literature (see, e.g. [Jevons et al., 2005](#), for a meta-analysis). Note however that using aggregated quantities and prices to estimate price elasticity, as done by most of these studies, is likely to produce estimates that are substantially biased towards

¹For a detailed review of that practice and other revenue management techniques, see [Talluri and van Ryzin \(2005\)](#).

zero. Second, we show that even though based on weak conditions, the moment inequalities we use are sufficient to produce tight bounds on most counterfactual revenues. We find that the actual revenue management generates a gain of between 3.3 and 5.7% compared to optimal uniform pricing under incomplete information. However, we also estimate a loss of around 16% compared to the optimal pricing strategy under the same restriction of 12 ascending fare classes as those actually used. These results suggest that the actual revenue management practice was effective but still sub-optimal.

Third, we emphasize the key role of demand uncertainty on revenues. Compared to the optimal uniform pricing under incomplete information, the actual revenue management entails a gain of between 8 and 22%. But the actual revenue is also between 6 and 16% lower than what could have been achieved under uniform pricing with complete information on the demand. Compared to fully dynamic strategies, we estimate the loss to be between 8 and 18%. Interestingly, these two figures remain nearly the same under incomplete information. In other words, using the optimal quantity-based dynamic pricing strategy mitigates almost entirely the loss entailed by demand uncertainty.

Related Literature. Our paper relates to several theoretical and empirical literatures in operational research and economics. Theoretical literature on revenue management has investigated optimal quantity-based revenue managements, where firms segment demand by choosing either once for all or dynamically the allocation of, say, seats into fare classes in which prices are predetermined. We refer in particular to [Littlewood \(1972\)](#) and [Brumelle and McGill \(1993\)](#) for static solutions, and to [Gallego and Van Ryzin \(1994\)](#), [Feng and Gallego \(1995\)](#), [Feng and Xiao \(2000\)](#) for dynamic solutions. These last papers have studied optimal pricing strategies assuming that consumers arrive under some homogeneous Poisson process. We allow for flexible non-homogeneous Poisson arrival process, as [Bitran and Mondschein \(1997\)](#), [Zhao and Zheng \(2000\)](#), and [McAfee and te Velde \(2008a\)](#). Our demand model is close to this latter paper, but with one key difference. Whereas they assume that the firm has a complete information on the demand parameters, we also consider an incomplete information set-up where only the distribution of these parameters is known. The firm then updates this distribution as consumers arrive. Such an incomplete information set-up seems more plausible when, as here, aggregate demand is quite volatile. We thus extend [McAfee and te Velde \(2008a\)](#)'s theoretical results on optimal revenues to this case of incomplete information. We also extend it by studying constrained pricing strategies close to those implemented in practice.

Our results underline the important role of information and demand learning to explain the gains and losses of revenue management. Such a point has already been made in the theoretical literature but to our knowledge, we are the first to quantify these roles using real data. [Lin \(2006\)](#) studies similar models to ours (see his sections 5.1 and 5.2) and allows for firm's Bayesian learning from the observed purchases (or arrivals). Instead of deriving the optimal policy, his paper focuses on a specific policy (variable-rate) that is shown to be nearly optimal in simulations. [Aviv and Pazgal \(2002\)](#) studies different learning models featured by the way firm incorporates updated information. Finally, in contrast to all these papers and ours, [den Boer and Zwart \(2015\)](#) consider an updating approach based on maximizing the likelihood. We refer to [den Boer \(2015\)](#) for a complete survey on demand learning in dynamic pricing.

In the empirical literature on revenue management, the closest papers to ours are [Lazarev \(2013\)](#)

and Williams (2017), both of which study dynamic airline pricing in a monopolistic market.² While both papers accentuate price discrimination and its welfare effects, the main goal of our paper is to quantify the potential gains and loss due to revenue management in practice. As a result, different from their models, ours explicitly incorporates firm’s learning behaviour from the realised demand. Moreover, we do not impose the strong optimality of the observed prices. In Lazarev (2013), since sales are only observed quarterly, he uses a simpler model of traveller behaviour to approximate the expected aggregate demand functions and allows travellers to be forward-looking. In contrast, we model explicitly the demand functions at train level and reasonably assume travellers to be myopic (buy or leave) due to the restriction of increasing prices in iDTGV’s revenue management. Williams (2017) also considers the demand at flight level and dynamic adjustments of prices. Differently, he assumes that the airline monopoly’s complete information on demand and the price adjustment is only due to the randomness in the demand process, while we allow for incomplete information and, consequently, price adjustments can be additionally due to updated information on demand.

The rest of the paper is organized as follows. In Section 2, we present the context and our data. Section 3 displays the demand model and our assumptions on the supply side. Section 4 is devoted to the identification and estimation of demand under our assumptions and given the data at our disposal. Section 5 presents the results. The proofs of our identification results are gathered in the appendix.

3.2 Institutional Background and Data

3.2.1 Revenue Management at iDTGV in 2007-2009

iDTGV was a low-cost subsidiary of the French railway monopoly, SNCF, which was created in 2004 and disappeared in December 2017.³ It owned its trains and had a pricing strategy independent from SNCF. Its prices were generally lower than the full-rate prices of SNCF, but also associated with a slightly lower quality of services. Namely, tickets could only be bought on Internet, were nominative and could not be cancelled. They could be exchanged under conditions and at some cost.

Table 3.1: Routes with intermediate and final destinations

Line name	Final stop(s)	Intermediate stop(s)	Nb. of trains
Côte d’Azur	Cannes,St Raphael,Nice	Avignon	452
Marseille	Marseille	Aix-en-Provence/Avignon	453
Perpignan	Perpignan	Nîmes, Montpellier	689
Côte Basque	St Jean de Luz,Bayonne, Biarritz,Hendaye	Bordeaux	405
Toulouse	Toulouse	Bordeaux	411
Mulhouse	Mulhouse	Strasbourg	499
Total			2,909

Notes: we have different number of observations for the different lines because the period we cover varies slightly from one line to another.

²Another recent empirical paper is Cho et al. (2018) that studies revenue management under oligopoly in hospitality industry. Their analysis focuses on the pricing behavior of “hotel 0” (from which the demand data is obtained) in a competing environment.

³Its disappearance was not due to poor economic results, but rather to internal strategic considerations at SNCF.

The lines of iDTGV were all between Paris and other towns. For each of those towns and each day, one train was leaving Paris and another coming to Paris. Table 3.1 presents the lines we observe in our data on our period, namely from October 2007 till February 2009. These lines have several stops, but to simplify the analysis below, we gather them so as to form a single intermediate stop and a single final stop. We did this aggregation according to the price schedule. For instance, we aggregate Aix-en-Provence and Avignon together in the Paris-Marseille line since the corresponding prices are always the same. This gathering is consistent with Assumption 1 below, as our demand model remains valid after aggregation of cities.

Different lines may also share the same intermediate destination. For instance, Bordeaux is also the intermediate destination of Paris-Côte Basque. Finally, no tickets are sold between the intermediate and the final destination, e.g. some Bordeaux-Toulouse tickets in the Paris-Toulouse line. Our understanding is that this was done to avoid controlling people in intermediate destinations, as there were no ticket inspectors in the trains.

As usually, the trains are split into economy class cars and business class cars of fixed sizes. Revenue management is done almost independently between business and economic classes, i.e. under the sole constraint that prices in economy class are always lower than in business class. We focus hereafter on the economy class, which represents roughly 73% of the seats. In this category, there are 12 fare classes within which prices are constant and sorted by increasing order in terms of prices. The price corresponding to a given origin-destination trip (e.g. Paris-Bordeaux), at a peak time or off peak and for a certain fare class remained constant for several months (e.g. from 03/01/2007 to 10/31/2007) before being adjusted marginally, mostly to account for inflation. Contrary to SNCF, iDTGV did not make any third-degree price discrimination, so there was no discount for young people, old people or families.

In this context, revenue management consists in deciding at each moment to maintain the current fare class or to close it and move to the next one, resulting in an increase in the price. Coming back to a previous fare is impossible, so there are for instance no last minute drops in ticket prices for trains that have still several empty seats. Also, revenue managers could decide to close from the inception the first fare classes and begin to sell directly tickets in a higher fare class. Symmetrically, the 12th fare class may never be reached. In practice, a Computerized Reservation System (CRS) was used to manage reservation, which serves the basis for revenue management. Before the beginning of sales, it fixes a seat allocation planning for all fare classes, using the history of purchases on past trains. During sales, the CRS uses the number of tickets sold up to now to make recommendations on the size of subsequent fare classes. Revenue management manager can nevertheless always intervene, both on the initial and on subsequent seat allocations, according to his experience on past trains.⁴

Finally, and crucially for our identification strategy, the revenue management did not use separate fare classes for a given train with several destinations. For instance, in a Paris-Toulouse train, the closure of the first fare class occurred exactly at the same moment for both Bordeaux and Toulouse. Hence, price changes of Paris-Bordeaux and Paris-Toulouse tickets happened exactly at the same time, for all trains. According to discussions with people in the revenue management department, this was done to limit the quantities of decisions to be taken at each moment.

⁴Manager intervention in automatized revenue management also exists in other industries. See [Cho et al. \(2018\)](#) for an example in hospitality industry.

3.2.2 Data and descriptive statistics

We have data on iDTGV trains between October 2007 and February 2009 in economy class and for journeys from Paris only. We first observe basic characteristics of the trains: all the stops, departure and arrival time, day of departure (e.g. May 2, 2008) and whether it corresponds to a peak time or not. We also observe the price grid used for that train for each fare class. For each route and type of period (peak time or off peak), there are a limited number of such grids, as they change these grids only a few times during the period we observe (e.g. 3 times for the Paris-Toulouse). We also observe sales for each fare classes of all trains. On the other hand, we do not observe the purchasing dates, nor the opening moments of each fare class.

Table 3.2 presents some descriptive statistics on our data. We observe a substantial amount of price dispersion within trains. For instance on the Côte d’Azur line, the minimal price paid by consumers on average over the different trains (19.3€) was three times lower than the average maximal price (68.4€). We also observe substantial variations on the average load across lines. While trains in Paris-Marseille were always nearly full, with an average load above 95%, this was far from being the case on the Côte Basque line, with an average load of only 65.4%. This suggests that the actual pricing may not be fully optimal, at least for some lines.

Table 3.2: Descriptive statistics: economy class, from Paris

Line	Capacity	Average	% final	Prices		
		Load	destination	Avg	Avg min.	Avg max.
Côte d’Azur	324	85.4%	81.5%	50.3	19.3	68.4
Marseille	328	95.5%	60.0%	49.5	19.0	70.5
Perpignan	326	88.6%	27.4%	50.0	20.2	72.6
Côte Basque	350	65.4%	64.1%	37.3	19.7	53.3
Toulouse	351	87.3%	55.3%	43.6	19.4	67.2
Mulhouse	238	79.4%	24.1%	35.0	19.4	50.0

Notes: Avg min. and max. are the average of the minimal and maximal prices charged for each train, for the final destination. Capacity is the observed maximal number of sold places for each line.

3.3 Theoretical Model

3.3.1 Demand side

We consider a demand model that is close to McAfee and te Velde (2008a). A train T is defined by its line $\ell(T)$ (e.g. Paris-Toulouse) and its day of departure (e.g. May 2, 2008). For each line ℓ , we denote by a_ℓ the intermediate destination and by b_ℓ the final destination. To simplify notation and in the absence of ambiguity, we just denote the destinations of a train T by a and b instead of $a_{\ell(T)}$ and $b_{\ell(T)}$. For any train T , tickets are sold between the normalized dates $t = 0$ and $t = 1$. We denote the fare classes by $k \in \{1, \dots, K\}$. Within fare class k , tickets for train T and destination $d \in \{a_{\ell(T)}, b_{\ell(T)}\}$ are sold at price p_{dkT} . We recall that p_{dkT} belongs to a grid of K prices that remains fixed for several months and depends only on the destination d and whether the train leaves at a peak time or not. Finally, we denote by $D_{dT}(t, t'; p_d)$ the demand for destination d in train T between dates t and t' (with $(t, t') \in [0, 1]^2$) when the price is constant and equal to p_d . We then assume the following condition on this function.

Assumption 1. (*Consumers' demand*) For all d, T and $0 \leq t < t' \leq 1$,

1. $p \mapsto D_{dT}(t, t'; p)$ is non-increasing;
2. $D_{dT}(t, t'; p_d) \sim \mathcal{P}(B_T(t, t')\xi_d p_d^{-\varepsilon})$, where $\varepsilon > 1$, $t \mapsto B_T(0, t)$ is continuously differentiable and $\inf_{t \in [0, 1]} dB_T/dt(0, t) > 0$. Moreover, $\xi_a = 1$ and conditional on $B_T(t, t')$, $D_{dT}(t, t'; p_d)$ are independent across d .

The first condition is a weak rationality assumption on consumers. The second condition, even though it does not imply the first, is more substantial. Following the literature on revenue management (see, e.g. Gallego and van Ryzin, 1994), it first imposes that $D_{dT}(t, t'; p_d)$ follows a non-homogeneous Poisson process, with a constant price elasticity. As shown in Appendix C.1, this can be rationalized by combining a non-homogeneous process of consumers' arrival and a decision rule of purchase characterized by a Pareto distribution. The separability between train-time effect, $B_T(t, t')$, destination effect, ξ_d , and price effect, $p_d^{-\varepsilon}$ is key to identify of price elasticity ε and relative demand effect ξ_b . Importantly however, this condition can be tested, a point on which we come back in Section 3.4.2.

We do not impose any functional restrictions on the function $(t, t') \mapsto B_T(t, t')$. This is important because consumers arriving at different time may differ in their willingness to pay and the conditions of competing offers from other trains, airlines etc. may also vary with time. On the other hand, and in line with McAfee and te Velde (2008a), we assume that price elasticities ε are time invariant. This assumption rules out for instance, late arrival of more price inelastic consumers. Note however that we focus here on the economy class, and we can expect part of these consumers to chose the business class. Importantly also, we test implications of this assumption in Section 3.5.1 below. Finally, remark that $\xi_a = 1$ is a mere normalization, since we can arbitrarily multiply ξ_d by any positive constant, and modify $B_T(t, t')$ accordingly.

Assumption 1 together with a supply-side restriction (Assumption 3) turns out to be sufficient to identify ε and destination effects ξ_b , see Theorem 11 below. However, because of censorship, it is not sufficient to recover the distribution of total demand $B_T := B_T(0, 1)$. To this end, we consider the following restriction. Hereafter, X_T denotes the vector of train T 's characteristics that we observe, namely the route, week day, month and whether it operates on a rush hour or not.

Assumption 2. $B_T := B_T(0, 1)$ satisfies $B_T = \exp\{X_T' \beta_0\} \eta_T$ with $\eta_T | X_T \sim \Gamma(\lambda_0, \mu_0)$.

Assumption 2 assumes that the train "fixed effect" B_T is the product of a function of X_T and a term η_T that is independent of X_T and follows a gamma distribution. As detailed below, the assumption of a gamma distribution is not essential for most of our identification results, but is primarily made for computational convenience.

3.3.2 Supply side

We now formalize the features of revenue management already discussed in section 3.2.1. First, recall that the revenue management is operated at a line level (e.g. Paris-Toulouse) rather than for each route of this line (e.g. Paris-Bordeaux and Paris-Toulouse for the line Paris-Toulouse). We thus make the following assumption.

Assumption 3. (*revenue management at the line level*) The opening times of fare class $k \in \{1, \dots, K\}$, τ_k , is a stopping time with respect to the process $t \mapsto \sum_{d \in \{a, b\}} N_{dT}(t)$, where $N_{dT}(t)$ is the number of purchases for d made before t .

Assumption 3 states that the decision of opening a new fare class will depend only on past total purchases, rather than on the repartition between purchases for a and for b . Such an assumption is fully in line with the fact that a single fare class is used for the two destinations of each line. It was also confirmed by discussions we had with the revenue management department.

Our second assumption on the supply side is a weak optimality condition. To introduce it, let $R_T(p_a, p_b)$ denote the revenue for train T under a uniform pricing of (p_a, p_b) for destinations a and b respectively. Let also p_{dkT} denote the price in train T and fare class $k \in \{1, \dots, K\}$ for destination $d \in \{a, b\}$. The weak optimality condition we consider is the following:

Assumption 4. (*Weak optimality of actual revenue management*) We have

$$\max_{k=1, \dots, K} \mathbb{E} [R_T(p_{akT}, p_{bkT}) | X_T] \leq \mathbb{E} [R_T^{obs} | X_T]. \quad (3.1)$$

By averaging over X_T , which only includes rough proxies of the true demand, we allow for the possibility that revenue managers uses only incomplete information for its pricing strategy. In reality, it seems credible that it has access to additional signals on the true demand for a specific train. For instance, it could use the past number of purchases in each fare class on previous years for the same exact train. If so, we would expect that Inequalities (3.1) would also hold conditional on this information.

Importantly, Assumption 4 does not imply that the revenue management performs better than the optimal uniform pricing, as we compare the uniform pricing strategy within the grid (12 predetermined fare classes). Moreover, we do not impose any optimality vis-à-vis all dynamic strategies. We refrain to do so for several reasons. First, such an assumption would conflict with our very objective to quantify the gains or losses of the actual revenue management, compared to alternative scenarios. By definition, assuming a strong form of optimality would result in gains against most simpler pricing strategies.

Second and related to this first point, it seems very restrictive in our setting to assume that the optimal dynamic strategy was used among all such strategies, and given a perfect knowledge of the demand parameters (ε , ξ_b and the function $B_T(\cdot, \cdot)$). First, as we discussed in Section 3.2.1, the revenue management applied simplified rules (increasing fares from 12 predetermined fare classes) in the implementation of the revenue management and this only provides at most an approximation to the optimal solution. Moreover, seat allocation decisions were also subject to the manager's manual intervention, which could also be a source of suboptimality. Second, even if it knew such parameters, computing the optimal dynamic strategy is a very complicated dynamic programming problem. While [Feng and Xiao \(2000\)](#) and [McAfee and te Velde \(2008a\)](#) have proposed an algorithm for computing the solution for a homogeneous Poisson process, nothing has been done so far for the non-homogeneous case, to our knowledge. Third, given that iDTGV has been merely created in 2004, we can doubt that it knows perfectly the demand parameters, and in particular all train-specific effects B_T .

3.4 Identification and Estimation

In this section, we first detail our counterfactuals and corresponding parameters of interests. We also clarify in Theorem 10 which parameters of the demand function are needed to recover these parameters of interest. We then proceed to studying the identification of these relevant elements of the demand function. Finally, we show how to estimate the parameters and perform inference on them.

3.4.1 Counterfactuals and parameters of interests

We aim at comparing the current revenues with several counterfactual revenues, depending on the type of revenue management and the information it has access to. We consider several possible pricing strategies, from the most basic ones to the most sophisticated ones. The first, uniform pricing, simply corresponds to fixing the price of each route in a given train once for all. We let R_u denote optimal counterfactual revenues (averaged over all trains) under this pricing regime. At the other extreme, in “full” dynamic pricing, prices can be changed at any time. R_f then corresponds to optimal counterfactual revenues in this set-up. We also study pricing strategies, called stopping-time strategies hereafter, where prices can be changed only after a ticket is sold. The corresponding optimal revenues are then R_s . Finally, we consider constrained stopping-time strategies close to what was implemented in practice, by assuming that only M number of fares, or M increasing fares, are allowed. The corresponding optimal revenues are denoted by R_{sM} and R_{sM+} respectively.

Hereafter, we consider two scenarios in terms of information available to the revenue managers. Recalling that $B_T = B_T(0, 1)$, we use hereafter $g_T(t, t') = B_T(t, t')/B_T$ for all $t < t'$.

1. (Complete information) Revenue managers fully know the expected demand for each train. Thus, they observe $\xi_a, \xi_b, \varepsilon$ but also $B_T(t, t')$ for all (t, t') and each train T .
2. (Incomplete information) At $t = 0$, revenue managers know ξ_d, X_T and $g_T(t, t')$ for all $t < t'$, but only $f_{B_T|X_T}$. As time goes by, revenue managers update their information on B_T according to Bayes' rule.

The complete information case should be seen as a benchmark. It is useful in particular to quantify the value of information and contrast the gains of revenue management in complete and incomplete information set-ups. The case of incomplete information is probably more realistic. In this scenario, revenue managers know, for each train, the pattern of consumers' arrival over time ($g_T(., .)$) but does not know exactly the aggregate demand (B_T). The assumption that $g_T(., .)$ is known makes especially sense if $g_T(., .)$ does not depend on T , in which case revenue managers can have learned from previous trains how consumers arrive through time. If the scenario of incomplete information holds in practice, the differences between the counterfactual revenues and the observed ones can be interpreted as the potential gains or losses of the optimal revenue management under different constraints compared to the actual ones.

Hereafter, we use the exponents c and i to specify the two information set-ups. Hence, R_u^c denotes for instance is the counterfactual optimal revenue under uniform pricing and complete information. The following theorem clarifies which parameters of the demand are required to identify all the counterfactual revenues we consider hereafter.

Theorem 10. *Suppose that Assumptions 1 and 3 hold. Then, R_r^I is a function of $(f_{B_T|X_T}, \varepsilon, \xi_b)$, for $I = c$ and $r \in \{u, f, s, sM, sM+\}$ or $I = i$ and $r \in \{u, s, sM, sM+\}$. The same is true for R_f^i if Assumption 2 also holds.*

We do not specify here the exact forms of the counterfactual revenues, as they have no closed forms, but more details are provided in Appendix C.3. The key result of Theorem 10 is that counterfactual revenues do not depend on $g_T(\cdot, \cdot)$. This turns out to be crucial, since we have no information on purchasing dates and thus no way to recover this function. Instead, we only have to recover the price elasticity ε , the destination effects ξ_b and the conditional distribution of the total demand B_T in order to identify the counterfactual revenues.

McAfee and te Velde (2008b) obtains the result for the “full” dynamic pricing strategy under complete information and a similar demand model. We extend their results in two directions. First, we consider other types of pricing strategies, and in particular possibly constrained stopping-time strategies, which are very common in practice and correspond to the actual revenue management. Second, we also show a similar result in an incomplete information set-up.

Challenges in identification We face two main issues for recovering the demand parameters. First, demand is actually unobserved; only bounds on it can be obtained. Let n_{dkT} denote the number of sales for train T , fare class $k \in \{1, \dots, K\}$ and destination $d \in \{a, b\}$. Then

$$D_{dT}(p_{dkT}) \geq D_{dT}(\tau_{k,T}, \tau_{k+1,T}; p_{dkT}) = n_{dkT},$$

where $\tau_{k,T}$ is the (random) time at which the k th fare class opens, which we do not observe. Hence, without further assumptions, we only observe a crude lower bound on the total demand at price p_{dkT} . This point was already made in similar contexts by Swan (1990), Lee (1990), and Stefanescu (2012).

The second issue we face is the absence of usual instruments for prices. Prices only vary within the grid specified by revenue managers, and to our knowledge, fare classes did not close for exogenous reasons unrelated to demand. In other words, there is no exogenous variations of prices in our context. The bottom line is that usual strategies to identify the demand function do not apply here.

We now show that despite these limitations, it is possible to point or partially identify the parameters $(\varepsilon, \xi_b, \beta_0, \lambda_0, \mu_0)$, where $(\beta_0, \lambda_0, \mu_0)$ are defined in Assumption 2. Then, in view of Theorem 10, we obtain bounds on the counterfactual revenues. We proceed in two steps hereafter, by first showing point identification of (ε, ξ_b) and then partial identification of $(\beta_0, \lambda_0, \mu_0)$.

3.4.2 Identification of (ε, ξ_b)

Our strategy to identify (ε, ξ_b) is to exploit variations in the relative prices p_{bkT}/p_{akT} between the two destinations and from one fare class to another. We start from $n_{dkT} = D_{dT}(\tau_k, \tau_{k+1}; p_{dkT})$. Now, let us assume first that τ_k and τ_{k+1} are deterministic. Then, by Assumption 1, $D_{aT}(\tau_k, \tau_{k+1}; p_{ak})$ and $D_{bT}(\tau_k, \tau_{k+1}; p_{bk})$ are independent conditional on $B_T(\tau_k, \tau_{k+1})$ and follow Poisson distributions. As a result,

$$n_{bkT}|n_{akT} + n_{bkT} = n \sim \text{Binomial}(n, \Lambda(\ln(\xi_b) - \varepsilon \ln(p_{bkT}/p_{akT}))), \quad (3.2)$$

where $\Lambda(x) = 1/(1 + \exp(-x))$. Hence, as long as there are variations through fare classes k in the relative prices p_{bkT}/p_{akT} , we can identify ξ_b and ε . In the data, we do observe such variations. In Paris-Toulouse for instance, p_{bkT}/p_{akT} vary from 1 for $k = 1$ to 1.18 for $k = 12$.

To obtain (3.2), we have assumed so far that the stopping times $(\tau_k)_{k=1,\dots,K}$ were fixed, which is not realistic. Nonetheless, the previous result shows that (3.2) still holds provided that these stopping times satisfy Assumption 3.

Theorem 11. *Suppose that Assumptions 1 and 3 hold and $k \mapsto p_{akT}/p_{bkT}$ is not constant. Then (3.2) holds and (ξ_b, ε) are point identified.*

Equation (3.2) does not hold for any possible random stopping times. We can easily build counterexamples by making $(\tau_k)_{k=1,\dots,K_T}$ depend solely on $N_{aT}(\cdot)$, for instance. Such situations are however ruled out by Assumption 3. Under this condition, intuitively, the stopping times will be independent of the proportion of consumers buying tickets for a (versus b).

Beyond the identification and estimation of ξ_b and ε , Equation (3.2) can be the basis of testing some of the conditions we have imposed. First, the separability between train effect $B_T(t, t')$ and destination effect ξ_d in Assumption 1 implies that if $p_{bkT} = p_{akT}$ for several fare classes k , we should observe similar proportions $n_{bkT}/(n_{akT} + n_{bkT})$ for the corresponding k . Second, we can also test for the fact that price elasticities do not evolve over time, by considering more general specifications than (3.2). Finally, we have imposed that the price elasticity was constant for all lines. Though we could in fact identify a different price elasticity for each line, we made that restriction for consistency, because several lines share common origin-destination trips (e.g. Paris-Toulouse and Paris-Côte Basque share the Paris-Bordeaux). But we can at least separate the lines that have no such trips in common into different groups and identify a different price elasticity for each of these groups. We consider all these robustness checks in Section 3.5.1 below.

3.4.3 Partial identification of $\theta_0 = (\beta_0, \lambda_0, \mu_0)$

As mentioned earlier, the total demand is not directly observed. Thus, unless we impose very restrictive conditions, we cannot point identify the parameters $\theta_0 = (\beta_0, \lambda_0, \mu_0)$ governing under Assumption 2 the conditional distribution of $B_T|X_T$. We then partially identify θ_0 by building moment inequalities based on consumers' rationality (Assumption 1.1) and weak optimality of the actual revenue management (Assumption 4).

Consumers' rationality First, by Assumption 1.1, all consumers who bought a ticket for d at price p_{djT} for $j \geq k$ would have also bought it at price p_{dkT} . Therefore, for all $k = 1, \dots, K$ and $d \in \{a, b\}$,

$$D_{dT}(p_{dkT}; \theta_0) \geq \sum_{j=k}^K n_{djT},$$

where we now index total demand $D_{dT}(p_{dk})$ by θ_0 . Let C_T denote the capacity of train T . Then we also have $C_T \geq \sum_{j=k}^K n_{djT}$. Combining these inequalities and integrating over η_T , we obtain, for all $k = 1, \dots, K$ and $d \in \{a, b\}$,

$$\mathbb{E} \left[\sum_{j=k}^K n_{djT} - C_T \wedge D_{dT}(p_{dk}; \theta_0) \middle| X_T \right] \leq 0. \quad (3.3)$$

While $\mathbb{E}[C_T \wedge D_{dT}(p_{dk}; \theta_0)|X_T]$ does not have any closed form, we can compute it easily through simulations.

Next, let K_T denote the last fare class that was open. When the train is not full,

$$\sum_{j=1}^K n_{djT} \geq D_{dT}(p_{dK_T T}; \theta_0),$$

since some of the consumers who purchased their ticket at the price p_{dkT} , with $k < K_T$, would not have purchased it at price $p_{dK_T T}$. Therefore, when the train is not full,

$$\sum_{j=1}^{K_T} (n_{ajT} + n_{bjT}) \geq C_T \wedge (D_{aT}(p_{aK_T T}; \theta_0) + D_{bT}(p_{bK_T T}; \theta_0)). \quad (3.4)$$

When the train is full, this inequality also holds since $\sum_{j=1}^K n_{ajT} + n_{bjT} = C_T$. Hence, (3.4) holds in all cases. Integrating over η_T , we obtain

$$\mathbb{E} \left[C_T \wedge (D_{aT}(p_{aK_T T}; \theta_0) + D_{bT}(p_{bK_T T}; \theta_0)) - \sum_{j=1}^K (n_{ajT} + n_{bjT}) \middle| X_T \right] \leq 0. \quad (3.5)$$

Similarly to above, even if the first term in the expectation does not take a closed form, it can be approximated easily using simulations.

Weak optimality condition We now rely on Assumption 4 to form additional moment inequalities. To exploit them, we show in the proof of Theorem 10 that under Assumptions 1-2,⁵

$$\mathbb{E}[R_T(p_a, p_b; \theta_0) | X_T] = \frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} \int_0^\infty \mathbb{E}[D([\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}] e^{X_T \beta_0} z) \wedge C_T] g_{\lambda_0, \mu_0}(z) dz,$$

where $D(u) \sim \mathcal{P}(u)$ and g_{λ_0, μ_0} is the density of a $\Gamma(\lambda_0, \mu_0)$. In other words,

$$\max_{k=1, \dots, K} \mathbb{E}[R_T(p_{akT}, p_{bkT}) | X_T]$$

is an identified function $h(X_T, \theta_0)$. Hence, (3.1) can be rewritten as:

$$\mathbb{E}[h(X_T, \theta_0) - R_T^{\text{obs}} | X_T] \leq 0. \quad (3.6)$$

To summarize, (3.3), (3.5) and (3.6) can be stacked together, producing

$$\mathbb{E}(m(U_T, \theta_0, \nu_0) | X_T) \leq 0, \quad (3.7)$$

for some known function $m(\cdot, \cdot, \cdot) \in \mathbb{R}^L$ and $\nu_0 = (\xi_b, \varepsilon)$. Here U_T is the vector of data corresponding to train T and the inequality sign should be understood componentwise. In general, the inequalities (3.7) are not sufficient to point identify θ_0 . In turn, this implies that the counterfactual revenues are partially identified under Assumption 1-4. Note that we let the dependence of $m(\cdot, \cdot, \cdot)$ on ν_0 explicit in (3.7). Even if this is immaterial for identification, it matters for inference, to which we now turn.

⁵See (C.27) in section C.7.2.

3.4.4 Inference on counterfactual revenues

Because counterfactual revenues are only partially identified, we directly focus here on constructing confidence sets for them. For that purpose, we first construct a joint confidence set on (ν_0, θ_0) (with $\nu_0 = (\xi_b, \varepsilon)$) and then apply the projection method, using Theorem 10. The projection method is conservative but simple and as such, has been often used in empirical literature, see e.g. [Ciliberto and Tamer \(2009\)](#) and [Grieco \(2014\)](#) for examples in empirical industrial organization.⁶

To construct a joint confidence set on (ν_0, θ_0) , we first construct a confidence set on ν_0 . Equation (3.2) gives a straightforward way to estimate ν_0 , by a logit of $Y = \mathbf{1}\{\text{destination} = a\}$ on $Z = \log(p_{bk}/p_{ak})$ for each (k, T) . Then, for any $\alpha \in (0, 1)$, we can construct a confidence set $\text{CI}_{1-\alpha}^1$ on ν_0 with coverage $1 - \alpha$ by using the asymptotic distribution of the MLE.

Next, assuming first that ν_0 is known, we can construct confidence sets on $\theta_0 = (\beta_0, \lambda_0, \mu_0)$ using the conditional moment inequalities (3.7). Specifically, because X_T is discrete in our case, we can first transform these conditional moment inequalities into unconditional moment inequalities. We then use the Generalized Moment Selection (GMS) of [Andrews and Soares \(2010\)](#). We let $\text{CI}_{\nu, 1-\alpha}^2$ denote a confidence set on θ_0 obtained by this procedure if $\nu_0 = \nu$.

Now, ν_0 actually enters into these moment inequalities. To account for this additional source of uncertainty and construct a joint confidence set on (ν_0, θ_0) , we propose to consider, for any $\alpha_1 \in (0, \alpha)$, the confidence sets

$$\text{CI}_{1-\alpha} = \cup_{\nu \in \text{CI}_{1-\alpha_1}^1} \{(\nu, \theta) : \theta \in \text{CI}_{\nu, 1-\alpha+\alpha_1}^2\}.$$

We then have

$$\begin{aligned} \Pr((\nu_0, \theta_0) \in \text{CI}_{1-\alpha}) &= \Pr(\theta_0 \in \text{CI}_{\nu_0, 1-\alpha+\alpha_1}^2, \nu_0 \in \text{CI}_{1-\alpha_1}^1) \\ &\geq \Pr(\theta_0 \in \text{CI}_{\nu_0, 1-\alpha+\alpha_1}^2) + \Pr(\nu_0 \in \text{CI}_{1-\alpha_1}^1) - 1 \end{aligned}$$

Hence, because $\text{CI}_{1-\alpha_1}^1$ (resp. $\text{CI}_{\nu_0, 1-\alpha+\alpha_1}^2$) has asymptotic coverage of $1 - \alpha_1$ (resp. $1 - \alpha + \alpha_1$), $\text{CI}_{1-\alpha}$ will asymptotically cover (ν_0, θ_0) with probability $1 - \alpha$.

3.5 Results

3.5.1 Demand estimation

3.5.1.1 Estimation of ν_0

We first consider the estimation of the relative demand (ξ_b) and the price elasticity ($-\varepsilon$). The results are displayed in Column I of Table 3.3. We obtain a price elasticity of -2.07. This result is slightly larger but in line with the literature on the transportation industry. We refer for instance to the meta-analysis by [Jevons et al. \(2005\)](#) and the studies of [Wardman \(1997\)](#), [Wardman \(2006\)](#) and [Wardman et al. \(2007\)](#), which point to price elasticities in the range $[-1.3; -2.2]$. Unlike ours, most of the studies rely on aggregated data. This is likely to bias upwards price-elasticity estimates, a point that we illustrate in Appendix C.4 by running regressions based on our data aggregated at different levels.

⁶An alternative to projection is to conduct direct inference on functions of parameters, see in particular [Bugni et al. \(2017\)](#); [Kaido et al. \(2019\)](#). But our parameters of interest are non-linear with respect to (ν_0, θ_0) and their computation is costly. This makes the direct approaches either not applicable (as the parameters are not linear in (μ_0, θ_0)) or computationally prohibitive.

Table 3.3: Binomial model of demand

	All fare classes			First 11 fare classes	
	I	II	III	I'	II'
Price elasticity ($-\varepsilon$)					
Constant	-2.07 (0.06)	-2.20 (0.06)	-1.98 (0.06)	-1.94 (0.07)	-2.32 (0.07)
Southwest			0.34 (0.07)		
Weekend/national holidays			-0.74 (0.08)		
Peak hour			0.009 (0.08)		
Final (vs intermediate)					
Destination effect ($\ln \xi_b$)					
Côte d'Azur (vs Avignon)	1.53 (0.007)			1.54 (0.007)	
Marseille (vs Aix-Avignon)	0.41 (0.005)			0.42 (0.005)	
Perpignan (vs Nîmes-Montpellier)	-0.87 (0.005)			-0.89 (0.005)	
Côte Basque (vs Bordeaux)	0.59 (0.007)			0.58 (0.007)	
Toulouse (vs Bordeaux)	0.27 (0.006)			0.27 (0.006)	
Mulhouse (vs Strasbourg)	-1.15 (0.008)			-1.15 (0.008)	
Dest. characteristics W_d					
Population		0.941 (0.006)	0.951 (0.006)		0.958 (0.006)
Regional capital		0.44 (0.004)	0.44 (0.004)		0.45 (0.004)
Trav. time by train		-1.90 (0.001)	-1.90 (0.001)		-1.88 (0.001)
Trav. time by train squared		0.25 (0.0002)	0.25 (0.0002)		0.25 (0.0002)
Trav. Dist. by car		0.32 (0.002)	0.34 (0.002)		0.33 (0.002)
Trav. Dist. by car, squared		-0.025 (0.0001)	-0.027 (0.0001)		-0.026 (0.0001)

Notes: The total number of trains is 2,909. In specifications with all fare classes, the total number of observations (fare classes \times trains) is 21,988. Southwest correspond to the lines to Côte Basque, Toulouse and Perpignan. In the specifications including W_d , we replace $\ln(\xi_d)$ by $W_d'\zeta$. Population is measured in 1 million inhabitants, traveling time is measured in hours and traveling distance is measured in 100 kilometers.

Regarding our estimates of relative demand coefficients ($\ln(\xi_b)$), a positive coefficient should be read as a higher demand for the terminal than for the intermediate stop, and its exponential corresponds to the ratio of demand. For instance, the demand for Mulhouse (the terminal) is equal to $\exp(-1.15) \approx 0.32$ times the demand for Strasbourg, which is the intermediate stop. To check whether the results on

relative demand make sense, we re-estimate the model adding some structure on these relative demand parameters. Namely, instead of letting $\ln(\xi_d)$ unspecified, we replace it by $W_d'\zeta$, where W_d includes the population of the town, the traveling time from Paris by train and its square, and the traveling distance by car from Paris and its square. The latter captures the outside option of traveling by car. Note that with such a specification, we do not have to impose the normalization $\xi_a = 1$ anymore. Instead, we just replace $\ln(\xi_b/\xi_a)$ by $(W_b - W_a)'\zeta$, and estimate ζ .

The results are presented in Column II. The estimated price elasticity remains similar as in Column I (-2.20). The effect of the population size, traveling time by train and distance by car are as expected. Larger cities lead to higher demand and a longer traveling time (resp. distance) by train (resp. car) leads to a lower (resp. higher) demand for train tickets. The effect of traveling time is nonetheless attenuated for long journeys, as the positive coefficient of the square of traveling time indicates.

We also estimate the demand model by allowing price elasticity to vary across lines and across trains and the results are shown in Column III. The estimated price elasticities are quantitatively close to the estimates in Columns I-II. We find that travelers of lines from Paris to the southwest of France (namely, the lines to Côte Basque, Toulouse and Perpignan) are less sensitive to price than those of other lines. Travelers on weekend or national holidays have a larger price elasticity (in absolute value) than those on other days. This may stem from a larger proportion of people traveling for business reason during weekdays. Finally, once controlling for weekend and national holidays, individuals traveling during peak hours appear to have a similar elasticity than the others.

We now check the plausibility of Assumption 1, on which the identification of ν_0 relies. First, we investigate the separability between train-time effect ($B_T(t, t')$), destination effect (ξ_d) and price effect ($p_d^{-\varepsilon}$). This separability implies that the proportions $n_{bkT}/(n_{akT} + n_{bkT})$ remain constant through fare classes k satisfying $p_{bkT} = p_{akT}$. A convenient way to check this is to restrict ourselves on two lines, Paris-Marseille and Paris-Mulhouse, for which $p_{bkT} = p_{akT}$ for all $k \in \{1, \dots, K\}$. By taking the first fare class as a reference, we simply regress $n_{bkT}/(n_{akT} + n_{bkT})$ over the other 11 fare class dummies and test whether the corresponding coefficients are all equal to zero.

Table 3.4: Test of the separability in Assumption 1

Fare class	Paris-Marseille	Paris-Mulhouse
	Coefficient estimates	
2	0.019	-0.003
3	0.042***	0.009
4	0.022	0.010
5	0.009	0.010
6	0.008	0.027**
7	0.004	0.021
8	0.016	0.004
9	-0.023	-0.004
10	0.018	-0.002
11	0.011	-0.063**
12	0.020	-0.100***
Joint test of nullity	p-values	
2-12	0.09	0.0047
2-11	0.05	0.08

Notes: Coefficient estimates of the regression of $n_{bkT}/(n_{akT} + n_{bkT})$ on fare class dummies, with Fare class 1 taken as the reference. The intermediate stops for Paris-Marseille are Avignon and Aix-en-Provence. The intermediate stop for Paris-Mulhouse is Strasbourg.

The results are presented in Table 3.4. As the first part of the table emphasizes, most coefficients are not significant, despite the large number of observations (453 and 499 for the two lines). For Paris-Marseille, the p-value of the joint test is larger than 0.05. For Paris-Mulhouse, the p-value is lower, but it appears that this result is mostly driven by the last fare class. The coefficient of the 12th fare class is indeed negative and quite large for this line, indicating that there would be more “late consumers” traveling to Strasbourg. To see whether this pattern could influence our results beyond this specific line, we re-estimate ν_0 using only the first 11 fare classes. The results are shown in columns I' and II' of Table 3.3. We obtain respectively a price elasticity of -1.94 and -2.32, both of which are very close to their baseline estimates in columns I and II. The relative demand parameters and parameters in destination effects are also very close. Considering only the first 10 fare classes also lead to almost identical results, with in particular price elasticities of -1.94 and -2.27.

Finally, we test that the price elasticity remains constant over time. To this end, we replace $\Lambda(\ln(\xi_b) - \varepsilon \ln(p_{bkT}/p_{akT}))$ in (3.2) by

$$\Lambda[\ln(\xi_b) - (\varepsilon_{\text{early}}1\{k < S\} + \varepsilon_{\text{late}}1\{k \geq S\}) \ln(p_{bkT}/p_{akT})], \quad (3.8)$$

for some threshold S that we vary. We can then compare $\varepsilon_{\text{early}}$ to $\varepsilon_{\text{late}}$ to assess the credibility of this assumption of constant price elasticities. We consider threshold values S equal to 3, 11 and 12.

The results are displayed in the first lines of Table 3.5. The resulting price elasticities are close to the baseline elasticities (-2.07 and -2.20), except for $S = 3$, where $\varepsilon_{\text{early}}$, while still including in its 95% confidential interval the estimate of ε , is lower in magnitude. Note that intertemporal price discrimination would suggest the opposite, namely that price elasticity of early birds would be higher in absolute value. Next, we consider three classes of elasticities instead of two, by separating early,

middle and late travelers. In the specification 1-2, 3-10, 11-12 for instance, we separate the first two fare classes, the third to tenth fare classes and the last two classes. Again, in this specification and the other 1-2, 3-11, 12, we obtain again similar results. As before, the baseline elasticity is still within the 95% confidential intervals of $\varepsilon_{\text{early}}$ and one can not reject that they are equal at 95% confidential level. Note that when early travelers are defined as those who arrive before the opening of the 2nd fare class, the prices for both destinations of the 1st fare class are the same, $p_{a1T} = p_{b1T}$. Then, we can not identify $\varepsilon_{\text{early}}$ in this case.

Table 3.5: Test of constant elasticity over time

Segmentation Early(/Middle)/Late travelers	without structure on ξ_d			with structure on ξ_d		
	$\varepsilon_{\text{early}}$	$\varepsilon_{\text{middle}}$	$\varepsilon_{\text{late}}$	$\varepsilon_{\text{early}}$	$\varepsilon_{\text{middle}}$	$\varepsilon_{\text{late}}$
1-2, 3-12	-1.65 (0.37)	-	-2.07 (0.06)	-2.17 (0.41)	-	-2.23 (0.06)
1-10, 11-12	-2.13 (0.07)	-	-1.92 (0.13)	-2.35 (0.07)	-	-2.15 (0.11)
1-11, 12	-2.15 (0.06)	-	-1.47 (0.20)	-2.33 (0.07)	-	-2.17 (0.13)
1, 2-10, 11-12	-	-2.12 (0.07)	-1.99 (0.11)	-	-2.41 (0.07)	-2.00 (0.11)
1, 2-11, 12	-	-2.06 (0.07)	-2.10 (0.13)	-	-2.33 (0.07)	-2.17 (0.13)
1-2, 3-11, 12	-1.46 (0.41)	-2.06 (0.07)	-2.09 (0.13)	-2.20 (0.41)	-2.33 (0.07)	-2.17 (0.13)
1-2, 3-10, 11-12	-1.49 (0.41)	-2.13 (0.07)	-1.98 (0.11)	-2.16 (0.41)	-2.44 (0.07)	-2.00 (0.11)

Notes: In the three first lines, we consider only two different elasticities, in line with (3.8). In the fourth and fifth lines, we consider three different elasticities for fare classes 1, 2-10, 11-12 and 1, 2-11, 12. Because prices for all destinations in the first fare class are the same for all lines, the corresponding price elasticity is not identified. Standard errors are reported under parenthesis.

3.5.2 Counterfactuals

We now turn to the counterfactual optimal revenues under different pricing strategies, namely uniform, stopping-time, and full dynamic pricing. Table 3.6 summarizes average revenues over all lines under different pricing strategies. Our results first show that the actual revenue management achieves an expected revenue of 12.2K€, which is between 3.3 and 5.7% higher than the optimal uniform pricing strategy under incomplete information (line u.2 of Table 3.6). Note that Assumption 4 did not impose this condition, since we only suppose that the actual revenue management leads to a higher revenue than uniform pricing with prices constrained to belong to the grid set by revenue managers (i.e. line u.1). Actually, this grid seems fine enough to approximate correctly optimal prices, as the revenue under uniform pricing with such constrained prices (line u.1) is very close to that without price constraints (line u.2). More precisely, we estimate a loss of only 200€, or 1.7%, by imposing such constraints on prices.

Table 3.6: Optimal revenues under different pricing strategies

Scenario	Estimate (in thousands of €)	95% CI
Observed pricing strategy	12.21	[12.05, 12.36]
Optimal uniform pricing strategy		
u.1 incomplete information, constrained prices		[11.35, 11.59]
u.2 incomplete information, unconstrained prices		[11.55, 11.82]
u.3 complete information, constrained prices		[13.48, 13.76]
u.4 complete information, unconstrained prices		[14.48, 14.81]
Optimal stopping-time pricing strategy		
s.1 incomplete information, 2 increasing fares		[13.73, 14.04]
s.2 incomplete information, 2 fares		[14.16, 14.48]
s.3 incomplete information		[14.72, 15.05]
s.4 complete information, 2 increasing fares		[14.61, 14.95]
s.5 complete information, 2 fares		[14.68, 15.01]
s.6 complete information		[14.74, 15.09]
Optimal “full” dynamic pricing strategy		
f.1 incomplete information		[14.73, 15.07]
f.2 complete information		[14.76, 15.10]

Notes: With “constrained prices” (resp. “unconstrained prices”), optimization is conducted over the actual price grid (resp. over all positive real numbers). Revenues are averaged over all lines. We use bootstrap (5000) to compute CI for observed revenue. To compute 95% CIs of other revenues under different counterfactuals, we use the GMS procedure and projection method as described in Section 3.4.4.

Our results also suggest that even compared to the incomplete information benchmark, the actual revenue management was suboptimal. In practice, revenue managers could use stopping-time strategies with up to 12 increasing fares chosen from a predetermined set. By using stopping-time strategies with 2 increasing fares, revenue managers could have already obtained around 13.93K€, a gain around 14.1% relative to the actual practice; with 12 increasing fares, they would have obtained a gain around 19.0% (see Table 3.7).⁷ These results justify our weak optimality assumption on the supply side.

⁷We also computed the optimal revenue under 2 increasing fares chosen from the actual price grid and also obtained that this pricing strategy achieves on average a better revenue than the observed one.

Table 3.7: Optimal revenues under constrained stopping-time pricing

Number of fares (unif. pricing)	Complete information		Incomplete information	
	No constraint	increasing fares	No constraint	increasing fares
1	14.685		11.751	
2	14.884	14.822	14.368	13.929
3	14.925	14.856	14.729	14.324
4	14.939	14.869	14.824	14.440
5	14.946	14.876	14.864	14.485
6	14.949	14.880	14.885	14.507
7	14.951	14.882	14.899	14.518
8	14.953	14.884	14.908	14.525
9	14.954	14.885	14.915	14.528
10	14.954	14.886	14.920	14.530
11	14.955	14.886	14.924	14.531
12	14.955	14.887	14.928	14.532

Notes: the revenues are computed at a set of parameters “at the middle” of confidential region, i.e., the set of parameters satisfies the GMS criterion and is not rejected by the data, as described in Sections 3.4.4 and ??.

A related finding is that restricting to stopping-time pricing strategies incur virtually no loss, compared to “full” dynamic pricing. By changing prices only when a purchase is observed, the firm can secure around 99% of the revenue gain from uniform pricing to dynamic pricing regimes (compare lines s.4 and f.1, or similarly lines s.6 and f.2 in Table 3.6). Hence, the optimal stopping-time pricing strategy provides a simple yet very good approximation to the optimal dynamic pricing strategy. This finding seems valuable in particular when changing prices incurs a non-negligible cost.

We now turn to the comparison with the complete information set-up. Table 3.6 first shows that compared to its actual revenue management, the firm could have expected at most a gain between 17.3% and 19.1% (see line f.2), which corresponds to the full dynamic pricing case under complete information. Also, the difference in pricing strategies are notably modest in this complete information case. In particular, the difference in revenue between uniform pricing with unconstrained prices and full dynamic pricing is only around 2%.⁸ This figure sharply contrasts with the 28% gain we estimate under incomplete information (cf. line f.1 v.s. u.2). Intuitively, dynamic pricing still helps in the complete information case because of the randomness of the demand process. But this randomness plays a much more minor role than the variations in overall demand (B_T). The only substantial loss we observe in the complete information case (around 7%) is when prices are constrained to belong to the grid, see lines u.3 and u.4. Intuitively, this is because demand may be very high occasionally, and the maximal price in the grid is still too cheap to be close to profit maximization

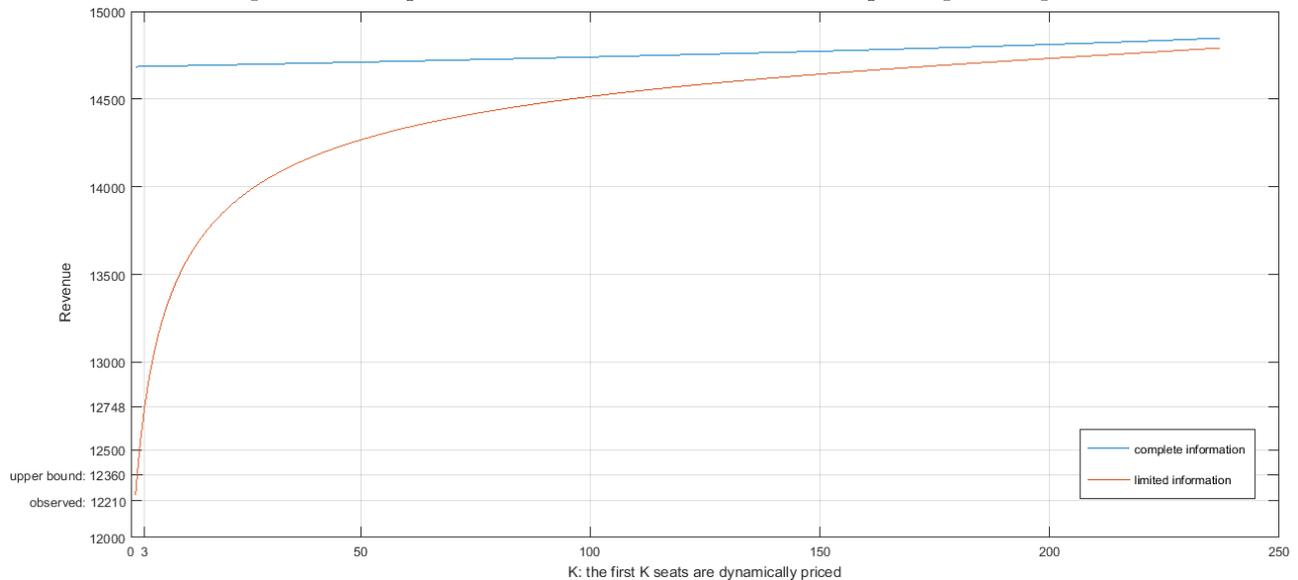
The difference between the gains of full dynamic pricing under complete and incomplete information shows that revenue management is an effective instrument for demand learning. By learning from consumers’ purchases, it can gradually pin down the uncertainty on B_T in a Bayesian way. Pricing

⁸Our results under complete information are consistent with simulation results in operational research and empirical results in economics. For example, [Zhao and Zheng \(2000\)](#) shows a similar improvement by between 2.4% and 7.3%. [Williams \(2017\)](#) estimates a revenue improvement due to optimal dynamic pricing of around 2% in airline industry.

decision then take this renewed information into account, improving total revenue. And actually, this demand learning can compensate almost all revenue loss due to ex ante uncertainty on demand. The difference in revenues under optimal uniform pricing between incomplete and complete information is around 3K€ (lines u.4 and u.2), while this difference decreases to around 0.03K€ only (see lines f.2 and f.1 in Table 3.6) under optimal dynamic pricing.

The reason of this very modest loss compared to the complete information set-up is that information accumulates quickly. To illustrate this, we computed expected revenues under a class of intermediate stopping-time pricing strategies, where the firm is only allowed to dynamically price the first K seats while uniform pricing is applied to the remaining seats. Thus, $K = 0$ and $K = C$ correspond respectively to the optimal uniform and stopping-time pricing strategies. Even if such a class of pricing strategy is not covered by Theorem 10, we show in Appendix C.8 that the same reasoning as for other strategies apply, and we can partially identify the corresponding optimal revenues in a very similar way. By quantifying the revenue gain from K to $K + 1$, we can characterize how much can be marginally gained from being able to update its information on one additional purchase and optimally adjust its pricing.

Figure 3.1: Expected revenues under intermediate pricing strategies



The optimal revenues under these intermediate pricing strategies are displayed in Figure 3.1. Under incomplete information, demand learning is rather quick, as we can see from the important concavity of the red line. With just $K = 3$, the firm can already achieve a revenue equal to the observed one, while by learning from the first 50 purchases, it can already achieve a revenue only 3% lower than that of the complete information. On the other hand, the blue line shows that the revenue gains under complete information is small. The incremental revenue from K to $K + 1$ is almost constant and barely reaches ~ 1 €. This latter result could be expected, given that the difference between uniform pricing and the full stopping-time pricing is small under complete information (1.8%). The comparison between the incremental gains of the two curves allows one to identify the pure effect of the learning on B_T , once the effect of learning on the demand process (for a given B_T) has been removed. In line with our previous results, it appears that the former effect largely dominates the latter.

Finally, as a robustness check, we conduct the same counterfactual exercises with a lognormal

specification on η_T instead of a gamma distribution. The results are overall similar, see Table C.2 in Appendix C.5.

3.6 Conclusion

In this paper, we quantify the gains and losses of revenue management and identify their underlying sources in the context of French railway transportation. We first clarify that many counterfactual revenues only depend on price elasticity and total demand, and not on the precise timing of consumers' arrival. This is convenient when, as here, no details on the dates of purchases are available. Still, demand estimation is complicated in such a context by the absence of exogenous variations in prices, and censorship. Such problems are likely to appear in many cases where perishable goods are sold. To tackle such issues, we develop an original identification strategy combining exogenous variations in relative prices and moment inequalities built on basic rationality on consumer's side and weak optimality conditions on the firm's pricing strategy.

Even though such conditions only yield partial identification in theory, they turn out to be quite informative. Our results show that the actual revenue management enhances revenue with respect to the optimal uniform pricing strategy, but that it incurs a loss between 17.1 and 19.0% compared to the optimal dynamic pricing under incomplete information. Dynamic pricing appears to be in particular effective under demand uncertainty, as information cumulates quickly.

Our work has some limitations, that are in part common to the literature and in part related to the nature of our data. First, we assume that price elasticity is constant. This assumption seems reasonable in our context but may be less in others. Observing consumers' purchasing dates could help alleviate this restriction, and price changes resulting from the closure of a fare class could be used to identify demand in a flexible way. Second, while we take into account demand uncertainty in our counterfactuals, we still assume that the firm knows the pattern of consumers' arrival. If consumers' purchasing dates were observed, it could be possible to relax this assumption. Ex ante uncertainty in this dimension could however substantially complicate the computation of counterfactuals.

Appendices

Appendix to Chapter 1

A.1 Hendel (1999) and Dubé (2004) as Special Cases of Model (2.3)

In this Appendix, we illustrate that the model of preference for variety originally proposed by Hendel (1999) in the context of demand for computers and then applied by Dubé (2004) in the context of demand for soft drinks is a special case of model (2.3). In particular, Hendel (1999)'s model is a version of model (2.3) in which each demand synergy parameter $\Gamma_{it\mathbf{b}}$ is restricted to be negative in a special way. Hendel (1999)'s model is about individuals who go shopping less often than they consume. During any purchase occasion, individuals may buy several units of different products in anticipation of the various consumption occasions they will face before the next shopping trip. Suppose there are J different products and denote by \mathbf{J} their collection. Denote by 0 the outside option, the choice of consuming none of the J products. Denote by $R_i \in \mathbb{N}$ the maximal number of units of any product that individual i can consume during any consumption occasion, and by K_i the number of consumption occasions in between any two shopping trips. On any consumption occasion, Hendel (1999) assumes that different products are perfect substitutes, so that each individual will effectively choose a certain number of units of at most one product j . As a consequence, the actual choice set faced by individual i on any consumption occasion can be defined as:

$$\mathbf{A}_i = \left\{ \underbrace{(j, \dots, j)}_q : \text{for } j \in \mathbf{J}, q = 1, \dots, R_i \right\} \cup \{0\},$$

where q is the number of units of any product j that could be consumed on this consumption occasion and 0 is the outside option. Then, individual i 's choice set during any purchase occasion is:

$$\mathbf{C}_i = \underbrace{\mathbf{A}_i \times \dots \times \mathbf{A}_i}_{K_i},$$

where each element of \mathbf{C}_i is a bundle of size up to $R_i \times K_i$. To ease exposition, we represent each bundle $\mathbf{b} \in \mathbf{C}_i$ by $\mathbf{b} = (j_k, q_k)_{k=1}^{K_i}$, where (j_k, q_k) refers to the chosen product and to the corresponding number of units on consumption occasion k . Denote by $(j_k, q_k) = (0, 0)$ the decision of not consuming anything on consumption occasion k .

For the rest of this Appendix, we focus on Dubé (2004)'s notation, which specializes Hendel (1999)'s model to the case of demand for bundles in grocery shopping. Following Dubé (2004)'s equation (2) at page 68, denote by $(\Psi_{ij_k k} q_k)^\alpha S_i$ the indirect utility of individual i from choosing (j_k, q_k) on consumption occasion k : $\Psi_{ij_k k}$ is i 's perceived quality for product j_k on consumption occasion k , S_i

is an i -specific scaling factor, and $\alpha \in (0, 1)$ captures the curvature of the utility function.¹ Moreover, denote by p_{j_k} the price of one unit of product j_k and by y_i the income of individual i . Then, from [Dubé \(2004\)](#)'s equation (6) at page 69, the indirect utility of individual i from purchasing bundle $\mathbf{b} = ((j_1, q_1), \dots, (j_{K_i}, q_{K_i})) \in \mathbf{C}_i$ is:

$$\begin{aligned}
U_{i\mathbf{b}} &= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i - \sum_{k=1}^{K_i} p_{j_k} q_k + y_i \\
&= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i + \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} p_{j_k} q_k + y_i \\
&= \sum_{k=1}^{K_i} (\Psi_{ij_k k} q_k)^\alpha S_i + \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} (\Psi_{ij_k k})^\alpha S_i - \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i q_k - \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} p_{j_k} + y_i \quad (\text{A.1}) \\
&= \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} [(\Psi_{ij_k k})^\alpha S_i - p_{j_k}] + \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i [q_k^\alpha - q_k] + y_i \\
&= \sum_{k=1}^{K_i} \sum_{q=1}^{q_k} u_{ij_k k} + \Gamma_{i\mathbf{b}} + y_i,
\end{aligned}$$

where $u_{ij_k k} = (\Psi_{ij_k k})^\alpha S_i - p_{j_k}$ and $\Gamma_{i\mathbf{b}} = \sum_{k=1}^{K_i} (\Psi_{ij_k k})^\alpha S_i [q_k^\alpha - q_k]$. The sum over q_k on the right hand side of (A.1) is zero when $q_k = 0$. Note that [Dubé \(2004\)](#) assumes $\Psi_{ij_k k} \geq 0$. As a consequence, the demand synergy $\Gamma_{i\mathbf{b}}$ will be constrained to be strictly negative as long as $\Psi_{ij_k k} > 0$. [Dubé \(2004\)](#)'s demand model is therefore a special case of model (2.3) with non-positive demand synergies and without the i.i.d. Gumbel error terms.

A.2 Proof of Lemma 1

To prove the first statement, we show that given a distribution function for β_{it} , $F(\cdot; \Sigma'_F)$, there exists a unique $\delta'_t \in \mathbb{R}^{\mathbf{C}_{t1}}$ for $t = 1, \dots, T$ that solves $s_t(\delta'_t; \Sigma'_F) = \mathcal{J}_t$. This is equivalent to showing that given $F(\cdot; \Sigma'_F)$, the market share function $s_t(\cdot; \Sigma'_F)$ is invertible for $t = 1, \dots, T$. Because our arguments with regard to the first statement do not depend on whether F is parametric or non-parametric, hereafter we denote $F(\cdot; \Sigma'_F)$ simply by F .

Given a distribution F , for market $t = 1, \dots, T$, define the Jacobian matrix of the market share function $s_t(\cdot; F)$ from (2.3) by:

$$\mathbb{J}_t(\delta'_t; F) = \frac{\partial s_t}{\partial \delta'_t}(\delta'_t; F) = \left(\frac{\partial s_{t\mathbf{b}}}{\partial \delta'_{t\mathbf{b}'}}(\delta'_t; F) \right)_{\mathbf{b}, \mathbf{b}' \in \mathbf{C}_{t1}}. \quad (\text{A.2})$$

Corollary 2 from [Berry et al. \(2013\)](#) provides sufficient conditions for the invertibility of differentiable market share functions. We now verify that market share function (2.3) satisfies the two sufficient conditions of Corollary 2 from [Berry et al. \(2013\)](#): (a) weak substitutes (Assumption 2 in [Berry](#)

¹Note that [Dubé \(2004\)](#)'s equation (2) at page 68 reports the direct utility function defined over the entire vector $(q_{jk})_{j=1}^J$ of possible units for each product $j \in \mathbf{J}$ on consumption occasion k . However, because of the assumption of perfect substitutes mentioned earlier, positive units $q_{jk} > 0$ will be chosen for at most one product j on any consumption occasion k . For this reason, here we simplify the discussion and immediately consider the indirect utility of choosing (j_k, q_k) with $q_{j_k k} = q_k$.

et al. (2013)) and (b) non-singularity of the Jacobian matrix $\mathbb{J}_t(\delta'_t; F)$. We first compute $\mathbb{J}_t(\delta'_t; F)$ for $\mathbf{b}, \mathbf{b}' \in \mathbf{C}_{t1}, \mathbf{b} \neq \mathbf{b}'$:

$$\begin{aligned}\frac{\partial s_{t\mathbf{b}}}{\partial \delta'_{t\mathbf{b}}}(\delta'_t; F) &= \int s_{t\mathbf{b}}(\delta'_t; \beta_{it})(1 - s_{t\mathbf{b}}(\delta'_t; \beta_{it}))dF(\beta_{it}) \\ \frac{\partial s_{t\mathbf{b}}}{\partial \delta'_{t\mathbf{b}'}}(\delta'_t; F) &= - \int s_{t\mathbf{b}}(\delta'_t; \beta_{it})s_{t\mathbf{b}'}(\delta'_t; \beta_{it})dF(\beta_{it}).\end{aligned}\tag{A.3}$$

As discussed by Berry et al. (2013), the weak substitutes condition does not rule out complementarity in a discrete choice model in which alternatives are defined as bundles, as in demand model (2.3). In practice, the weak substitutes condition requires that for all $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_{t1}$, $s_{t\mathbf{b}}(\delta'_t; F)$ be weakly decreasing in $\delta'_{t\mathbf{b}'}$ for any $\mathbf{b}' \neq \mathbf{b}, \mathbf{b}' \in \mathbf{C}_{t1}$. This is immediate from the second equation in (A.3). In what follows, we verify that $\mathbb{J}_t(\delta'_t; F)$ is non-singular.

Define the $C_{t1} \times 1$ vector $s_{it}(\delta'_t; \beta_{it}) = (s_{it\mathbf{b}}(\delta'_t; \beta_{it}))_{\mathbf{b} \in \mathbf{C}_{t1}}$. By using (A.3), we can re-write $\mathbb{J}_t(\delta'_t; F)$ as:

$$\mathbb{J}_t(\delta'_t; F) = \int [\text{Diag}(s_{it}(\delta'_t; \beta_{it})) - s_{it}(\delta'_t; \beta_{it})s_{it}(\delta'_t; \beta_{it})^T] dF(\beta_{it}),\tag{A.4}$$

where $\text{Diag}(s_{it}(\delta'_t; \beta_{it}))$ is a diagonal matrix with the elements of $s_{it}(\delta'_t; \beta_{it})$ on the main diagonal. We first show that the symmetric matrix $\text{Diag}(s_{it}(\delta'_t; \beta_{it})) - s_{it}(\delta'_t; \beta_{it})s_{it}(\delta'_t; \beta_{it})^T$ is positive-definite. This is equivalent to showing that its eigenvalues are all positive. Note that every element of $s_{it}(\delta'_t; \beta_{it})$ is strictly positive and that their sum is strictly less than one:

$$\begin{aligned}s_{it\mathbf{b}}(\delta'_t; \beta_{it}) &> 0, \\ \sum_{\mathbf{b} \in \mathbf{C}_{t1}} s_{it\mathbf{b}}(\delta'_t; \beta_{it}) &< 1.\end{aligned}$$

Denote any of the eigenvalues of $\text{Diag}(s_{it}(\delta'_t; \beta_{it})) - s_{it}(\delta'_t; \beta_{it})s_{it}(\delta'_t; \beta_{it})^T$ by λ and its corresponding (non-degenerate) eigenvector by \mathbf{x} . Without loss of generality, suppose that the maximal element of vector \mathbf{x} in absolute value is its first element $x_1 \neq 0$:

$$|x_1| \geq |x_{\mathbf{b}}| \text{ for any } \mathbf{b} \in \mathbf{C}_{t1}.$$

Then, we have:

$$\begin{aligned}
& [\text{Diag}(s_{it}(\delta'_t; \beta_{it})) - s_{it}(\delta'_t; \beta_{it})s_{it}(\delta'_t; \beta_{it})^\top] \boldsymbol{x} = \lambda \boldsymbol{x} \\
\implies & s_{it\mathbf{b}}(\delta'_t; \beta_{it})x_{\mathbf{b}} - s_{it\mathbf{b}}(\delta'_t; \beta_{it}) \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta'_t; \beta_{it})x_{\mathbf{b}'} = \lambda x_{\mathbf{b}}, \text{ for all } \mathbf{b} \in \mathbf{C}_{t1} \\
\implies & s_{it1}(\delta'_t; \beta_{it})x_1 - s_{it1}(\delta'_t; \beta_{it}) \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta'_t; \beta_{it})x_{\mathbf{b}'} = \lambda x_1 \\
\implies & \lambda = s_{it1}(\delta'_t; \beta_{it}) \left(1 - \frac{\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta'_t; \beta_{it})x_{\mathbf{b}'}}{x_1} \right) \\
& \geq s_{it1}(\delta'_t; \beta_{it}) \left(1 - \left| \frac{\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta'_t; \beta_{it})x_{\mathbf{b}'}}{x_1} \right| \right) \\
& \geq s_{it1}(\delta'_t; \beta_{it}) \left(1 - \frac{\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta'_t; \beta_{it})|x_{\mathbf{b}'}}{|x_1|} \right) \\
& \geq s_{it1}(\delta'_t; \beta_{it}) \left(1 - \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'}(\delta'_t; \beta_{it}) \right) \\
& > 0.
\end{aligned}$$

Any eigenvalue of $\text{Diag}(s_{it}(\delta'_t; \beta_{it})) - s_{it}(\delta'_t; \beta_{it})s_{it}(\delta'_t; \beta_{it})^\top$ is thus strictly positive: for any $v \in \mathbb{R}^{\mathbf{C}_{t1}}$,

$$v^\top [\text{Diag}(s_{it}(\delta'_t; \beta_{it})) - s_{it}(\delta'_t; \beta_{it})s_{it}(\delta'_t; \beta_{it})^\top] v > 0.$$

As a consequence,

$$\begin{aligned}
v^\top \mathbb{J}_t(\delta'_t; F)v &= \int v^\top [\text{Diag}(s_{it}(\delta'_t; \beta_{it})) - s_{it}(\delta'_t; \beta_{it})s_{it}(\delta'_t; \beta_{it})^\top] v dF(\beta_{it}) \\
&> 0.
\end{aligned}$$

Thus, given F , for any $\delta'_t \in \mathbb{R}^{\mathbf{C}_{t1}}$, $\mathbb{J}_t(\delta'_t; F)$ is positive-definite and non-singular. Because both conditions (a) and (b) of Corollary 2 by [Berry et al. \(2013\)](#) are satisfied, then the market share function $s_t(\delta'_t; F)$ is invertible with respect to δ'_t , for $t = 1, \dots, T$. This completes the proof of the first statement.

We now prove the second statement of the Lemma. According to Assumption 1, the density function $\frac{dF(\beta_{it}; \Sigma'_F)}{d\beta_{it}}$ is continuously differentiable with respect to Σ'_F . As a consequence, $s_t(\delta'_t; \Sigma'_F) - \mathcal{J}'_t$ is continuously differentiable with respect to $(\delta'_t, \mathcal{J}'_t, \Sigma'_F)$. As we showed above, the Jacobian matrix $\frac{\partial [s_t(\delta'_t; \Sigma'_F) - \mathcal{J}'_t]}{\partial \delta'_t} \Big|_{(\delta'_t, \mathcal{J}'_t, \Sigma'_F) = (\delta_t, \mathcal{J}_t, \Sigma_F)} = \mathbb{J}_t(\delta_t; F(\cdot; \Sigma_F))$ is invertible. Then, according to the Implicit Function Theorem, in a neighbourhood of $(\delta_t, \mathcal{J}_t, \Sigma_F)$, for any $(\mathcal{J}'_t, \Sigma'_F)$ there exists a unique δ'_t such that $s_t(\delta'_t; \Sigma'_F) = \mathcal{J}'_t$ and $s_t^{-1}(\mathcal{J}'_t; \Sigma'_F) = \delta'_t$ is continuously differentiable with respect to $(\mathcal{J}'_t, \Sigma'_F)$. This completes the proof of the second statement.

A.3 Proof of Rank Regularity Property

Without loss of generality, suppose that $\Theta_\Sigma \subset \Upsilon$ is a compact set, where $\Upsilon \subset \mathbb{R}^{P+D}$ is a topological space of \mathbb{R}^{P+D} . Moreover, according to Assumption 1, $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is continuous with respect to $\Sigma' \in \Upsilon$. According to Property 4 from [Lewis \(2009\)](#), the set of rank regular points for $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is open and dense in Υ . This completes the proof.

A.4 Preliminaries for Theorem 1

Here we report a preliminary Lemma useful to prove Theorem 1.

Lemma 2. *If Assumptions 1 and 2 hold, and the Jacobian matrix $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma'=\Sigma}$ is of full column rank, then Σ is locally uniquely determined by moment conditions (1.9).*

Proof. The differentiability of moment conditions (1.9) with respect to Σ' follows from the second statement of Lemma 1 and the differentiability of $g(\Sigma_g)$ with respect to Σ_g in Assumption 2. It then suffices to show that the true Σ is the unique local solution to $m(\Sigma') = 0$. From the definition of model (2.3), $m(\Sigma) = 0$. We prove the result by contradiction.

Suppose that Σ is not the unique local solution to $m(\Sigma') = 0$. As a consequence, there exists a sequence of Σ_N such that $\Sigma_N \rightarrow \Sigma$ as $N \rightarrow \infty$, and $m(\Sigma_N) = 0$. Because $m(\Sigma')$ is continuously differentiable in a neighbourhood of $\Sigma' = \Sigma$, by applying the first-order Taylor expansion, we have:

$$\begin{aligned} m(\Sigma_N) &= m(\Sigma) + \frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma'=\Sigma} (\Sigma_N - \Sigma) + o(|\Sigma_N - \Sigma|), \\ \frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma'=\Sigma} \frac{\Sigma_N - \Sigma}{|\Sigma_N - \Sigma|} &= -\frac{o(|\Sigma_N - \Sigma|)}{|\Sigma_N - \Sigma|}, \end{aligned} \quad (\text{A.5})$$

where $o(|\Sigma_N - \Sigma|)$ is such that $\lim_{N \rightarrow \infty} \frac{o(|\Sigma_N - \Sigma|)}{|\Sigma_N - \Sigma|} = 0$. Note that $\frac{\Sigma_N - \Sigma}{|\Sigma_N - \Sigma|}$ belongs to the unit sphere in \mathbb{R}^{P+D} , which is compact. Then, there exists a subsequence $\left\{ \frac{\Sigma_{N_\ell} - \Sigma}{|\Sigma_{N_\ell} - \Sigma|} \right\}$ and $v \in \mathbb{R}^{P+D}$ with $|v| = 1$, such that $\frac{\Sigma_{N_\ell} - \Sigma}{|\Sigma_{N_\ell} - \Sigma|} \rightarrow v$. By applying the second equation of (A.5) to the subsequence $\left\{ \frac{\Sigma_{N_\ell} - \Sigma}{|\Sigma_{N_\ell} - \Sigma|} \right\}$, and by combining $\Sigma_{N_\ell} \rightarrow \Sigma$ and the continuous differentiability of $m(\cdot)$ in a neighbourhood of Σ , we obtain $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma'=\Sigma} v = 0$. Because $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma'=\Sigma}$ is of full column rank, any vector $x \in \mathbb{R}^{P+D}$ that satisfies $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma'=\Sigma} x = 0$ must be zero. Then $v = 0$, which contradicts the fact that $|v| = 1$. As a consequence, Σ is the unique local solution to $m(\Sigma') = 0$. \square

A.5 Proof of Theorem 1

Sufficiency. We prove sufficiency by contradiction. Suppose that model (2.3) is not locally identified: there exists a sequence of solutions to system (1.4), $(\delta_{1\mathbf{J}_1}^N, \dots, \delta_{T\mathbf{J}_T}^N, \Gamma^N, \Sigma^N) \neq (\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ for any N , such that $(\delta_{1\mathbf{J}_1}^N, \dots, \delta_{T\mathbf{J}_T}^N, \Gamma^N, \Sigma^N) \rightarrow (\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ as $N \rightarrow \infty$. Applying (1.6) and (1.7) to each element of the sequence, one obtains:

$$\begin{aligned} \delta_{t\mathbf{b}}^N(\Gamma_{\mathbf{b}}^N) + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_F^N) &= s_{t\mathbf{b}}^{-1}(\mathcal{J}_t; \Sigma_F^N), \\ \delta_{tj}^N &= s_{tj}^{-1}(\mathcal{J}_t; \Sigma_F^N), \quad j \in \mathbf{b}, \\ \Gamma_{\mathbf{b}}^N &= s_{t\mathbf{b}}^{-1}(\mathcal{J}_t; \Sigma_F^N) - \sum_{j \in \mathbf{b}} s_{tj}^{-1}(\mathcal{J}_t; \Sigma_F^N) - g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g^N). \end{aligned} \quad (\text{A.6})$$

Then, by constructing moment conditions (1.9) for each element of the sequence, we have $m(\Sigma') \Big|_{\Sigma'=\Sigma^N} = 0$ for any N . Because the Jacobian matrix $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma'=\Sigma}$ is of full column rank, according to Lemma 2, then Σ is uniquely locally determined by moment conditions (1.9). Hence, there exists N_0 such that for all $N \geq N_0$, $\Sigma^N = \Sigma$. Because of the third equation of (A.6), then for all $N \geq N_0$, $\Gamma_{\mathbf{b}}^N = \Gamma_{\mathbf{b}}$. Moreover, because of the first two equations of (A.6), we have $\delta_{t\mathbf{b}}^N = \delta_{t\mathbf{b}}$, for all $N \geq N_0$, $t = 1, \dots, T$

and $\mathbf{b} \in \mathbf{C}_{t1}$. As a consequence, $(\delta_{1\mathbf{J}_1}^N, \dots, \delta_{T\mathbf{J}_T}^N, \Gamma^N, \Sigma^N) = (\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ for all $N \geq N_0$, which contradicts $(\delta_{1\mathbf{J}_1}^N, \dots, \delta_{T\mathbf{J}_T}^N, \Gamma^N, \Sigma^N) \neq (\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ for any N .

Necessity. To simplify notation, denote the number of moment conditions $\sum_{\mathbf{b} \in \mathbf{C}_2, |\mathbf{T}_b| \geq 2} (|\mathbf{T}_b| - 1)$ by Q and the rank of $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma' = \Sigma}$ by r . According to the definition of rank regularity in footnote 11, there exists a neighbourhood of the true Σ , U , such that $\text{rank}\left(\frac{\partial m(\Sigma')}{\partial \Sigma'}\right) = \text{rank}\left(\frac{\partial m(\Sigma')}{\partial \Sigma'}\right) \Big|_{\Sigma' = \Sigma} = r$ for each $\Sigma' \in U$. By applying the Constant Rank Theorem at $\Sigma' = \Sigma$, there are open sets $U_1, U_2 \subset \mathbb{R}^{P+D}$ and $U_3 \subset \mathbb{R}^Q$ and diffeomorphisms $\phi : U_1 \rightarrow U_2$, $\psi : U_3 \rightarrow U_3$ such that $\Sigma \in U_1 \subset U$ and $\psi \circ m \circ \phi^{-1}(x') = (x'_1, \dots, x'_r, 0, \dots, 0)$ for all $x' \in U_2$.²

Define $x = (x_1, \dots, x_{P+D}) = \phi(\Sigma)$ and a sequence $\{x^N = (x_1^N, \dots, x_{P+D}^N)\}$ such that $x_\ell^N = x_\ell$, for $\ell = 1, \dots, r$ and $x_\ell^N = x_\ell + \frac{1}{N}$, for N large enough so that $x^N \neq x$ and $x^N \in U_2$. Note that

$$\begin{aligned} \psi \circ m \circ \phi^{-1}(x) &= (x_1, \dots, x_r, 0, \dots, 0) \\ &= (x_1^N, \dots, x_r^N, 0, \dots, 0) \\ &= \psi \circ m \circ \phi^{-1}(x^N) \end{aligned} \tag{A.7}$$

and that

$$\begin{aligned} \psi \circ m \circ \phi^{-1}(x) &= \psi \circ m \circ \phi^{-1}(\phi(\Sigma)) \\ &= \psi \circ m(\Sigma) \\ &= \psi(0). \end{aligned} \tag{A.8}$$

As a consequence, $\psi \circ m \circ \phi^{-1}(x^N) = \psi(0)$. Because ψ is a diffeomorphism, we obtain $m(\phi^{-1}(x^N)) = 0$. Because ϕ and its inverse ϕ^{-1} are diffeomorphisms and $x \neq x^N \rightarrow x = \phi(\Sigma)$ as $N \rightarrow \infty$, we construct a sequence $\Sigma^N = \phi^{-1}(x^N) \rightarrow \phi^{-1}(x) = \Sigma$ with $\Sigma^N \neq \Sigma$ such that $m(\Sigma^N) = 0$ for each N . According to (A.6) from the proof of sufficiency, given Σ^N , we can construct a $(\delta_{1\mathbf{J}_1}^N, \dots, \delta_{T\mathbf{J}_T}^N, \Gamma^N, \Sigma^N)$ such that it is a solution to (1.4). Consequently, model (2.3) is not locally identified and this concludes the proof.

A.6 Proof of Corollary 1

Because $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma' = \Sigma}$ is of full row rank, then the positive definite matrix $\left[\frac{\partial m(\Sigma')}{\partial \Sigma'} \right] \left[\frac{\partial m(\Sigma')}{\partial \Sigma'} \right]^T \Big|_{\Sigma' = \Sigma}$ is not singular and its determinant $\text{Det} \left(\left[\frac{\partial m(\Sigma')}{\partial \Sigma'} \right] \left[\frac{\partial m(\Sigma')}{\partial \Sigma'} \right]^T \Big|_{\Sigma' = \Sigma} \right)$ is positive. Moreover, since $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is continuous with respect to Σ' , $\text{Det} \left(\left[\frac{\partial m(\Sigma')}{\partial \Sigma'} \right] \left[\frac{\partial m(\Sigma')}{\partial \Sigma'} \right]^T \right)$ is also continuous with respect to Σ' and is positive in a neighbourhood of $\Sigma' = \Sigma$. This implies that $\frac{\partial m(\Sigma')}{\partial \Sigma'}$ is of full row rank in a neighbourhood of $\Sigma' = \Sigma$, and its rank, $\text{rank} \left(\frac{\partial m(\Sigma')}{\partial \Sigma'} \right)$, is constant and equal to the number of rows in $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma' = \Sigma}$. Consequently, Σ is rank regular for $\frac{\partial m(\Sigma')}{\partial \Sigma'}$. Note that the number of rows in $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma' = \Sigma}$ is equal to the number of moment conditions $\sum_{\mathbf{b} \in \mathbf{C}_2, |\mathbf{T}_b| \geq 2} (|\mathbf{T}_b| - 1)$ and it is strictly smaller than the dimension of Σ . The latter is equal to the number of columns in $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma' = \Sigma}$. Then, $\frac{\partial m(\Sigma')}{\partial \Sigma'} \Big|_{\Sigma' = \Sigma}$ is not of full column rank. According to Theorem 1, model (2.3) is not locally identified and this concludes the proof.

²For the details of the Constant Rank Theorem, see Theorem 7.1 by Boothby (1986).

A.7 Testing Procedures for Assumption 2 and Assumption 3

In this section, we develop testing procedures for Assumptions 2 and 3 on the basis of partial identification methods. For a given subset of markets $\mathbf{T}_0 \subset \mathbf{T}$, the identification set of $\theta = ((\delta_{t\mathbf{J}_t})_{t \in \mathbf{T}_0}, \Gamma, \Sigma_F, \Sigma_g)$ is defined by the moment equalities:

$$s_{t\mathbf{b}}(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F) = s_{t\mathbf{b}}, \quad (\text{A.9})$$

for $t \in \mathbf{T}_0$ and $\mathbf{b} \in \mathbf{C}_{t1}$, where $g_t(\Sigma_g) = (g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g))_{\mathbf{b} \in \mathbf{C}_{t2}}$ and $\delta_t(\Gamma + g_t(\Sigma_g)) = (\delta_{t1}, \dots, \delta_{tJ_t}, (\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}} + g(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma_g)))_{\mathbf{b} \in \mathbf{C}_{t2}})$. We denote by $\Theta(\mathbf{T}_0)$ the identification set of θ defined by (A.9) and by $Q((s_t)_{t \in \mathbf{T}_0}, \theta')$ the following criterion function:

$$Q((s_t)_{t \in \mathbf{T}_0}, \theta') = \sum_{t \in \mathbf{T}_0} (s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - s_t)^T \Omega_t^{-1} (s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - s_t). \quad (\text{A.10})$$

Note that $Q((s_t)_{t \in \mathbf{T}_0}, \theta') = 0$ if and only if $\theta' \in \Theta(\mathbf{T}_0)$. Denote by $I_{t\mathbf{b}}$ the number of individuals in market t observed to choose \mathbf{b} and by $\hat{s}_{t\mathbf{b}} = \frac{I_{t\mathbf{b}}}{I}$ the corresponding observed market share. As I increases to infinity, $\hat{s}_{t\mathbf{b}} \xrightarrow{p} s_{t\mathbf{b}}$ and $\sqrt{I}(\hat{s}_t - s_t) \xrightarrow{p} \mathcal{N}(0, \Omega_t)$ for $t = 1, \dots, T$, $\mathbf{b} \in \mathbf{C}_t$, where $\Omega_t = (\omega_{t\mathbf{b}\mathbf{b}'}))_{\mathbf{b}, \mathbf{b}' \in \mathbf{C}_{t1}}$ with $\omega_{t\mathbf{b}\mathbf{b}'} = s_{t\mathbf{b}}(1 - s_{t\mathbf{b}})$ when $\mathbf{b} = \mathbf{b}'$ and $\omega_{t\mathbf{b}\mathbf{b}'} = -s_{t\mathbf{b}}s_{t\mathbf{b}'}$ otherwise. Denote by $\hat{\Omega}_t$ an estimator of Ω_t that satisfies $\hat{\Omega}_t \xrightarrow{p} \Omega_t$ and $\sqrt{I}(\hat{\Omega}_t - \Omega_t) = O_p(1)$.³ We then define the sample counterpart of criterion function $Q(\cdot)$ as:

$$Q_I((\hat{s}_t)_{t \in \mathbf{T}_0}, \theta') = \sum_{t \in \mathbf{T}_0} (s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - \hat{s}_t)^T \hat{\Omega}_t^{-1} (s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - \hat{s}_t). \quad (\text{A.11})$$

Testing Procedure for Assumption 2. In this section, we maintain $\mathbf{T}_0 = \mathbf{T}$. Note that Assumption 2 holds if and only if $\Theta(\mathbf{T}) \neq \emptyset$, i.e., there is at least a profile of θ that satisfies moment equalities (A.9). Hence, we propose a specification test on the basis of the following hypotheses:

$$H_0 : \Theta(\mathbf{T}) \neq \emptyset \text{ versus } H_1 : \Theta(\mathbf{T}) = \emptyset. \quad (\text{A.12})$$

Denote by $q_{\sum_{t \in \mathbf{T}} C_{t1}}^{1-\alpha}$ the $1 - \alpha$ quantile of $\chi^2(\sum_{t \in \mathbf{T}} C_{t1})$ and define the following random set:

$$\Theta_I(\mathbf{T}) = \{\theta' \in \Theta : I \cdot Q_I((\hat{s}_t)_{t \in \mathbf{T}}, \theta') \leq q_{\sum_{t \in \mathbf{T}} C_{t1}}^{1-\alpha}\}. \quad (\text{A.13})$$

If $\Theta_I(\mathbf{T}) = \emptyset$, then we reject H_0 from (A.12).

Proposition 3. Under H_0 from (A.12), $\limsup_{I \rightarrow \infty} \sup_{\theta' \in \Theta(\mathbf{T})} \Pr[\Theta_I(\mathbf{T}) = \emptyset] \leq \alpha$.

Proof. Under H_0 from (A.12), for any $\theta' \in \Theta(\mathbf{T})$, we have:

$$\begin{aligned} \Pr[\Theta_I(\mathbf{T}) = \emptyset] &\leq \Pr[\theta' \notin \Theta_I(\mathbf{T})] \\ &= \Pr\left[I \cdot Q_I((\hat{s}_t)_{t \in \mathbf{T}}, \theta') > q_{\sum_{t \in \mathbf{T}} C_{t1}}^{1-\alpha}\right] \\ &= \Pr\left[\sum_{t \in \mathbf{T}} \left[\sqrt{I}(\hat{s}_t - s_t)\right]^T \hat{\Omega}_t^{-1} \left[\sqrt{I}(\hat{s}_t - s_t)\right] > q_{\sum_{t \in \mathbf{T}} C_{t1}}^{1-\alpha}\right]. \end{aligned} \quad (\text{A.14})$$

³Such an estimator can be $\hat{\Omega}_t = (\hat{\omega}_{t\mathbf{b}\mathbf{b}'}))_{\mathbf{b}, \mathbf{b}' \in \mathbf{C}_{t1}}$, where $\hat{\omega}_{t\mathbf{b}\mathbf{b}'} = \hat{s}_{t\mathbf{b}}(1 - \hat{s}_{t\mathbf{b}})$ when $\mathbf{b} = \mathbf{b}'$ and $\hat{\omega}_{t\mathbf{b}\mathbf{b}'} = -\hat{s}_{t\mathbf{b}}\hat{s}_{t\mathbf{b}'}$ otherwise.

Since $\hat{\Omega}_t \xrightarrow{p} \Omega_t$ (and hence $\hat{\Omega}_t^{-1} \xrightarrow{p} \Omega_t^{-1}$) and $\sqrt{I}(\hat{\beta}_t - \beta_t) \xrightarrow{p} \mathcal{N}(0, \Omega_t)$, for $t \in \mathbf{T}$, we obtain:

$$\sum_{t \in \mathbf{T}} \left[\sqrt{I}(\beta_t - \hat{\beta}_t) \right]^T \hat{\Omega}_t^{-1} \left[\sqrt{I}(\beta_t - \hat{\beta}_t) \right] \xrightarrow{d} \chi^2 \left(\sum_{t \in \mathbf{T}} C_{t1} \right).$$

Note that the probability on the right-hand side of (A.14) converges to α and does not depend on θ' . Then,

$$\limsup_{I \rightarrow \infty} \sup_{\theta' \in \Theta(\mathbf{T})} \Pr [\Theta_I(\mathbf{T}) = \emptyset] \leq \alpha.$$

The proof is completed. \square

Testing Procedure for Assumption 3. In this section, we assume that $\Theta(\mathbf{T}) \neq \emptyset$ and therefore that $\Theta(\mathbf{T}_0) \neq \emptyset$ for any $\mathbf{T}_0 \subset \mathbf{T}$. Given \mathbf{T}_0 , we first derive a consistent estimator for $\Theta(\mathbf{T}_0)$:

Lemma 3. *Define a sequence of random sets:*

$$\mathcal{E}(a_I) = \{\theta' \in \Theta : I \cdot Q_I((\hat{\beta}_t)_{t \in \mathbf{T}_0}, \theta') \leq a_I\},$$

where $a_I \geq 0$ satisfies $a_I \rightarrow \infty$ and $\frac{a_I}{I} \rightarrow 0$. Then,

$$\lim_{I \rightarrow \infty} \inf_{\theta' \in \Theta(\mathbf{T}_0)} \Pr [\Theta(\mathbf{T}_0) \subset \mathcal{E}(a_I)] = 1$$

and

$$\lim_{I \rightarrow \infty} \sup_{\theta' \in \Theta(\mathbf{T}_0)} d_H(\Theta(\mathbf{T}_0), \mathcal{E}(a_I)) = 0,$$

where $d_H(\cdot, \cdot)$ is the Hausdorff metric: $d_H(A, B) = \sup_{a \in A} \inf_{b \in B} |a - b| + \sup_{b \in B} \inf_{a \in A} |b - a|$.

Proof. See Theorem 3.1 by Chernozhukov et al. (2007). \square

The choice of a_I is up to the econometrician. For example, one can choose $a_I = \ln I$ (see Chernozhukov et al. (2007) for a detailed discussion). In what follows, we focus on situations in which $\Theta(\mathbf{T}_0)$ contains only interior points of Θ . While Assumption 3 is abstract and not easy to test directly, we propose the following Condition and show that it implies Assumption 3:

Condition 1. *There exists $\mathbf{T}_0 \subsetneq \mathbf{T}$ such that $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}$ is of full column rank when evaluated at any of the solutions to $m(\Sigma'; \mathbf{T}_0) = 0$.*

Remark 8. *Denote by $\Theta_\Sigma(\mathbf{T}_0)$ the set of solutions to $m(\Sigma'; \mathbf{T}_0) = 0$. Since the true parameters $\Sigma_0 \in \Theta_\Sigma(\mathbf{T}_0)$, Condition 1 implies that $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}$ is of full column rank when evaluated at $\Sigma' = \Sigma_0$. As a consequence, model (2.3) is locally identified according to Theorem 1.*

Remark 9. *As shown in section 2.3, $\theta' \in \Theta(\mathbf{T}_0)$ holds if and only if $m(\Sigma'; \mathbf{T}_0) = 0$. Then, $\Theta_\Sigma(\mathbf{T}_0)$ is the projection of $\Theta(\mathbf{T}_0)$ along the dimensions of Σ . Moreover, because of Lemma 3, the projection of $\mathcal{E}(a_n)$ along the dimensions of Σ also defines a consistent estimator for $\Theta_\Sigma(\mathbf{T}_0)$ which covers asymptotically $\Theta_\Sigma(\mathbf{T}_0)$ with probability 1 and that we denote by $\mathcal{E}_\Sigma(a_n)$.*

The next Proposition shows that Condition 1 is sufficient for Assumption 3:

Proposition 4. *If Condition 1 holds, then Assumption 3 holds.*

Proof. We prove this by contradiction. Denote the solution set of $m(\Sigma'; \mathbf{T}_0) = 0$ in Θ_Σ by S . Suppose that S contains infinitely many elements. Because S is a closed subset of the compact set Θ_Σ , S is itself compact. Consequently, because S has infinitely many elements, then there exists an accumulation point $\Sigma'_0 \in S$: in any neighbourhood of Σ'_0 , we can find another $\Sigma''_0 \in S$, i.e. another solution to $m(\Sigma'; \mathbf{T}_0) = 0$. Due to Assumption 3, we know that at $\Sigma'_0 \in S$, the corresponding Jacobian matrix $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}|_{\Sigma'=\Sigma'_0}$ is of full column rank. Then, locally, $\Sigma' = \Sigma'_0$ must be the unique solution to $m(\Sigma'; \mathbf{T}_0) = 0$. This contradicts Σ'_0 being an accumulation point in S . \square

We then propose a test for Assumption 3 on the basis of the following hypotheses:

$$H_0 : \text{Condition 1 does not hold. versus } H_1 : \text{Condition 1 holds.} \quad (\text{A.15})$$

H_0 from (A.15) is equivalent to the hypothesis that there exists some $\theta' \in \Theta(\mathbf{T}_0)$ such that $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}$ is rank deficient when evaluated at $\Sigma' = (\Sigma'_F, \Sigma'_g)$. Define the following function:

Assumption 9. Suppose that $J((s'_t)_{t \in \mathbf{T}_0}, \Sigma') : \times_{t \in \mathbf{T}_0} \mathcal{S}_t \times \Theta_\Sigma \rightarrow \mathbb{R}_+ \cup \{0\}$ satisfies:

- $J \geq 0$.
- $J = 0$ if and only if $\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'}$ is rank deficient.

where $\mathcal{S}_t = \{s'_t \in \mathbb{R}^{C_{t1}} : s'_{t\mathbf{b}} > 0 \text{ and } \sum_{\mathbf{b} \in C_{t1}} s'_{t\mathbf{b}} < 1, \mathbf{b} \in C_{t1}\}$.

Example 1. The determinant function

$$J((s'_t)_{t \in \mathbf{T}_0}, \Sigma') = \text{Det} \left(\left[\left(\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'} \right)^T \left(\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'} \right) \right] \right)$$

Example 2. The minimal eigenvalue function

$$J((s'_t)_{t \in \mathbf{T}_0}, \Sigma') = \inf_{|\lambda|=1} \lambda^T \left[\left(\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'} \right)^T \left(\frac{\partial m(\Sigma'; \mathbf{T}_0)}{\partial \Sigma'} \right) \right] \lambda,$$

where λ is unit vector of dimension $P + D$.

Define the criterion function $J^*((s'_t)_{t \in \mathbf{T}_0}) = \inf_{\Sigma' \in \Theta_\Sigma(\mathbf{T}_0)} J((s'_t)_{t \in \mathbf{T}_0}, \Sigma')$. Note that $J^*((s_t)_{t \in \mathbf{T}_0}) = 0$ if and only if H_0 from (A.15) holds. We then propose the following test statistic:

$$J_I^*((\hat{s}_t)_{t \in \mathbf{T}_0}) = \inf_{\Sigma' \in \mathcal{C}_\Sigma(a_n)} J((\hat{s}_t)_{t \in \mathbf{T}_0}, \Sigma') \quad (\text{A.16})$$

and the next two Theorems establish its properties.

Theorem 12. Suppose Assumptions 1, 2 and 9 hold. Moreover, $\sqrt{I}(\hat{s}_t - s_t) \xrightarrow{d} \mathcal{N}(0, \Omega_t)$ for $t \in \mathbf{T}$.

- If J is continuous in $\times_{t \in \mathbf{T}_0} \mathcal{S}_t \times \Theta_\Sigma$, then $J_I^*((\hat{s}_t)_{t \in \mathbf{T}_0}) \xrightarrow{P} J^*((s_t)_{t \in \mathbf{T}_0})$, uniformly for $\theta \in \Theta(\mathbf{T}_0)$.
- If J is Lipschitz continuous in $\times_{t \in \mathbf{T}_0} \mathcal{S}_t \times \Theta_\Sigma$, then under H_0 from (A.15), $\sqrt{I} \cdot J_I^*((\hat{s}_t)_{t \in \mathbf{T}_0}) = O_p(1)$.

Proof. For any $\theta \in \Theta(\mathbf{T})$, on the stochastic event $\{\Theta(\mathbf{T}_0) \subset \mathcal{C}(a_I)\}$, we can write:

$$\begin{aligned} J^*((\mathcal{J}_t)_{t \in \mathbf{T}_0}) &= J((\mathcal{J}_t)_{t \in \mathbf{T}_0}, \Sigma^*), \\ J_I^*((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}) &= J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \Sigma_I^*), \end{aligned} \quad (\text{A.17})$$

where $\Sigma^* \in \arg \min_{\Sigma' \in \Theta_\Sigma(\mathbf{T}_0)} J((\mathcal{J}_t)_{t \in \mathbf{T}_0}, \Sigma')$ and $\Sigma_I^* \in \arg \min_{\Sigma' \in \mathcal{C}_\Sigma(a_I)} J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \Sigma')$. Note that $\Sigma^* \in \Theta(\mathbf{T}_0) \subset \mathcal{C}_\Sigma(a_I)$. Then, we have:

$$J_I^*((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}) - J^*((\mathcal{J}_t)_{t \in \mathbf{T}_0}) \leq J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \Sigma^*) - J((\mathcal{J}_t)_{t \in \mathbf{T}_0}, \Sigma^*). \quad (\text{A.18})$$

By construction, $\Theta_\Sigma(\mathbf{T}_0)$ is a compact set. Then, there exists $\tilde{\Sigma}_I^* \in \Theta_\Sigma(\mathbf{T}_0)$ such that $d(\Sigma_I^*, \Theta_\Sigma(\mathbf{T}_0)) = d(\Sigma_I^*, \tilde{\Sigma}_I^*)$. Hence, we obtain:

$$\begin{aligned} J_I^*((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}) - J^*((\mathcal{J}_t)_{t \in \mathbf{T}_0}) &= \left[J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \Sigma_I^*) - J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \tilde{\Sigma}_I^*) \right] + \left[J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \tilde{\Sigma}_I^*) - J((\mathcal{J}_t)_{t \in \mathbf{T}_0}, \tilde{\Sigma}_I^*) \right] \\ &\quad + \left[J((\mathcal{J}_t)_{t \in \mathbf{T}_0}, \tilde{\Sigma}_I^*) - J((\mathcal{J}_t)_{t \in \mathbf{T}_0}, \Sigma^*) \right] \\ &\geq \left[J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \Sigma_I^*) - J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \tilde{\Sigma}_I^*) \right] + \left[J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \tilde{\Sigma}_I^*) - J((\mathcal{J}_t)_{t \in \mathbf{T}_0}, \tilde{\Sigma}_I^*) \right]. \end{aligned} \quad (\text{A.19})$$

According to Lemma 3, $d_H(\Theta(\mathbf{T}_0), \mathcal{C}(a_I)) \rightarrow 0$ uniformly for $\theta \in \Theta(\mathbf{T}_0)$. We then obtain $d(\Sigma_I^*, \tilde{\Sigma}_I^*) \rightarrow 0$ uniformly for $\theta \in \Theta(\mathbf{T}_0)$.

Suppose that J is continuous in $\times_{t \in \mathbf{T}_0} \mathcal{S}_t \times \Theta_\Sigma$. Then, in a compact set $\mathcal{S}^* \times \Theta$, where \mathcal{S}^* is a compact neighbourhood of $(\mathcal{J}_t)_{t \in \mathbf{T}_0}$, J is uniformly continuous. Together with $d(\Sigma_I^*, \tilde{\Sigma}_I^*) \rightarrow 0$ uniformly for $\theta \in \Theta(\mathbf{T}_0)$, we obtain that the right-hand side of (A.18) and that of (A.19) converge to 0 on $\{\Theta(\mathbf{T}_0) \subset \mathcal{C}(a_I)\}$, uniformly for $\theta \in \Theta(\mathbf{T}_0)$. Note that $\{\Theta(\mathbf{T}_0) \subset \mathcal{C}(a_I)\}$ holds asymptotically with probability 1, uniformly for $\theta \in \Theta(\mathbf{T}_0)$. This proves the first statement.

Suppose that J is Lipschitz continuous in $\times_{t \in \mathbf{T}_0} \mathcal{S}_t \times \Theta_\Sigma$. Under H_0 from (A.15), we have $J^*((\mathcal{J}_t)_{t \in \mathbf{T}_0}) = 0$. Then, by applying the Mean Value Theorem on the right-hand side of (A.18), we obtain that on $\{\Theta(\mathbf{T}_0) \subset \mathcal{C}(a_I)\}$:

$$\begin{aligned} 0 \leq J_I^*((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}) &= J_I^*((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}) - J^*((\mathcal{J}_t)_{t \in \mathbf{T}_0}) \\ &\leq J((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}, \Sigma^*) - J((\mathcal{J}_t)_{t \in \mathbf{T}_0}, \Sigma^*) \leq L|\hat{\mathcal{J}} - \mathcal{J}|, \end{aligned} \quad (\text{A.20})$$

where L is the Lipschitz constant of $J(\cdot)$. Then, by using $\sqrt{I}(\hat{\mathcal{J}} - \mathcal{J}) \xrightarrow{d} \mathcal{N}(0, \Omega_t)$, we obtain that:

$$0 \leq \sqrt{I} \cdot J_I^*((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0}) \leq L|\sqrt{I}(\hat{\mathcal{J}} - \mathcal{J})| = O_p(1) \quad (\text{A.21})$$

and the second statement is proved. \square

We now illustrate how to approximate the quantiles of $J_I^*((\hat{\mathcal{J}}_t)_{t \in \mathbf{T}_0})$ under H_0 from (A.15) by bootstrap methods building on Romano and Shaikh (2012). Denote by $\mathbf{P}_R^I = \{P_s^I : s \in \mathbb{R}^R, s_r > 0, \sum_{r=1}^R s_r = 1\}$ the set of multinomial distributions with R outcomes out of I trials. Define the distance ρ on \mathbf{P}_R^I as $\rho(P_s^I, P_{s'}^I) = \sum_{r=1}^R |s_r - s'_r|$ and $\mathbf{P}^I = \times_{t \in \mathbf{T}_0} \mathbf{P}_{C_{t1}}^I$ as the set of joint distributions $P = (P_{s_t}^I)_{t \in \mathbf{T}_0}$, where each $P_{s_t}^I$ is independently distributed across $t \in \mathbf{T}_0$. Note that ρ can be extended to any $P^I = (P_{s_t}^I)_{t \in \mathbf{T}_0}$ and $Q^I = (P_{s'_t}^I)_{t \in \mathbf{T}_0}$ in \mathbf{P}^I as: $\rho(P, Q) = \sum_{t \in \mathbf{T}_0} \rho(P_{s_t}^I, P_{s'_t}^I)$. For any I and any realization $\omega = (i_t)_{t \in \mathbf{T}_0}$ of P , we can define the non-negative random variable $J_I^*(\omega; P) = J_I^*((\frac{i_t}{I})_{t \in \mathbf{T}_0})$. Denote the distribution

function of $J_I^*(\omega; P)$ evaluated at $x \geq 0$ by $G_I(x, P)$.

Theorem 13. *Suppose the same Assumptions of Theorem 12 hold. If J is Lipschitz continuous in $\times_{t \in \mathbf{T}_0} \mathcal{S}_t \times \Theta_\Sigma$, then under H_0 from (A.15), for any α_1, α_2 such that $\alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_1 + \alpha_2 < 1$,*

$$\liminf_{I \rightarrow \infty} \inf_{\theta \in \Theta(\mathbf{T}_0)} \Pr \left[G_I^{-1}(\alpha_1, \hat{P}^I) < \sqrt{I} \cdot J_I^*((\hat{\jmath}_t)_{t \in \mathbf{T}_0}) \leq G_I^{-1}(1 - \alpha_2, \hat{P}^I) \right] \geq 1 - \alpha_1 - \alpha_2,$$

where $\hat{P}^I = (P_{\hat{\jmath}_t}^I)_{t \in \mathbf{T}_0}$.

Proof. Our proof builds on Theorem 2.4 of Romano and Shaikh (2012). It suffices to verify two conditions. First, we show that for any sequences Q^I and P^I in \mathbf{P}^I satisfying $\rho(Q^I, P^I) \rightarrow 0$, we have:

$$\lim_{I \rightarrow \infty} \sup_{x \geq 0} \{|G_I(x, Q^I) - G_I(x, P^I)|\} \rightarrow 0.$$

This can be seen from the construction of $J_I^*(\omega; P)$. For any $x \geq 0$, we have:

$$\begin{aligned} G_I(x, P) &= \Pr [J_I^*(\omega; P) \leq x] \\ &= \Pr_P \left[(i_t)_{t \in \mathbf{T}_0} : J_I^* \left(\left(\frac{i_t}{I} \right)_{t \in \mathbf{T}_0} \right) \leq x \right] \\ &= \Pr_P \left[(i_t)_{t \in \mathbf{T}_0} : \inf_{\Sigma' \in \mathcal{C}_\Sigma(a_I)} J \left(\left(\frac{i_t}{I} \right)_{t \in \mathbf{T}_0}, \Sigma' \right) \leq x \right] \\ &= \Pr_P \left[(i_t)_{t \in \mathbf{T}_0} : (i_t)_{t \in \mathbf{T}_0} \in \mathcal{N}(x, J(\cdot), a_I, I, \mathbf{T}_0) \right], \end{aligned} \tag{A.22}$$

where $\mathcal{N}(x, J(\cdot), a_I, I, \mathbf{T}_0)$ is the set of realizations for which $\inf_{\Sigma' \in \mathcal{C}_\Sigma(a_I)} J \left(\left(\frac{i_t}{I} \right)_{t \in \mathbf{T}_0}, \Sigma' \right) \leq x$ holds. Then, given $(x, J(\cdot), a_I, I, \mathbf{T}_0)$, we obtain:

$$\begin{aligned} |G_I(x, Q^I) - G_I(x, P^I)| &= |\Pr_{Q^I} [(i_t)_{t \in \mathbf{T}_0} : (i_t)_{t \in \mathbf{T}_0} \in \mathcal{N}(x, J(\cdot), a_I, I, \mathbf{T}_0)] \\ &\quad - \Pr_{P^I} [(i_t)_{t \in \mathbf{T}_0} : (i_t)_{t \in \mathbf{T}_0} \in \mathcal{N}(x, J(\cdot), a_I, I, \mathbf{T}_0)]| \\ &\leq \rho(Q^I, P^I). \end{aligned}$$

Because $\sup_{x \geq 0} \{|G_I(x, Q^I) - G_I(x, P^I)|\} \leq \rho(Q^I, P^I)$ and $\rho(Q^I, P^I) \rightarrow 0$, the first condition is verified. We now move on to verifying the second condition. For any sequence $P^I \in \{(P_{\hat{\jmath}_t}^I)_{t \in \mathbf{T}_0} : s_t(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F) = \hat{\jmath}_t, (\delta_{\mathbf{J}}, \Gamma, \Sigma_F, \Sigma_g) \in \Theta(\mathbf{T}_0)\}$, we have $\rho(\hat{P}^I, P^I) \xrightarrow{P} 0$. This condition holds because $\hat{\jmath}_t$ converges in probability to \jmath_t for any $t \in \mathbf{T}_0$. This completes the proof. \square

Finally, for $0 < \alpha < 1$, we propose the following rejection region for test (A.16):

$$\{\sqrt{I} \cdot J_I^*((\hat{\jmath}_t)_{t \in \mathbf{T}_0}) > G_I^{-1}(1 - \alpha, \hat{P}^I)\}.$$

According to Theorem 12, test (A.16) has asymptotically unit power, uniformly for $\theta \in \Theta(\mathbf{T}_0)$. Moreover, according to Theorem 13, the size of test (A.16) is controlled by α , uniformly for $\theta \in \Theta(\mathbf{T}_0)$.

A.8 Proof of Theorem 2

For this result, our arguments do not depend on whether the distribution of random coefficients is parametric or non-parametric and we then denote $F(\cdot; \Sigma_F)$ simply by F . Remember that

$$\begin{aligned} s_{t\mathbf{b}}(\delta_t; F) &= \int s_{it\mathbf{b}}(\delta_t; \beta_{it}) dF(\beta_{it}) \\ &= \int \frac{e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})}}{\sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})}} dF(\beta_{it}). \end{aligned}$$

To prove the real analytic property of the market share function $s_{t\mathbf{b}}(\delta_t; F)$, it suffices to study $\frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^l}$, where l is an integer and $\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = l$. We first prove the following Lemma.

Lemma 4. *For any non-negative integer l ,*

$$\sup_{\delta_t, \beta_{it}} \left| \frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\partial \delta_{t\mathbf{b}}^l} \right| \leq A_l l!,$$

where $A_l = (e-1)^l \sum_{k=0}^l \frac{1}{(e-1)^k k!}$.

Proof. Define $a_l = \sup_{\delta_t, \beta_{it}} \left| \frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\partial \delta_{t\mathbf{b}}^l} \right|$. Note that:

$$\begin{aligned} e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})} &= s_{it\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} \\ e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})} &= \frac{\partial^l e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})}}{\partial \delta_{t\mathbf{b}}^l} \\ &= \frac{\partial^l (s_{it\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})})}{\partial \delta_{t\mathbf{b}}^l} \\ &= \sum_{k=0}^l C_l^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k} \frac{\partial^{l-k} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})}}{\partial \delta_{t\mathbf{b}}^{l-k}} \\ &= \frac{\partial^l s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^l} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} + \sum_{k=0}^{l-1} C_l^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k} e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})}, \tag{A.23} \\ \frac{\partial^l s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^l} &= s_{it\mathbf{b}} \left(1 - \sum_{k=0}^{l-1} C_l^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k} \right), \\ \left| \frac{\partial^l s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^l} \right| &\leq 1 + \sum_{k=0}^{l-1} C_l^k \left| \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k} \right|, \\ a_l &\leq 1 + \sum_{k=0}^{l-1} C_l^k a_k, \\ \frac{a_l}{l!} &\leq \frac{1}{l!} + \sum_{k=0}^{l-1} \frac{a_k}{k!} \frac{1}{(l-k)!}. \end{aligned}$$

We now show that $\frac{a_l}{l!} \leq A_l$ by induction. For $l = 0$, the result holds trivially. For $l = 1$, we

have $a_1 = \sup_{\delta_t, \beta_{it}} \left| \frac{\partial s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\partial \delta_{t\mathbf{b}}} \right| = \sup_{\delta_t, \beta_{it}} |s_{it\mathbf{b}}(1 - s_{it\mathbf{b}})| \leq \frac{1}{4} < e = A_1$. Suppose that $\frac{a_k}{k!} \leq A_k$ holds for $k = 1, \dots, l-1$. Note that $A_l = \frac{1}{l!} + (e-1)A_{l-1} > A_{l-1}$, for any $l \geq 0$. Then,

$$\begin{aligned} \frac{a_l}{l!} &\leq \frac{1}{l!} + \sum_{k=0}^{l-1} \frac{a_k}{k!} \frac{1}{(l-k)!} \\ &\leq \frac{1}{l!} + \sum_{k=0}^{l-1} A_k \frac{1}{(l-k)!} \\ &\leq \frac{1}{l!} + A_{l-1} \sum_{k=0}^{l-1} \frac{1}{(l-k)!} \\ &\leq \frac{1}{l!} + A_{l-1}(e-1) \\ &= A_l. \end{aligned} \tag{A.24}$$

As a consequence, the inequality holds for any $l > 0$ and $a_l = \sup_{\delta_t, \beta_{it}} \left| \frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\partial \delta_{t\mathbf{b}}^l} \right| \leq A_l l!$. This completes the proof. \square

The next Lemma controls the size of $\frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^l}$.

Lemma 5. *Suppose $C_{t1} \geq 2$. For any $\mathbf{b} \in \mathbf{C}_{t1}$ and $l \geq 0$,*

$$\left| \frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^l} \right| \leq [C_{t1}(e-1)]^l \prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'!},$$

where $\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = l$.

Proof. We prove the result by induction. For $l = 0$, the result holds trivially. For $l = 1$, the result follows directly from Lemma 4 with $l = 1$. For $l = 2$ and $l_{\mathbf{b}'} = 2$, according to Lemma 4, we have $\left| \frac{\partial^2 s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}^2} \right| \leq A_2 2!$. For $l = 2$ and $l_{\mathbf{b}'} = l_{\mathbf{b}''} = 1$, $\mathbf{b}' \neq \mathbf{b}''$:

$$\begin{aligned} e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})} &= s_{it\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})}, \\ 0 &= \frac{\partial^2 s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'} \partial \delta_{t\mathbf{b}''}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} + e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}''}} + e^{\delta_{t\mathbf{b}''} + \mu_{it\mathbf{b}''}(\beta_{it})} \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}} \tag{A.25} \\ \frac{\partial^2 s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'} \partial \delta_{t\mathbf{b}''}} &= -s_{it\mathbf{b}'} \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}''}} - s_{it\mathbf{b}''} \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}} \end{aligned}$$

By using $\left| \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}} \right| \leq \frac{1}{4} < 1$ and $\left| \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}} \right| \leq 1$, we have $\left| \frac{\partial^2 s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'} \partial \delta_{t\mathbf{b}''}} \right| \leq \left| \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}} \right| + \left| \frac{\partial s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}'}} \right| \leq 2 \leq [C_{t1}(e-1)]^2$. Note that $A_2 = (e-1)^2 \left(1 + \frac{1}{e-1} + \frac{1}{2(e-1)^2}\right) \leq [C_{t1}(e-1)]^2$ for $C_{t1} \geq 2$. As a consequence, the conclusion holds for $l = 2$ and $\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = 2$.

Suppose that for $k = 0, \dots, l-1$ the inequality holds for any $\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = k$. First, remember that $A_l = (e-1)^l \sum_{k=0}^l \frac{1}{(e-1)^k k!}$, as defined in Lemma 4, is smaller than $[C_{t1}(e-1)]^l$ because $C_{t1} \geq 2$. Then,

the conclusion holds for any $l > 0$ with $l_{\mathbf{b}'} = l$ and $l_{\mathbf{b}''} = 0$, $\mathbf{b}'' \neq \mathbf{b}'$. It remains to show that the conclusion holds when there exist \mathbf{b}' and \mathbf{b}'' such that $l_{\mathbf{b}'} > 0$ and $l_{\mathbf{b}''} > 0$.

By taking $l_{\mathbf{b}}$ -th derivatives of both sides of the first equation in (A.23) with respect to $\delta_{t\mathbf{b}}$, we obtain:

$$\begin{aligned}
e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})} &= \frac{\partial^{l_{\mathbf{b}}} e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}}} \\
&= \frac{\partial^{l_{\mathbf{b}}} (s_{it\mathbf{b}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})})}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}}} \\
&= \sum_{k=0}^{l_{\mathbf{b}}} C_{l_{\mathbf{b}}}^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k} \frac{\partial^{l_{\mathbf{b}}-k} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}-k}} \\
&= \frac{\partial^{l_{\mathbf{b}}} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}}} \sum_{\mathbf{b}' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} + e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})} \sum_{k=0}^{l_{\mathbf{b}}-1} C_{l_{\mathbf{b}}}^k \frac{\partial^k s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k}.
\end{aligned} \tag{A.26}$$

Note that, by taking derivatives of both sides of equation (A.26) with respect to $\delta_{t\mathbf{b}'}$, $\mathbf{b}' \neq \mathbf{b}$, the left hand-side vanishes and we obtain:

$$0 = \frac{\partial^{l_{\mathbf{b}}+l_{\mathbf{b}'}} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \sum_{\mathbf{b}'' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}''} + \mu_{it\mathbf{b}''}(\beta_{it})} + e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l_{\mathbf{b}'}}^k \frac{\partial^{l_{\mathbf{b}}+k} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^{l_{\mathbf{b}}} \partial \delta_{t\mathbf{b}'}^k} + e^{\delta_{t\mathbf{b}} + \mu_{it\mathbf{b}}(\beta_{it})} \sum_{k=0}^{l_{\mathbf{b}}-1} C_{l_{\mathbf{b}}}^k \frac{\partial^{k+l_{\mathbf{b}'}} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}}. \tag{A.27}$$

By taking $l_{\mathbf{b}'}$ -th derivatives with respect to $\delta_{t\mathbf{b}'}$, for all $\mathbf{b}' \in \mathbf{C}_{t1}$:

$$\begin{aligned}
0 &= \frac{\partial^{l_{\mathbf{b}'}} s_{it\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \sum_{\mathbf{b}'' \in \mathbf{C}_t} e^{\delta_{t\mathbf{b}''} + \mu_{it\mathbf{b}''}(\beta_{it})} + \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} e^{\delta_{t\mathbf{b}'} + \mu_{it\mathbf{b}'}(\beta_{it})} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l_{\mathbf{b}'}}^k \frac{\partial^{l_{\mathbf{b}}-l_{\mathbf{b}'}+k} s_{it\mathbf{b}}}{\partial \delta_{t\mathbf{b}}^k \prod_{\mathbf{b}'' \neq \mathbf{b}'} \partial \delta_{t\mathbf{b}''}^{l_{\mathbf{b}''}}}, \\
\frac{\partial^{l_{\mathbf{b}'}} s_{it\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} &= - \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'} \sum_{k=0}^{l_{\mathbf{b}'}-1} C_{l_{\mathbf{b}'}}^k \frac{\partial^{l_{\mathbf{b}}-l_{\mathbf{b}'}+k} s_{it\mathbf{b}'}}{\partial \delta_{t\mathbf{b}}^k \prod_{\mathbf{b}'' \neq \mathbf{b}'} \partial \delta_{t\mathbf{b}''}^{l_{\mathbf{b}''}}}, \\
\frac{\partial^{l_{\mathbf{b}'}} s_{it\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} &= - \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} s_{it\mathbf{b}'} \sum_{k=0}^{l_{\mathbf{b}'}-1} \frac{1}{(l_{\mathbf{b}'} - k)!} \frac{\partial^{l_{\mathbf{b}}-l_{\mathbf{b}'}+k} s_{it\mathbf{b}'}}{\partial \delta_{t\mathbf{b}}^k \prod_{\mathbf{b}'' \neq \mathbf{b}'} \partial \delta_{t\mathbf{b}''}^{l_{\mathbf{b}''}}}, \\
\sup_{\delta_t, \beta_{it}} \left| \frac{\partial^{l_{\mathbf{b}'}} s_{it\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}^{l_{\mathbf{b}'}}} \right| &\leq \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} \sum_{k=0}^{l_{\mathbf{b}'}-1} \frac{1}{(l_{\mathbf{b}'} - k)!} \sup_{\delta_t, \beta_{it}} \left| \frac{\partial^{l_{\mathbf{b}}-l_{\mathbf{b}'}+k} s_{it\mathbf{b}'}}{\partial \delta_{t\mathbf{b}}^k \prod_{\mathbf{b}'' \neq \mathbf{b}'} \partial \delta_{t\mathbf{b}''}^{l_{\mathbf{b}''}}} \right|.
\end{aligned} \tag{A.28}$$

Then, applying the conclusion for any $k = 0, \dots, l - 1$ on the last equation in (A.28), we obtain:

$$\begin{aligned}
\sup_{\delta_t, \beta_{it}} \left| \frac{\frac{\partial^l s_{it\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'!}} \right| &\leq \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} \sum_{k=0}^{l_{\mathbf{b}'}-1} \frac{1}{(l_{\mathbf{b}'} - k)!} [C_{t1}(e-1)]^{l-l_{\mathbf{b}'}+k} \\
&= [C_{t1}(e-1)]^l \sum_{\mathbf{b}' \in \mathbf{C}_{t1}} \sum_{k=1}^{l_{\mathbf{b}'}} \frac{1}{k!} [C_{t1}(e-1)]^{-k} \\
&\leq [C_{t1}(e-1)]^l C_{t1} (e^{[C_{t1}(e-1)]^{-1}} - 1) \\
&\leq [C_{t1}(e-1)]^l [e^{(e-1)^{-1}} - 1] \\
&< [C_{t1}(e-1)]^l.
\end{aligned} \tag{A.29}$$

Hence, the conclusion holds for $\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = l$, and $\sup_{\delta_t, \beta_{it}} \left| \frac{\partial^l s_{it\mathbf{b}}}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}} \right| \leq [C_{t1}(e-1)]^l \prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'!}$ for any $l > 0$ and $\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'} = l$. The proof is completed. \square

The size of the l -th derivative of $s_{t\mathbf{b}}(\delta_t; F)$ with respect to δ_t can then be controlled as:

$$\begin{aligned}
\left| \frac{\partial^l s_{t\mathbf{b}}(\delta_t; F)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}} \right| &\leq \int \left| \frac{\partial^l s_{it\mathbf{b}}(\delta_t; \beta_{it})}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}} \right| dF(\beta_{it}) \\
&\leq [C_{t1}(e-1)]^l \prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'!}
\end{aligned} \tag{A.30}$$

and, consequently, the Taylor expansion of $s_{t\mathbf{b}}(\cdot; F)$ at some δ'_t around δ_t as:

$$\begin{aligned}
\left| \sum_{L=0}^{\infty} \frac{1}{L!} \left[\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} (\delta'_{t\mathbf{b}'} - \delta_{t\mathbf{b}'}) \frac{\partial}{\partial \delta_{t\mathbf{b}'}} \right]^L s_{t\mathbf{b}}(\delta_t; F) \right| &\leq \sum_{L=0}^{\infty} \frac{1}{L!} d^L \sum_{\sum l_{\mathbf{b}'}=L} \frac{L!}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} l_{\mathbf{b}'!}} \left| \frac{\partial^L s_{t\mathbf{b}}(\delta_t; F)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}} \right| \\
&\leq \sum_{L=0}^{\infty} d^L C_{t1}^L [C_{t1}(e-1)]^L,
\end{aligned} \tag{A.31}$$

where $d = |\delta'_t - \delta_t|$. Consequently, whenever $d < d^* = \frac{1}{C_{t1}^2(e-1)}$, the Taylor expansion (A.31) converges. Finally, by applying Taylor's Theorem to the multivariate function $s_{t\mathbf{b}}(\delta'_t; F)$, we obtain for any $R > 0$ and uniformly for $|\delta'_t - \delta_t| < \frac{d^*}{2}$:

$$\begin{aligned}
&\left| s_{t\mathbf{b}}(\delta'_t; F) - \sum_{L=0}^R \frac{1}{L!} \left[\sum_{\mathbf{b}' \in \mathbf{C}_{t1}} (\delta'_{t\mathbf{b}'} - \delta_{t\mathbf{b}'}) \frac{\partial}{\partial \delta_{t\mathbf{b}'}} \right]^L s_{t\mathbf{b}}(\delta_t; F) \right| \\
&\leq d^{R+1} \sum_{\sum l_{\mathbf{b}'}=R+1} \frac{1}{\prod l_{\mathbf{b}'!}} \sup_{|\delta'_t - \delta_t| < d} \left| \frac{\partial^{R+1} s_{t\mathbf{b}}(\delta'_t; F)}{\prod_{\mathbf{b}' \in \mathbf{C}_{t1}} \partial \delta_{t\mathbf{b}'}} \right| \\
&\leq d^{R+1} [C_{t1}(e-1)]^{R+1} C_{t1}^{R+1} \\
&\rightarrow 0.
\end{aligned}$$

In conclusion, the market share function $s_{\mathbf{t}\mathbf{b}}(\delta'_t; F)$ is equal to its Taylor expansion and therefore real analytic with respect to δ'_t . This completes the proof.

A.9 Proof of Theorem 3

The necessity part of the first statement is immediate. To prove sufficiency, note that when $(\delta_{t\mathbf{J}_t}^0)_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{\text{ID}} = \cap_{r=1}^R \Delta_r^{\text{ID}}$, for any $r = 1, \dots, R$ there exists some market $t \in \mathbf{T} \setminus \mathbf{T}_0$ such that $M_t s_t^{-1}(s_t; \Sigma_F^r) \neq \Gamma^r + g_t(\Sigma_g^r)$ and therefore $m(\Sigma^r; \mathbf{T}) \neq 0$ for $r = 1, \dots, R$. Remember that the set of solutions to $m(\Sigma'; \mathbf{T}_0) = 0$ in Θ_Σ is $S = \{\Sigma^r : r = 0, \dots, R\}$. Consequently, the set of solutions to $m(\Sigma'; \mathbf{T}) = 0$ is a subset of S . Given that $m(\Sigma^r; \mathbf{T}) \neq 0$ for $r = 1, \dots, R$, and that $m(\Sigma^0; \mathbf{T}) = 0$, $\Sigma' = \Sigma^0$ is the unique solution to system (1.9) in Θ_Σ . The remaining parameters of model (2.3) can then be uniquely pinned down by the demand inverse from Lemma 1 and model (2.3) is globally identified.

To prove the second statement, we first note that

$$\times_{t \in \mathbf{T} \setminus \mathbf{T}_0} \mathbb{R}^{J_t} \setminus \Delta^{\text{ID}} = \cup_{r=1}^R [\times_{t \in \mathbf{T} \setminus \mathbf{T}_0} \mathbb{R}^{J_t} \setminus \Delta_r^{\text{ID}}].$$

It is then sufficient to show that the Lebesgue measure of $\times_{t \in \mathbf{T} \setminus \mathbf{T}_0} \mathbb{R}^{J_t} \setminus \Delta_r^{\text{ID}}$ is zero. Note that

$$\begin{aligned} \times_{t \in \mathbf{T} \setminus \mathbf{T}_0} \mathbb{R}^{J_t} \setminus \Delta_r^{\text{ID}} &= \{(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} : \text{for any } t \in \mathbf{T} \setminus \mathbf{T}_0, M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) = \Gamma^r + g_t(\Sigma_g^r)\} \\ &= \times_{t \in \mathbf{T} \setminus \mathbf{T}_0} \{\delta_{t\mathbf{J}_t} : M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) = \Gamma^r + g_t(\Sigma_g^r)\} \\ &= \times_{t \in \mathbf{T} \setminus \mathbf{T}_0} Z_t^r, \end{aligned}$$

where Z_t^r is the zero set of function $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$. Because $\Delta_r^{\text{ID}} \neq \emptyset$, there exists some $t \in \mathbf{T} \setminus \mathbf{T}_0$ for which the zero set $Z_t^r \subsetneq \mathbb{R}^{J_t}$, i.e. $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is not equal to zero for some $\delta_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. It is then enough to show that, for this specific $Z_t^r \subsetneq \mathbb{R}^{J_t}$, the Lebesgue measure is zero.

For any Γ and Σ_F , because $s_t(\delta_t(\Gamma); \Sigma_F)$ is a composition of two real analytic functions, $\delta_t(\Gamma) : \mathbb{R}^{J_t} \rightarrow \mathbb{R}^{C_{t1}}$ and $s_t(\cdot; \Sigma_F) : \mathbb{R}^{C_{t1}} \rightarrow (0, 1)^{C_{t1}}$ (from Theorem 2), it is itself a real analytic function from \mathbb{R}^{J_t} to $(0, 1)^{C_{t1}}$. Moreover, because $s_t(\cdot; \Sigma_F^r)$ is real analytic with respect to $\delta_t \in \mathbb{R}^{C_{t1}}$, the inverse market share function from Lemma 1, $s_t^{-1}(\cdot; \Sigma_F^r)$, is also real analytic with respect to $s'_t \in (0, 1)^{C_{t1}}$. Then, the composition of $M_t s_t^{-1}(s'_t; \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ and $s'_t = s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0)$ is also real analytic with respect to $\delta_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. Consequently, Z_t^r is the zero set of the real analytic function $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$. There are two cases to be considered. When $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is a constant different from zero, $Z_t^r = \emptyset$ and it has zero Lebesgue measure. Similarly, when $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0)); \Sigma_F^0); \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is not a constant, according to Mityagin (2015), Z_t^r has also zero Lebesgue measure.⁴ This completes the proof.

⁴More generally, the zero set of a non-constant real analytic function defined on a P -dimensional domain can be written as the union of j -dimensional sub-manifolds, with j ranging from 0 to $P - 1$. As a consequence, the zero set has zero Lebesgue measure. For details, see the second statement of Theorem 6.3.3 (Lojasiewicz's Structure Theorem for Varieties) from Krantz and Parks (2002).

A.10 Price-Setting Models Consistent with Assumption 4

Here we show that Assumption 4 is consistent with commonly employed pure components pricing models with any profile of demand synergies (substitutability and/or complementarity).

To simplify notation, in this Appendix we drop the market index t . Denote by \mathbf{J}_f the collection of products owned by firm f and by \mathbf{J}_{-f} the set of products owned by the other firms, where $\mathbf{J} = \mathbf{J}_f \cup \mathbf{J}_{-f} = \{1, \dots, J\}$ is the collection of all products available in the market. Let c_j denote the constant marginal cost of product $j \in \mathbf{J}$, $p_f = (p_j)_{j \in \mathbf{J}_f}$ the vector of prices chosen by firm f for the products it owns, and $p_{-f} = (p_k)_{k \in \mathbf{J}_{-f}}$ the vector of prices chosen by the other firms. With pure components pricing, the price of a bundle \mathbf{b} is given by the sum of the prices of its components $p_{\mathbf{b}} = \sum_{j \in \mathbf{b}} p_j$, where each p_j is chosen by the firm that owns it. Then, the profit function of firm f takes the following form:

$$\pi_f(p_f, p_{-f}) = \sum_{j \in \mathbf{J}_f} s_j(p_{\mathbf{J}})(p_j - c_j), \quad (\text{A.32})$$

where $s_j(p_{\mathbf{J}}) = \sum_{\mathbf{b}: \mathbf{b} \ni j} s_{\mathbf{b}}(p_{\mathbf{J}})$ is the product-level market share function of product j and $p_{\mathbf{J}} = (p_1, \dots, p_J)$. Denote the ownership matrix $\Omega = (a_{jj'})_{j, j'=1, \dots, J}$ where $a_{jj'} = 1$ if j and j' are owned by the same firm and 0 otherwise. Under complete information, the necessary first-order conditions for a Bertrand-Nash equilibrium in pure components are:

$$\left[\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} \odot \Omega \right] (p_{\mathbf{J}} - c_{\mathbf{J}}) + s_{\mathbf{J}}(p_{\mathbf{J}}) = 0, \quad (\text{A.33})$$

where \odot denotes the Hadamard product, or element-by-element multiplication, $s_{\mathbf{J}} = (s_j(p_{\mathbf{J}}))_{j \in \mathbf{J}}$ is the vector of product-level market share functions, $p_{\mathbf{J}} = (p_j)_{j \in \mathbf{J}}$, and $c_{\mathbf{J}} = (c_j)_{j \in \mathbf{J}}$. Given different configurations of the ownership matrix, (A.33) specialize to different market structures such as monopoly, duopoly, or oligopoly.

The identifiability of $c_{\mathbf{J}}$ is determined by the invertibility of the matrix $\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} \odot \Omega$. As long as $\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} \odot \Omega$ is invertible, we obtain:

$$c_{\mathbf{J}} = p_{\mathbf{J}} + \left[\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} \odot \Omega \right]^{-1} s_{\mathbf{J}}(p_{\mathbf{J}}).$$

We now show that for any ownership matrix, $\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} \odot \Omega$ is invertible. Let $p = (p_{\mathbf{J}}, (p_{\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2})$ denote the vector of prices for all single products and bundles in the choice set. Moreover, we assume that p_j enters linearly in $u_{ij} = \delta_j + \mu_{ij}(\beta_i)$ with individual-specific coefficient $\alpha_i < 0$, which is part of the vector of random coefficients β_i . Then, by using the notation M_t^1 introduced prior to Theorem 3, we can write:

$$\begin{aligned} \frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} &= \int [\mathbf{I} \ M_t^{1\top}] \frac{\partial s_i(\beta_i)}{\partial p_{\mathbf{J}}} dF(\beta_i) \\ &= \int \alpha_i [\mathbf{I} \ -M_t^{1\top}] \frac{\partial s_i(\beta_i)}{\partial u_i} [\mathbf{I} \ -M_t^{1\top}]^{\top} dF(\beta_i), \end{aligned} \quad (\text{A.34})$$

where $u_i = (\delta_{\mathbf{b}} + \mu_{i\mathbf{b}}(\beta_i))_{\mathbf{b} \in \mathbf{C}_1}$. As shown in the proof of Lemma 1 (see Appendix A.2), $\frac{\partial s_i(\beta_i)}{\partial u_i}$ is positive-definite for any β_i . Moreover, $[\mathbf{I} \ -M_t^{1\top}]$ is of full row rank and therefore $[\mathbf{I} \ -M_t^{1\top}]^{\top}$ is of full column rank. Consequently, $[\mathbf{I} \ -M_t^{1\top}] \frac{\partial s_i(\beta_i)}{\partial u_i} [\mathbf{I} \ -M_t^{1\top}]^{\top}$ is positive-definite for any β_i . Because $\alpha_i < 0$, $\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}}$ is negative-definite. Note that Ω is a symmetric block diagonal matrix that contains only 1's and 0's. Then, $\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} \odot \Omega$ is also block diagonal. Because each block is a principal sub-matrix of $\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}}$, these

blocks are also negative-definite. Then, $\frac{\partial s_{\mathbf{J}}}{\partial p_{\mathbf{J}}} \odot \Omega$ is negative-definite and thus invertible.

A.11 Proof of Corollary 2

Take $\Sigma = \Sigma^r$ and $\Gamma = \Gamma^r$. Because $s_t(\delta'_t(\Gamma^r + g_t(\Sigma_g^r)); p'_{t\mathbf{J}_t}, \Sigma_F^r)$ is real analytic with respect to $(\delta'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t})$, then the inverse market share function, $s_t^{-1}(s'_t; p'_{t\mathbf{J}_t}, \Sigma_F^r)$, is real analytic with respect to $(s'_t, p'_{t\mathbf{J}_t})$. Consequently, $M_t s_t^{-1}(s'_t; p'_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is real analytic with respect to $(s'_t, p'_{t\mathbf{J}_t})$. Moreover, for $\Sigma = \Sigma^0$ and $\Gamma = \Gamma^0$, $s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p'_{t\mathbf{J}_t}, \Sigma_F^0)$ is real analytic with respect to $(\delta'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t})$. Then, the composition $M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is real analytic with respect to $(\delta'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t})$.

We now prove Corollary 2 by contradiction. Suppose that Assumption 5 does not hold. Then, for some $r = 1, \dots, R$ and $t \in \mathbf{T} \setminus \mathbf{T}_0$, there exists a set $\tilde{\mathbf{P}}_t \subset \mathbf{P}_t$ such that $\tilde{\mathbf{P}}_t$ has positive Lebesgue measure and

$$\Gamma^r + g_t(\Sigma_g^r) = M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r)$$

for any $\xi'_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. We then obtain that the zero set of the real analytic function $M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is at least $\tilde{\mathbf{P}}_t \times \mathbb{R}^{J_t}$. Because the Lebesgue measure of $\tilde{\mathbf{P}}_t$ is positive, then the Lebesgue measure of $\tilde{\mathbf{P}}_t \times \mathbb{R}^{J_t}$ is also positive. According to Mityagin (2015), $M_t s_t^{-1}(s_t(\delta'_t(\Gamma^0 + g_t(\Sigma_g^0)); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is then constant and equal to zero on $\mathbf{P}_t \times \mathbb{R}^{J_t}$. This contradicts Ξ_r^{ID} being non-empty. This completes the proof.

A.12 Proof of Theorem 4

Here we rely on the same notation for M_t as introduced prior to Theorem 3. M_t is a matrix of dimension $C_{t2} \times C_{t1}$. Remember that C_{t2} is the number of bundles and C_{t1} the number of inside options (bundles plus single products). M_t is made of two sub-matrices: $M_t = [M_t^1, M_t^2]$. M_t^1 is a matrix of -1 's and 0 's of dimension $C_{t2} \times J_t$, where the columns represent individual products and the rows represent bundles. Each row of M_t^1 identifies with -1 's the product composition of the corresponding bundle. M_t^2 is instead an identity matrix \mathbf{I} of dimension $C_{t2} \times C_{t2}$, with the rows corresponding to bundles.

The proof of the first statement is similar to that of Theorem 3. On the one hand, when $(\xi_{t\mathbf{J}_t}^0, c_{t\mathbf{J}_t}^0)_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Xi^{\text{ID}} = \bigcap_{r=1}^R \Xi_r^{\text{ID}}$, for any $r = 1, \dots, R$ there exists some market $t \in \mathbf{T} \setminus \mathbf{T}_0$ such that $M_t s_t^{-1}(s_t; p_{t\mathbf{J}_t}, \Sigma_F^r) \neq \Gamma^r + g_t(\Sigma_g^r)$ and therefore $m(\Sigma^r; \mathbf{T}) \neq 0$ for $r = 1, \dots, R$.

Remember that the set of solutions to $m(\Sigma'; \mathbf{T}_0) = 0$ in Θ_Σ is $S = \{\Sigma^r : r = 0, \dots, R\}$. Consequently, the set of solutions to $m(\Sigma'; \mathbf{T}) = 0$ is a subset of S . Given that $m(\Sigma^r; \mathbf{T}) \neq 0$ for $r = 1, \dots, R$, and that $m(\Sigma^0; \mathbf{T}) = 0$, $\Sigma' = \Sigma^0$ is the unique solution to system (1.9) in Θ_Σ . The remaining parameters of model (2.3) can then be uniquely pinned down by the demand inverse from Lemma 1 and model (2.3) is globally identified.

To prove the second statement, we first note that

$$\times_{t \in \mathbf{T} \setminus \mathbf{T}_0} [D_{t\xi} \times D_{tc}] \setminus \Xi^{\text{ID}} = \bigcup_{r=1}^R [\times_{t \in \mathbf{T} \setminus \mathbf{T}_0} [D_{t\xi} \times D_{tc}] \setminus \Xi_r^{\text{ID}}].$$

It is then sufficient to show that the Lebesgue measure of $\times_{t \in \mathbf{T} \setminus \mathbf{T}_0} [D_{t\xi} \times D_{tc}] \setminus \Xi_r^{\text{ID}}$ is zero. Note that

$$\begin{aligned} \times_{t \in \mathbf{T} \setminus \mathbf{T}_0} [D_{t\xi} \times D_{tc}] \setminus \Xi_r^{\text{ID}} &= \{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} : \\ &\quad \text{for any } t \in \mathbf{T} \setminus \mathbf{T}_0, M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) = \Gamma^r + g_t(\Sigma_g^r) \\ &\quad \text{for some } p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})\} \\ &= \times_{t \in \mathbf{T} \setminus \mathbf{T}_0} \{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) : \Gamma^r + g_t(\Sigma_g^r) \in M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_F^0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_F^r)\} \\ &= \times_{t \in \mathbf{T} \setminus \mathbf{T}_0} Z_t^{+r}, \end{aligned}$$

where Z_t^{+r} is the zero set of $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ such that $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r) = 0$ for some $p_{t\mathbf{J}_t} \in p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$. It then suffices to show that there exists a $t \in \mathbf{T} \setminus \mathbf{T}_0$ such that the Lebesgue measure of Z_t^{+r} is zero.

The Lebesgue measure of Z_t^{+r} in $D_{t\xi} \times D_{tc}$ is

$$\begin{aligned} me(Z_t^{+r}) &= \int_{D_{t\xi} \times D_{tc}} \mathbf{1}\{Z_t^{+r}\} d(c_{t\mathbf{J}_t}, \xi_{t\mathbf{J}_t}) \\ &= \int_{D_{t\xi} \times D_{tc}} \mathbf{1}\{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) : \Gamma^r + g_t(\Sigma_g^r) \in M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_F^0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_F^r)\} d(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \end{aligned}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Define $\Phi : (\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \rightarrow (\xi_{t\mathbf{J}_t}, \phi(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}))$. According to Assumption 4, Φ is a C^1 mapping from $(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \in \{(\xi'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t}) : \xi'_{t\mathbf{J}_t} \in D_{t\xi}, p'_{t\mathbf{J}_t} \in \mathbf{P}_t(\xi_{t\mathbf{J}_t})\}$ to $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \in D_{t\xi} \times D_{tc}$ and onto. Let $\text{Card}[\Phi^{-1}](\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ denote the cardinality of the inverse image of Φ at $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$. Note that $\text{Card}[\Phi^{-1}](\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ is equal to the number of Nash equilibria of the pricing game at $(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t})$ and therefore belongs to $\mathbb{N}_+ \cup \{\infty\}$ according to Assumption 4. Then, by Theorem 1.16-2 of Ciarlet (2013) and Fubini's Theorem, we obtain:

$$\begin{aligned} me(Z_t^{+r}) &\leq \int_{D_{t\xi} \times D_{tc}} \mathbf{1}\{(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) : \Gamma^r + g_t(\Sigma_g^r) \in M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_F^0); p_{t\mathbf{J}_t}(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}), \Sigma_F^r)\} \text{Card}[\Phi^{-1}](\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) d(\xi_{t\mathbf{J}_t}, c_{t\mathbf{J}_t}) \\ &= \int_{\{(\xi'_{t\mathbf{J}_t}, p'_{t\mathbf{J}_t}) : \xi'_{t\mathbf{J}_t} \in D_{t\xi}, p'_{t\mathbf{J}_t} \in \mathbf{P}_t(\xi_{t\mathbf{J}_t})\}} \mathbf{1}\{(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) : \Gamma^r + g_t(\Sigma_g^r) = M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r)\} \left| \frac{\partial \Phi_{t\mathbf{J}_t}}{\partial (\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t})}(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \right| d(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \\ &= \int_{\mathbf{P}_t} \left[\int_{\mathbf{P}_t(\xi_{t\mathbf{J}_t}) \ni p_{t\mathbf{J}_t}} \mathbf{1}\{\xi_{t\mathbf{J}_t} : \Gamma^r + g_t(\Sigma_g^r) = M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r)\} \left| \frac{\partial \phi_{t\mathbf{J}_t}}{\partial p_{t\mathbf{J}_t}}(\xi_{t\mathbf{J}_t}, p_{t\mathbf{J}_t}) \right| d\xi_{t\mathbf{J}_t} \right] dp_{t\mathbf{J}_t}. \end{aligned}$$

Suppose that Assumption 5 holds. Denote by $\Delta_{t\mathbf{J}_t} = (\Delta_{tj}(x_{tj}, p_{tj}))_{j \in \mathbf{J}_t}$. Because $\delta_{t\mathbf{J}_t}(\Delta_{t\mathbf{J}_t}, \xi_{t\mathbf{J}_t}) = \Delta_{t\mathbf{J}_t} + \xi_{t\mathbf{J}_t}$, given $p_{t\mathbf{J}_t}$ (and therefore $\Delta_{t\mathbf{J}_t}$) and by applying Theorem 2, we obtain that the market share function $s_t(\delta_t(\Gamma^r + g_t(\Sigma_g^r))); p_{t\mathbf{J}_t}, \Sigma_F^r)$ is also real analytic with respect to $\xi_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$. Then, given $p_{t\mathbf{J}_t}$, by the Inverse Function Theorem for real analytic functions, $s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r)$ is real analytic with respect to $\xi_{t\mathbf{J}_t}$, and therefore $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is real analytic with respect to $\xi_{t\mathbf{J}_t}$. For each $r = 1, \dots, R$, we focus on the market $t \in \mathbf{T} \setminus \mathbf{T}_0$ that satisfies Assumption 5: for any $p_{t\mathbf{J}_t} \in \mathbf{P}_t$, there exists $\xi_{t\mathbf{J}_t} \in D_{t\xi}$ such that $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) \neq \Gamma^r + g_t(\Sigma_g^r)$, i.e., $M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) - \Gamma^r - g_t(\Sigma_g^r)$ is not always equal to zero in $D_{t\xi}$. Similar to the proof of the second statement of Theorem 3, $\{\xi_{t\mathbf{J}_t} : M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) = \Gamma^r + g_t(\Sigma_g^r)\}$ has thus zero Lebesgue measure in $D_{t\xi}$ and

$$\mathbf{1}\{\xi_{t\mathbf{J}_t} : M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) = \Gamma^r + g_t(\Sigma_g^r)\} \left| \frac{\partial \phi_{t\mathbf{J}_t}}{\partial p_{t\mathbf{J}_t}}(p_{t\mathbf{J}_t}; \xi_{t\mathbf{J}_t}) \right| = 0 \text{ almost everywhere.}$$

It then follows that

$$\int_{\mathbf{P}_t(\xi_{t\mathbf{J}_t}) \ni p_{t\mathbf{J}_t}} \mathbf{1}\{\xi_{t\mathbf{J}_t} : M_t s_t^{-1}(s_t(\delta_t(\Gamma^0 + g_t(\Sigma_g^0))); p_{t\mathbf{J}_t}, \Sigma_F^0); p_{t\mathbf{J}_t}, \Sigma_F^r) = \Gamma^r + g_t(\Sigma_g^r)\} \left| \frac{\partial \phi_{t\mathbf{J}_t}}{\partial p_{t\mathbf{J}_t}}(p_{t\mathbf{J}_t}; \xi_{t\mathbf{J}_t}, \Gamma^0, \Sigma^0) \right| d\xi_{t\mathbf{J}_t} = 0,$$

and finally $me(Z_t^{+r}) \leq 0$. Consequently, $me(Z_t^{+r}) = 0$. This completes the proof.

A.13 Proof of Theorem 7

We first introduce some notation. Denote the collection of demand synergies that can rationalize the observed product-level market shares $s_{t\mathbf{J}_t}$ in market t by $\bar{\Theta}_\Gamma^t(\Sigma') = \{\Gamma' : \exists \delta'_{t\mathbf{J}_t} \in \mathbb{R}^{J_t}$ such that $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = s_{t\mathbf{J}_t}\}$ and across all T markets by $\bar{\Theta}_\Gamma = \cup_{\Sigma' \in \Theta_\Sigma} \cap_{t=1}^T \bar{\Theta}_\Gamma^t(\Sigma')$. Define also $\bar{\Theta}_\Sigma = \{\Sigma' : \cap_{t=1}^T \bar{\Theta}_\Gamma^t(\Sigma') \neq \emptyset\}$. $\bar{\Theta}_\Gamma^t$ and $\bar{\Theta}_\Sigma$ collect the values of Γ' and of Σ' that can rationalize the observed product-level market shares. When $\Sigma' \notin \bar{\Theta}_\Sigma$ or $\Gamma' \notin \cap_{t=1}^T \bar{\Theta}_\Gamma^t(\Sigma')$, then there exists no $\delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = s_{t\mathbf{J}_t}$ for any $t = 1, \dots, T$ (i.e., the demand inverse is not defined at (Γ', Σ')). The remainder of the proof focuses on the case of $\Sigma' \in \bar{\Theta}_\Sigma$ and $\Gamma' \in \cap_{t=1}^T \bar{\Theta}_\Gamma^t(\Sigma')$ (i.e., the demand inverse is defined at (Γ', Σ')), and in particular on showing the uniqueness of the corresponding $\delta'_{t\mathbf{J}_t}$.

We rely on the same notation for M_t as introduced prior to Theorem 3 and in the proof of Theorem 4. Note that M_t is of full row rank and therefore M_t^\top is of full column rank. Without loss of generality, we prove Theorem 7 for market t .

Denote by $\mathbf{1}$ a vector of 1's and define $\mathbf{S}_{t2}(s_{t\mathbf{J}_t}) = \{s'_{t\mathbf{C}_{t2}} : s'_{t\mathbf{C}_{t2}} = (s'_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t2}}, s'_{t\mathbf{b}} > 0, -M_t^{1\top} s'_{t\mathbf{C}_{t2}} < s_{t\mathbf{J}_t}, (M_t^{1\top} s'_{t\mathbf{C}_{t2}} + s_{t\mathbf{J}_t}, s'_{t\mathbf{C}_{t2}})^\top \mathbf{1} < 1\}$, as the collection of admissible vectors of market shares of bundles consistent with the observed product-level market shares, $s_{t\mathbf{J}_t}$. Given any $s'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(s_{t\mathbf{J}_t})$ and observed product-level market shares $s_{t\mathbf{J}_t}$, we can construct an admissible vector of market shares $s'_t = ((s'_{tj})_{j \in \mathbf{J}_t}, s'_{t\mathbf{C}_{t2}})$, where $s'_{tj} = s_{tj} - \sum_{\mathbf{b} \in \mathbf{C}_{t2}: j \in \mathbf{b}} s'_{t\mathbf{b}}$. Because of Lemma 1, given Σ' we can invert s'_t and obtain the corresponding $\delta'_t \in \mathbb{R}^{C_{t1}}$:

$$\begin{aligned} \delta'_t &= ((\delta'_{tj})_{j \in \mathbf{J}_t}, (\delta'_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t2}})^\top \\ &= s_t^{-1}(s'_t; \Sigma'_F) \\ &= s_t^{-1}((s_{tj} - \sum_{\mathbf{b} \in \mathbf{C}_{t2}: j \in \mathbf{b}} s'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, s'_{t\mathbf{C}_{t2}}; \Sigma'_F), \text{ where} \end{aligned} \tag{A.35}$$

$$\begin{aligned} \delta'_{tj} &= s_t^{-1}((s_{tj} - \sum_{\mathbf{b} \in \mathbf{C}_{t2}: j \in \mathbf{b}} s'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, s'_{t\mathbf{C}_{t2}}; \Sigma'_F), \\ \delta'_{t\mathbf{b}} &= s_t^{-1}((s_{tj} - \sum_{\mathbf{b} \in \mathbf{C}_{t2}: j \in \mathbf{b}} s'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, s'_{t\mathbf{C}_{t2}}; \Sigma'_F). \end{aligned}$$

Using the matrix M_t , we can recover an admissible Γ'_t from δ'_t by:

$$\begin{aligned} \Gamma'_t + g_t(\Sigma'_g) &= M_t \delta'_t, \\ \Gamma'_t &= M_t \delta'_t - g_t(\Sigma'_g) \\ &= M_t s_t^{-1}((s_{tj} - \sum_{\mathbf{b} \in \mathbf{C}_{t2}: j \in \mathbf{b}} s'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, s'_{t\mathbf{C}_{t2}}; \Sigma'_F) - g_t(\Sigma'_g) \\ &= M_t s_t^{-1}(M_t^\top s'_{t\mathbf{C}_{t2}} + (s_{t\mathbf{J}_t}^\top, 0, \dots, 0)^\top; \Sigma'_F) - g_t(\Sigma'_g). \end{aligned} \tag{A.36}$$

Consequently, for any $s'_{t\mathbf{C}_{t2}}$ there exists a $\Gamma'_t = \Gamma_t(s'_{t\mathbf{C}_{t2}}; s_{t\mathbf{J}_t}, \Sigma')$ such that (A.35) holds. We now

compute from (A.36) the derivative of $\Gamma'_t = \Gamma_t(\mathcal{J}'_{t\mathbf{C}_{t2}}; \mathcal{J}_{t\mathbf{J}_t}, \Sigma')$ with respect to $\mathcal{J}'_{t\mathbf{C}_{t2}}$:

$$\begin{aligned} \frac{d\Gamma_t}{d\mathcal{J}'_{t\mathbf{C}_{t2}}} &= M_t \frac{\partial s_t^{-1}}{\partial \mathcal{J}'_t} (M_t^\top \mathcal{J}'_{t\mathbf{C}_{t2}} + (\mathcal{J}_{t\mathbf{J}_t}^\top, 0, \dots, 0)^\top; \Sigma'_F) M_t^\top \\ &= M_t \left[\frac{\partial s_t}{\partial \delta'_t}(\delta'_t; \Sigma'_F) \right]^{-1} M_t^\top. \end{aligned} \quad (\text{A.37})$$

Because $\frac{\partial s_t}{\partial \delta'_t}(\delta'_t; \Sigma'_F)$ is positive-definite and M_t^\top is of full column rank, $\frac{d\Gamma_t}{d\mathcal{J}'_{t\mathbf{C}_{t2}}}$ is also positive-definite and therefore positive quasi-definite for any $\mathcal{J}'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(\mathcal{J}_{t\mathbf{J}_t})$.⁵ Note that $\mathbf{S}_{t2}(\mathcal{J}_{t\mathbf{J}_t})$ is convex. According to Theorem 6 by Gale and Nikaido (1965), p. 88, $\Gamma'_t = \Gamma_t(\mathcal{J}'_{t\mathbf{C}_{t2}}; \mathcal{J}_{t\mathbf{J}_t}, \Sigma')$ is globally invertible as a function of $\mathcal{J}'_{t\mathbf{C}_{t2}} \in \mathbf{S}_{t2}(\mathcal{J}_{t\mathbf{J}_t})$ and therefore we can express $\mathcal{J}'_{t\mathbf{C}_{t2}}$ as a function of $\Gamma'_t \in \bar{\Theta}_\Gamma^t$, given $\mathcal{J}_{t\mathbf{J}_t}$ and Σ' : $\mathcal{J}'_{t\mathbf{C}_{t2}} = \tilde{s}_{t\mathbf{C}_{t2}}(\Gamma'_t; \mathcal{J}_{t\mathbf{J}_t}, \Sigma')$. Then, by plugging $\mathcal{J}'_{t\mathbf{C}_{t2}} = \tilde{s}_{t\mathbf{C}_{t2}}(\Gamma'_t; \mathcal{J}_{t\mathbf{J}_t}, \Sigma')$ into (A.35), we can express each δ'_{tj} from $\delta'_{t\mathbf{J}_t} = (\delta'_{tj})_{j \in \mathbf{J}_t}$ as a function of the observed product-level market shares $\mathcal{J}_{t\mathbf{J}_t}$:

$$\begin{aligned} \delta'_{tj} &= s_{tj}^{-1} \left((\mathcal{J}_{tj} - \sum_{\mathbf{b} \in \mathbf{C}_{t2}: j \in \mathbf{b}} \mathcal{J}'_{t\mathbf{b}})_{j \in \mathbf{J}_t}, \mathcal{J}'_{t\mathbf{C}_{t2}}; \Sigma'_F \right) \\ &= s_{tj}^{-1} \left((\mathcal{J}_{tj} - \sum_{\mathbf{b} \in \mathbf{C}_{t2}: j \in \mathbf{b}} \tilde{s}_{t\mathbf{b}}(\Gamma'_t; \mathcal{J}_{t\mathbf{J}_t}, \Sigma'))_{j \in \mathbf{J}_t}, \tilde{s}_{t\mathbf{C}_{t2}}(\Gamma'_t; \mathcal{J}_{t\mathbf{J}_t}, \Sigma'); \Sigma'_F \right) \\ &= s_{tj}^{-1}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma') \end{aligned}$$

and determine the remaining $\delta'_{t\mathbf{b}}$ for each $\mathbf{b} \in \mathbf{C}_{t2}$ by $\delta'_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} \delta'_{tj} + \Gamma'_{\mathbf{b}} + g_{t\mathbf{b}}(x_{t\mathbf{b}}, p_{t\mathbf{b}}; \Sigma'_g)$, so that $s_{t\mathbf{b}}(\delta'_{t\mathbf{b}}(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) = \mathcal{J}'_{t\mathbf{b}}$ for each $\mathbf{b} \in \mathbf{C}_{t1}$. Then, for any $j \in \mathbf{J}_t$, we obtain $s_{tj}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathcal{J}_{tj}$ and finally:

$$s_{t\mathbf{J}_t}(\delta'_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathcal{J}_{t\mathbf{J}_t}.$$

This shows existence. To prove uniqueness, suppose that there exists another $\delta''_{t\mathbf{J}_t} \neq \delta'_{t\mathbf{J}_t}$ such that $s_{t\mathbf{J}_t}(\delta''_{t\mathbf{J}_t}; \Gamma', \Sigma') = \mathcal{J}_{t\mathbf{J}_t}$. Then, $\delta''_t \neq \delta'_t$. Because Σ' is given, according to Lemma 1, $s_t(\delta''_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) \neq s_t(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$. Moreover, because also $\mathcal{J}_{t\mathbf{J}_t}$ is given, then there must exist some $\mathbf{b} \in \mathbf{C}_{t2}$ for which $s_{t\mathbf{b}}(\delta''_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) \neq s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$. This contradicts $\tilde{s}_{t\mathbf{C}_{t2}}(\Gamma'; \mathcal{J}_{t\mathbf{J}_t}, \Sigma')$ being a function of Γ' .

A.14 Proof of Theorem 6

We start by proving a useful Lemma. Denote the log-likelihood function evaluated at the market shares observed without sampling error by:

$$\ell(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma') = \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_t} \mathcal{J}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F). \quad (\text{A.38})$$

Lemma 6. *If Assumptions 1–3 hold and the true $(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{ID}$, then the true $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ is the unique maximizer of $\ell(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ in Θ .*

Proof. We first show that $\ell(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ is maximized at the true $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$. Note that for any $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_t$, $\mathcal{J}_{t\mathbf{b}} = s_{t\mathbf{b}}(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F)$. Then, by using Jensen's inequality,

⁵A square matrix B is positive quasi-definite if $\frac{1}{2}(B + B^\top)$ is positive-definite.

for any $(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ we have:

$$\begin{aligned} \ell(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma') - \ell(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma) &= \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_t} \jmath_{t\mathbf{b}} \log \frac{s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)}{s_{t\mathbf{b}}(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F)} \\ &\leq \sum_{t=1}^T \log \sum_{\mathbf{b} \in \mathbf{C}_t} \jmath_{t\mathbf{b}} \frac{s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)}{s_{t\mathbf{b}}(\delta_t(\Gamma + g_t(\Sigma_g)); \Sigma_F)} \quad (\text{A.39}) \\ &\leq 0. \end{aligned}$$

We now show the uniqueness by contradiction. Suppose that there exists a $(\tilde{\delta}_{1\mathbf{J}_1}, \dots, \tilde{\delta}_{T\mathbf{J}_T}, \tilde{\Gamma}, \tilde{\Sigma}) \neq (\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ such that $(\tilde{\delta}_{1\mathbf{J}_1}, \dots, \tilde{\delta}_{T\mathbf{J}_T}, \tilde{\Gamma}, \tilde{\Sigma})$ is also a maximizer of $\ell(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$. According to Jensen's inequality (A.39), this is equivalent to having $s_{t\mathbf{b}}(\tilde{\delta}_t(\tilde{\Gamma} + g_t(\tilde{\Sigma}_g)); \tilde{\Sigma}_F) = \jmath_{t\mathbf{b}}$ for each $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_{t1}$. As a consequence, we have $m_{\mathbf{b}}(\tilde{\Sigma}; \mathbf{T}) = 0$ and hence $m(\tilde{\Sigma}; \mathbf{T}) = 0$ in addition to $m(\Sigma; \mathbf{T}) = 0$. Note that $\tilde{\Sigma} \neq \Sigma$. Otherwise, by Lemma 1, $\tilde{\delta}_{t\mathbf{J}_t} = \delta_{t\mathbf{J}_t}$ and $\tilde{\Gamma} = \Gamma$ and this would be inconsistent with $(\tilde{\delta}_{1\mathbf{J}_1}, \dots, \tilde{\delta}_{T\mathbf{J}_T}, \tilde{\Gamma}, \tilde{\Sigma}) \neq (\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$. However, because the true $(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{\text{ID}}$, Theorem 3 rules out the possibility of having any other $\tilde{\Sigma}$ different from Σ for which system (1.9) holds, giving rise to a contradiction. \square

We assume the following regularity conditions.

1. θ is an interior point of Θ ;
2. $g_t(\Sigma'_g)$ is twice continuously differentiable with respect to Σ'_g , and the market share function $s_{t\mathbf{b}}(\delta'_t; \Sigma'_F)$, $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_t$, is twice continuously differentiable with respect to (δ'_t, Σ'_F) ;
3. $\sqrt{I}(\hat{\jmath}_t - \jmath_t) \xrightarrow{d} \mathcal{N}(0, \Omega_t)$ independently for $t = 1, \dots, T$, where Ω_t is positive-definite;
4. $\sum_{t=1}^T G_t \Omega_t G_t^T$ is positive-definite, where $G_t = \left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'} \right] \Big|_{\theta'=\theta} \right)_{\mathbf{b} \in \mathbf{C}_{t1}}$;
5. $\frac{\partial^2 \ell(\theta')}{\partial \theta'^2} \Big|_{\theta'=\theta}$ is non-singular.

Condition 3 is compatible with cases in which the individuals in market t make independent purchase decisions. Condition 4 can be obtained when G_t is a full row rank matrix for each $t = 1, \dots, T$. Define $\ell^c(\Gamma', \Sigma')$ on the basis of (A.38):

$$\ell^c(\Gamma', \Sigma') = \ell((\delta_{t\mathbf{J}_t}(\jmath_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T; \Gamma', \Sigma').$$

Throughout the proof, we assume that $\delta_{t\mathbf{J}_t}(\jmath_{t\mathbf{J}_t}; \Gamma', \Sigma')$ and $\delta_{t\mathbf{J}_t}(\hat{\jmath}_{t\mathbf{J}_t}; \Gamma', \Sigma')$ exist. As discussed in the main text, existence can always be verified numerically during estimation. Provided existence, then Theorem 7 guarantees that $\delta_{t\mathbf{J}_t}(\cdot; \Gamma', \Sigma')$ is a global bijection. Our proof for the consistency statement is mainly based on Theorem 2.1 by Newey and McFadden (1994), according to which we need to verify four conditions.

1. (Γ, Σ) is the unique maximizer of $\ell^c(\Gamma', \Sigma')$ in $\Theta_\Gamma \times \Theta_\Sigma$. Given Assumptions 1–3 and that the true $(\delta_{t\mathbf{J}_t})_{t \in \mathbf{T} \setminus \mathbf{T}_0} \in \Delta^{\text{ID}}$, Lemma 6 guarantees that the true $(\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$ is the unique maximizer of $\ell(\delta'_{1\mathbf{J}_1}, \dots, \delta'_{T\mathbf{J}_T}, \Gamma', \Sigma')$ in Θ . Theorem 7 then implies that (Γ, Σ) is the unique maximizer of $\ell^c(\Gamma', \Sigma')$ in $\Theta_\Gamma \times \Theta_\Sigma$.

2. $\Theta_\Gamma \times \Theta_\Sigma$ is compact. This is guaranteed by the definition of Θ .

3. $\ell^c(\Gamma', \Sigma')$ is continuous with respect to (Γ', Σ') in $\Theta_\Gamma \times \Theta_\Sigma$. According to regularity condition 2, for any $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_t$, the market share function $s_{t\mathbf{b}}(\delta'_t; \Sigma'_F)$ is twice continuously differentiable with respect to (δ'_t, Σ'_F) . Remember that $\theta = (\delta_{1\mathbf{J}_1}, \dots, \delta_{T\mathbf{J}_T}, \Gamma, \Sigma)$. Then, $\ell(\theta')$ in (A.38) is twice continuously differentiable in Θ . Moreover, the inverse market share function, $s_t^{-1}(\mathcal{J}_t; \Sigma'_F)$ is continuous with respect to $(\mathcal{J}_t, \Sigma'_F)$, and therefore continuous with respect to $((\mathcal{J}_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t2}}, \Sigma'_F)$. In addition, $g_t(\Sigma'_g)$ is continuously differentiable with respect to Σ'_g . Then, $\Gamma_t((\mathcal{J}_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t2}}, \Sigma')$, as defined in the proof of Theorem 7, is continuous with respect to $((\mathcal{J}_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_{t2}}, \Sigma')$. By applying the invertibility result from Theorem 7 and the continuous dependence with respect to Σ' , we obtain that $\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma')$ is continuous with respect to (Γ', Σ') . Combining this with the continuity of $\ell(\theta')$ in (A.38), we obtain the desired condition.

4. $\sup_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} |\ell_I^c(\Gamma', \Sigma'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T) - \ell^c(\Gamma', \Sigma')| \xrightarrow{P} 0$. Note that

$$\begin{aligned} & \sup_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} |\ell_I^c(\Gamma', \Sigma'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T) - \ell^c(\Gamma', \Sigma')| \leq \sup_{\theta' \in \Theta} |\ell_I(\theta'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T) - \ell(\theta')| \\ & \quad + \sup_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} |\ell((\delta_{t\mathbf{J}_t}(\hat{\mathcal{J}}_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma') - \ell((\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma')| \end{aligned}$$

First, we prove that $\sup_{\theta' \in \Theta} |\ell_I(\theta'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T) - \ell(\theta')| \xrightarrow{P} 0$. To see this, note that:

$$\begin{aligned} & \sup_{\theta' \in \Theta} |\ell_I(\theta'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T) - \ell(\theta')| \\ &= \sup_{\theta' \in \Theta} \left| \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_t} \hat{\mathcal{J}}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_t} \mathcal{J}_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F) \right| \quad (\text{A.40}) \\ &\leq \sup_{\substack{\theta' \in \Theta \\ t=1, \dots, T, \mathbf{b} \in \mathbf{C}_t}} |\log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)| \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_t} |\hat{\mathcal{J}}_{t\mathbf{b}} - \mathcal{J}_{t\mathbf{b}}|. \end{aligned}$$

Because $\log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$ is continuous in Θ and Θ is compact, $\log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$ is uniformly bounded in Θ . Moreover, because both the number of markets, T , and C_t are finite,

$$\sup_{\substack{\theta' \in \Theta \\ t=1, \dots, T, \mathbf{b} \in \mathbf{C}_t}} |\log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)| < \infty.$$

Note that $\hat{\mathcal{J}}_{t\mathbf{b}} \xrightarrow{P} \mathcal{J}_{t\mathbf{b}}$ for $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_t$. Then, the right-hand side of (A.40) converges to zero in probability. Consequently, $\sup_{\theta' \in \Theta} |\ell_I(\theta'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T) - \ell(\theta')| \xrightarrow{P} 0$.

Second, we prove

$$\sup_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} |\ell((\delta_{t\mathbf{J}_t}(\hat{\mathcal{J}}_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma') - \ell((\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma')| \xrightarrow{P} 0. \quad (\text{A.41})$$

Note that for each t , $\delta_{t\mathbf{J}_t}(\mathcal{J}'_{t\mathbf{J}_t}; \Gamma', \Sigma')$ is uniformly continuous with respect to $(\mathcal{J}'_{t\mathbf{J}_t}, \Gamma', \Sigma')$ in a compact set $U_{\mathcal{J}_{t\mathbf{J}_t}} \times \Theta_\Gamma \times \Theta_\Sigma$, where $U_{\mathcal{J}_{t\mathbf{J}_t}}$ is a compact neighbourhood of $\mathcal{J}_{t\mathbf{J}_t}$. Moreover, $\ell(\theta')$ is uniformly

continuous with respect to $\theta' \in \Theta$. Consequently, $\ell((\delta_{t\mathbf{J}_t}(\mathcal{J}'_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma')$ is uniformly continuous with respect to $((\mathcal{J}'_{t\mathbf{J}_t})_{t=1}^T, \Gamma', \Sigma')$. Because $\hat{\mathcal{J}}_{t\mathbf{J}_t} \xrightarrow{P} \mathcal{J}_{t\mathbf{J}_t}$ for $t = 1, \dots, T$, we obtain (A.41) and finally proved the desired condition.

According to Theorem 2.1 by Newey and McFadden (1994), the four conditions verified above guarantee the consistency of $(\hat{\Gamma}, \hat{\Sigma})$. By applying the invertibility result from Theorem 7 and Slutsky's Theorem, $\hat{\theta}_\delta$ is also consistent. This completes the proof of consistency. The proof of asymptotic normality is based on Theorem 3.1 by Newey and McFadden (1994), according to which we need to verify the following six conditions.

1. $(\hat{\Gamma}, \hat{\Sigma}) \xrightarrow{P} (\Gamma, \Sigma)$. This has just been shown above.
2. (Γ, Σ) is an interior point of $\Theta_\Gamma \times \Theta_\Sigma$. This is guaranteed by regularity condition 1.
3. $\ell_I^c(\Gamma', \Sigma'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T)$ is twice continuously differentiable in $\Theta_\Gamma \times \Theta_\Sigma$. According to regularity condition 2, the market share function $s_{t\mathbf{b}}(\delta'_t; \Sigma'_F)$, $t = 1, \dots, T$ and of $\mathbf{b} \in \mathbf{C}_t$, is twice continuously differentiable with respect to (δ'_t, Σ'_F) , the inverse market share function $s_t^{-1}(\mathcal{J}'_t; \Sigma'_F)$ is thus twice continuously differentiable with respect to $(\mathcal{J}'_t; \Sigma'_F)$. Moreover, $g_t(\Sigma'_g)$ is twice continuously differentiable with respect to Σ'_g . As a consequence, by applying the invertibility result from Theorem 7, we obtain that $\delta_{t\mathbf{J}_t}(\mathcal{J}'_{t\mathbf{J}_t}; \Gamma', \Sigma')$ is twice continuously differentiable with respect to $(\mathcal{J}'_{t\mathbf{J}_t}, \Gamma', \Sigma')$. Because $\ell_I^c(\Gamma', \Sigma'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T)$ is a composition of $\ell_I(\delta_{t\mathbf{J}_t}, \Gamma', \Sigma'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T)$ and of $\delta_{t\mathbf{J}_t}(\mathcal{J}'_{t\mathbf{J}_t}; \Gamma', \Sigma')$, and both functions are twice continuously differentiable, $\ell_I^c(\Gamma', \Sigma'; \hat{\mathcal{J}}_1, \dots, \hat{\mathcal{J}}_T)$ is also twice continuously differentiable with respect to (Γ', Σ') .
4. $\sqrt{I} \frac{\partial \ell_I^c}{\partial (\Gamma', \Sigma')} \Big|_{(\Gamma', \Sigma') = (\Gamma, \Sigma)}$ converges to a centered normal distribution. We can write:

$$\begin{aligned} \sqrt{I} \frac{\partial \ell_I^c}{\partial (\Gamma', \Sigma')} &= \sqrt{I} \sum_{t=1}^T \frac{\partial \delta_{t\mathbf{J}_t}}{\partial (\Gamma', \Sigma')} \frac{\partial \ell_I}{\partial \delta_{t\mathbf{J}_t}} + \frac{\partial \ell_I}{\partial (\Gamma', \Sigma')} \\ &= \left[\left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial (\Gamma', \Sigma')} \right)_{t=1, \dots, T} \quad \mathbf{I} \right] \sqrt{I} \frac{\partial \ell_I}{\partial \theta'} \end{aligned} \tag{A.42}$$

It suffices to prove that $\sqrt{I} \frac{\partial \ell_I}{\partial \theta'}$ converges to a centered normal distribution at $\theta' = \theta$.

Define $\ell^t(\theta') = \sum_{\mathbf{b} \in \mathbf{C}_t} s_{t\mathbf{b}} \log s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma'_F)$. Note that $\ell^t(\theta')$ is maximized at $\theta' = \theta$, for $t = 1, \dots, T$. As a consequence, $\frac{\partial \ell^t}{\partial \theta'} \Big|_{\theta'=\theta} = 0$ for $t = 1, \dots, T$. Then,

$$\begin{aligned}
\sqrt{I} \frac{\partial \ell_I}{\partial \theta'} \Big|_{\theta'=\theta} &= \sqrt{I} \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_t} \hat{s}_{t\mathbf{b}} \frac{\log \partial s_{t\mathbf{b}}}{\partial \theta'} \Big|_{\theta'=\theta} \\
&= \sqrt{I} \sum_{t=1}^T \left[\sum_{\mathbf{b} \in \mathbf{C}_t} \hat{s}_{t\mathbf{b}} \frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} \Big|_{\theta'=\theta} - \frac{\partial \ell^t}{\partial \theta'} \Big|_{\theta'=\theta} \right] \\
&= \sqrt{I} \sum_{t=1}^T \left[\sum_{\mathbf{b} \in \mathbf{C}_{t1}} [\hat{s}_{t\mathbf{b}} - s_{t\mathbf{b}}] \frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} \Big|_{\theta'=\theta} + [\hat{s}_{t0} - s_{t0}] \frac{\partial \log s_{t0}}{\partial \theta'} \Big|_{\theta'=\theta} \right] \quad (\text{A.43}) \\
&= \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_{t1}} \sqrt{I} [\hat{s}_{t\mathbf{b}} - s_{t\mathbf{b}}] \left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'} \right] \Big|_{\theta'=\theta} \\
&= \sum_{t=1}^T \left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'} \right] \Big|_{\theta'=\theta} \right)_{\mathbf{b} \in \mathbf{C}_{t1}} \sqrt{I} [\hat{s}_t - s_t].
\end{aligned}$$

where \mathbf{I} denotes the identity matrix. According to regularity condition 3, $\sqrt{I} [\hat{s}_t - s_t] \xrightarrow{d} \mathcal{N}(0, \Omega_t)$ independently for $t = 1, \dots, T$. By using Slutsky's Theorem, we obtain that $\sqrt{I} \frac{\partial \ell_I}{\partial \theta'} \Big|_{\theta'=\theta} \xrightarrow{d} \mathcal{N}(0, \sum_{t=1}^T G_t \Omega_t G_t^T)$, where $\sum_{t=1}^T G_t \Omega_t G_t^T$ is positive-definite according to regularity condition 4. As a consequence, $\sqrt{I} \frac{\partial \ell_I}{\partial (\Gamma', \Sigma')} \Big|_{(\Gamma', \Sigma')=(\Gamma, \Sigma)}$ converges to a centered normal distribution.

5. $\sup_{(\Gamma', \Sigma') \in \Theta_{\Gamma} \times \Theta_{\Sigma}} \left| \frac{\partial^2 \ell_I^c}{\partial (\Gamma', \Sigma')^2} (\Gamma', \Sigma') - H(\Gamma', \Sigma') \right| \xrightarrow{p} 0$, where

$$\begin{aligned}
H(\Gamma', \Sigma') &= \frac{\partial^2 \ell((\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma')}{\partial (\Gamma', \Sigma')^2} \\
&= \sum_{t=1}^T \sum_{\mathbf{b} \in \mathbf{C}_t} s_{t\mathbf{b}} \frac{\partial^2 \log s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'), \Gamma' + g_t(\Sigma'_g)); \Sigma'_F)}{\partial (\Gamma', \Sigma')^2}, \quad (\text{A.44})
\end{aligned}$$

where $\delta_t(\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'), \Gamma' + g_t(\Sigma'_g)) = (\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'), (\sum_{j \in \mathbf{b}} \delta_{tj}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma') + \Gamma'_{\mathbf{b}} + g_{t\mathbf{b}}(\Sigma'_g))_{\mathbf{b} \in \mathbf{C}_{t2}})$. Under regularity condition 2, $H(\Gamma', \Sigma')$ is continuous in $\Theta_{\Gamma} \times \Theta_{\Sigma}$. Note that, similarly to (A.40), we have:

$$\begin{aligned}
&\sup_{(\Gamma', \Sigma') \in \Theta_{\Gamma} \times \Theta_{\Sigma}} \left| \frac{\partial^2 \ell_I^c}{\partial (\Gamma', \Sigma')^2} (\Gamma', \Sigma') - H(\Gamma', \Sigma') \right| \\
&\leq \sum_{t=1, \dots, T, \mathbf{b} \in \mathbf{C}_t} \sup_{(\Gamma', \Sigma') \in \Theta_{\Gamma} \times \Theta_{\Sigma}} \left| \frac{\partial^2 \log s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'), \Gamma' + g_t(\Sigma'_g)); \Sigma'_F)}{\partial (\Gamma', \Sigma')^2} \right| |\hat{s}_{t\mathbf{b}} - s_{t\mathbf{b}}| \\
&\quad + \sum_{t=1, \dots, T, \mathbf{b} \in \mathbf{C}_t} \hat{s}_{t\mathbf{b}} \sup_{(\Gamma', \Sigma') \in \Theta_{\Gamma} \times \Theta_{\Sigma}} \left| \frac{\partial^2 \log [s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'), \Gamma' + g_t(\Sigma'_g)); \Sigma'_F) - s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t}(\mathcal{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'), \Gamma' + g_t(\Sigma'_g)); \Sigma'_F)]}{\partial (\Gamma', \Sigma')^2} \right| \quad (\text{A.45})
\end{aligned}$$

Due to the twice continuous differentiability of $s_{t\mathbf{b}}(\delta'_t; \Sigma')$ and of $\delta_{t\mathbf{J}_t}(\mathfrak{J}_{t\mathbf{J}_t}; \Gamma', \Sigma')$ in the compact set $\Theta_\Gamma \times \Theta_\Sigma$, for $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_t$, we have:

$$\sup_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} \left| \frac{\partial^2 \log s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t}(\mathfrak{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'), \Gamma' + g_t(\Sigma'_g)); \Sigma'_F)}{\partial(\Gamma', \Sigma')^2} \right| < \infty.$$

Because $\hat{\mathfrak{J}}_t \xrightarrow{p} \mathfrak{J}_t$ for $t = 1, \dots, T$, then the first part on the right-hand side of (A.45) converges to zero in probability. For the second part, note that $s_{t\mathbf{b}}(\delta'_t(\Gamma' + g_t(\Sigma'_g)); \Sigma_F)$ is twice continuously differentiable with respect to $(\delta'_{\mathbf{J}_t}, \Gamma', \Sigma')$ and that $\delta_{t\mathbf{J}_t}(\mathfrak{J}'_{t\mathbf{J}_t}; \Gamma', \Sigma')$ is twice continuously differentiable with respect to $(\mathfrak{J}'_{t\mathbf{J}_t}, \Gamma', \Sigma')$ in a compact set $U_{\mathfrak{J}_{t\mathbf{J}_t}} \times \Theta_\Gamma \times \Theta_\Sigma$, where $U_{\mathfrak{J}_{t\mathbf{J}_t}}$ is a compact neighbourhood of $\mathfrak{J}_{t\mathbf{J}_t}$, for $t = 1, \dots, T$ and $\mathbf{b} \in \mathbf{C}_{t1}$, we then obtain that $\frac{\partial^2 \log s_{t\mathbf{b}}(\delta_t(\delta_{t\mathbf{J}_t}(\mathfrak{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'), \Gamma' + g_t(\Sigma'_g)); \Sigma'_F)}{\partial(\Gamma', \Sigma')^2}$ is uniformly continuous in $U_{\mathfrak{J}_{t\mathbf{J}_t}} \times \Theta_\Gamma \times \Theta_\Sigma$. Combining this with $\hat{\mathfrak{J}}_t \xrightarrow{p} \mathfrak{J}_t$ for $t = 1, \dots, T$, we obtain that the second part on the right-hand side of (A.45) also converges to zero in probability. Consequently,

$$\sup_{(\Gamma', \Sigma') \in \Theta_\Gamma \times \Theta_\Sigma} \left| \frac{\partial^2 \ell_f}{\partial(\Gamma', \Sigma')^2}(\Gamma', \Sigma') - H(\Gamma', \Sigma') \right| \xrightarrow{p} 0.$$

6. $H(\Gamma, \Sigma) = \frac{\partial^2 \ell^c(\Gamma', \Sigma')}{\partial(\Gamma', \Sigma')^2} \Big|_{(\Gamma', \Sigma') = (\Gamma, \Sigma)}$ is non-singular. Note that

$$\begin{aligned} H(\Gamma', \Sigma') &= \frac{\partial^2 \ell((\delta_{t\mathbf{J}_t}(\mathfrak{J}_{t\mathbf{J}_t}; \Gamma', \Sigma'))_{t=1}^T, \Gamma', \Sigma')}{\partial(\Gamma', \Sigma')^2} \\ &= \sum_{t=1}^T \left[\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} \frac{\partial^2 \ell}{\partial \delta_{t\mathbf{J}_t}^2} \left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} \right)^\top + \frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} \frac{\partial^2 \ell}{\partial \delta'_{t\mathbf{J}_t} \partial(\Gamma', \Sigma')} \right] \\ &\quad + \sum_{t=1}^T \sum_{j \in \mathbf{J}_t} \frac{\partial \ell}{\partial \delta'_{tj}} \frac{\partial^2 \delta_{tj}}{\partial(\Gamma', \Sigma')^2} \\ &\quad + \sum_{t=1}^T \frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} \frac{\partial^2 \ell}{\partial(\Gamma', \Sigma') \partial \delta'_{t\mathbf{J}_t}} + \frac{\partial^2 \ell}{\partial(\Gamma', \Sigma')^2}. \end{aligned}$$

At $(\Gamma', \Sigma') = (\Gamma, \Sigma)$, $\delta'_{t\mathbf{J}_t} = \delta_{t\mathbf{J}_t}(\mathfrak{J}_{t\mathbf{J}_t}; \Gamma, \Sigma) = \delta_{t\mathbf{J}_t}$ and $\frac{\partial \ell}{\partial \delta'_{t\mathbf{J}_t}} = 0$. Then,

$$H(\Gamma, \Sigma) = \left[\left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} \right)_{t=1, \dots, T} \quad \mathbf{I} \right] \left[\frac{\partial^2 \ell(\theta')}{\partial \theta'^2} \right]_{\theta' = \theta} \left[\left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} \right)_{t=1, \dots, T} \quad \mathbf{I} \right]^\top,$$

Because $\left[\left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} \right)_{t=1, \dots, T} \quad \mathbf{I} \right]$ is of full row rank and $\left[\frac{\partial^2 \ell(\theta')}{\partial \theta'^2} \right]_{\theta' = \theta}$ is non-singular according to regularity condition 5, $H(\Gamma, \Sigma)$ is therefore non-singular.

All the six conditions of Theorem 3.1 by [Newey and McFadden \(1994\)](#) are satisfied and there exists W_2 such that $\sqrt{I}[(\hat{\Gamma}, \hat{\Sigma}) - (\Gamma, \Sigma)] \xrightarrow{d} \mathcal{N}(0, W_2)$. By applying the invertibility result from Theorem 7, we have:

$$\begin{aligned} \sqrt{I}(\hat{\delta}_{t\mathbf{J}_t} - \delta_{t\mathbf{J}_t}) &= \sqrt{I}(\delta_{t\mathbf{J}_t}(\hat{\beta}_{t\mathbf{J}_t}, \hat{\Gamma}, \hat{\Sigma}) - \delta_{t\mathbf{J}_t}(\hat{\beta}_{t\mathbf{J}_t}, \Gamma, \Sigma) + \delta_{t\mathbf{J}_t}(\hat{\beta}_{t\mathbf{J}_t}, \Gamma, \Sigma) - \delta_{t\mathbf{J}_t}(\beta_{t\mathbf{J}_t}, \Gamma, \Sigma)) \\ &= \frac{\partial \delta_{t\mathbf{J}_t}(\hat{\beta}_{t\mathbf{J}_t}; \Gamma', \Sigma')}{\partial(\Gamma', \Sigma')} \Big|_{(\Gamma', \Sigma')=(\hat{\Gamma}, \hat{\Sigma})} \sqrt{I}[(\hat{\Gamma}, \hat{\Sigma}) - (\Gamma, \Sigma)] \\ &\quad + \frac{\partial \delta_{t\mathbf{J}_t}(\beta_{t\mathbf{J}_t}; \Gamma, \Sigma)}{\partial \beta_{t\mathbf{J}_t}} \Big|_{\beta_{t\mathbf{J}_t}=\hat{\beta}_{t\mathbf{J}_t}} \sqrt{I}(\hat{\beta}_{t\mathbf{J}_t} - \beta_{t\mathbf{J}_t}). \end{aligned}$$

Using the following Taylor expansion of $\frac{\partial \ell_I^c}{\partial(\Gamma', \Sigma')}$ around (Γ, Σ) :

$$0 = \frac{\partial \ell_I^c}{\partial(\Gamma', \Sigma')} \Big|_{(\Gamma', \Sigma')=(\hat{\Gamma}, \hat{\Sigma})} = \frac{\partial \ell_I^c}{\partial(\Gamma', \Sigma')} \Big|_{(\Gamma', \Sigma')=(\Gamma, \Sigma)} + \frac{\partial^2 \ell_I^c}{\partial(\Gamma', \Sigma')^2} \Big|_{(\Gamma', \Sigma')=(\hat{\Gamma}, \hat{\Sigma})} [(\hat{\Gamma}, \hat{\Sigma}) - (\Gamma, \Sigma)],$$

we obtain

$$\begin{aligned} \sqrt{I}[(\hat{\Gamma}, \hat{\Sigma}) - (\Gamma, \Sigma)] &= - \left[\frac{\partial^2 \ell_I^c}{\partial(\Gamma', \Sigma')^2} \Big|_{(\Gamma', \Sigma')=(\hat{\Gamma}, \hat{\Sigma})} \right]^{-1} \sqrt{I} \frac{\partial \ell_I^c}{\partial(\Gamma', \Sigma')} \Big|_{(\Gamma', \Sigma')=(\Gamma, \Sigma)} \\ &= - \left[\frac{\partial^2 \ell_I^c}{\partial(\Gamma', \Sigma')^2} \Big|_{(\Gamma', \Sigma')=(\hat{\Gamma}, \hat{\Sigma})} \right]^{-1} \\ &\quad \left[\left(\frac{\partial \delta_{t\mathbf{J}_t}}{\partial(\Gamma', \Sigma')} \right)_{t=1, \dots, T} \quad \mathbf{I} \right] \sum_{t=1}^T \left(\left[\frac{\partial \log s_{t\mathbf{b}}}{\partial \theta'} - \frac{\partial \log s_{t0}}{\partial \theta'} \right] \Big|_{\theta'=\theta} \right)_{\mathbf{b} \in \mathbf{C}_{t1}} \sqrt{I} [\hat{\beta}_t - \beta_t]. \end{aligned}$$

Since $\sqrt{I}(\hat{\beta}_t - \beta_t)$ converges to a centered normal distribution, by using Slutsky's Theorem and the consistency of $\hat{\delta}_{t\mathbf{J}_t}$ and $(\hat{\Gamma}, \hat{\Sigma})$, we conclude that $\sqrt{I}(\hat{\delta}_{t\mathbf{J}_t} - \delta_{t\mathbf{J}_t})$ converges to a centered normal distribution. This completes the proof.

A.15 Appendix Tables

Table A.1: RTE Cereal Brands and Market Shares

RTE Cereal Brands		Average Market Shares
General Mills	Fiber/Whole Grain	34.99%
Kellogg's	Regular	8.46%
	Fiber/Whole Grain	17.30%
	Added Sugar	4.45%
Quaker	Regular	1.42%
	Fiber/Whole Grain	9.09%
	Added Sugar	0.76%
Post	Regular	0.04%
	Fiber/Whole Grain	8.31%
	Added Sugar	0.69%
Private Labels	Regular	3.21%
	Fiber/Whole Grain	3.12%
	Added Sugar	2.01%
Small Producers	Regular	0.14%
	Fiber/Whole Grain	4.30%
	Added Sugar	1.71%

Notes: The Table lists the 16 RTE cereal brands obtained by aggregating UPCs as described in the text. For each brand, we report the average market share across the 83256 shopping trips with *some* RTE cereal purchases. Market shares are computed over the shopping trips observed in each store-week combination.

Appendix to Chapter 2

B.1 Cross-Price Elasticities

For the cross-price elasticity between j and r :

$$\begin{aligned}
\frac{\partial s_j}{\partial p_r} &= \int \sum_{\mathbf{b}: \mathbf{b} \ni j} \frac{\partial s_{i\mathbf{b}}}{\partial p_r} dF(\theta_i) \\
&= - \int \alpha_i \sum_{\mathbf{b}: \mathbf{b} \ni j} \frac{\partial s_{i\mathbf{b}}}{\partial \delta_{ir}} dF(\theta_i) \\
&= - \int \alpha_i \left[- \sum_{\mathbf{b}: \mathbf{b} \ni j, r \notin \mathbf{b}} s_{ir} \cdot s_{i\mathbf{b}} + s_{ijr} - s_{ir} \cdot s_{ijr} \right] dF(\theta_i) \\
&= \int \alpha_i [s_{ij} \cdot s_{ir} - s_{ijr}] dF(\theta_i).
\end{aligned}$$

B.2 Proof of Proposition 1

Without loss of generality, I fix $x_{t\mathbf{J}_t} = x$ and ignore the notation t , re-writing the product-level (bundle-level) market share functions as $s_{\mathbf{J}}(\cdot; \Gamma, F)$ ($s_{\mathbf{b}}(\cdot; F)$). Denote the ownership matrix in the factual by Ω and that the after-merger by Ω_m .

First statement. For the price elasticity ε_{jr} at $p_{\mathbf{J}}$ between j and r , I obtain:

$$\varepsilon_{jr} = -\alpha \frac{p_j}{s_j} \frac{s_j(\delta_{\mathbf{J}}; \Gamma, F)}{\partial \delta_r}.$$

If $s_{\mathbf{J}}(\cdot; \Gamma, F)$ is identified, then $\frac{s_j(\delta_{\mathbf{J}}; \Gamma, F)}{\partial \delta_r}$ is also identified. Moreover, because of the invertibility of $s_{\mathbf{J}}(\cdot; \Gamma, F)$ in Theorem 7, $\delta_{\mathbf{J}}$ is also identified from $s_{\mathbf{J}}$. Consequently, ε_{jr} is identified.

For the marginal costs $c_{\mathbf{J}}$, I first derive the first-order conditions (FOCs) of the Bertrand game in the factual:

$$-\alpha \left[\Omega \odot \frac{\partial s_{\mathbf{J}}}{\partial \delta_{\mathbf{J}}} \right] (p_{\mathbf{J}} - c_{\mathbf{J}}) + s_{\mathbf{J}}(\delta_{\mathbf{J}}; \Gamma, F) = 0.$$

As shown in [Iaria and Wang \(2019a\)](#), $\Omega \odot \frac{\partial s_{\mathbf{J}}}{\partial \delta_{\mathbf{J}}}$ is a positive-definite symmetric matrix and therefore invertible. Then, $c_{\mathbf{J}} = p_{\mathbf{J}} - \frac{1}{\alpha} \left[\Omega \odot \frac{\partial s_{\mathbf{J}}}{\partial \delta_{\mathbf{J}}} \right]^{-1} s_{\mathbf{J}}(\delta_{\mathbf{J}}; \Gamma, F)$ is identified.

Given the uniqueness of the prices after the merger, it suffices to examine the FOCs of the Bertrand pricing game after the merger that uniquely determine the prices. In the case of mergers under linear

pricing, the FOCs are:

$$-\alpha \left[\Omega_m \odot \frac{\partial s_{\mathbf{J}}}{\partial \delta_{\mathbf{J}}} \right] (p_{\mathbf{J}}^m - c_{\mathbf{J}}) + s_{\mathbf{J}}(\delta_{\mathbf{J}}^m; \Gamma, F) = 0, \quad (\text{B.1})$$

where $p_{\mathbf{J}}^m$ denote the prices after merger and $\delta_{\mathbf{J}}^m = -\alpha p_{\mathbf{J}}^m + \beta x + \eta + \xi_{\mathbf{J}}$ denote the mean utilities of products after merger. Note that $s_{\mathbf{J}}(\cdot; \Gamma, F)$ (and its derivatives) and (α, β, η) are known. Moreover, given $s_{\mathbf{J}}$ and $p_{\mathbf{J}}$, $\delta_{\mathbf{J}}$ and $\xi_{\mathbf{J}}$ are identified due to the invertibility of $s_{\mathbf{J}}(\cdot; \Gamma, F)$ in Theorem 7. In addition, $c_{\mathbf{J}}$ are already identified. According to (B.1), $p_{\mathbf{J}}^m$ is uniquely determined by $s_{\mathbf{J}}(\cdot; \Gamma, F)$, $(\alpha, \beta, \eta, \xi_{\mathbf{J}})$, and $c_{\mathbf{J}}$. Because the profit after merger is a function of the prices and product-level market shares, the profit change is also identified. Finally, denote the consumer surplus function by

$$V(\delta_{\mathbf{J}}; \Gamma, F) = \mathbb{E}[u_{it\mathbf{b}}] = E_{\theta_i} \left[\ln \left(1 + \sum_{\mathbf{b} \in \text{JUC}_2} e^{\delta_{\mathbf{b}}(\Gamma_{\mathbf{b}}) + \mu(\theta_{i\mathbf{b}})} \right) \right].$$

By using Roy's identity, I obtain $\frac{\partial V(\delta_{\mathbf{J}}; \Gamma, F)}{\partial \delta_{\mathbf{J}}} = s_{\mathbf{J}}(\delta_{\mathbf{J}}; \Gamma, F)$. Because $s_{\mathbf{J}}(\cdot; \Gamma, F)$ is known and $\delta_{\mathbf{J}}$ and $\delta_{\mathbf{J}}^m$ are also identified, then the consumer surplus change before and after merger $\Delta V = \int_{\delta_{\mathbf{J}}}^{\delta_{\mathbf{J}}^m} s_{\mathbf{J}}(\delta'_{\mathbf{J}}; \Gamma, F) d\delta'_{\mathbf{J}}$ is identified.

Second statement. Suppose that Γ and bundle-level market share functions are further identified. Moreover, because $c_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} c_{tj}$, then bundle-level marginal costs are also identified. Then, one can apply similar arguments to those in the proof of the first statement to mergers under nonlinear pricing.

B.3 Proof of Theorem 8

To start with, I plug the definition of $\xi_{\mathbf{J}}(\cdot)$ into (2.7):

$$\mathbb{E}[s_{\mathbf{J}}^{-1}(s_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) + \alpha p_{t\mathbf{J}} - \beta x_{t\mathbf{J}} - \eta | z_{t\mathbf{J}} = z, x_{t\mathbf{J}} = x] = 0, \quad (\text{B.2})$$

where $z \in \mathbf{D}_z$ and $x \in \mathbf{D}_x$ and \mathbf{D}_z and \mathbf{D}_x denotes the support of $z_{t\mathbf{J}}$ and $x_{t\mathbf{J}}$, respectively. In what follows, I fix $x_{t\mathbf{J}} = x$.¹ For a given z , (B.2) defines a set of moment restrictions: $(\alpha, \beta, \eta, \Gamma, F)$ should be such that the left-hand side of (B.2) is equal to zero. I assume the following regularity condition:

Condition 2. For any (Γ', F') and any $z \in \mathbf{D}_z$, there exists $M_z > 0$, such that

$$\mathbb{E} \left[\left| s_{\mathbf{J}}^{-1}(s_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) \right| | z \right], \mathbb{E} [|p_{t\mathbf{J}}| | z] \leq M_z.$$

Sketch of the proof. The proof is proceeded in three steps. In the first step, I prove that under conditions 1-3 in Theorem 8, the identification by moment restrictions (2.7) is equivalent to uniquely solving a convolution equation. This convolution equation is generated by the distribution of the demand and shocks and the translation in the convolution equation is defined by $z_{t\mathbf{J}} \in \mathbb{R}^J$. In the second step, by further using condition 4, I prove that the property that the zero function is the unique solution to the convolution equation is sufficient for the identification of α and $s_{t\mathbf{J}}(\cdot; x^{(2)}, \Gamma, F)$. In the final step, by leveraging the completeness of location families in [Mattner \(1992\)](#), I demonstrate that

¹All the conditions and results in this section should be considered as being conditioned on $x_{t\mathbf{J}} = x$.

when $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$ is Gaussian distributed (or their joint distribution satisfies some “fat-tail” conditions), the property that the zero function is the unique solution to the convolution equation will hold; under regularity condition 3 of Appendix B.3, by leveraging the polynomial completeness in [D’Haultfoeuille \(2011\)](#), I prove that the same property will hold when F has compact support and the data generating process is a model of multiple choice of products across categories.

B.3.1 Conditional Moment Restrictions and Convolution Equation

The following theorem equivalently transforms (B.2) to a convolution equation:

Theorem 14. *Suppose that Assumptions 6-8 and regularity condition 2 hold. Moreover, the following conditions hold:*

1. $z_{t\mathbf{J}}$ is independent of $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$.
2. $\alpha_i = \alpha \neq 0$
3. Given $x^{(2)}$, $p_{t\mathbf{J}} = p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}; x^{(2)})$ is a continuous function of $(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$.

Then, for any $z \in \mathbf{D}_z$, $(\alpha', \beta', \eta', \Gamma', F')$ satisfies moment conditions (B.2) if and only if the following convolution equation

$$\int G(t; \alpha', \beta', \eta', \Gamma', F') \Lambda_G(t - z; f_{\xi, w}) dt = 0, \quad (\text{B.3})$$

holds, where

$$G(t; \alpha', \beta', \eta', \Gamma', F') = s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, t; x^{(2)}); x^{(2)}, \Gamma, F); x^{(2)}, \Gamma', F') + \alpha' p_{\mathbf{J}}(0, t; x^{(2)}) + \left(\frac{\alpha'}{\alpha} \beta - \beta'\right) x + \frac{\alpha'}{\alpha} \eta - \eta',$$

$$\Lambda_G(\lambda; f_{\xi, w}) = \int \alpha f_{\xi, w}(\alpha(w - \lambda) - \beta x - \eta, w) dw,$$

and $f_{\xi, w}$ is the density function of (ξ, w) .

Proof. Since x is fixed, I drop this notation in this proof and also the dependence of $p_{\mathbf{J}}(\cdot)$ and $s_{\mathbf{J}}(\cdot; \Gamma', F')$ on $x_{t\mathbf{J}}^{(2)}$. Because $\alpha_i = \alpha$, moment restrictions (B.2) can be further simplified to:

$$\mathbb{E}[s_{\mathbf{J}}^{-1}(\mathcal{J}_{t\mathbf{J}}; \Gamma, F) + \alpha p_{t\mathbf{J}} - \beta x - \eta | z_{t\mathbf{J}} = z] = 0. \quad (\text{B.4})$$

To start with, I prove the following Lemma:

Lemma 7. *Suppose that $\alpha_i = \alpha$ and $p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$ is a function of $(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$. Then, for any $\Delta \in \mathbb{R}^J$,*

$$p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}} + \alpha \Delta, c_{t\mathbf{J}} + \Delta) = p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \Delta.$$

Proof. Denote by Ω the factual ownership matrix. Then, I can derive the FOCs of the simultaneous Bertrand pricing game:

$$-\alpha \left[\Omega \odot \frac{\partial s_{\mathbf{J}}}{\partial \delta_{t\mathbf{J}}} \right] (p_{t\mathbf{J}} - c_{t\mathbf{J}}) + s_{\mathbf{J}}(\delta_{t\mathbf{J}}; \Gamma, F) = 0, \quad (\text{B.5})$$

where $\delta_{t\mathbf{J}} = -\alpha p_{t\mathbf{J}} + \beta x + \eta + \xi_{t\mathbf{J}}$. Suppose that $c_{t\mathbf{J}}$ increases by Δ and $\beta x + \eta + \xi_{t\mathbf{J}}$ increases by $\alpha \Delta$. Then, the FOCs (B.5) with $p_{\mathbf{J}}(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \Delta$, $c_{t\mathbf{J}} + \Delta$ and $\beta x + \eta + \xi_{t\mathbf{J}} + \alpha \Delta$ still hold because $\delta_{t\mathbf{J}}$ and

$p_{t\mathbf{J}} - c_{t\mathbf{J}}$ remain unchanged. Due to the uniqueness of $p_{t\mathbf{J}}$ as function of $(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$, I obtain that $p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}} + \alpha\Delta, c_{t\mathbf{J}} + \Delta) = p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \Delta$. \square

First, I prove the sufficiency of Theorem 14. For any $\Delta \in \mathbb{R}^J$, by using Lemma 7, I obtain:

$$\begin{aligned} \mathbb{E}[p_{t\mathbf{J}}|z_{t\mathbf{J}} = z] &= \int p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, z + w_{t\mathbf{J}}) f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) \\ &= \int \left[p_{\mathbf{J}}\left(0, z + w_{t\mathbf{J}} - \frac{\beta x + \eta + \xi_{t\mathbf{J}}}{\alpha}\right) + \frac{\beta x + \eta + \xi_{t\mathbf{J}}}{\alpha} \right] f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) \quad (\text{B.6}) \\ &= \int p_{\mathbf{J}}\left(0, z + w_{t\mathbf{J}} - \frac{\beta x + \eta + \xi_{t\mathbf{J}}}{\alpha}\right) f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) + \frac{\beta x + \eta}{\alpha} \end{aligned}$$

Similarly, for $(\alpha', \beta', \eta', \Gamma', F')$ satisfying (B.4), I compute

$$\begin{aligned} &\mathbb{E}[s_{\mathbf{J}}^{-1}(\mathcal{J}_{t\mathbf{J}}; \Gamma', F')|z_{t\mathbf{J}} = z] \\ &= \int s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}} - \alpha p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, z_{t\mathbf{J}} + w_{t\mathbf{J}}); \Gamma, F); \Gamma', F') f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) \quad (\text{B.7}) \\ &= \int s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, z + w_{t\mathbf{J}} - \frac{\beta x + \eta + \xi_{t\mathbf{J}}}{\alpha}); \Gamma, F); \Gamma', F') f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) \end{aligned}$$

I now plug (B.6) and (B.7) in (B.4) evaluated at $(\alpha', \beta', \eta', \Gamma', F')$, and make the transformation $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$ to $(z + w_{t\mathbf{J}} - \frac{\beta x + \eta + \xi_{t\mathbf{J}}}{\alpha}, w_{t\mathbf{J}})$:

$$\begin{aligned} &\mathbb{E}[s_{\mathbf{J}}^{-1}(\mathcal{J}_{t\mathbf{J}}; \Gamma', F') + \alpha' p_{t\mathbf{J}} - \beta' x - \eta'|z_{t\mathbf{J}} = z] \\ &= \int [s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, t); \Gamma, F); \Gamma', F') + \alpha' p_{\mathbf{J}}(0, t) + \left(\frac{\alpha'}{\alpha}\beta - \beta'\right)x + \frac{\alpha'}{\alpha}\eta - \eta'] \alpha f_{\xi,w}(\alpha(z + w_{t\mathbf{J}} - t) - \beta x - \eta, w_{t\mathbf{J}}) d(t, w_{t\mathbf{J}}) \\ &= \int [s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, t); \Gamma, F); \Gamma', F') + \alpha' p_{\mathbf{J}}(0, t) + \left(\frac{\alpha'}{\alpha}\beta - \beta'\right)x + \frac{\alpha'}{\alpha}\eta - \eta'] \Lambda_G(t - z; f_{\xi,w}) dt. \end{aligned}$$

The proof of sufficiency is completed. For the necessity part, one can reverse the argument of the sufficiency. \square

B.3.2 Unique Solution for the Convolution Equation (B.3) and Identification of Product-Level Market Share Functions

Denote the identification set defined by moment conditions (B.2) as

$$\Theta = \{(\alpha', \beta', \eta', \Gamma', F') : (\text{B.2}) \text{ holds at } (\alpha', \beta', \eta', \Gamma', F') \text{ for any } z \in \mathbf{D}_z\},$$

and that by (B.3) as

$$\Theta_G = \{(\alpha', \beta', \eta', \Gamma', F') : (\text{B.3}) \text{ holds for any } z \in \mathbf{D}_z\}$$

Theorem 14 establishes $\Theta = \Theta_G$. Define $\Theta_G^0 = \{(\alpha', \beta', \eta', \Gamma', F') : G(\cdot; \alpha', \beta', \eta', \Gamma', F') = 0\}$, the set of parameters that delivers $G(\cdot) = 0$. Note that $\Theta_G^0 \subset \Theta_G$ and the true parameters $(\alpha, \beta, \eta, \Gamma, F) \in \Theta_G^0 \subset \Theta_G = \Theta$. Then, a necessary condition for the identification of $(\alpha, \beta, \eta, \Gamma, F)$ by moment conditions (B.2), i.e. $\Theta = \{(\alpha, \beta, \eta, \Gamma, F)\}$, is $\Theta_G^0 = \Theta_G$. Equivalently, $G = 0$ is the unique solution of convolution equation (B.3), i.e. the completeness of the location families generated by $\Lambda_G(\cdot; f_{\xi,w})$. The next theorem characterizes the implications of this completeness on identification:

Theorem 15. *Suppose that conditions of Theorem 14 hold.*

1. *If $\Theta = \{(\alpha, \beta, \eta, \Gamma, F)\}$, then $\Theta_G^0 = \Theta_G$.*
2. *Suppose that for any $(\Gamma', F') \neq (\Gamma, F)$, there exists $\delta'_J \neq \delta''_J$ such that $s_J(\delta'_J; x^{(2)}, \Gamma, F) = s_J(\delta'_J; x^{(2)}, \Gamma', F')$ and $s_J(\delta''_J; x^{(2)}, \Gamma, F) = s_J(\delta''_J; x^{(2)}, \Gamma', F')$. If $\Theta_G^0 = \Theta_G$, then α and $s_J(\cdot; x^{(2)}, \Gamma, F)$ are identified.*

Remark 10. *According to Theorem 15, the completeness of the location families $(\Theta_G^0 = \Theta_G)$ is necessary for the identification of the full model by moment conditions (2.7). Moreover, as shown in the second statement, it is also sufficient for the identification of α and $s_J(\cdot; x^{(2)}, \Gamma, F)$ under mild conditions.*

Proof. For the first statement, if $\Theta = \{(\alpha, \beta, \eta, \Gamma, F)\}$, then $\Theta_G^0 = \Theta_G = \{(\alpha, \beta, \eta, \Gamma, F)\}$.

To prove the second statement, note that if $G(t; \alpha', \beta', \eta', \Gamma', F') = 0$ for any $t \in \mathbb{R}^J$, then we have

$$s_J(-\alpha p_J(0, t); \Gamma, F) = s_J(-\alpha' p_J(0, t) + v; \Gamma', F'), \quad (\text{B.8})$$

for any $t \in \mathbb{R}^J$, where $v = \left(\frac{\alpha'}{\alpha}\beta - \beta'\right)x + \frac{\alpha'}{\alpha}\eta - \eta'$. As shown in Iaria and Wang (2019a) (Appendix 8.10), marginal costs c_{tJ} are identifiable: for any p'_{tJ} , there exists a unique c'_{tJ} such that the FOCs of the Bertrand pricing game hold. Moreover, this mapping from p'_{tJ} to c'_{tJ} is C^1 . Because $p_J(0, c'_{tJ})$ is continuous, then $p'_{tJ} = p_J(0, c'_{tJ})$ defines a continuous bijection between prices and marginal costs. Consequently, $\mathbf{D}_p = \{p' \in \mathbb{R}^J : p' = p_J(0, c_{tJ'}), c_{tJ'} \in \mathbb{R}^J\}$ is an open set in \mathbb{R}^J and (B.8) holds in \mathbf{D}_p . Finally, according to Iaria and Wang (2019a) (Theorem Real Analytic Property), given any (Γ'', F'') , $s_b(\delta'_{tJ}; \Gamma'', F'')$ is real analytic with respect to δ'_{tJ} . Then, $s_J(\delta'_{tJ}; \Gamma'', F'')$ is real analytic with respect to δ'_{tJ} . Consequently, $s_J(-\alpha p'_{tJ}; \Gamma, F)$ and $s_J(-\alpha' p'_{tJ} + v; \Gamma', F')$ are both real analytic with respect to p'_{tJ} . Because these two real analytic functions coincide on an open set $\mathbf{D}_p \subset \mathbf{R}^J$ according to (B.8), then they coincide for all $p'_{tJ} \in \mathbf{R}^J$: (B.8) holds for any $p_{tJ} \in \mathbf{R}^J$. Moreover, there exist at least two $\delta'_J \neq \delta''_J$ such that $s_J(\delta'_J; \Gamma, F) = s_J(\delta'_J; \Gamma', F')$ and $s_J(\delta''_J; \Gamma, F) = s_J(\delta''_J; \Gamma', F')$. Setting $p_{tJ} = -\frac{\delta'_J}{\alpha}$ and plugging this into (B.8), we obtain:

$$s_J(\delta'_J; \Gamma, F) = s_J\left(\frac{\alpha'}{\alpha}\delta'_J + v; \Gamma', F'\right) = s_J(\delta'_J; \Gamma', F').$$

Similarly, setting $p_{tJ} = -\frac{\delta''_J}{\alpha}$, we obtain:

$$s_J(\delta''_J; \Gamma, F) = s_J\left(\frac{\alpha'}{\alpha}\delta''_J + v; \Gamma', F'\right) = s_J(\delta''_J; \Gamma', F').$$

Because $s_J(\cdot; \Gamma', F')$ is bijective, we obtain:

$$\begin{aligned} \frac{\alpha'}{\alpha}\delta'_J + v &= \delta'_J, \\ \frac{\alpha'}{\alpha}\delta''_J + v &= \delta''_J. \end{aligned}$$

Then, $(\frac{\alpha'}{\alpha} - 1)(\delta'_J - \delta''_J) = 0$, and therefore $\alpha' = \alpha$ and $v = 0$. Finally, we obtain $s_J(\delta'_{tJ}; \Gamma, F) = s_J(\delta'_{tJ}; \Gamma', F')$ for any $\delta'_{tJ} \in \mathbf{R}^J$. The proof of the second statement is completed. \square

B.3.3 Sufficient Conditions for the Completeness of Location Families

In general, depending on the regularity of $G(\cdot)$ (bounded, polynomially bounded, integrable with respect to $\Lambda_G(\cdot)$, etc.), the completeness of location families can be achieved with different sufficient conditions on $\Lambda_G(\cdot)$ (and hence on $f_{\xi,w}$).² In the setting of mixed logit models, the next theorem establishes two sets of sufficient conditions for the completeness of location families ($\Theta_G^0 = \Theta_G$):

Theorem 16. *Suppose that (B.3) holds for $z \in \mathbb{R}^J$.*

- *If $f_{\xi,w}$ is Gaussian, then $\Theta_G^0 = \Theta_G$.*
- *Suppose that the data generating process is a model of multiple choice of product across categories (see section 2.2.2). If F and F' have compact support. Then under regularity condition 3, $\Theta_G^0 = \Theta_G$.*

Proof. Note that the location families are generated by $\Lambda_G(\cdot; f_{\xi,w})$, which is the density function of a translation of demand and supply shocks in model (2.4). For the first statement, when $f_{\xi,w}$ is Gaussian, $\Lambda_G(\lambda; f_{\xi,w})$ is also Gaussian. Then, the first statement follows directly from Theorem 2.4 of [Mattner \(1993\)](#).

For the second statement, I leverage Theorem 2.1 of [D'Haultfoeulle \(2011\)](#). To do so, I require the following regularity conditions:

Condition 3.

- (i). *$(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$ are continuous random variables with finite moments.*
- (ii). *The characteristics function of Λ_G is infinitely often differentiable in \mathbb{R}^J except for some finite set. Moreover, the characteristics function of Λ_G does not vanish on \mathbb{R}^J .*
- (iii). *There exists B and l , such that $|p_{t\mathbf{J}}(0, c_{t\mathbf{J}})| \leq B|c_{t\mathbf{J}}|^l$, where $|\cdot|$ refers to Euclidean norm.*

Condition 3(i) implies Assumption A3 of [D'Haultfoeulle \(2011\)](#) and quite standard. Condition 3(ii) implies his Assumption A4. The differentiability requirement and the zero-freeness requirement are satisfied by many commonly used distributions. Condition 3(iii) restricts pricing behaviors to be controlled by a polynomial of marginal costs and is satisfied at least by the mixed logit demand models of single products. Moreover, together with Condition 3(i), it implies Assumption A5 of [D'Haultfoeulle \(2011\)](#).

First, I re-write G as a function of $p_{t\mathbf{J}}(0, t)$:

$$G = G(p_{t\mathbf{J}}) = s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{t\mathbf{J}}; \Gamma, F); \Gamma', F') + \alpha' p_{t\mathbf{J}} + \left(\frac{\alpha'}{\alpha} \beta - \beta' \right) x + \frac{\alpha'}{\alpha} \eta - \eta'.$$

To apply statement (ii) of Theorem 2.1 in [D'Haultfoeulle \(2011\)](#), it is enough to prove that G can be polynomially controlled by $p_{t\mathbf{J}}$:

Lemma 8. *There exists $A, M > 0$, such that $|G(p_{t\mathbf{J}})| \leq A|p_{t\mathbf{J}}| + M$, for any $p_{t\mathbf{J}} \in \mathbb{R}^J$.*

Combining this lemma with Conditions 3(i)-(iii), I can apply the P-completeness result in Theorem 2.1 of [D'Haultfoeulle \(2011\)](#): if G satisfies convolution equation (B.3) for any $z \in \mathbb{R}^J$, then $G \equiv 0$. In the remaining part, I prove Lemma 8.

²For different concepts of completeness, see [Mattner \(1992, 1993\)](#), [D'Haultfoeulle \(2011\)](#), and [Andrews \(2017\)](#).

Without loss of generality, normalize the support of F and F' to $[0, 1]^K$, where K is the dimension of random coefficients. Note that it is sufficient to prove

$$s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{t\mathbf{J}}; \Gamma, F); \Gamma', F') \leq A' |p_{t\mathbf{J}}| + M'$$

for some constant A' and M' . First, consider demand models of single products. For any $\delta'_{\mathbf{J}} \in \mathbb{R}^J$, denote $s_{\mathbf{J}} = s_{\mathbf{J}}(\delta_{\mathbf{J}}; F) = s_{\mathbf{J}}(\delta'_{\mathbf{J}}; F')$. Then, we have for any $j \in \mathbf{J}$:

$$\begin{aligned} \ln s_j - \ln s_0 &= \delta_j + \ln \frac{\int \frac{e^{\mu_{tj}(\theta_i)}}{1 + \sum_{j \in \mathbf{J}} e^{\delta_j + \mu_{tj}(\theta_i)}} dF(\theta_i)}{\int \frac{1}{1 + \sum_{j \in \mathbf{J}} e^{\delta_j + \mu_{tj}(\theta_i)}} dF(\theta_i)} \\ &= \delta_j + \mu_{tj}(\tilde{\theta}), \end{aligned}$$

where $\tilde{\theta}$ is some value in $[0, 1]^K$. We apply the same arguments to F' and obtain:

$$\ln s_j - \ln s_0 = \delta'_j + \mu_{tj}(\tilde{\theta}'),$$

where $\tilde{\theta}'$ is some value in $[0, 1]^K$. Then, we have $\delta_j - \delta'_j = \mu_{tj}(\tilde{\theta}') - \mu_{tj}(\tilde{\theta})$. Because both $\tilde{\theta}$ and $\tilde{\theta}'$ are bounded by 1, then $\mu_{tj}(\tilde{\theta})$ and $\mu_{tj}(\tilde{\theta}')$ are also bounded. Then, we obtain that there exists a constant M_j that does not depend on $\delta_{\mathbf{J}}$, such that $|\delta_j - \delta'_j| \leq M_j$. Consequently, $|\delta - \delta'| \leq M' = \sqrt{\sum_{j=1}^J M_j^2}$, or equivalently, $|s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(\delta_{\mathbf{J}}; \Gamma, F); \Gamma', F') - \delta_{\mathbf{J}}| \leq M'$ for any $\delta_{\mathbf{J}} \in \mathbb{R}^J$. Plug $\delta_{\mathbf{J}} = -\alpha p_{t\mathbf{J}}$ into this inequality, we obtain $|s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{t\mathbf{J}}; \Gamma, F); \Gamma', F') + \alpha p_{t\mathbf{J}}| \leq M'$ and therefore $|s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{t\mathbf{J}}; \Gamma, F); \Gamma', F')| \leq \alpha |p_{t\mathbf{J}}| + M'$.

For models of multiple choice of products across K categories, for any $\delta'_{\mathbf{J}} \in \mathbb{R}^J$, denote $s_{\mathbf{J}} = s_{\mathbf{J}}(\delta_{\mathbf{J}}; \Gamma, F) = s_{\mathbf{J}}(\delta'_{\mathbf{J}}; \Gamma', F')$. Take product category \mathbf{J}_1 and define $\tilde{s}_0 = 1 - \sum_{j \in \mathbf{J}_1} s_j$. Note that

$$\tilde{s}_0 = \int \frac{\sum_{\mathbf{b}=(j_k)_{j_k \in \mathbf{J}_k, k=2, \dots, K}} e^{\sum_{k=2}^K \delta_{j_k} + \Gamma_{\mathbf{b}} + \sum_{k=2}^K \mu_{tj_k}(\theta_i)}}{\sum_{\mathbf{b}=(j_k)_{j_k \in \mathbf{J}_k, k=1, \dots, K}} e^{\sum_{k=1}^K \delta_{j_k} + \Gamma_{\mathbf{b}} + \mu_{t\mathbf{b}}(\theta_i)}} dF(\theta_i)$$

and for $j \in \mathbf{J}_1$,

$$s_j = \int \frac{\sum_{\mathbf{b}=(j_k)_{j_k \in \mathbf{J}_k, k=2, \dots, K}} e^{\delta_j + \mu_{tj}(\theta_i)} e^{\sum_{k=2}^K \delta_{j_k} + \Gamma_{\mathbf{b} \cup \{j\}} + \sum_{k=2}^K \mu_{tj_k}(\theta_i)}}{\sum_{\mathbf{b}=(j_k)_{j_k \in \mathbf{J}_k, k=1, \dots, K}} e^{\sum_{k=1}^K \delta_{j_k} + \Gamma_{\mathbf{b}} + \mu_{t\mathbf{b}}(\theta_i)}} dF(\theta_i).$$

Then, similar to demand models of single products, we obtain:

$$\ln s_j - \ln \tilde{s}_0 = \delta_j + \mu_{tj}(\tilde{\theta}) + \Gamma_{\tilde{\mathbf{b}} \cup \{j\}} - \Gamma_{\tilde{\mathbf{b}}},$$

where $\tilde{\theta}$ is some value in $[0, 1]^K$ and $\tilde{\mathbf{b}}$ is some bundle without $j \in \mathbf{J}_1$. We apply the same arguments to (Γ', F') and obtain:

$$\ln s_j - \ln s_0 = \delta'_j + \mu_{tj}(\tilde{\theta}') + \Gamma_{\tilde{\mathbf{b}}' \cup \{j\}} - \Gamma_{\tilde{\mathbf{b}}'},$$

where $\tilde{\theta}'$ is some value in $[0, 1]^K$ and $\tilde{\mathbf{b}}'$ is some bundle without $j \in \mathbf{J}_1$. Then, similar arguments in demand models of single product apply and $|\delta_j - \delta'_j|$ is bounded by some constant that only depends on the support of $\tilde{\theta}$ and $\tilde{\theta}'$ and the bounds of Γ . The proof of the second statement is completed. \square

Combining Theorems 14-16, we obtain Theorem 8.

B.4 Proof of Theorem 9

The proof is proceeded in two steps. In the first step, I prove $\Gamma_{(j,r)}$ is identified for $j, r \in \mathbf{J}$, $j \neq r$. In the second step, I prove that the $s_{\mathbf{b}}(\cdot; x_t^{(2)}, F)$ is identified for $\mathbf{b} \in \mathbf{C}_1$. To simplify exposition, I fix $x_{t\mathbf{J}}^{(2)}$ and drop this notation. Suppose that there exist (Γ', F') such that

$$s_j(\delta_{\mathbf{J}}; \Gamma', F') = s_j(\delta_{\mathbf{J}}; \Gamma, F)$$

for any $j \in \mathbf{J}$ and $\delta \in \mathbb{R}^J$. Throughout the proof, denote the Fourier transformation of function ϕ by $\mathcal{F}(\phi)$.

Step 1: $\Gamma' = \Gamma$. Without loss of generality, we show $\Gamma_{(1,2)} = \Gamma'_{(1,2)}$. First, note that by letting all δ'_{it} 's, $l \neq 1, 2$ tend to $-\infty$, market shares of single products l and of all bundles that contain any product $l \neq 1, 2$ converge to zero. Consequently, we obtain that:

$$\begin{aligned} s_1(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F') &= s_1(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F), \\ s_2(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F') &= s_2(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F), \end{aligned} \tag{B.9}$$

for any $\delta'_{t\{1,2\}} = (\delta'_{t1}, \delta'_{t2}) \in \mathbb{R}^2$. Take the first equation in (B.9) and compute the partial derivatives with respect to δ'_{t2} :

$$\begin{aligned} \frac{\partial s_1(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F')}{\partial \delta'_{t2}} &= \int \frac{(e^{\delta'_{t1} + \delta'_{t2} + \mu_{t1}(\theta_{it}) + \mu_{t2}(\theta_{it})})(e^{\Gamma'_{(1,2)}} - 1)}{(1 + e^{\delta'_{t1} + \mu_{t1}(\theta_{it})} + e^{\delta'_{t2} + \mu_{t2}(\theta_{it})} + e^{\delta'_{t1} + \mu_{t1}(\theta_{it}) + \delta'_{t1} + \mu_{t1}(\theta_{it}) + \Gamma'_{(1,2)}})^2} dF'(\theta_{it}), \\ \frac{\partial s_1(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F)}{\partial \delta'_{t2}} &= \int \frac{(e^{\delta'_{t1} + \delta'_{t2} + \mu_{t1}(\theta_{it}) + \mu_{t2}(\theta_{it})})(e^{\Gamma_{(1,2)}} - 1)}{(1 + e^{\delta'_{t1} + \mu_{t1}(\theta_{it})} + e^{\delta'_{t2} + \mu_{t2}(\theta_{it})} + e^{\delta'_{t1} + \mu_{t1}(\theta_{it}) + \delta'_{t1} + \mu_{t1}(\theta_{it}) + \Gamma_{(1,2)}})^2} dF(\theta_{it}), \\ \frac{\partial s_1(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F')}{\partial \delta'_{t2}} &= \frac{\partial s_1(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F)}{\partial \delta'_{t2}}. \end{aligned}$$

I can then cancel out $e^{\delta'_{t1} + \delta'_{t2}}$ in the nominators of $\frac{\partial s_1(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F')}{\partial \delta'_{t2}}$ and $\frac{\partial s_1(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F)}{\partial \delta'_{t2}}$. Letting $\delta_{t2} \rightarrow -\infty$, I obtain:

$$[e^{\Gamma'_{(1,2)}} - 1] \int \frac{e^{\mu_{t1}(\theta_{it}) + \mu_{t2}(\theta_{it})}}{(1 + e^{\delta'_{t1} + \mu_{t1}(\theta_{it})})^2} dF'(\theta_{it}) = [e^{\Gamma_{(1,2)}} - 1] \int \frac{e^{\mu_{t1}(\theta_{it}) + \mu_{t2}(\theta_{it})}}{(1 + e^{\delta'_{t1} + \mu_{t1}(\theta_{it})})^2} dF(\theta_{it}). \tag{B.10}$$

From (B.10), if $\Gamma_{(1,2)} = 0$, then $\Gamma'_{(1,2)} = \Gamma_{(1,2)} = 0$.³ Suppose $\Gamma_{(1,2)} \neq 0$. Denote the density functions of $\mu_{it} = (\mu_{it1}, \mu_{it2}) = (\mu_{t1}(\theta_{it}), \mu_{t2}(\theta_{it}))$ for $\theta_{it} \sim F$ and $\theta_{it} \sim F'$ by f_{μ} and f'_{μ} , respectively. Then, I can re-write (B.10) as:

$$[e^{\Gamma'_{(1,2)}} - 1] \int \frac{e^{\mu_{it1} + \mu_{it2}}}{(1 + e^{\mu_{it1}})^2} f'_{\mu}(\mu_{it1} - \delta'_{t1}, \mu_{it2}) d\mu_{it} = [e^{\Gamma_{(1,2)}} - 1] \int \frac{e^{\mu_{it1} + \mu_{it2}}}{(1 + e^{\mu_{it1}})^2} f_{\mu}(\mu_{it1} - \delta'_{t1}, \mu_{it2}) d\mu_{it}.$$

³In fact, the sign of $\Gamma_{(1,2)}$ is already identified from (B.10).

Define $g(\lambda) = \frac{e^\lambda}{(1+e^\lambda)^2}$. Then,

$$[e^{\Gamma(1,2)} - 1] \int g(\mu_{it1}) \tilde{f}'_\mu(\mu_{it1} - \delta'_{t1}) d\mu_{it1} = [e^{\Gamma(1,2)} - 1] \int g(\mu_{it1}) \tilde{f}_\mu(\mu_{it1} - \delta'_{t1}) d\mu_{it1}, \quad (\text{B.11})$$

where $\tilde{f}'_\mu(\mu_{it1}) = \int e^{\mu_{it2}} f'_\mu(\mu_{it1}, \mu_{it2}) d\mu_{it2}$ and $\tilde{f}_\mu(\mu_{it1}) = \int e^{\mu_{it2}} f_\mu(\mu_{it1}, \mu_{it2}) d\mu_{it2}$. Either side of (B.11) defines a convolution system. Note that $g(\cdot), \tilde{f}'_\mu, \tilde{f}_\mu \in L^1(\mathbb{R})$. Consequently, I apply Fourier transformation on both sides of (B.11) and obtain:

$$[e^{\Gamma(1,2)} - 1] \mathcal{F}(g)(t) \mathcal{F}(\tilde{f}'_\mu)(t) = [e^{\Gamma(1,2)} - 1] \mathcal{F}(g)(t) \mathcal{F}(\tilde{f}_\mu)(t),$$

for any $t \in \mathbb{R}$. Particularly, at $t = 0$, $\mathcal{F}(g)(0) > 0$. Then,

$$[e^{\Gamma(1,2)} - 1] \mathcal{F}(\tilde{f}'_\mu)(0) = [e^{\Gamma(1,2)} - 1] \mathcal{F}(\tilde{f}_\mu)(0), \quad (\text{B.12})$$

Note that $\mathcal{F}(\tilde{f}'_\mu)(t) = E_{\tilde{f}'}[e^{\mu_{it2}}]$ and $\mathcal{F}(\tilde{f}_\mu)(t) = E_{\tilde{f}}[e^{\mu_{it2}}]$. If they are equal, then $\Gamma'_{(1,2)} = \Gamma_{(1,2)}$. In particular, if $\Gamma_{(1,2)} = -\infty$, i.e. bundle (1, 2) is not in the choice set, I obtain that $\Gamma'_{(1,2)} = -\infty$ and therefore identify that bundle (1, 2) is not in the choice set. In what follows, we prove $E_{\tilde{f}'}[e^{\mu_{it2}}] = E_{\tilde{f}}[e^{\mu_{it2}}]$.

Take the second equation of (B.9) and let $\delta'_{it1} \rightarrow -\infty$. I then obtain:

$$\int \frac{e^{\delta'_{t2} + \mu_{it2}}}{1 + e^{\delta'_{t2} + \mu_{it2}}} f'_\mu(\mu_{it}) = \int \frac{e^{\delta'_{t2} + \mu_{it2}}}{1 + e^{\delta'_{t2} + \mu_{it2}}} f_\mu(\mu_{it}). \quad (\text{B.13})$$

I cancel out $e^{\delta'_{t2}}$ from the nominators on both sides of (B.13) and let $\delta'_{2t} \rightarrow -\infty$. I then obtain $\int e^{\mu_{it2}} f'_\mu(\mu_{it}) = \int e^{\mu_{it2}} f_\mu(\mu_{it})$, i.e. $E_{\tilde{f}'}[e^{\mu_{it2}}] = E_{\tilde{f}}[e^{\mu_{it2}}]$.

Step 2: $s_{\mathbf{b}}(\delta_{\mathbf{J}}; F') = s_{\mathbf{b}}(\delta_{\mathbf{J}}; F)$ for any $\mathbf{b} \in \mathbf{C}_1$. I prove this result for for the model of multiple choice of products within category. The proof is similar for the model of multiple choice of products across two categories.

Recall that the density function of $\mu_{it\mathbf{J}} = \mu_{t\mathbf{J}}(\theta_{it})$ for $\theta_{it} \sim F'$ and $\theta_{it} \sim F$ are f'_μ and f_μ , respectively. It suffices to prove that $f'_\mu = f_\mu$ almost everywhere. In the model of multiple choice of products within category in Theorem 9, plug $\Gamma' = \Gamma$ into the product-level market share function of j . I then have for any $\delta'_{t\mathbf{J}} \in \mathbb{R}^J$, $s_j(\delta'_{t\mathbf{J}}; \Gamma, F) = s_j(\delta'_{t\mathbf{J}}; \Gamma, F')$. According to the arguments in Appendix 8.13 of [Iaria and Wang \(2019a\)](#), given the product-level market share functions and Γ , one can uniquely determine the bundle-level market shares, as function of $\delta_{t\mathbf{J}}$. Because both the product-level market share functions and Γ are identified, then $s_{\mathbf{b}}(\delta_t(\Gamma); F)$, where $\delta_t(\Gamma) = (\delta_{t1}, \dots, \delta_{tJ}, (\delta_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2} + \Gamma)$, is identified as a function of $\delta_{t\mathbf{J}}$, for any $\mathbf{b} \in \mathbf{C}_1$. Consequently, the market share function of the outside option, $s_0(\delta_t(\Gamma); F)$, is identified as a function of $\delta_{t\mathbf{J}}$: for any $\delta'_{t\mathbf{J}} \in \mathbb{R}^J$

$$s_0(\delta'_{t\mathbf{J}}(\Gamma); F) = s_0(\delta'_{t\mathbf{J}}(\Gamma); F'). \quad (\text{B.14})$$

I compute the higher-order cross derivative of both sides of (B.14):

$$\begin{aligned}\frac{\partial^J s_0(\delta'_{t\mathbf{J}}(\Gamma); F)}{\partial \delta'_{t1}, \dots, \partial \delta'_{tJ}} &= \int \frac{P_\Gamma(\delta_{t1} + \mu_{it1}, \dots, \delta_{tJ} + \mu_{itJ})}{Q_\Gamma(\delta_{t1} + \mu_{it1}, \dots, \delta_{tJ} + \mu_{itJ})} f_\mu(\mu_{it}) d\mu_{it}, \\ \frac{\partial^J s_0(\delta'_{t\mathbf{J}}(\Gamma); F')}{\partial \delta'_{t1}, \dots, \partial \delta'_{tJ}} &= \int \frac{P_\Gamma(\delta_{t1} + \mu_{it1}, \dots, \delta_{tJ} + \mu_{itJ})}{Q_\Gamma(\delta_{t1} + \mu_{it1}, \dots, \delta_{tJ} + \mu_{itJ})} f'_\mu(\mu_{it}) d\mu_{it}, \\ \frac{\partial^J s_0(\delta'_{t\mathbf{J}}(\Gamma); F)}{\partial \delta'_{t1}, \dots, \partial \delta'_{tJ}} &= \frac{\partial^J s_0(\delta'_{t\mathbf{J}}(\Gamma); F')}{\partial \delta'_{t1}, \dots, \partial \delta'_{tJ}},\end{aligned}\tag{B.15}$$

where

$$Q_\Gamma(u_{it1}, \dots, u_{itJ}) = 1 + \sum_{j \in \mathbf{J}} e^{u_{itj}} + \sum_{j < j'} e^{\Gamma_{(j,j')}} e^{u_{itj} + u_{itj'}}.$$

and

$$\begin{aligned}P_\Gamma(u_{it1}, \dots, u_{itJ}) &= \sum_{S \in \mathbf{S}} A(S) \prod_{(j,j') \in S} \frac{e^{u_{it(j,j')}}}{Q_\Gamma(u_{it1}, \dots, u_{itJ})} \prod_{j \in S} \frac{\sum_{\mathbf{b}: \mathbf{b} \ni j} e^{u_{it\mathbf{b}}}}{Q_\Gamma(u_{it1}, \dots, u_{itJ})} \\ &= \prod_{j \in \mathbf{J}} e^{u_{itj}} \sum_{S \in \mathbf{S}} A(S) \prod_{(j,j') \in S} \frac{e^{\Gamma_{(j,j')}}}{Q_\Gamma(u_{it1}, \dots, u_{itJ})} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} e^{u_{itj'} + \Gamma_{(j,j')}}}{Q_\Gamma(u_{it1}, \dots, u_{itJ})}\end{aligned}\tag{B.16}$$

where S is a partition of $\{1, \dots, J\}$ with each part being at most size 2, \mathbf{S} collects all such partitions which are the results of the higher-order cross derivative $\frac{\partial^J}{\partial \delta'_{t1}, \dots, \partial \delta'_{tJ}}$, and $A(S)$ is a constant depending on the partition $S \in \mathbf{S}$. An example of S is $\{\{1\}, \{2, 5\}, \{4\}, \{3, 6\}\}$. Each term in the products of P_Γ corresponds to the choice probability of either bundle (j, j') or the product-level choice probability of product j , evaluated at $u_{it\mathbf{J}}$ and Γ , and bounded by 1. From (B.15), I obtain:

$$\int \frac{P_\Gamma((e^{\lambda_{itj}})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_{itj}})_{j \in \mathbf{J}})} [f_\mu(\lambda_{it\mathbf{J}} - \delta'_{t\mathbf{J}}) - f'_\mu(\lambda_{it\mathbf{J}} - \delta'_{t\mathbf{J}})] d\lambda_{it\mathbf{J}} = 0,\tag{B.17}$$

for any $\delta'_{t\mathbf{J}} \in \mathbb{R}^J$. I prove the following lemma:

Lemma 9.

- $\frac{P_\Gamma((e^{\lambda_{itj}})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_{itj}})_{j \in \mathbf{J}})} \in L^1(\mathbb{R}^J)$.
- The zero set of $\mathcal{F} \left(\frac{P_\Gamma((e^{\lambda_{itj}})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_{itj}})_{j \in \mathbf{J}})} \right)$ in \mathbb{R}^J is of zero Lebesgue measure.

Note that the right-hand side of (B.17) is a convolution. Because of the first statement of Lemma 9, I can apply Fourier transformation on both sides and obtain:

$$\mathcal{F} \left(\frac{P_\Gamma((e^{\lambda_{itj}})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_{itj}})_{j \in \mathbf{J}})} \right) \mathcal{F}(f_\mu - f'_\mu) = 0.$$

Applying the second statement of Lemma 9, I obtain $\mathcal{F}(f_\mu) = \mathcal{F}(f'_\mu)$ almost everywhere. Due to the continuity of characteristics functions, $\mathcal{F}(f_\mu) = \mathcal{F}(f'_\mu)$ everywhere and hence the distribution of μ_{it} is identified. In the remaining part, I prove Lemma 9.

Proof. First, we make the transformation of variables $\lambda_{\mathbf{J}}$ to $e^{\lambda_{\mathbf{J}}}$:

$$\int \left| \frac{P_{\Gamma}((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_{\Gamma}((e^{\lambda_j})_{j \in \mathbf{J}})} \right| d\lambda_{\mathbf{J}} = \int_{\mathbb{R}_+^J} \left| \frac{P_{\Gamma}(y_1, \dots, y_J)}{Q_{\Gamma}(y_1, \dots, y_J)} \frac{1}{\prod_{j=1}^J y_j} \right| dy_{\mathbf{J}}.$$

For the first statement, because of (B.16), it suffices to prove that for each $S \in \mathbf{S}$,

$$\left| \frac{1}{Q_{\Gamma}(y_1, \dots, y_J)} \prod_{(j,j') \in S} \frac{e^{\Gamma(j,j')}}{Q_{\Gamma}(y_1, \dots, y_J)} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma(j,j')}}{Q_{\Gamma}(y_1, \dots, y_J)} \right| \quad (\text{B.18})$$

is integrable in \mathbb{R}_+^J . To show this, I divide \mathbb{R}_+^J into 2^J regions: $\mathbb{R}_+^J = \times_{j=1}^J I_j$, where $I_j = (0, 1], (1, +\infty)$. Then, it is enough to prove that (B.18) is integrable in each of these regions. Without loss of generality, suppose that the region is $R_k = \{(y_1, \dots, y_J) : y_j \in (0, 1), j = 1, \dots, k; y_{j'} \geq 1, j' = k+1, \dots, J\}$. Then, for a given j , we have four cases to control:

1. $j \leq k$ and j appears in S as (j, j') .
2. $j \leq k$ and j appears in S as j .
3. $j > k$ and j appears in S as (j, j') .
4. $j > k$ and j appears in S as j .

Note that for cases 1 and 2, the corresponding terms in (B.18) can be controlled by e^{Γ_m} with $\Gamma_m = \max\{0, (\Gamma(j, j'))_{j \leq j'}\}$. For case 3, $\frac{e^{\Gamma(j, j')}}{Q_{\Gamma}(y_1, \dots, y_J)} \leq \frac{e^{\Gamma_m}}{y_j}$. For case 4,

$$\frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma(j, j')}}{Q_{\Gamma}(y_1, \dots, y_J)} \leq \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma(j, j')}}{y_j + \sum_{j' \neq j} y_j y_{j'} e^{\Gamma(j, j')}} = \frac{1}{y_j} \leq \frac{e^{\Gamma_m}}{y_j}.$$

Moreover,

$$\begin{aligned} Q_{\Gamma}(y_1, \dots, y_J) &\leq \frac{1}{\sum_{j > k} y_j + \sum_{k < j < j'} y_j y_{j'} e^{\Gamma(j, j')}} \\ &\leq \frac{2}{(J-k)(J-k+1) (\prod_{k < j < j'} e^{\Gamma(j, j')})^{\frac{2}{(J-k)(J-k+1)}} \prod_{j=k+1}^J y_j^{\frac{2}{J-k+1}}}. \end{aligned}$$

The last step is due to the inequality of arithmetic and the geometric means. Then, for all the four cases, we have:

$$\left| \frac{1}{Q_{\Gamma}(y_1, \dots, y_J)} \prod_{(j,j') \in S} \frac{e^{\Gamma(j,j')}}{Q_{\Gamma}(y_1, \dots, y_J)} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma(j,j')}}{Q_{\Gamma}(y_1, \dots, y_J)} \right| \leq A(J, k, \Gamma) \prod_{j=k+1}^J y_j^{-1 - \frac{2}{J-k+1}}, \quad (\text{B.19})$$

where $A(J, k, \Gamma) = \frac{2e^{J\Gamma_m}}{(J-k)(J-k+1) (\prod_{k < j < j'} e^{\Gamma(j, j')})^{\frac{2}{(J-k)(J-k+1)}}}$. Note that $\prod_{j=k+1}^J y_j^{-1 - \frac{2}{J-k+1}}$ is integrable

in R_k and $A(J, k, \Gamma)$ is a constant. Then,

$$\left| \frac{1}{Q_\Gamma(y_1, \dots, y_J)} \prod_{(j,j') \in S} \frac{e^{\Gamma(j,j')}}{Q_\Gamma(y_1, \dots, y_J)} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma(j,j')}}{Q_\Gamma(y_1, \dots, y_J)} \right|$$

is integrable in $\{(y_1, \dots, y_J) : y_j \in (0, 1), j = 1, \dots, k; y_{j'} \geq 1, j' = k + 1, \dots, J\}$. The proof of the first statement is completed.

To prove the second statement, according to Mityagin (2015), it suffices to show that the real (or imaginary) part of $\mathcal{F} \left(\frac{P_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})} \right)$ is non-constant real analytic function. In order to prove the real analytic property, the key is to control the higher order derivatives of $\mathcal{F} \left(\frac{P_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})} \right) (y)$:

$$\frac{\partial^L \mathcal{F} \left(\frac{P_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})} \right) (y)}{\prod_{j=1}^J \partial y_j^{l_j}} = \mathcal{F} \left(\prod_{j=1}^J (-i \lambda_j)^{l_j} \frac{P_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})} \right) (y), \quad (\text{B.20})$$

where $\sum_{j=1}^J l_j = L$ and i is the imaginary unit. I now prove that this higher order derivative can be controlled by $(\frac{J+1}{2})^L \prod_{j=1}^J l_j!$. This result will then imply that for any $y \in \mathbb{R}^J$, there exist $0 < \varepsilon < \frac{2}{J+1}$ such that for $y' \in \mathbb{R}^J$ and $|y' - y| < \varepsilon$, the Taylor expansion of $\mathcal{F} \left(\frac{P_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})} \right) (y')$ around y uniformly converges to $\mathcal{F} \left(\frac{P_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})} \right) (y')$. Consequently, $\mathcal{F} \left(\frac{P_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})} \right) (y)$ is everywhere real analytic in \mathbb{R}^J . It is not constantly zero because $\frac{P_\Gamma}{Q_\Gamma}$ is not constantly zero. In the remaining part of the proof, I prove (B.20) can be controlled by $(\frac{J+1}{2})^L \prod_{j=1}^J l_j!$.

It suffices to study $\int \left| \prod_{j=1}^J (|\lambda_j|)^{l_j} \frac{P_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j \in \mathbf{J}})} \right| d\lambda$, or equivalently,

$$\int_{\mathbb{R}_+^J} \left| \prod_{j=1}^J (|\ln y_j|)^{l_j} \frac{P_\Gamma((y_j)_{j \in \mathbf{J}})}{Q_\Gamma((y_j)_{j \in \mathbf{J}})} \frac{1}{\prod_{j=1}^J y_j} \right| dy_{\mathbf{J}}.$$

I follow the same technique as in the proof of the first statement and evaluate, for each $S \in \mathbf{S}$,

$$\left| \prod_{j=1}^J (|\ln y_j|)^{l_j} \frac{1}{Q_\Gamma(y_1, \dots, y_J)} \prod_{(j,j') \in S} \frac{e^{\Gamma(j,j')}}{Q_\Gamma(y_1, \dots, y_J)} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma(j,j')}}{Q_\Gamma(y_1, \dots, y_J)} \right| \quad (\text{B.21})$$

in each of the 2^J regions. Without loss of generality, for region R_k , using (B.19), we have:

$$\begin{aligned} & \left| \prod_{j=1}^J (|\ln y_j|)^{l_j} \frac{1}{Q_\Gamma(y_1, \dots, y_J)} \prod_{(j,j') \in S} \frac{e^{\Gamma(j,j')}}{Q_\Gamma(y_1, \dots, y_J)} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma(j,j')}}{Q_\Gamma(y_1, \dots, y_J)} \right| \\ & \leq A(J, k, \Gamma) \prod_{j=1}^k |\ln y_j|^{l_j} \prod_{j=k+1}^J |\ln y_j|^{l_j} y_j^{-1 - \frac{2}{J-k+1}}. \end{aligned}$$

Finally,

$$\begin{aligned}
& \int_0^1, \dots, \int_0^1 \int_1^\infty \dots \int_1^\infty A(J, k, \Gamma) \prod_{j=1}^k |\ln y_j|^{l_j} \prod_{j=k+1}^J |\ln y_j|^{l_j} y_j^{-1 - \frac{2}{J-k+1}} dy_{\mathbf{J}} \\
&= A(J, k, \Gamma) \left(\frac{J-k+1}{2}\right)^{L+J-k} \prod_{j=1}^J l_j! \\
&\leq A(J, k, \Gamma) \left(\frac{J+1}{2}\right)^J \left(\frac{J+1}{2}\right)^L \prod_{j=1}^J l_j!
\end{aligned}$$

Consequently, when I sum over all the integrals in the 2^J regions and all $S \in \mathbf{S}$,

$$\int_{\mathbb{R}_+^J} \left| \prod_{j=1}^J (|\ln y_j|)^{l_j} \frac{P_\Gamma((y_j)_{j \in \mathbf{J}})}{Q_\Gamma((y_j)_{j \in \mathbf{J}})} \frac{1}{\prod_{j=1}^J y_j} \right| dy_{\mathbf{J}}$$

will be bounded by $\left(\frac{J+1}{2}\right)^L \prod_{j=1}^J l_j!$ multiplied by some constant only depending on J and Γ . The proof is completed. \square

B.5 Proof of Corollary 3

In this proof, I will construct (Γ_0, F'_0) and (Γ'_0, F_0) such that $\Gamma_0 \neq \Gamma'_0$ and $F_0 \neq F'_0$, while $s_1(\cdot; \Gamma_0, F'_0) = s_1(\cdot; \Gamma'_0, F_0)$. Because $F_0 \neq F'_0$, then $s_{(1,1)}(\cdot; F_0) \neq s_{(1,1)}(\cdot; F'_0)$ in (2.8).

First, I compute the derivative of $s_1(\delta; \Gamma, F)$ with respect to δ :

$$\begin{aligned}
\frac{\partial s_1(\delta; \Gamma, F)}{\partial \delta} &= \int \frac{e^{\delta+\mu} + 4e^{2\delta+2\mu+\Gamma}}{(1 + e^{\delta+\mu} + e^{2\delta+2\mu+\Gamma})^2} dF(\mu) \\
&= \int R(\delta + \mu; \Gamma) dF(\mu),
\end{aligned} \tag{B.22}$$

where $R(x; \Gamma) = \frac{e^x + 4e^{2x+\Gamma}}{(1 + e^x + e^{2x+\Gamma})^2}$. Note that $R(\cdot; \Gamma) \in L^1(\mathbb{R})$. Define $\gamma = e^\Gamma$ and

$$V(\gamma) = \int_{\mathbb{R}} R(x; \Gamma) dx = \int_{\mathbb{R}_+} \frac{1 + 4\gamma t}{(1 + t + \gamma t^2)^2} dt$$

$V(\gamma)$ is a continuous function of $\gamma \in [0, \infty)$, with $V(0) = 1 > V(\infty) = 0$. Moreover, $\lim_{\gamma \rightarrow 0^+} \frac{dV}{d\gamma} = +\infty > 0$. As a consequence, there exist $\gamma_0 \neq \gamma'_0$ and $\gamma_0, \gamma'_0 > 0$, such that $V(\gamma_0) = V(\gamma'_0)$. Therefore, there exists $\Gamma_0 = \ln \gamma_0 > -\infty$ and $\Gamma'_0 = \ln \gamma'_0 > -\infty$, such that $\Gamma_0 \neq \Gamma'_0$ and $V_0 = \int_{\mathbb{R}} R(x; \Gamma_0) = \int_{\mathbb{R}} R(x; \Gamma'_0) dx$. Note that $\frac{R(\cdot; \Gamma_0)}{V_0}$ and $\frac{R(\cdot; \Gamma'_0)}{V_0}$ are both well-defined but different density functions. Denote the corresponding distribution functions as F_0 and F'_0 , respectively: $\frac{dF_0}{d\mu} = \frac{R(\mu; \Gamma_0)}{V_0}$ and $\frac{dF'_0}{d\mu} = \frac{R(\mu; \Gamma'_0)}{V_0}$. Based on (B.22), consider the Fourier transformation of $\frac{\partial s_1(\delta; \Gamma_0, F'_0)}{\partial \delta}$ and $\frac{\partial s_1(\delta; \Gamma'_0, F_0)}{\partial \delta}$:

$$\begin{aligned}
\mathcal{F} \left(\frac{\partial s_1(\delta; \Gamma_0, F'_0)}{\partial \delta} \right) (t) &= \mathcal{F} (R(\cdot; \Gamma_0)) (t) \mathcal{F} \left(\frac{dF'_0}{d\mu} \right) (t), \\
\mathcal{F} \left(\frac{\partial s_1(\delta; \Gamma'_0, F_0)}{\partial \delta} \right) (t) &= \mathcal{F} (R(\cdot; \Gamma'_0)) (t) \mathcal{F} \left(\frac{dF_0}{d\mu} \right) (t).
\end{aligned}$$

Then,

$$\mathcal{F} \left(\frac{\partial s_1(\delta; \Gamma_0, F'_0)}{\partial \delta} \right) (t) = \mathcal{F} \left(\frac{\partial s_1(\delta; \Gamma'_0, F_0)}{\partial \delta} \right) (t),$$

Consequently, $\frac{\partial s_1(\delta; \Gamma_0, F'_0)}{\partial \delta} = \frac{\partial s_1(\delta; \Gamma'_0, F_0)}{\partial \delta}$ for $\delta \in \mathbb{R}$. Then, $s_1(\delta; \Gamma'_0, F_0) - s_1(\delta; \Gamma_0, F'_0)$ is a constant function in \mathbb{R} . Taking $\delta = +\infty$, we obtain that this constant is zero and hence $s_1(\delta; \Gamma'_0, F_0) = s_1(\delta; \Gamma_0, F'_0)$ for $\delta \in \mathbb{R}$. The construction is completed.

B.6 Proof of Property 2

The proof directly follows from that of Lemma 8.

B.7 Construction of Product-Level Market Shares

I provide computational details of the product level market shares from the market-level sales data. I suppress t to simplify the exposition. Suppose that there are I households and the size of household $i = 1, \dots, I$ is $n_i \in \{1, \dots, N\}$. Denote by q_k the weekly per capita consumption of the relevant products of category k (breakfast cereals or milk). Then, for product j in category k , the total consumption D_{jk} , is:

$$\begin{aligned} D_{jk} &= \sum_{i=1}^I \sum_{\mathbf{b}: \mathbf{b} \ni j} \mathbf{1}\{i \text{ chooses } \mathbf{b}\} n_i q_k \\ &= q_k \sum_{n=1}^N n \sum_{\mathbf{b}: \mathbf{b} \ni j} \sum_{i=1}^I \mathbf{1}\{i \text{ chooses } \mathbf{b}, n_i = n\} \\ &= I q_k \sum_{n=1}^N n \sum_{\mathbf{b}: \mathbf{b} \ni j} \frac{\sum_{i=1}^I \mathbf{1}\{i \text{ chooses } \mathbf{b}, n_i = n\}}{I}. \end{aligned}$$

Denote by $\mathcal{J}_{\mathbf{b}}^n$ the average choice probability of bundle \mathbf{b} among households of size n . Then, when I is very large,

$$\begin{aligned} \frac{D_{jk}}{I q_k} &= \sum_{n=1}^N n \sum_{\mathbf{b}: \mathbf{b} \ni j} \frac{\sum_{i=1}^I \mathbf{1}\{i \text{ chooses } \mathbf{b}, n_i = n\}}{I} \\ &\approx \sum_{n=1}^N n \pi_n \sum_{\mathbf{b}: \mathbf{b} \ni j} \mathcal{J}_{\mathbf{b}}^n \\ &= \sum_{n=1}^N n \pi_n \mathcal{J}_j^n, \\ \frac{D_{jk}}{I \bar{N} q_k} &\approx \sum_{n=1}^N \bar{\pi}_n \mathcal{J}_j^n = \mathcal{J}_j, \end{aligned}$$

where $\bar{N} = \sum_{n=1}^N n \pi_n$ is the average household size and $\{\bar{\pi}_n\}_{n=1}^N$ is the distribution of household sizes weighted by size. Note that when computing the product-level market shares, one should use the weighted distribution $\{\bar{\pi}_n\}_{n=1}^N$ rather than $\{\pi_n\}_{n=1}^N$ to properly take into account heterogeneous consumption across households of different sizes.

Under the assumptions in section 2.6, D_{jk} is equal to the sales in lbs of product j of category k .

Moreover, the IRI dataset contains information on the set of sampled households with which we can infer the number of households I and the distribution of their demographic characteristics. Finally, for d_k , $k \in \mathbf{K}$, I use external sources: the weekly per capita consumption of breakfast cereals is 0.19 lbs and that of fluid milk is 3.4 lbs.⁴ Based on these pieces of information, we construct the observed product-level market shares s_j 's.⁵

B.8 Main Tables

Table B.1: RTE Cereal Products

Brand	Flavour	Fortification	Grain
General Mills Cheerios	Toasted	Missing	WHOLE GRAIN OAT
General Mills Cinnamon TST CR	Cinnamon Toast	12 ESSNTL VTMN&MNRL	WHOLE WHEAT AND RICE
General Mills Cinnamon TST CR	Cinnamon Toast	Missing	WHOLE WHEAT AND RICE
General Mills Honey Nut Cheer	Honey Nut	Missing	WHL GRAIN OAT & BRLY
General Mills Honey Nut Cheer	Honey Nut	Missing	WHOLE GRAIN OAT
General Mills Lucky Charms	Toasted	CALCIUM & VITAMIN D	WHOLE GRAIN OAT
General Mills Lucky Charms	Toasted	Missing	WHOLE GRAIN OAT
General Mills Multi Grain Che.	Regular	10 VITAMINS&MINERALS	MULTI GRAIN
Kashi Go Lean Crunch	Regular	Missing	MULTI GRAIN
Kellogg's Apple Jacks	Apple Cinnamon	Missing	3 GRAIN
Kellogg's Corn Flakes	Regular	Missing	CORN
Kellogg's Frosted Flakes	Regular	VITAMIN D	CORN
Kellogg's Frosted Mini Wheats	Regular	Missing	WHOLE GRAIN WHEAT
Kellogg's Raisin Bran	Regular	Missing	WHL GRN WHT WHT BRN
Kellogg's Rice Krispies	Toasted	Missing	RICE
Kellogg's Special K	Toasted	Missing	RICE
Kellogg's Special K Fruit & Yo	Regular	Missing	OAT RICE WHEAT
Kellogg's Special K Red Berrie	Regular	Missing	RICE AND WHEAT
Kellogg's Special K Vanilla AL	Regular	Missing	RICE AND WHEAT
Post Grape Nuts	Regular	Missing	WHOLE GRN WHT & BRLY
Post Honey Bunches of Oats	Honey	Missing	WHOLE GRAIN OAT
Post Honey Bunches of Oats	Honey Roasted	Missing	WHOLE GRAIN OAT
Post Raisin Bran	Regular	Missing	WHOLE GRAN WHT & BRN
Post Selects Great Grains	Regular	Missing	MULTI GRAIN
Private Label	Regular	Missing	GRANOLA

⁴See <https://hypertextbook.com/facts/2006/LauraFalci.shtml> for a collection of these reports.

⁵Another implicit assumption is that the ratio of consumption between breakfast cereals and fluid milk is the same across households of different sizes. However, it is possible that this is not true. For example, households with children may consume relatively more fluid milk.

Table B.2: Milk Products

Brand	Flavour	Fortification	Fat Content	Type of Milk
GARELICK FARMS	WHITE	VITAMIN A & D	skimmed	dairy
GARELICK FARMS	WHITE	VITAMIN A & D	low fat	dairy
GARELICK FARMS	WHITE	VITAMIN D	whole fat	dairy
GARELICK FARMS TRUMOO	CHOCOLATE	MISSING	low fat	dairy
GUIDAS	WHITE	VITAMIN A & D	skimmed	dairy
GUIDAS	WHITE	VITAMIN A & D	low fat	dairy
GUIDAS	WHITE	VITAMIN D	whole fat	dairy
HIGH LAWN FARM	WHITE	VITAMIN A & D	whole fat	dairy
HIGH LAWN FARM	WHITE	VITAMIN A & D	skimmed	dairy
HIGH LAWN FARM	WHITE	VITAMIN A & D	low fat	dairy
HOOD	WHITE	VITAMIN A C D W CLCM	skimmed	dairy
HOOD	WHITE	VITAMIN A C D W CLCM	low fat	dairy
HOOD	WHITE	VIT C D CALCIUM	whole fat	dairy
HOOD LACTAID	WHITE	VITAMIN A & D	low fat	dairy
HOOD SIMPLY SMART	WHITE	VIT A & D W/CALC&PROTN	skimmed	dairy
PRIVATE LABEL	CHOCOLATE	MISSING	low fat	dairy
PRIVATE LABEL	WHITE	VITAMIN A & D	skimmed	dairy
PRIVATE LABEL	WHITE	VITAMIN A & D W/CALC	skimmed	dairy
PRIVATE LABEL	WHITE	VITAMIN A & D	low fat	dairy
PRIVATE LABEL	WHITE	VITAMIN D	whole fat	dairy

Table B.3: Average Estimated Own- and Cross-Price Elasticities (Model II): Grain Type and Fat Content

	RTE cereals			Milk		
	uni-grain	multi-grain	granola	skimmed	low fat	whole fat
RTE cereals, uni-grain	-1.512	0.231	0.009	-0.007	-0.017	-0.008
multi-grain	0.318	-1.642	0.010	-0.007	-0.018	-0.009
granola	0.214	0.185	-1.518	-0.009	-0.019	-0.009
Milk, skimmed	-0.078	-0.061	-0.005	-0.327	0.056	0.029
low fat	-0.079	-0.061	-0.005	0.022	-0.280	0.029
whole fat	-0.078	-0.061	-0.005	0.022	0.056	-0.319

Table B.4: Average Estimated Own- and Cross-Price Elasticities (Model II): Flavours

	RTE cereals		Milk	
	unflavoured	flavoured	unflavoured	chocolate
RTE cereals, unflavoured	-1.504	0.179	-0.036	-0.0003
flavoured	0.137	-1.425	-0.028	-0.001
Milk, unflavoured	-0.071	-0.072	-0.235	0.002
chocolate	-0.036	-0.185	0.096	-0.366

Table B.5: Average Estimated Own- and Cross-Price Elasticities (Model II): Fortification and Flavours

	RTE cereals		Milk	
	unfortified	fortified	unflavoured	chocolate
RTE cereals, unfortified	-1.355	0.070	-0.032	-0.001
fortified	0.454	-1.804	-0.033	-0.001
Milk, unflavoured	-0.123	-0.019	-0.235	0.002
chocolate	-0.197	-0.023	0.096	-0.366

Table B.6: Average Estimated Own- and Cross-Price Elasticities (Model II): Brands

	RTE cereals					Milk				
	General Mills	Kashi	Kellogg's	Post	Private Label	Garelick Farms	Guidas	High Lawn Farm	Hood	Private Label
RTE cereals, General Mills	-1.556	0.009	0.188	0.052	0.008	-0.006	-0.026	-0.002	-0.002	-0.016
Kashi	0.151	-1.829	0.178	0.038	0.013	-0.007	-0.028	-0.003	-0.002	-0.020
Kellogg's	0.187	0.011	-1.572	0.038	0.011	-0.007	-0.031	-0.003	-0.002	-0.019
Post	0.222	0.011	0.176	-1.710	0.011	-0.007	-0.029	-0.003	-0.002	-0.020
Private Label	0.159	0.015	0.185	0.039	-1.518	-0.008	.	-0.003	-0.002	-0.023
Milk, Garelick Farms	-0.072	-0.006	-0.074	-0.018	-0.005	-0.356	.	0.006	0.006	0.059
Guidas	-0.029	-0.002	-0.033	-0.006	.	.	-0.189	0.013	0.002	0.015
High Lawn Farm	-0.059	-0.005	-0.064	-0.015	-0.005	0.017	0.132	-0.528	0.005	0.050
Hood	-0.058	-0.005	-0.063	-0.014	-0.005	0.017	0.126	0.007	-0.562	0.050
Private Label	-0.063	-0.005	-0.068	-0.016	-0.005	0.017	0.129	0.007	0.005	-0.273

Notes: The . refers to the situation where the products of the row and column brands are not available at the same time in any market in the data.

B.9 Identification of Product-Level Market Share Functions Using Other IVs.

In this Appendix, I develop similar identification arguments with other types of IVs. I will focus on BLP-type instruments and exogenous product characteristics.

BLP-type instruments Cost shifters are not always available to the econometrician. Moreover, the validity of Hausman-type instruments requires independence of demand shocks across markets of the same region or of the same time period. This can be violated whenever there is unobserved correlated demand shock across markets, such as national advertisement. In demand models of single products, [Berry et al. \(1995\)](#) proposed to use characteristics (and their functions) of other products in the same market as instruments. Their validity follows from the intuition that products with similar characteristics are closer substitutes. Then, “distance” in the space of product characteristics will be a good proxy of substitution among products.

Because such variables for product j , denoted as $x_{t,-j}$, are excluded from indirect utility of j , then, they can provide useful variation in price p_{tj} via the markup of product j that identifies (Γ, F) . Formally, in (B.2), for the equation of product j , one can fix $x_{tj} = x_j$ and let $x_{t,-j} = (x_{tr})_{r \neq j}$ varies in \mathbb{R}^{J-1} .

It is worth noting that, different from cost shifters, BLP instruments may not always be able to provide useful variation even though they vary exogenously. For example, if prices p_{tj} 's are not responsive to $x_{t,-j}$, then there is no variation in p_{tj} due to the variation of $x_{t,-j}$. The unresponsiveness of prices with respect to BLP instruments can occur in a *large-market* setting when the number of products increases to infinity and therefore the competition between two products becomes very weak.⁶ Asymptotically, product prices are no more functions of characteristics of other products, but only of their own characteristics. Then, BLP instruments (say, x_{tj} for product different from j) does not enter pricing functions of any other product ($p_{t,-j}$) and hence do not produce any exogenous variation in prices. In this paper, because I focus on *many-market* settings and the number of products is fixed (see Assumption 6(iii)), BLP instruments are still valid for the identification of the price coefficient, demand synergy parameters, and the distribution of the random coefficients.

B.9.1 Exogenous product characteristics

I focus on x_{tjk} , the k^{th} element in vector x_{tj} , as example and treat other product characteristics $x_{t\mathbf{J},-k}$ as fixed. Suppose that prices are generated from a linear pricing simultaneous Bertrand game under complete information with constant marginal cost c_{tj} for $j \in \mathbf{J}$. I abstract from cost shifters in c_{tj} and denote the joint density function of $(\xi_{t\mathbf{J}}, c_{t\mathbf{J}})$ by $f_{\xi,c}$. In this section, I assume the following regularity condition:

Condition 4. For any (Γ', F') and any $x_k \in \mathbf{D}_{x_k}$, there exists $M_z > 0$, such that

$$\mathbb{E} \left[\left| s_j^{-1}(\partial_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) \right| \middle| z_{t\mathbf{J}} = z \right], \mathbb{E}[|p_{t\mathbf{J}}| \mid z_{t\mathbf{J}} = z] \leq M_z,$$

⁶Assuming that the distribution of random coefficients is priorly identified, [Armstrong \(2016b\)](#) provides conditions under which BLP instruments are weak for prices and therefore invalid for the identification of price coefficient. Intuitively, because prices and BLP instruments are correlated via markups, in his large-market setting, when the number of products increases fast enough, markups converge to constants fast enough that this correlation disappears.

$$\mathbb{E} \left[\left[\frac{\partial s_{\mathbf{J}}^{-1}(s_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma', F')}{\partial s_{t\mathbf{J}}} \frac{\partial s_{\mathbf{J}}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma, F)}{\partial \delta_{t\mathbf{J}}} \frac{\partial p_{\mathbf{J}}(0, t; x_{t\mathbf{J}}^{(2)})}{\partial c_{t\mathbf{J}}} \right] \Big| z_{t\mathbf{J}} = z \right] \leq M_z,$$

$$\mathbb{E} \left[\left[\frac{\partial p_{\mathbf{J}}(0, t; x_{t\mathbf{J}}^{(2)})}{\partial c_{t\mathbf{J}}} \right] \Big| z_{t\mathbf{J}} = z \right] \leq M_z.$$

I propose the following identification result:

Theorem 17. *Suppose that Assumptions 6-8 and regularity condition 4 hold. If the following conditions hold:*

1. $z_{t\mathbf{J}}$ is independent of $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$ and $z \in \mathbb{R}^J$.
 2. $\alpha_i = \alpha \neq 0$, $\beta_{ik} = \beta_k$.
 3. Given $x^{(2)}$, $p_{t\mathbf{J}} = p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}; x^{(2)})$ is a C^1 function of $(\beta x + \eta + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$.
 4. The following condition holds:
 - (i). For any $(\Gamma', F') \neq (\Gamma, F)$, there exists $\delta'_{\mathbf{J}} \neq \delta''_{\mathbf{J}}$ such that $s_{\mathbf{J}}(\delta'_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma, F) = s_{\mathbf{J}}(\delta'_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma', F')$ and $s_{\mathbf{J}}(\delta''_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma, F) = s_{\mathbf{J}}(\delta''_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma', F')$.
 - (ii). For such feasible $\delta'_{\mathbf{J}}$ and $\delta''_{\mathbf{J}}$ in Condition (i), there exist a pair (δ'_J, δ''_J) such that $\delta'_J - \delta''_J$ and $t' - t''$ are not collinear, where t' and t'' are defined as $-\alpha p_{\mathbf{J}}(0, t^r) = \delta_{\mathbf{J}}^r$ for $t^r = t', t''$.
- If (ξ, c) is Gaussian distributed, then (α, β_k) and $s_{\mathbf{J}}(\cdot; x_{t\mathbf{J}}^{(2)}, \Gamma, F)$ are identified.
 - Suppose that the data generating process is a model of multiple choice of products across categories (see section 2.2.2). Then, under regularity condition 5, (α, β_k) is identified and $s_{\mathbf{J}}(\cdot; x_{t\mathbf{J}}^{(2)}, \Gamma, F)$ is identified.

Remark 11. Once α , β_k , and $s_{\mathbf{J}}(\cdot; x_{t\mathbf{J}}^{(2)}, \Gamma, F)$ are identified for any $x_{t\mathbf{J}}^{(2)}$, by combining the demand inverse in Theorem 7, we can recover $\delta_{t\mathbf{J}}$. Then, identification of β_{-k} and η follows from standard linear IV arguments.

As the proof of Theorem 8, the proof is proceeded in three steps.

Conditional Moment Restrictions and Convolution Equation. Differently from cost-type instruments, x_{tjk} is not excluded from demand equation. Consequently, function G defined in (B.3) also directly depends on $x_{t\mathbf{J}}$ and hence (B.3) is no more a convolution equation. To solve this problem, instead of using G in (B.3), I use its derivative with respect to $x_{t\mathbf{J}}$.

Fixing a reference point $x_{t\mathbf{J}} = (x_j)_{j \in \mathbf{J}}$ and varying $x_{t\mathbf{J}k}$ with $\Delta_k = (\Delta_{jk})_{j \in \mathbf{J}}$, we obtain:

$$\begin{aligned} & \mathbb{E}[s_{tj}^{-1}(s_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) | x_{t\mathbf{J}k} = x_k + \Delta_k] - \mathbb{E}[s_{tj}^{-1}(s_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) | x_{t\mathbf{J}k} = x_k] \\ &= -\alpha(\mathbb{E}[p_{tj} | x_{t\mathbf{J}k} = x_k + \Delta_k] - \mathbb{E}[p_{tj} | x_{t\mathbf{J}k} = x_k]) + \beta_k \Delta_{jk}, \end{aligned} \quad (\text{B.23})$$

where β_k is the coefficient for x_{tjk} . Given a Δ_k , (B.23) defined a moment restriction of $(\alpha, \beta_k, \Gamma, F)$.

Theorem 18. *Suppose that Assumptions 6-8, regularity condition 4 hold. Suppose the following conditions hold:*

1. $x_{t\mathbf{J}k}$ is independent of $(\xi_{t\mathbf{J}}, c_{t\mathbf{J}})$, and the domain of $x_{t\mathbf{J}k}$, \mathbf{D}_{x_k} , is an open set in \mathbb{R}^J .
2. $\alpha_i = \alpha \neq 0, \beta_{ik} = \beta_k$.
3. Given $x_{t\mathbf{J}}^{(2)}$, $p_{t\mathbf{J}} = p_{\mathbf{J}}(\beta x + \eta + \xi_{t\mathbf{J}}, c'_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)})$ is a C^1 function of $(\beta x + \eta + \xi_{t\mathbf{J}}, c'_{t\mathbf{J}})$.

Then, for any $x_k \in \mathbf{D}_{x_k}$, $(\alpha', \beta'_k, \Gamma', F')$ satisfies moment conditions (B.23) if and only if the following equation

$$\int H(t; \alpha', \beta'_k, \Gamma', F') \Lambda_H(t + \frac{\beta_k}{\alpha} x_k; f_{\xi, c}) dt = 0, \quad (\text{B.24})$$

holds, where

$$\begin{aligned} & H(t; \alpha', \beta'_k, \Gamma', F') \\ &= \left(\left[\frac{\partial s_{\mathbf{J}}^{-1}(s_{t\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, t); x_{t\mathbf{J}}^{(2)}, \Gamma, F); x_{t\mathbf{J}}^{(2)}, \Gamma', F')}{\partial \beta_{t\mathbf{J}}} \right] \left[\frac{\partial s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, t); x_{t\mathbf{J}}^{(2)}, \Gamma, F)}{\partial \delta_{t\mathbf{J}}} \right] - \frac{\alpha'}{\alpha} \mathbf{I} \right) \frac{\partial p_{\mathbf{J}}}{\partial c_{t\mathbf{J}}}(0, t) + \left(\frac{\alpha'}{\alpha} - \frac{\beta'_k}{\beta_k} \right) \mathbf{I}, \end{aligned}$$

and

$$\Lambda_H(\lambda; f_{\xi, c}) = \int \alpha f_{\xi, c}(-\alpha(\lambda + \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta}{\alpha} - c), c) dc.$$

Proof. Since x_{-k} is fixed, I drop this notation in this proof and also the dependence of $p_{\mathbf{J}}(\cdot)$ and $s_{\mathbf{J}}(\cdot; \Gamma', F')$ on $x_{t\mathbf{J}}^{(2)}$. For $\Delta_k \in \mathbb{R}^J$, by using Lemma 7, we obtain:

$$\begin{aligned} \mathbb{E}[p_{t\mathbf{J}}|x_{t\mathbf{J}k} = x_k + \Delta_k] &= \int p_{\mathbf{J}}(\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}} + \beta_k(x_{t\mathbf{J}k} + \Delta_k), c_{t\mathbf{J}}) f(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) \\ &= \int \left[p_{\mathbf{J}}(0, c_{t\mathbf{J}} - \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}} + \beta_k(x_{t\mathbf{J}k} + \Delta_k)}{\alpha}) + \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}} + \beta_k(x_{t\mathbf{J}k} + \Delta_k)}{\alpha} \right] f(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) \\ &= \int p_{\mathbf{J}}(0, c_{t\mathbf{J}} - \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}} + \beta_k(x_{t\mathbf{J}k} + \Delta_k)}{\alpha}) f(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \frac{\mathbb{E}[\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}}]}{\alpha} + \frac{\beta_k}{\alpha} (x_{t\mathbf{J}k} + \Delta_k) \end{aligned}$$

Then,

$$\begin{aligned} & \mathbb{E}[p_{t\mathbf{J}}|x_{t\mathbf{J}k} = x_k + \Delta_k] - \mathbb{E}[p_{t\mathbf{J}}|x_{t\mathbf{J}k} = x_k] \\ &= \int \left[p_{\mathbf{J}}(0, c_{t\mathbf{J}} - \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}} + \beta_k(x_{t\mathbf{J}k} + \Delta_k)}{\alpha}) - p_{\mathbf{J}}(0, c_{t\mathbf{J}} - \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}} + \beta_k x_{t\mathbf{J}k}}{\alpha}) \right] f(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \frac{\beta_k}{\alpha} \Delta_k. \end{aligned}$$

As $\Delta_k \rightarrow 0$, under regularity condition 4, we obtain:

$$\begin{aligned} \frac{\partial \mathbb{E}[p_{t\mathbf{J}}|x_{t\mathbf{J}k} = x_k]}{\partial x_{t\mathbf{J}k}} &= - \int \frac{\partial p_{\mathbf{J}}}{\partial c_{t\mathbf{J}}}(0, c_{t\mathbf{J}} - \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}} + \beta_k x_{t\mathbf{J}k}}{\alpha}) \frac{\beta_k}{\alpha} f(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \frac{\beta_k}{\alpha} \\ &= -\alpha \int \frac{\partial p_{\mathbf{J}}}{\partial c_{t\mathbf{J}}}(0, \lambda) \frac{\beta_k}{\alpha} f(-\alpha(\lambda + \frac{\beta_k}{\alpha} x_{t\mathbf{J}k} + \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta}{\alpha} - c_{t\mathbf{J}}), c_{t\mathbf{J}}) d(\lambda, c_{t\mathbf{J}}) + \frac{\beta_k}{\alpha} \\ &= -\frac{\beta_k}{\alpha} \int \frac{\partial p_{\mathbf{J}}}{\partial c_{t\mathbf{J}}}(0, \lambda) \Lambda_H(\lambda + \frac{\beta_k}{\alpha} x_{t\mathbf{J}k}; f) d\lambda + \frac{\beta_k}{\alpha}, \end{aligned}$$

where $\Lambda_H(\cdot)$ is defined in Theorem 18. Similarly, we compute

$$\begin{aligned} & \mathbb{E}[s_{\mathbf{J}}^{-1}(s_{t\mathbf{J}}; \Gamma', F')|x_{t\mathbf{J}k} = x_k + \Delta_k] \\ &= \int s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, c_{t\mathbf{J}} - \frac{\beta_{-k} x_{t\mathbf{J}, -k} + \eta + \xi_{t\mathbf{J}} + \beta_k(x_{t\mathbf{J}k} + \Delta_k)}{\alpha}); \Gamma, F); \Gamma', F') f(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) \end{aligned}$$

Then, as $\Delta_k \rightarrow 0$,

$$\begin{aligned} & \frac{\partial \mathbb{E}[s_j^{-1}(s_{t\mathbf{J}}; \Gamma', F') | z_{t\mathbf{J}} = z]}{\partial x_{t\mathbf{J}k}} \\ &= \beta_k \int \left[\frac{\partial s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, \lambda); \Gamma, F); \Gamma', F')}{\partial s_{t\mathbf{J}}} \right] \left[\frac{\partial s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, \lambda); \Gamma, F)}{\partial \delta_{t\mathbf{J}}} \right] \frac{\partial p_{\mathbf{J}}}{\partial c_{t\mathbf{J}}}(0, \lambda) \Lambda_H\left(\lambda + \frac{\beta_k}{\alpha} x_{t\mathbf{J}k}; f\right) d\lambda. \end{aligned}$$

Because moment conditions (B.23) holds for all $x_k \in \mathbf{D}_{x_k}$ and \mathbf{D}_{x_k} is an open set, we obtain that:

$$\begin{aligned} & \frac{\partial \mathbb{E}[s_j^{-1}(s_{t\mathbf{J}}; \Gamma', F') | x_{t\mathbf{J}k} = x_k]}{\partial x_{t\mathbf{J}k}} = -\alpha' \frac{\partial \mathbb{E}[p_{t\mathbf{J}} | x_{t\mathbf{J}k} = x_k]}{\partial x_{t\mathbf{J}k}} + \beta'_k \mathbf{I} \\ \iff & \int \left(\left[\frac{\partial s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, \lambda); \Gamma, F); \Gamma', F')}{\partial s_{t\mathbf{J}}} \frac{\partial s_{\mathbf{J}}(-\alpha p_{\mathbf{J}}(0, \lambda); \Gamma, F)}{\partial \delta_{t\mathbf{J}}} - \frac{\alpha'}{\alpha} \mathbf{I} \right] \frac{\partial p_{\mathbf{J}}}{\partial c_{t\mathbf{J}}}(0, \lambda) + \left(\frac{\alpha'}{\alpha} - \frac{\beta'_k}{\beta_k} \right) \mathbf{I} \right) \Lambda_H\left(\lambda + \frac{\beta_k}{\alpha} x_{t\mathbf{J}k}; f\right) d\lambda = 0, \end{aligned}$$

for $x_k \in \mathbf{D}_{x_k}$. The proof of the sufficiency is completed. One can reverse the arguments of the sufficiency to obtain the necessity part. \square

Unique Solution for the Convolution Equation (B.24) and Identification of Product-Level Market Share Functions. Define

$$\Theta = \{(\alpha', \beta'_k, \Gamma', F') : (\text{B.23}) \text{ hold at } (\alpha', \beta'_k, \Gamma', F') \text{ for } x_k \in \mathbf{D}_{x_k}\},$$

and

$$\Theta_H = \{(\alpha', \beta'_k, \Gamma', F') : (\text{B.24}) \text{ holds for } x_k \in \mathbf{D}_{x_k}\}$$

Theorem 18 establishes $\Theta = \Theta_H$. Define $\Theta_H^0 = \{(\alpha', \beta'_k, \Gamma', F') : H(\cdot; \alpha', \beta'_k, \Gamma', F') = 0\}$. Note that the true parameters $(\alpha, \beta_k, \Gamma, F) \in \Theta_H^0 \subset \Theta_H = \Theta$. Then, a necessary condition for the identification of $(\alpha, \beta_k, \Gamma, F)$ by moment conditions (B.23), i.e. $\Theta = \{(\alpha, \beta_k, \Gamma, F)\}$, is $\Theta_H^0 = \Theta_H$, i.e. $H = 0$ is the unique solution for convolution equation (B.24). This is the completeness of the location families generated by $\Lambda_H(\cdot; f_{\xi, w})$. Similar to Theorem 15, the next theorem characterizes the implications of this completeness:

Theorem 19. *Suppose that conditions of Theorem 18 hold.*

1. *If $\Theta = \{(\alpha, \beta_k, \Gamma, F)\}$, then $\Theta_H^0 = \Theta_H$.*

2. *Suppose the following conditions hold:*

(i). *For any $(\Gamma', F') \neq (\Gamma, F)$, there exists $\delta'_{\mathbf{J}} \neq \delta''_{\mathbf{J}}$ such that $s_{\mathbf{J}}(\delta'_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma, F) = s_{\mathbf{J}}(\delta'_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma', F')$ and $s_{\mathbf{J}}(\delta''_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma, F) = s_{\mathbf{J}}(\delta''_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma', F')$.*

(ii). *For such feasible $\delta'_{\mathbf{J}}$ and $\delta''_{\mathbf{J}}$ in Condition (i), there exist a pair $(\delta'_{\mathbf{J}}, \delta''_{\mathbf{J}})$ such that $\delta'_{\mathbf{J}} - \delta''_{\mathbf{J}}$ and $t' - t''$ are not collinear, where t' and t'' are defined as $-\alpha p_{\mathbf{J}}(0, t^r) = \delta_{\mathbf{J}}^r$ for $t^r = t', t''$.*

If $\Theta_H^0 = \Theta_H$, then α , β_k , and $s_{\mathbf{J}}(\cdot; x_{t\mathbf{J}}^{(2)}, \Gamma, F)$ are identified.

Remark 12. *While Condition (i) of Theorem 19 is the same as that in the second statement of Theorem 15, Condition (ii) further requires non-collinearity between $\delta'_{\mathbf{J}} - \delta''_{\mathbf{J}}$ and $t' - t''$, where t' and t'' satisfy $-\alpha p_{\mathbf{J}}(0, t^r) = \delta_{\mathbf{J}}^r$, $t^r = t', t''$. Typically, $p_{\mathbf{J}}(\cdot)$ is highly nonlinear and this additional requirement is satisfied.*

Proof. For the first statement, if $\Theta = \{(\alpha, \beta_k, \Gamma, F)\}$, then $\Theta_H^0 = \Theta_H = \{(\alpha, \beta_k, \Gamma, F)\}$.

For the second statement, note that if $H(\cdot; \alpha', \beta'_k, \Gamma', F') = 0$, then there exists a constant vector v such that

$$s_{\mathbf{J}}(-\alpha p_{t\mathbf{J}}(0, t); \Gamma, F) = s_{\mathbf{J}}\left(-\alpha' p_{t\mathbf{J}}(0, t) + \left(\alpha' - \frac{\beta'_k}{\beta_k} \alpha\right)t + v; \Gamma', F'\right), \quad (\text{B.25})$$

for any $t \in \mathbb{R}^J$. Similar to the arguments in the proof of Theorem 15, (B.25) holds for all $p'_{t\mathbf{J}} \in \mathbb{R}^J$:

$$s_{\mathbf{J}}(-\alpha p'_{t\mathbf{J}}; \Gamma, F) = s_{\mathbf{J}}\left(-\alpha' p'_{t\mathbf{J}} + \left(\alpha' - \frac{\beta'_k}{\beta_k} \alpha\right)t + v; \Gamma', F'\right), \quad (\text{B.26})$$

According to Condition (i) of Theorem 19, given (Γ', F') , there exist at least two $\delta'_{\mathbf{J}} \neq \delta''_{\mathbf{J}}$ such that $s_{\mathbf{J}}(\delta'_{\mathbf{J}}; \Gamma, F) = s_{\mathbf{J}}(\delta'_{\mathbf{J}}; \Gamma', F')$ and $s_{\mathbf{J}}(\delta''_{\mathbf{J}}; \Gamma, F) = s_{\mathbf{J}}(\delta''_{\mathbf{J}}; \Gamma', F')$. Then, combining these with (B.26), we obtain:

$$s_{\mathbf{J}}(\delta'_{\mathbf{J}}; \Gamma, F) = s_{\mathbf{J}}\left(\frac{\alpha'}{\alpha} \delta'_{\mathbf{J}} + \left(\alpha' - \frac{\beta'_k}{\beta_k} \alpha\right)t' + v; \Gamma', F'\right) = s_{\mathbf{J}}(\delta'_{\mathbf{J}}; \Gamma', F'),$$

$$s_{\mathbf{J}}(\delta''_{\mathbf{J}}; \Gamma, F) = s_{\mathbf{J}}\left(\frac{\alpha'}{\alpha} \delta''_{\mathbf{J}} + \left(\alpha' - \frac{\beta'_k}{\beta_k} \alpha\right)t'' + v; \Gamma', F'\right) = s_{\mathbf{J}}(\delta''_{\mathbf{J}}; \Gamma', F').$$

Because $s_{\mathbf{J}}(\cdot; \Gamma', F')$ is bijective, we obtain that

$$\begin{aligned} \frac{\alpha'}{\alpha} \delta'_{\mathbf{J}} + \left(\alpha' - \frac{\beta'_k}{\beta_k} \alpha\right)t' + v &= \delta'_{\mathbf{J}}, \\ \frac{\alpha'}{\alpha} \delta''_{\mathbf{J}} + \left(\alpha' - \frac{\beta'_k}{\beta_k} \alpha\right)t'' + v &= \delta''_{\mathbf{J}}. \end{aligned}$$

Then, $(\frac{\alpha'}{\alpha} - 1)(\delta'_{\mathbf{J}} - \delta''_{\mathbf{J}}) + (\alpha' - \frac{\beta'_k}{\beta_k} \alpha)(t' - t'') = 0$. Due to the non-collinearity of Condition (ii) in Theorem 19, we can find $(\delta'_{\mathbf{J}}, \delta''_{\mathbf{J}})$ such that $\delta'_{\mathbf{J}} - \delta''_{\mathbf{J}}$ and $t' - t''$ are not collinear. Consequently, for such a pair, we must have $\alpha = \alpha'$ and $\beta_k = \beta'_k$. Then, $v = 0$ and the identification of the product-level market share functions follows from (B.26), $v = 0$, $\alpha = \alpha'$, and $\beta_k = \beta'$. \square

Sufficient Conditions for the Completeness of Location Families. In general, depending on regularities of $H(\cdot)$ (bounded, polynomially bounded, integrable with respect to $\Lambda_H(\cdot)$, etc.), the completeness of location families can be achieved with different sufficient conditions on $\Lambda_H(\cdot)$ (and hence on $f_{\xi, w}$) and large-support condition $\mathbf{D}_{x_k} = \mathbb{R}^J$. The following theorem establishes two sets of sufficient conditions for $H(\cdot) = 0$ being the unique solution to (B.24):

Theorem 20. *Suppose that (B.24) holds for $x_k \in \mathbb{R}^J$.*

- If $f_{\xi, c}$ is Gaussian, then $\Theta_H^0 = \Theta_H$.
- Suppose that the data generating process is a model of multiple choice of products across K categories (see section 2.2.2). Then, under regularity condition 5, $\Theta_H^0 = \Theta_H$.

Proof. The proof of the first statement is the same as the first statement of Theorem 16. For the second statement, note that H defined in equation (B.24) only depends on functional $p_{\mathbf{J}}(0, t)$.⁷ Therefore, (B.24) defines a convolution equation. In this proof, I assume the following regularity conditions:

⁷Note that $\frac{\partial p_{\mathbf{J}}}{\partial c_{t\mathbf{J}}}(0, t)$ can be written as $\left[\frac{\partial c}{\partial p_{t\mathbf{J}}}(0, p_{\mathbf{J}}(0, t))\right]^{-1}$.

Condition 5.

(i). The characteristics function of Λ_H do not vanish on \mathbb{R}^J .

(ii). $\left| \frac{\partial p_{\mathbf{J}}}{\partial c_{t\mathbf{J}}}(0, t) \right| \in L^\infty(\mathbb{R}^J)$.

(iii). Suppose that (Γ, F) is such that for each category $k = 1, \dots, K$, there exists $l_k > 0$ such that

$$\left| \frac{\int (1 - \sum_{j \in \mathbf{J}_k} s_{ij.}) \prod_{j \in \mathbf{J}_k} s_{ij.} dF(\theta_i)}{(1 - \sum_{j \in \mathbf{J}_k} \delta_j.) \prod_{j \in \mathbf{J}_k} \delta_j.} \right| \geq l_k$$

uniformly for $\delta_{\mathbf{J}} \in \mathbb{R}^J$.

As argued by D'Haultfoeuille (2011), Condition 5(i), i.e., zero-freeness, is an usual assumption in deconvolution problem and many commonly used distributions satisfy this condition. Moreover, classic results on bounded completeness show that the location families generated by $\Lambda_H(\cdot; f_{\xi, w})$ is bounded complete if and only if this condition holds (see Theorem 2.1 of Mattner (1993) and Ghosh et al. (1966)). Consequently, under this zero-freeness condition, as long as $H(\cdot) \in L^\infty(\mathbb{R}^J)$ and (B.24) holds for $x_k \in \mathbb{R}^J$, we obtain $H(\cdot) = 0$ and the proof will be completed.

To prove that H is bounded, i.e. $H(\cdot) \in L^\infty(\mathbb{R}^J)$, I require Condition 5(ii). Condition 5(iii) restricts the family of (Γ, F) . This family includes many distributions with compact support (e.g., F is degenerated). In what follows, given Conditions 5(i)-(iii), I prove $H(\cdot) \in L^\infty(\mathbb{R}^J)$. It suffices to prove the following lemma:

Lemma 10.

$$\frac{\partial s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma, F); x_{t\mathbf{J}}^{(2)}, \Gamma', F')}{\partial \delta_{t\mathbf{J}}} \frac{\partial s_{\mathbf{J}}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma, F)}{\partial \delta_{t\mathbf{J}}} \in L^\infty(\mathbb{R}^J).$$

Proof. Since $x_{t\mathbf{J}}^{(2)}$ is fixed, I drop this notation in the proof. For product category k , denote the market share of the outside option in category k by $\delta_0^k = 1 - \sum_{j \in \mathbf{J}_k} \delta_j$. The key of the proof relies on the following “pseudo” demand inverse defined in each category k : for $j \in \mathbf{J}_k$,

$$\tilde{s}_{kj.}^{-1}(\delta_{\mathbf{J}_k.}) = \ln \frac{\delta_j.}{\delta_0^k},$$

$$\tilde{s}^{-1}(\delta_{\mathbf{J}_1.}, \dots, \delta_{\mathbf{J}_K.}) = \left(\ln \frac{\delta_{\mathbf{J}_1.}}{\delta_0^1}, \dots, \ln \frac{\delta_{\mathbf{J}_K.}}{\delta_0^K} \right).$$

and

$$\tilde{s}(\delta_{\mathbf{J}_1}, \dots, \delta_{\mathbf{J}_K}) = \left(\frac{e^{\delta_{\mathbf{J}_1}}}{1 + \sum_{j \in \mathbf{J}_1} e^{\delta_j}}, \dots, \frac{e^{\delta_{\mathbf{J}_K}}}{1 + \sum_{j \in \mathbf{J}_K} e^{\delta_j}} \right).$$

Note that

$$\frac{\partial s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(\delta_{t\mathbf{J}}; \Gamma, F); \Gamma', F')}{\partial \delta_{t\mathbf{J}}} \frac{\partial s_{\mathbf{J}}(\delta_{t\mathbf{J}}; \Gamma, F)}{\partial \delta_{t\mathbf{J}}} = \left[\frac{\partial \tilde{s}^{-1}(\delta_{\mathbf{J}_J.})}{\partial \delta_{\mathbf{J}_J.}} \frac{\partial s_{\mathbf{J}}(\delta'_{t\mathbf{J}}; \Gamma', F')}{\partial \delta_{t\mathbf{J}}} \right]^{-1} \left[\frac{\partial \tilde{s}^{-1}(\delta_{\mathbf{J}_J.})}{\partial \delta_{\mathbf{J}_J.}} \frac{\partial s_{\mathbf{J}}(\delta_{t\mathbf{J}}; \Gamma, F)}{\partial \delta_{t\mathbf{J}}} \right],$$

where $\delta'_{t\mathbf{J}} = s_{\mathbf{J}}^{-1}(s_{\mathbf{J}}(\delta_{t\mathbf{J}}; \Gamma, F); \Gamma', F')$. In what follows, I show that both $\left[\frac{\partial \tilde{s}^{-1}(\delta_{\mathbf{J}_J.})}{\partial \delta_{\mathbf{J}_J.}} \frac{\partial s_{\mathbf{J}}(\delta_{t\mathbf{J}}; \Gamma, F)}{\partial \delta_{t\mathbf{J}}} \right]$ and $\left[\frac{\partial \tilde{s}^{-1}(\delta_{\mathbf{J}_J.})}{\partial \delta_{\mathbf{J}_J.}} \frac{\partial s_{\mathbf{J}}(\delta'_{t\mathbf{J}}; \Gamma', F')}{\partial \delta_{t\mathbf{J}}} \right]^{-1}$ are uniformly bounded for $\delta_{t\mathbf{J}} \in \mathbb{R}^J$.

First, $\frac{\partial \bar{s}^{-1}(\mathcal{J}_{\mathbf{J}_k})}{\partial \mathcal{J}_{\mathbf{J}_k}}$ is block-wise diagonal: $\frac{\partial \bar{s}^{-1}(\mathcal{J}_{\mathbf{J}_k})}{\partial \mathcal{J}_{\mathbf{J}_k}} = \text{Diag}(D_1, \dots, D_K)$, where $D_k = \text{Diag}(\frac{1}{\mathcal{J}_j})_{j \in \mathbf{J}_k} + \frac{1}{\mathcal{J}_0^k} (1, \dots, 1)^\top (1, \dots, 1)$. Consequently, $\left[\frac{\partial \bar{s}^{-1}(\mathcal{J}_{\mathbf{J}_k})}{\partial \mathcal{J}_{\mathbf{J}_k}} \frac{\partial s_{\mathbf{J}_k}(\delta_{t\mathbf{J}_k}; \Gamma, F)}{\partial \delta_{t\mathbf{J}_k}} \right]$ (and $\left[\frac{\partial \bar{s}^{-1}(\mathcal{J}_{\mathbf{J}_k})}{\partial \mathcal{J}_{\mathbf{J}_k}} \frac{\partial s_{\mathbf{J}_k}(\delta'_{t\mathbf{J}_k}; \Gamma', F')}{\partial \delta'_{t\mathbf{J}_k}} \right]$) is block-wise diagonal:

$$\left[\frac{\partial \bar{s}^{-1}(\mathcal{J}_{\mathbf{J}_k})}{\partial \mathcal{J}_{\mathbf{J}_k}} \frac{\partial s_{\mathbf{J}_k}(\delta_{t\mathbf{J}_k}; \Gamma, F)}{\partial \delta_{t\mathbf{J}_k}} \right] = \text{Diag}(E_1, \dots, E_K),$$

where $E_k = D_k \int (\text{Diag}(s_{ij})_{j \in \mathbf{J}_k} - (s_{ij})_{j \in \mathbf{J}_k}^\top (s_{ij})_{j \in \mathbf{J}_k}) dF(\theta_i)$. Then, it suffices to prove that each block E_k is uniformly bounded to prove $\left[\frac{\partial \bar{s}^{-1}(\mathcal{J}_{\mathbf{J}_k})}{\partial \mathcal{J}_{\mathbf{J}_k}} \frac{\partial s_{\mathbf{J}_k}(\delta_{t\mathbf{J}_k}; \Gamma, F)}{\partial \delta_{t\mathbf{J}_k}} \right]$ is uniformly bounded. I will also prove that the inverse of E_k is uniformly bounded and the bound does not depend on (Γ, F) . As a consequence, $\left[\frac{\partial \bar{s}^{-1}(\mathcal{J}_{\mathbf{J}_k})}{\partial \mathcal{J}_{\mathbf{J}_k}} \frac{\partial s_{\mathbf{J}_k}(\delta'_{t\mathbf{J}_k}; \Gamma', F')}{\partial \delta'_{t\mathbf{J}_k}} \right]^{-1}$ is also uniformly bounded.

For diagonal elements in e_k^{jj} in E_k , take $j = 1$ in category k as example:

$$\begin{aligned} e_k^{11} &= \frac{1}{\mathcal{J}_1} \left(\mathcal{J}_1 - \int s_{i1}^2 dF \right) + \frac{1}{\mathcal{J}_0^k} \left(\mathcal{J}_1 - \int s_{i1} \left(\sum_{j \in \mathbf{J}_k} s_{ij} \right) dF \right) \\ &= 1 - \frac{1}{\mathcal{J}_1} \int s_{i1}^2 dF + \frac{\mathcal{J}_1 - \int (s_{i1} (1 - s_{i0}^k) dF)}{\mathcal{J}_0^k} \\ &= 1 - \frac{1}{\mathcal{J}_1} \int s_{i1}^2 dF + \frac{\int s_{i1} s_{i0}^k dF}{\mathcal{J}_0^k}. \end{aligned}$$

It is bounded by 2, with the sum of first two terms bounded by 1 and the last term by 1. Similarly, for off-diagonal elements e_k^{jr} in E_k , without loss of generality, take $j = 1, r = 2$. Then, I obtain $e_k^{12} = -\frac{\int s_{i1} s_{i2} dF}{\mathcal{J}_1} - \frac{\int s_{i2} s_{i0}^k dF}{\mathcal{J}_0^k}$. It is also bounded by 2. Therefore, E_k is uniformly bounded.

I now prove that the inverse of E_k is also uniformly bounded. I make use of the adjugate form of matrix inverse $E_k^{-1} = \frac{1}{\text{Det}(E_k)} \text{adj}(E_k)$, where $\text{adj}(E_k)$ is the adjugate matrix of E_k . Because E_k is uniformly bounded, then $\text{adj}(E_k)$ is also uniformly bounded. Then, it suffices to show that $\frac{1}{\text{Det}(E_k)}$ is uniformly bounded away from zero. Recall that $E_k = D_k \int (\text{Diag}(s_{ij})_{j \in \mathbf{J}_k} - (s_{ij})_{j \in \mathbf{J}_k}^\top (s_{ij})_{j \in \mathbf{J}_k}) dF(\theta_i)$. Then,

$$\text{Det}(E_k) = \text{Det}(D_k) \text{Det} \left(\int (\text{Diag}(s_{ij})_{j \in \mathbf{J}_k} - (s_{ij})_{j \in \mathbf{J}_k}^\top (s_{ij})_{j \in \mathbf{J}_k}) dF(\theta_i) \right).$$

Moreover, by Sylvester's determinant theorem, I obtain

$$\begin{aligned} \text{Det}(D_k) &= \frac{1}{\prod_{j \in \mathbf{J}_k} \mathcal{J}_j} \text{Det} \left(I_{J_k \times J_k} + \left(\frac{\mathcal{J}_1}{\mathcal{J}_0^k}, \dots, \frac{\mathcal{J}_{J_k}}{\mathcal{J}_0^k} \right)^\top (1, \dots, 1) \right) \\ &= \frac{1}{\prod_{j \in \mathbf{J}_k} \mathcal{J}_j} \text{Det} \left(1 + (1, \dots, 1) \left(\frac{\mathcal{J}_1}{\mathcal{J}_0^k}, \dots, \frac{\mathcal{J}_{J_k}}{\mathcal{J}_0^k} \right)^\top \right) \\ &= \frac{1}{\mathcal{J}_0^k \prod_{j \in \mathbf{J}_k} \mathcal{J}_j}. \end{aligned}$$

For the determinant of $\int (\text{Diag}(s_{ij})_{j \in \mathbf{J}_k} - (s_{ij})_{j \in \mathbf{J}_k}^\top (s_{ij})_{j \in \mathbf{J}_k}) dF(\theta_i)$, note that every square matrix inside the integral is positive-definite. Then, due to the super-additivity of determinant for positive-

definite matrices and Sylvester's determinant theorem:

$$\begin{aligned} & \text{Det} \left(\int (\text{Diag}(s_{ij})_{j \in \mathbf{J}_k} - (s_{ij})_{j \in \mathbf{J}_k}^{\text{T}} (s_{ij})_{j \in \mathbf{J}_k}) dF(\theta_i) \right) \\ & \geq \int \text{Det} \left((\text{Diag}(s_{ij})_{j \in \mathbf{J}_k} - (s_{ij})_{j \in \mathbf{J}_k}^{\text{T}} (s_{ij})_{j \in \mathbf{J}_k}) \right) dF(\theta_i) \\ & = \int s_{i0}^k \prod_{j \in \mathbf{J}_k} s_{ij} dF(\theta_i). \end{aligned}$$

Finally, combining the formula of $\text{Det}(D_k)$ and Condition 5(iii), I obtain:

$$\text{Det}(E_k) \geq \frac{\int s_{i0}^k \prod_{j \in \mathbf{J}_k} s_{ij} dF(\theta_i)}{\mathcal{J}_0^k \prod_{j \in \mathbf{J}_k} \mathcal{J}_j} \geq l_k.$$

This implies that each element of E_k^{-1} is bounded and the bound is independent of (Γ, F) . The proof is completed. \square

\square

Combining Theorems 18-20, I obtain Theorem 17.

B.10 Complementary Approach of Identification

In this appendix, I provide a complementary identification argument that applies to cost shifters, BLP instruments and exogenous product characteristics. Let $w_{t\mathbf{J}} = (x_{tj}, z_{tj})_{j \in \mathbf{J}} = (w_{tj})_{j \in \mathbf{J}} \in \mathbf{W} \subseteq \mathbb{R}^{(K+L) \times J}$ denote the collection of exogenous product characteristics x_{tj} and additional instruments z_{tj} , where \mathbf{W} denotes the support of $w_{t\mathbf{J}}$, K is the dimension of x_{tj} , and L is the dimension of z_{tj} . Moment conditions (B.2) can then be written as:

$$\mathbb{E}[s_j^{-1}(\mathcal{J}_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) | w_{t\mathbf{J}} = w] = x_j \beta - \alpha \mathbb{E}[p_{tj} | w_{t\mathbf{J}} = w] + \eta_j,$$

for $j \in \mathbf{J}$ and for any $w \in \mathbf{W}$. Define $w_0 = (\{x_{j0}^{\text{T}}, z_{j0}^{\text{T}}\}_{j=1, \dots, J}^{\text{T}}) \in \mathbf{W}$ as a reference point. For any $w \in \mathbf{W}$, we have:

$$\begin{aligned} & \mathbb{E}[s_j^{-1}(\mathcal{J}_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) | w_{t\mathbf{J}} = w] - \mathbb{E}[s_j^{-1}(\mathcal{J}_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) | w_{t\mathbf{J}} = w_0] \\ & = (x_j - x_{j0})\beta - \alpha (\mathbb{E}[p_{tj} | w_{t\mathbf{J}} = w] - \mathbb{E}[p_{tj} | w_{t\mathbf{J}} = w_0]). \\ & = (\beta^{\text{T}}, -\alpha) \begin{pmatrix} x_j - x_{j0} \\ \mathbb{E}[p_{tj} | w_{t\mathbf{J}} = w] - \mathbb{E}[p_{tj} | w_{t\mathbf{J}} = w_0] \end{pmatrix} \end{aligned}$$

and then:

$$(-1, \beta^{\text{T}}, -\alpha) \begin{pmatrix} \mathbb{E}[s_j^{-1}(\mathcal{J}_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) | w_{t\mathbf{J}} = w] - \mathbb{E}[s_j^{-1}(\mathcal{J}_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F) | w_{t\mathbf{J}} = w_0] \\ x_j - x_{j0} \\ \mathbb{E}[p_{tj} | w_{t\mathbf{J}} = w] - \mathbb{E}[p_{tj} | w_{t\mathbf{J}} = w_0] \end{pmatrix} = 0 \quad (\text{B.27})$$

Equation (B.27) provides a necessary condition for the identification of (Γ, F) : there exists some β and α , such that for any $w \in \mathbf{W}$, (Γ, F) satisfies (B.27). More in general, define

$$A(\Gamma, F; w) = \begin{pmatrix} \mathbb{E}[s_j^{-1}(\mathcal{J}_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F)|w_{t\mathbf{J}} = w] - \mathbb{E}[s_j^{-1}(\mathcal{J}_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F)|w_{t\mathbf{J}} = w_0] \\ x_j - x_{j0} \\ \mathbb{E}[p_{tj}|w_{t\mathbf{J}} = w] - \mathbb{E}[p_{tj}|w_{t\mathbf{J}} = w_0] \end{pmatrix}_{j \in \mathbf{J}} \in \mathbb{R}^{(K+2) \times J},$$

and

$$A(\Gamma, F; \mathbf{W}') = (A(\Gamma, F; w))_{w \in \mathbf{W}'} \in \mathbb{R}^{(K+2) \times (J|\mathbf{W}')}, \quad (\text{B.28})$$

for a finite subset \mathbf{W}' of \mathbf{W} . Then, the true parameters (Γ, F) satisfy that the minimal eigenvalue of the symmetric semi-positive definite matrix $A(\Gamma, F; \mathbf{W}')A(\Gamma, F; \mathbf{W}')^T \in \mathbb{R}^{(K+2) \times (K+2)}$ is zero for any finite subset \mathbf{W}' of \mathbf{W} . Let $\lambda_{\min}[P]$ denote the minimal eigenvalue of square matrix P . The following Lemma illustrates more variations in w help identification of (Γ, F) :

Lemma 11. $\lambda_{\min}[A(\Gamma', F'; \mathbf{W}_2)A(\Gamma', F'; \mathbf{W}_2)^T] \geq \lambda_{\min}[A(\Gamma', F'; \mathbf{W}_1)A(\Gamma', F'; \mathbf{W}_1)^T]$, for any (Γ', F') and $\mathbf{W}_1 \subseteq \mathbf{W}_2$.

Proof. Note that $A(\Gamma', F'; \mathbf{W}_2) = [A(\Gamma', F'; \mathbf{W}_1), A(\Gamma', F'; \mathbf{W}_2 \setminus \mathbf{W}_1)]$. Then,

$$\begin{aligned} A(\Gamma', F'; \mathbf{W}_2)A(\Gamma', F'; \mathbf{W}_2)^T &= A(\Gamma', F'; \mathbf{W}_1)A(\Gamma', F'; \mathbf{W}_1)^T \\ &\quad + A(\Gamma', F'; \mathbf{W}_2 \setminus \mathbf{W}_1)A(\Gamma', F'; \mathbf{W}_2 \setminus \mathbf{W}_1)^T. \end{aligned}$$

Because $A(\Gamma', F'; \mathbf{W}_2 \setminus \mathbf{W}_1)A(\Gamma', F'; \mathbf{W}_2 \setminus \mathbf{W}_1)^T$ is semi-positive definite, then for the unit eigenvector $v \in \mathbb{R}^{J \times 1}$ that corresponds to the minimal eigenvalue of $A(\Gamma', F'; \mathbf{W}_2)A(\Gamma', F'; \mathbf{W}_2)^T$, I obtain:

$$\begin{aligned} \lambda_{\min}[A(\Gamma', F'; \mathbf{W}_2)A(\Gamma', F'; \mathbf{W}_2)^T] &= v^T A(\Gamma', F'; \mathbf{W}_2)A(\Gamma', F'; \mathbf{W}_2)^T v \\ &\geq v^T A(\Gamma', F'; \mathbf{W}_2)A(\Gamma', F'; \mathbf{W}_2)^T v \\ &\geq \lambda_{\min}[A(\Gamma', F'; \mathbf{W}_1)A(\Gamma', F'; \mathbf{W}_1)^T]. \end{aligned}$$

□

The next theorem gives a sufficient condition for the identification of (Γ, F) :

Theorem 21. *If for any $(\Gamma', F') \neq (\Gamma, F)$, there exists a set \mathbf{W}' such that*

$$\lambda_{\min}[A(\Gamma', F'; \mathbf{W}')A(\Gamma', F'; \mathbf{W}')^T] > 0.$$

Then (Γ, F) is identified.

Lemma 11 illustrates the intuition that more variations in instruments help identification. Then, the sufficient condition in Theorem 21 requires useful variations in $w_{t\mathbf{J}}$ that can distinguish $(\Gamma', F') \neq (\Gamma, F)$ from (Γ, F) by shifting the minimal eigenvalue of $A(\Gamma', F'; \mathbf{W}')A(\Gamma', F'; \mathbf{W}')^T$ up to positive. This identification argument is not constructive. However, it does not require additional restrictions on the support of $w_{t\mathbf{J}}$.

Appendix to Chapter 3

C.1 A Micro-foundation of Assumption 1

In this section, we provide a set of assumptions that can rationalize Assumption 1. Suppose that consumers arrive randomly and the arrival process for destination d , $A_{dT}(\cdot)$, follows a non-homogeneous Poisson process:

$$\lim_{h \rightarrow 0} \frac{P(A_{dT}(t+h) - A_{dT}(t) = 1)}{h} = \lambda_{dT}(t),$$

where $\lambda_{dT}(t)$ is train-time-destination specific arrival rate.

Assumption 5. (*Consumers' arrival*) $\lambda_{dT}(t) = \xi_d \nu_T(t)$ for $(d, t) \in \{a, b\} \times [0, 1]$. Given $\nu_T(\cdot)$, $A_{dT}(\cdot)$ and $A_{bT}(\cdot)$ are independent. Also, $\xi_a = 1$.

Assumption 5 states that the temporal profile of consumers' arrival (namely, $\nu_T(\cdot)$) is the same for both destinations. This assumption, combined with our restriction on consumers' valuation, generates the stochastic consumers' demand in Assumption 1. Next, we impose the following condition on consumers' valuations. Hereafter, we let $x \wedge y$ denotes the minimum between x and y .

Assumption 6. (*Consumers' valuation*) The valuation V_t of a random consumer arriving at $t \in [0, 1]$ and intending to buy a ticket for $d \in \{a, b\}$ satisfies

$$P(V_t \geq v) = 1 \wedge \left(\frac{v}{\underline{v}_T(t)} \right)^{-\varepsilon},$$

for some $\varepsilon > 0$. Moreover, for all $t \in [0, 1]$ and all trains, $\underline{v}_T(t) \leq p_{d1T}$. Finally, the valuation of a given consumer are constant or decreasing over time.

Hence, valuations are supposed to be distributed according to a Pareto distribution, with parameters ε and $\underline{v}_T(t)$.

The final restriction in Assumption 6 is that consumers' valuation are weakly decreasing over time. Under this assumption, a consumer arriving at date t decides immediately either to buy the ticket or to leave forever. He will have no incentive to wait, since prices are increasing and his valuation is weakly decreasing over time.¹

Given the assumptions above, the demand $D_{dT}(t, t'; p_d)$ for destination $d \in \{a, b\}$ between $t \geq 0$ and $t' \in [t, 1]$ and served by train T satisfies

$$D_{dT}(t, t'; p_d) \sim \mathcal{P}(B_T(t, t') \xi_d p_d^{-\varepsilon}),$$

¹In this sense, this differs from that of Li et al. (2014), where prices may decrease over time, and therefore consumers have incentive to delay their purchases.

where $B_T(t, t') = \int_t^{t'} \underline{v}_T(u)^\varepsilon \nu_T(u) du$. Given $B_T(t, t')$, the independence of $D_{dT}(t, t'; p_d)$ across d comes from the independence between $A_{aT}(\cdot)$ and A_{bT} and the independent decision making of consumers arriving for destinations a and b .

C.2 Proof of Theorem 11

First, suppose that (3.2) holds for the stopping times satisfying Assumption 3. Because we can identify the conditional distribution of n_{bkT} , we identify the function $(p_a, p_b) \mapsto \Lambda(\ln(\xi_b) - \varepsilon \ln(p_b/p_a))$ is identified on the set $\mathcal{D} = \cup_{k=1}^{12} \text{Supp}(p_{akT}, p_{akT})$. By monotonicity of $\Lambda(\cdot)$, we then recover $\ln(\xi_b) - \varepsilon \ln(p_b/p_a)$ for $(p_a, p_b) \in \mathcal{D}$. Because the ratio $\ln(p_b/p_a)$ is not constant on \mathcal{D} , we separately identify $\ln(\xi_b)$ and ε .

We now prove (3.2). First, note that because the realization of τ_k is determined by the Poisson process before τ_k and is independent of $D_{dT}(\tau_k, \tau_{k+1}; p_{dkT})$ for $d \in \{a, b\}$, it suffices to show (3.2) if τ_k is replaced by any fixed number that we suppose equal to 0 without loss of generality. To ease the exposition, we often omit the index T hereafter and define $\lambda_d = \xi_d p_{dk}^{-\varepsilon}$, $\rho = \lambda_a / (\lambda_a + \lambda_b)$ and $b_t = \partial B_T(0, t) / \partial t$. We also introduce $D_{d, \tau_n} = D_{dT}(0, \tau_n; p_{dkT})$ for $d \in \{a, b\}$, $D_{\tau_n} = D_{a, \tau_n} + D_{b, \tau_n}$ and $\tau_n = \inf\{t > 0 : D_t \geq n\} \wedge 1$. We will show that for all $n \geq 1$,

$$D_{a, \tau_n} | D_{\tau_n}, B_T(\cdot, \cdot) \sim \text{Binomial}(D_{\tau_n}, \rho). \quad (\text{C.1})$$

Given the previous discussion and because the right-hand side of (C.1) does not depend on $B_T(\cdot, \cdot)$, (3.2) will follow from (C.1).

To prove (C.1), we introduce, for any $n \geq 1$, the hitting times $\sigma_n = \inf\{t \in [0, 1] : D_t \geq n\}$, with $\sigma_n = 2$ if $D_1 < n$. Let us also fix $\underline{t} \in (0, 1)$ and let us partition the interval $I = [\underline{t}, 1]$ into m intervals I_1, \dots, I_m of equal length $\Delta t = (1 - \underline{t})/m$. Finally, for all $c \leq n$, let

$$p_{c, n; k} = \Pr[D_{a, \sigma_n} = c | D_{\sigma_n} = n, \sigma_n \in I_k]. \quad (\text{C.2})$$

By Lemma 12, there exists (c_l, c_r) , independent of k and m , such that for all $k = 1, \dots, m$,

$$-c_l(1+n)\Delta t \leq p_{c, n; k} - \binom{n}{c} \rho^c (1-\rho)^{n-c} \leq c_r \Delta t.$$

Moreover, we have

$$\begin{aligned} \Pr[D_{a, \sigma_n} = c | D_{\sigma_n} = n, \sigma_n \in I] &= \frac{\sum_{k=1}^m \Pr[D_{a, \sigma_n} = c, D_{\sigma_n} = n, \sigma_n \in I_k]}{\sum_{k=1}^m \Pr[D_{\sigma_n} = n, \sigma_n \in I_k]} \\ &\in \left[\min_{k=1, \dots, m} p_{c, n; k}, \max_{k=1, \dots, m} p_{c, n; k} \right]. \end{aligned}$$

Consequently,

$$-c_l(1+n)\Delta t \leq \Pr[D_{a, \sigma_n} = c | D_{\sigma_n} = n, \sigma_n \in I] - \binom{n}{c} \rho^c (1-\rho)^{n-c} \leq c_r \Delta t.$$

By letting $m \rightarrow \infty$ and then let $\underline{t} \rightarrow 0$, we obtain

$$\Pr[D_{a,\sigma_n} = c | D_{\sigma_n} = n, \sigma_n \leq 1] = \binom{n}{c} \rho^c (1 - \rho)^{n-c}. \quad (\text{C.3})$$

Now, because $D_{\tau_n} = n$ if and only if $\sigma_n \leq 1$, we obtain (C.1) in this case. Further, because $D_{\tau_n} = n' < n$ if and only if $D_1 = n'$ and $\sigma_n = 2$, we have

$$\begin{aligned} \Pr[D_{a,\tau_n} = c | D_{\tau_n} = n'] &= \Pr[D_{a,1} = c | D_1 = n', \sigma_n = 2] \\ &= \Pr[D_{a,1} = c | D_1 = n'] \\ &= \binom{n'}{c} \rho^c (1 - \rho)^{n'-c}. \end{aligned}$$

Thus, (C.1) also holds when $D_{\tau_n} = n'$, $n' < n$. The result follows.

C.3 Expressions for the counterfactual revenues

In this appendix, we list the formulas for the counterfactual revenues. The proofs of these formulas can be found in section C.7 in the Online Appendix. The formulas are given conditional on X_T and for simplicity, we assume here that $C_T = C$, a constant; if not, the results should just be seen conditional on C_T . We both consider arbitrary distributions for B_T and the gamma distribution in Assumption 2. Finally, $D(q)$ denotes a random variable satisfying $D(q) \sim \mathcal{P}(q)$ and $g_{\lambda,\mu}$ is the density of the $\Gamma(\lambda, \mu)$ distribution.

C.3.1 Complete information

We display the general formulas; those under Assumption 2 are the same up to a single change, namely $\mathbb{E}[B_T^{\frac{1}{\varepsilon}} | X_T] = \exp(X_T' \beta_0 / \varepsilon) \mu_0^{-1/\varepsilon} \Gamma(\lambda_0 + 1/\varepsilon) / \Gamma(\lambda_0)$.

Uniform pricing $R_u^c = \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}} \mathbb{E}[D(q) \wedge C] \right\} [\xi_a + \xi_b]^{\frac{1}{\varepsilon}} \mathbb{E}[B_T^{\frac{1}{\varepsilon}} | X_T]$.

Full-dynamic pricing $R_f^c = \alpha_{C,f}^c [\xi_a + \xi_b]^{\frac{1}{\varepsilon}} \mathbb{E}[B_T^{\frac{1}{\varepsilon}} | X_T]$, where $\alpha_{0,f}^c = 0$ and for all $k \geq 1$, $\alpha_{k,f}^c = (\alpha_{k,f}^c - \alpha_{k-1,f}^c)^{1-\varepsilon} (1 - 1/\varepsilon)^{\varepsilon-1}$.

Stopping-time pricing $R_s^c = \alpha_{C,s}^c [\xi_a + \xi_b]^{\frac{1}{\varepsilon}} \mathbb{E}[B_T^{\frac{1}{\varepsilon}} | X_T]$, where $\alpha_{0,s}^c = 0$ and for all $k \geq 1$,

$$\alpha_{k,s}^c = \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}} (1 - e^{-q}) + \alpha_{k-1,s}^c \int_0^1 q e^{-sq} (1-s)^{\frac{1}{\varepsilon}} ds \right\}.$$

Stopping-time pricing with M fares $R_{sM}^c = \alpha_{C,sM}^c [\xi_a + \xi_b]^{\frac{1}{\varepsilon}} \mathbb{E}[B_T^{\frac{1}{\varepsilon}} | X_T]$, where $\alpha_{C,sM}^c = \max_{q>0} \alpha_{C,M}(q)$, $\alpha_{k,0}(q) = q^{-\frac{1}{\varepsilon}} \mathbb{E}[D(q) \wedge k]$ and for all $k \in \{1, \dots, C\}$,

$$\alpha_{k,m}(q) = \max \left\{ \int_0^1 q e^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m \wedge (k-1)}(q(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz, \right. \\ \left. \max_{q>0} \int_0^1 q e^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m-1}(q(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz \right\}.$$

Stopping-time pricing with M increasing fares $R_{sM}^c = \alpha_{C,sM+}^c [\xi_a + \xi_b]^{\frac{1}{\varepsilon}} \mathbb{E}[B_T^{\frac{1}{\varepsilon}} | X_T]$, where $\alpha_{C,sM+}^c = \max_{q>0} \alpha_{C,M}^+(q)$ with $\alpha_{k,0}^+(q) = \alpha_{k,0}(q)$ and

$$\alpha_{k,m}^+(q) = \max \left\{ q \int_0^1 e^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m \wedge (k-1)}^+(q(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz, \right. \\ \left. \max_{q' \in (0, q]} q' \int_0^1 e^{-q'z} \left[q'^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m-1}^+(q'(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz \right\}.$$

C.3.2 Incomplete Information

Uniform pricing $R_u^i = [\xi_a + \xi_b]^{\frac{1}{\varepsilon}} \max_{q>0} \left\{ \int_{\mathbb{R}^+} q^{-\frac{1}{\varepsilon}} \mathbb{E}[D(qz) \wedge C] f(z) dz \right\}$.

Full-dynamic pricing Under Assumption 2, $R_f^i = \alpha_{C,f}^i(\lambda_0) ((\xi_a + \xi_b) \exp(X_T \beta_0) / \mu_0)^{1/\varepsilon}$, where $\alpha_{0,f}^i(\lambda) = 0$ for any $\lambda > 0$ and for all $k \in \{1, \dots, C\}$,

$$\alpha_{k,f}^i(\lambda) = \lambda \left(1 - \frac{1}{\varepsilon} \right)^{\varepsilon-1} \left[-\alpha_{k-1,f}^i(\lambda+1) + \left(1 + \frac{1}{\lambda \varepsilon} \right) \alpha_{k,f}^i(\lambda) \right]^{1-\varepsilon}.$$

Stopping-time pricing $R_s^i = \alpha_{C,s}^i(f) [\xi_a + \xi_b]^{\frac{1}{\varepsilon}}$, where $\alpha_{0,s}^i(f) = 0$ and for any $k \in \{1, \dots, C\}$,

$$\alpha_{k,s}^i(f) = \max_{q>0} q \int_0^1 \left[q^{-1/\varepsilon} + (1-u)^{\frac{1}{\varepsilon}} \alpha_{k-1,s}^i(T(f; qu)) \right] \left[\int_0^\infty z e^{-quz} f(z) dz \right] du$$

and $T(f; q)$ is a transformation of density function f defined in Lemma 13 below. Under Assumption 2,

$$R_s^i = \alpha_{C,s}^i(\lambda_0) \left(\frac{(\xi_a + \xi_b) e^{X_T \beta_0}}{\mu_0} \right)^{\frac{1}{\varepsilon}},$$

where $\alpha_{0,s}^i(\lambda) = 0$ for $\lambda > 0$, and for all $k \in \{1, \dots, C\}$,

$$\alpha_{k,s}^i(\lambda) = \max_{q>0} q \int_0^1 \frac{\lambda}{(1+qs)^{\lambda+1}} \left[q^{-\frac{1}{\varepsilon}} + \left(\frac{1-s}{1+qs} \right)^{\frac{1}{\varepsilon}} \alpha_{k-1,s}^i(\lambda+1) \right] ds.$$

Stopping-time pricing with M fares $R_{M,s}^i(M, f) = \alpha_{sM}^i(f) [\xi_a + \xi_b]^{\frac{1}{\varepsilon}}$, where $\alpha_{sM}^i(f) = \max_{q>0} c_{C,M}(q, f)$ and for all k , $c_{k,0}(q, f) = q^{-\frac{1}{\varepsilon}} \int \mathbb{E}[D(qz) \wedge k] f(z) dz$ and

$$c_{k,m}(q, f) = \max \left\{ q \int_0^1 \int z e^{-qzu} f(z) dz \left[q^{-1/\varepsilon} + c_{k-1, m \wedge (k-1)}(q(1-u), T(f; qu)) \right. \right. \\ \left. \left. (1-u)^{\frac{1}{\varepsilon}} \right] du, \max_{q'>0} q' \int_0^1 \int z e^{-q'zu} f(z) dz \left[q'^{-1/\varepsilon} + c_{k-1, m-1}(q'(1-u), \right. \right. \\ \left. \left. T(f; q'u))(1-u)^{\frac{1}{\varepsilon}} \right] du \right\}$$

for any $m \in \{1, \dots, k\}$, T being the same transform as in the case of stopping-time pricing. Further under Assumption 2,

$$R_{sM}^i(M, \lambda_0, \mu_0) = \alpha_{sM}^i(\lambda_0) \left[\frac{(\xi_a + \xi_b) e^{X_T \beta_0}}{\mu_0} \right]^{\frac{1}{\varepsilon}},$$

where $\alpha_{sM}^i(\lambda) = \max_{q>0} c_{C,M}(q, \lambda)$ with, for all k , $c_{k,0}(q, \lambda) = q^{-\frac{1}{\varepsilon}} \int \mathbb{E}[D(qz) \wedge k] g_{\lambda,1}(z) dz$ and for all $k \in \{1, \dots, C\}$ and all $m \in \{1, \dots, k\}$,

$$c_{k,m}(q, \lambda) = \max \left\{ q \int_0^1 \frac{\lambda}{(1+qu)^{\lambda+1}} \left[q^{-\frac{1}{\varepsilon}} + c_{k-1, m \wedge (k-1)} \left(\frac{q(1-u)}{1+qu}, \lambda+1 \right) \left(\frac{1-u}{1+qu} \right)^{\frac{1}{\varepsilon}} \right] du, \right. \\ \left. \max_{q'>0} q' \int_0^1 \frac{\lambda}{(1+q'u)^{\lambda+1}} \left[q'^{-\frac{1}{\varepsilon}} + c_{k-1, m-1} \left(\frac{q'(1-u)}{1+q'u}, \lambda+1 \right) \left(\frac{1-u}{1+q'u} \right)^{\frac{1}{\varepsilon}} \right] du \right\}.$$

Stopping-time pricing with M increasing fares $R_{sM+}^i(M, f) = \alpha_{sM+}^i(C) [\xi_a + \xi_b]^{\frac{1}{\varepsilon}}$, where $\alpha_{sM+}^i = \max_{q>0} c_{C,M+}^+(q, f)$ with, for any $k \in \{0, \dots, C\}$, $c_{k,0}^+(q, f) = c_{k,0}(q, f)$ and for any $m \geq 1$,

$$c_{k,m}^+(q, f) = \max \left\{ q \int_0^1 \int z e^{-qzu} f(z) dz \left[q^{-1/\varepsilon} + c_{k-1, m \wedge (k-1)}^+(q(1-u), T(f; qu)) \right. \right. \\ \left. \left. (1-u)^{\frac{1}{\varepsilon}} \right] du, \max_{q' \in (0, q]} q' \int_0^1 \int z e^{-q'zu} f(z) dz \left[q'^{-1/\varepsilon} + c_{k-1, m-1}^+(q'(1-u), \right. \right. \\ \left. \left. T(f; q'u))(1-u)^{\frac{1}{\varepsilon}} \right] du \right\}$$

Under Assumption 2, we have

$$R_{sM+}^i(M, \lambda_0, \mu_0) = \alpha_{sM+}^i(\lambda_0) \left[\frac{(\xi_a + \xi_b) e^{X_T \beta_0}}{\mu_0} \right]^{\frac{1}{\varepsilon}},$$

where $\alpha_{sM+}^i(\lambda) = \max_{q>0} c_{C,M+}^+(q, \lambda)$ with $c_{k,m}^+(q, \lambda) = c_{k,m}^+(q, g_{\lambda,1})$ as defined above. Further, we have the following simplifications:

$$c_{k,m}^+(q, \lambda) = \max \left\{ q \int_0^1 \frac{\lambda}{(1+qu)^{\lambda+1}} \left[q^{-\frac{1}{\varepsilon}} + c_{k-1, m \wedge (k-1)}^+ \left(\frac{q(1-u)}{1+qu}, \lambda+1 \right) \left(\frac{1-u}{1+qu} \right)^{\frac{1}{\varepsilon}} \right] du, \right. \\ \left. \max_{q' \in (0, q]} q' \int_0^1 \frac{\lambda}{(1+q'u)^{\lambda+1}} \left[q'^{-\frac{1}{\varepsilon}} + c_{k-1, m-1}^+ \left(\frac{q'(1-u)}{1+q'u}, \lambda+1 \right) \left(\frac{1-u}{1+q'u} \right)^{\frac{1}{\varepsilon}} \right] du \right\}.$$

C.4 Demand Estimation with Aggregated Data

The difference between our results and those from studies relying on aggregated data comes precisely from the fact that we dispose of micro-level data. The approach based on aggregate data is likely to bias upwards the price-elasticity estimates. Average prices are endogenous, since the weights associated to each price or, equivalently, to each fare class, is fully driven by the demand. Basically, trains in high demand are likely to have a few number of seats available at a low price, resulting in a higher average price. To illustrate this point, we aggregate our micro data and estimate the corresponding price elasticities. For instance, we propose to aggregate data over fare classes at the train level, and thus to consider an average price for every train. Then we regress the logarithm of total purchases on the logarithm of this average price.

We first aggregate the data at the train and destination level. Let Q_{dT} be the total quantity of tickets purchased for destination d in the train T , $Q_{dT} = \sum_{k=1}^{K_T} n_{dkT}$. The corresponding average price \bar{p}_{dT} is given by:

$$\bar{p}_{dT} = \frac{\sum_{k=1}^{K_T} n_{dkT} p_{dkT}}{\sum_{k=1}^{K_T} n_{dkT}}.$$

We then consider a constant elasticity demand model with train fixed effects:

$$\ln(Q_{dT}) = -\varepsilon \ln(\bar{p}_T) + \delta_T + \xi_d + \nu_{dT}. \quad (\text{C.4})$$

As in our main specification, ξ_d accounts for a destination-specific component.

We then aggregate further our data at the train level, by considering $Q_T = Q_{aT} + Q_{bT}$ and defining the corresponding average price:

$$\bar{p}_T = \frac{\sum_{d \in \{a,b\}} \sum_{k=1}^{K_T} n_{dk} p_{dkT}}{\sum_{d \in \{a,b\}} \sum_{k=1}^{K_T} n_{dk}}.$$

We consider a similar model as (C.4), except that at that level of aggregation, we cannot include train and destination fixed effects. Instead, we include day of departure and route fixed effects:

$$\ln(Q_T) = -\varepsilon \ln(\bar{p}_T) + \delta_{t(T)} + \xi_{r(T)} + \nu_T, \quad (\text{C.5})$$

where $t(T)$ and $r(T)$ denote the day of departure and the route of train T . Finally, the most aggregated approach consists in aggregating these demands at a weekly or monthly level, either by train route or at the national level.

Results are given in Table C.1. The first line presents the price elasticity estimate for the less disaggregated specification. Strikingly, the estimate (-1.02) is already much higher than ours. It is close to the estimate of -0.70 obtained by [Sauvant \(2002\)](#) on SNCF aggregated data. By aggregating further at the train level, we exacerbate the bias and obtain already a positive coefficient (0.15). Aggregating further at the week or at the month level increases further the coefficient, up to 1.14 . Using data aggregated at the national level leads to somewhat lower coefficients, but still positive ones (0.14 and 0.56 for weekly and monthly data, respectively).

Table C.1: Estimated price elasticities with aggregated data

Model	Price elasticity
Train and destination level (Equation (C.4))	-1.02 (0.24)
Train level (Equation (C.5))	0.15 (0.03)
Week \times line level	0.29 (0.12)
Month \times line level	1.14 (0.40)
Week level (whole France)	0.14 (0.09)
Month (France)	0.56 (0.33)

Notes. We refer to the text for a detailed explanation of each model

C.5 Robustness check: lognormal distribution on η_T

Table C.2: Average revenues under optimal pricing strategies: robustness check

Scenarios	Estimate	95% CI
		(in thousands of €)
Observed pricing strategy	12.21	[12.06, 12.36]
Optimal uniform pricing strategy		
incomplete information, constrained prices		[9.67, 11.28]
incomplete information, unconstrained prices		[9.73, 11.35]
complete information, constrained prices		[12.33, 14.22]
complete information, unconstrained prices		[12.94, 15.10]
% of information needed to attain the observed revenue ²		[59.65%, 99.65%]
Optimal “full” dynamic pricing strategy		
complete information		[13.18, 15.38]

Notes: With “constrained prices” (resp. “unconstrained prices”), optimization is conducted over the actual price grid (resp. over all positive real numbers). Revenues are averaged over all lines. We use bootstrap (500) to compute CI for observed revenue. To compute 95% CIs of other expected revenues under different counterfactuals, we use the GMS procedure and projection method.

Online Appendix

C.6 A key lemma for Theorem 11

We use below the notation introduced in Theorem 11.

Lemma 12. *Suppose that Assumption 1 holds. Then, there exists c_l and c_r , independent of k and m , such that for all $k = 1, \dots, m$,*

$$-c_l(1+n)\Delta t \leq p_{c,n;k} - \binom{n}{c} \rho^c (1-\rho)^{n-c} \leq c_r \Delta t. \quad (\text{C.6})$$

Proof. First, observe that $\{\sigma_n \in I_k\} = \{D_{\underline{t}+(k-1)\Delta t} < n, D_{\underline{t}+k\Delta t} \geq n\}$. Then

$$\begin{aligned} & \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, \sigma_n \in I_k] \\ &= \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} < n, D_{\underline{t}+k\Delta t} \geq n] \\ &= \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{\underline{t}+k\Delta t} \geq n] \\ & \quad + \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} < n-1, D_{\underline{t}+k\Delta t} \geq n] \end{aligned} \quad (\text{C.7})$$

We first show that the second term in (C.7) is negligible, as being of order $(\Delta t)^2$. Simple algebra shows that if $U \sim \mathcal{P}(\lambda)$, then $\Pr(U \geq 2) \leq \lambda^2$. Hence,

$$\begin{aligned} & \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} < n-1, D_{\underline{t}+k\Delta t} \geq n] \\ & \leq \Pr[D_{\underline{t}+k\Delta t} - D_{\underline{t}+(k-1)\Delta t} \geq 2] \\ & \leq \left[(\lambda_a + \lambda_b) \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds \right]^2 \\ & \leq [(\lambda_a + \lambda_b) \bar{b} \Delta t]^2, \end{aligned}$$

where $\bar{b} = \sup_{t \in [0,1]} b_t$. Now, the first term in (C.7) satisfies:

$$\begin{aligned} & \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{\underline{t}+k\Delta t} \geq n] \\ &= \Pr[D_{a,\sigma_n} = n, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{\underline{t}+k\Delta t} = n] \\ & \quad + \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{\underline{t}+k\Delta t} > n], \end{aligned}$$

where the second term can be similarly controlled as above:

$$\begin{aligned} \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{\underline{t}+k\Delta t} > n] & \leq \Pr[D_{\underline{t}+k\Delta t} - D_{\underline{t}+(k-1)\Delta t} \geq 2] \\ & \leq [(\lambda_a + \lambda_b) \bar{b} \Delta t]^2. \end{aligned}$$

As a consequence,

$$\begin{aligned}
& \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{\underline{t}+k\Delta t} = n] \\
& < \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, \sigma_n \in I_k] \\
& \leq \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{\underline{t}+k\Delta t} = n] \\
& \quad + 2[(\lambda_a + \lambda_b)\bar{b}]^2(\Delta t)^2.
\end{aligned} \tag{C.8}$$

Now, we have

$$\begin{aligned}
& \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{\underline{t}+k\Delta t} = n] \\
& = \Pr[D_{a,\sigma_n} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1] \\
& \quad \Pr[D_{a,\sigma_n} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1] \\
& = \Pr[D_{a,\sigma_n} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{a,\underline{t}+(k-1)\Delta t} = c-1] \\
& \quad + \Pr[D_{a,\sigma_n} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{a,\underline{t}+(k-1)\Delta t} = c] \\
& = \Pr[D_{a,\underline{t}+k\Delta t} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{a,\underline{t}+(k-1)\Delta t} = c-1] \\
& \quad + \Pr[D_{a,\underline{t}+k\Delta t} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{a,\underline{t}+(k-1)\Delta t} = c].
\end{aligned} \tag{C.9}$$

Now, by independence between $(D_{a,t})_{t \geq 0}$ and $(D_{b,t})_{t \geq 0}$, and independence between $D_{d,t+s} - D_{d,t}$ and $D_{d,t}$ for all $s > 0$ and $d \in \{a, b\}$,

$$\begin{aligned}
& \Pr[D_{a,\underline{t}+k\Delta t} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{a,\underline{t}+(k-1)\Delta t} = c-1] \\
& = \Pr[D_{a,\underline{t}+(k-1)\Delta t} = c-1] \Pr[D_{a,\underline{t}+k\Delta t} = c | D_{a,\underline{t}+(k-1)\Delta t} = c-1] \\
& \quad \times \Pr[D_{b,\underline{t}+(k-1)\Delta t} = n-c] \Pr[D_{b,\underline{t}+k\Delta t} = n-c | D_{b,\underline{t}+(k-1)\Delta t} = n-c] \\
& = \frac{\lambda_a^c \lambda_b^{n-c} \left(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds \right)^{n-1}}{(n-c)!(c-1)!} \exp \left\{ -(\lambda_a + \lambda_b) \int_0^{\underline{t}+k\Delta t} b_s ds \right\} \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds.
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \Pr[D_{a,\underline{t}+k\Delta t} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1, D_{a,\underline{t}+(k-1)\Delta t} = c] \\
& = \frac{\lambda_a^c \lambda_b^{n-c} \left(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds \right)^{n-1}}{(n-c-1)!c!} \exp \left\{ -(\lambda_a + \lambda_b) \int_0^{\underline{t}+k\Delta t} b_s ds \right\} \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds.
\end{aligned}$$

By plugging the last two equalities into (C.9), we obtain

$$\begin{aligned}
& \Pr[D_{a,\sigma_n} = c, D_{\underline{t}+k\Delta t} = n, D_{\underline{t}+(k-1)\Delta t} = n-1] \\
& = \frac{n\lambda_a^c \lambda_b^{n-c} \left(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds \right)^{n-1}}{(n-c)!c!} \exp \left\{ -(\lambda_a + \lambda_b) \int_0^{\underline{t}+k\Delta t} b_s ds \right\} (\lambda_a + \lambda_b) \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds.
\end{aligned}$$

Inequality (C.8) then becomes

$$\begin{aligned}
& \frac{n\lambda_a^c \lambda_b^{n-c} (\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{(n-c)!c!} \exp\{-(\lambda_a + \lambda_b) \int_0^{\underline{t}+k\Delta t} b_s ds\} (\lambda_a + \lambda_b) \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds \\
& < P(D_{a,\sigma_n} = c, D_{\sigma_n} = n, \sigma_n \in I_k) \\
& \leq \frac{n\lambda_a^c \lambda_b^{n-c} (\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{(n-c)!c!} \exp\{-(\lambda_a + \lambda_b) \int_0^{\underline{t}+k\Delta t} b_s ds\} (\lambda_a + \lambda_b) \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds \\
& \quad + 2[(\lambda_a + \lambda_b)\bar{b}]^2 (\Delta t)^2.
\end{aligned} \tag{C.10}$$

By summing (C.10) over $c = 0, 1, \dots, n$, we obtain

$$\begin{aligned}
& \frac{n(\lambda_a + \lambda_b)^n (\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{n!} \exp\{-(\lambda_a + \lambda_b) \int_0^{\underline{t}+k\Delta t} b_s ds\} (\lambda_a + \lambda_b) \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds \\
& < P(D_{\sigma_n} = n, \sigma_n \in I_k) \\
& \leq \frac{n(\lambda_a + \lambda_b)^n (\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{n!} \exp\{-(\lambda_a + \lambda_b) \int_0^{\underline{t}+k\Delta t} b_s ds\} (\lambda_a + \lambda_b) \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds. \\
& \quad + 2(n+1)[(\lambda_a + \lambda_b)\bar{b}]^2 (\Delta t)^2.
\end{aligned} \tag{C.11}$$

By combining (C.10), (C.11), and (C.2), we obtain the following inequalities:

$$-c_{l,k}(1+n)(\Delta t)^2 \leq p_{c,n;k} - \binom{n}{c} \rho^c (1-\rho)^{n-c} \leq c_{r,k}(\Delta t)^2,$$

where

$$\begin{aligned}
c_{r,k} &= \frac{2(n+1)[(\lambda_a + \lambda_b)\bar{b}]^2}{\frac{n(\lambda_a + \lambda_b)^n (\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{n!} \exp\{-(\lambda_a + \lambda_b) \int_0^{\underline{t}+k\Delta t} b_s ds\} (\lambda_a + \lambda_b) \int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds}, \\
c_{l,k} &= c_{r,k} \binom{n}{c} \rho^c (1-\rho)^{n-c}.
\end{aligned}$$

Finally, note that $c_{r,k}\Delta t \leq c_r$ where

$$c_r = \frac{2(n+1)[(\lambda_a + \lambda_b)\bar{b}]^2 \exp\{(\lambda_a + \lambda_b) \int_0^1 b_s ds\}}{\frac{n(\lambda_a + \lambda_b)^n (\int_0^{\underline{t}} b_s ds)^{n-1}}{n!} (\lambda_a + \lambda_b) \inf_{s \in I} b_s}.$$

Moreover, c_r does not depend on k and m . Finally, defining $c_l = c_r \binom{n}{c} \rho^c (1-\rho)^{n-c}$, c_l does not depend on k and m either, and (C.6) holds for all $k = 1, \dots, m$. □

C.7 Proof of Theorem 10

We show the formulas in Appendix C.3, which also proves the theorem. Given the numerous formulas, the proof is long but two key properties that hold in all cases are worth mentioning. First, because the demands for the two destinations are independent (Assumption 1) and the revenue management

is implemented at the train rather than at the train-destination level (Assumption 3), revenues only depend on the total demand at the train level and on a weighted average of the two prices corresponding to the two destinations. Second, as a consequence of the first point, the optimal prices for both destinations can be proved to be equal. We then have only one control variable, i.e. the price for both destinations, in the Bellman equations. Finally, as in Appendix C.3, we always reason (implicitly) conditional on (X_T, C_T) and let $C = C_T$ and $D(q) \sim \mathcal{P}(q)$. Additionally, we let $\mathbf{p} = (p_a, p_b)$ and define $D_T(t, t'; \mathbf{p}) = D_{aT}(t, t'; p_a) + D_{bT}(t, t'; p_b)$.

C.7.1 Complete Information

Uniform pricing Given B_T , the revenue under uniform prices $\mathbf{p} = (p_a, p_b)$ for destinations a and b is

$$R_u^c(\mathbf{p}, B_T) = \mathbb{E}[p_a D_{aT}(0, \tau_C \wedge 1; p_a) + p_b D_{bT}(0, \tau_C \wedge 1; p_b) | B_T], \quad (\text{C.12})$$

where $\tau_C = \inf\{t : D_{aT}(0, t; p_a) + D_{bT}(0, t; p_b) \geq C\}$ is the stopping time of selling out all C seats. Then, from (C.12), we obtain

$$\begin{aligned} R_u^c(\mathbf{p}, B_T) &= \mathbb{E}[\mathbb{E}[p_a D_{aT}(0, \tau_C \wedge 1; p_a) + p_b D_{bT}(0, \tau_C \wedge 1; p_b) | B_T, D_T(0, \tau_C \wedge 1; \mathbf{p})] | B_T] \\ &= \mathbb{E}[p_a D_T(0, \tau_C \wedge 1; \mathbf{p}) + (p_b - p_a) \mathbb{E}[D_{bT}(0, \tau_C \wedge 1; p_b) | B_T, D_T(0, \tau_C \wedge 1; \mathbf{p})] | B_T]. \end{aligned}$$

Moreover, by Equation (C.1),

$$D_{bT}(0, \tau_C \wedge 1; p_b) | B_T, D_T(0, \tau_C \wedge 1; \mathbf{p}) \sim \text{Binomial} \left(D_T(0, \tau_C \wedge 1; \mathbf{p}), \frac{\xi_b p_b^{-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} \right).$$

Consequently, $\mathbb{E}[D_{bT}(0, \tau_C \wedge 1; p_b) | B_T, D_T(0, \tau_C \wedge 1; \mathbf{p})] = \frac{\xi_b p_b^{-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} D_T(0, \tau_C \wedge 1; \mathbf{p})$. Moreover, $D_T(0, \tau_C \wedge 1; \mathbf{p}) = D_T(0, 1; \mathbf{p}) \wedge C$. Then,

$$\begin{aligned} &\mathbb{E}[p_a D_{aT}(0, \tau_C \wedge 1; p_a) + p_b D_{bT}(0, \tau_C \wedge 1; p_b) | B_T, D_T(0, \tau_C \wedge 1; \mathbf{p})] \\ &= \frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} D_T(0, 1; \mathbf{p}) \wedge C. \end{aligned}$$

Hence,

$$R_u^c(\mathbf{p}, B_T) = \frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} \mathbb{E}[D_T(0, 1; \mathbf{p}) \wedge C | B_T],$$

Hence, given $\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}$, the maximum of $R_u^c(\mathbf{p}, B_T)$ is determined by and increasing with respect to $\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}$. Therefore, it suffices to maximize $\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}$ given $\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}$. The unique solution to this problem satisfies $p_a = p_b$. Then,

$$\begin{aligned} R_u^c(B_T) &= \max_{p>0} R_u^c(p, p, B_T) \\ &= \max_{p>0} p \mathbb{E}[D((\xi_a + \xi_b)p^{-\varepsilon} B_T) \wedge C | B_T]. \end{aligned}$$

We obtain the result by defining $q = (\xi_a + \xi_b)p^{-\varepsilon} B_T$ and integrate over B_T .

Full dynamic pricing. Denote by $V_k(t, \mathbf{p})$ the expected revenue when there remains k vacant seats before the departure and the current seat is priced at $\mathbf{p} = (p_a, p_b)$ at time $1 - t$. Let us denote its optimal value by $V_k^*(t)$. Define $b_T(t) = \lim_{t' \downarrow t} B_T(t, t') / (t' - t)$. From $1 - t$ to $1 - t + \Delta t$, the probability of selling one seat is $b_T(1 - t)(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon})\Delta t + o(\Delta t)$ and generates $\frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}}$ revenue if one seat is sold. With probability $o(\Delta t)$, more than one seats are sold. Then, following Gallego and Van Ryzin (1994) (Section 2.2.1 on page 1004), we have:

$$V_k^*(t) = \max_{p_a, p_b > 0} \left\{ b_T(1 - t)(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon})\Delta t \left(\frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} + V_{k-1}^*(t - \Delta t) \right) + [1 - b_T(1 - t)(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon})\Delta t] V_k^*(t - \Delta t) + o(\Delta t) \right\}. \quad (\text{C.13})$$

Letting $\Delta t \rightarrow 0$, this equation shows that V_k^* is continuous. Further, by considering $(V_k^*(t) - V_k^*(t - \Delta t)) / \Delta t$ and letting $\Delta t \rightarrow 0$, we obtain that V_k^* is differentiable, with

$$V_k^{*'}(t) = \max_{p_a, p_b > 0} b_T(1 - t)(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) \left[\frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} + V_{k-1}^*(t) - V_k^*(t) \right].$$

The maximum on the right-hand side is obtained with $p_a = p_b$. As a result,

$$V_k^{*'}(t) = \max_{p > 0} b_T(1 - t)(\xi_a + \xi_b) p^{-\varepsilon} [p + V_{k-1}^*(t) - V_k^*(t)], \quad (\text{C.14})$$

with boundary conditions $V_k^*(0) = 0$ for any $k = 1, \dots, C$ and $V^*(t, 0) = 0$ for any $t \in [0, 1]$. As a consequence, the optimal price p_{tk}^* can be obtained from the first-order condition of the right-hand side of (C.14):

$$p_{tk}^* = \frac{\varepsilon}{\varepsilon - 1} [V_k^*(t) - V_{k-1}^*(t)]. \quad (\text{C.15})$$

By plugging p_{tk}^* into (C.14) and using $B_T(t, 1) = \int_t^1 b_T(s) ds$, we obtain:

$$V_k^{*'}(t) = \partial_1 B_T(1 - t, 1) \frac{\xi_a + \xi_b}{\varepsilon - 1} \left(1 - \frac{1}{\varepsilon} \right)^\varepsilon [V_k^*(t) - V_{k-1}^*(t)]^{1-\varepsilon}, \quad (\text{C.16})$$

where $\partial_j B_T$ denotes the derivative of B_T with respect to its j -th argument. We now prove by induction on k that

$$V_k^*(t) = \alpha_{k,f}^c [(\xi_a + \xi_b) B_T(1 - t, 1)]^{\frac{1}{\varepsilon}} \quad (\text{C.17})$$

for all $k \in \{0, \dots, C\}$, with $\alpha_f^c(0) = 0$ and $\alpha_{k,f}^c = (\alpha_{k,f}^c - \alpha_{k-1,f}^c)^{1-\varepsilon} (1 - \frac{1}{\varepsilon})^{\varepsilon-1}$.

The result holds for $k = 0$ since $V_0^*(t) = 0$. Next, suppose that (C.17) holds for $k - 1 \geq 0$ and let us show that the result holds for k . By plugging this solution for $k - 1$ into the differential equation (C.14), we obtain:

$$V_k^{*'}(t) = \partial_1 B_T(1 - t, 1) \frac{\xi_a + \xi_b}{\varepsilon - 1} \left(1 - \frac{1}{\varepsilon} \right)^\varepsilon \left[V_k^*(t) - \alpha_{k-1,f}^c [(\xi_a + \xi_b) B_T(1 - t, 1)]^{\frac{1}{\varepsilon}} \right]^{1-\varepsilon}, \quad (\text{C.18})$$

with $V_k^*(0) = 0$. We can check that $V_k^*(t) = \alpha_{k,f}^c [(\xi_a + \xi_b) B_T(1 - t, 1)]^{1/\varepsilon}$ is a solution to (C.18).

To show uniqueness, let $\phi(v, z) = \frac{1}{\varepsilon - 1} (1 - \frac{1}{\varepsilon})^\varepsilon [v - \alpha_{k-1,f}^c z^{1/\varepsilon}]^{1-\varepsilon}$. Consider the diffeomorphism

$z(t) = (\xi_a + \xi_b)B_T(1-t, 1)$ and define $\bar{V}_k^*(z) = V_k^*(t(z))$. Then, (C.18) can be written as

$$\bar{V}_k^{*'}(z) = \phi(\bar{V}_k^*(z), z), \quad (\text{C.19})$$

with $\bar{V}_k^*(0) = 0$. It is enough to prove that \bar{V}_k^* is the unique solution of (C.19) and we prove this by contradiction. Suppose that there is another differentiable solution $\tilde{V}_k(\cdot)$ different from $\bar{V}_k^*(z) = \alpha_{k,f}^c z^{1/\varepsilon}$. Without loss of generality, $\tilde{V}_k(z_0) > \bar{V}_k^*(z_0)$ for some $z_0 > 0$. Because $\tilde{V}_k(0) = \bar{V}_k^*(0) = 0$, then $z_m = \sup\{z \leq z_0 : \tilde{V}_k(z) \leq \bar{V}_k^*(z)\}$ exists and $z_m < z_0$. Moreover, $\tilde{V}_k(z_m) = \bar{V}_k^*(z_m)$. Then, (C.19) implies the contradiction

$$0 < \tilde{V}_k(z_0) - \bar{V}_k^*(z_0) = \int_{z_m}^{z_0} [\phi(\tilde{V}_k(z), z) - \phi(\bar{V}_k^*(z), z)] dz \leq 0,$$

where the second inequality follows from the fact that ϕ is a decreasing function of z and $\tilde{V}_k(s) > \bar{V}_k^*(s)$ for all $s \in (z_m, z_0]$. Finally, we conclude that $\bar{V}_k^*(\cdot)$ is the unique solution. Hence, the result holds for k , and (C.17) holds. By taking $t = 1, k = C$ and integrating over B_T , we obtain the formula in Section C.3.

Stopping-time pricing. Denote by $V_k(t, \mathbf{p})$ the expected optimal revenue at time $1-t$ when pricing the next seat at \mathbf{p} and with k remaining seats. In this scenario, prices do not change until the next seat is sold. Define $\tau_{1-t, \mathbf{p}} = \inf\{s > 0 : D_T(1-t, 1-t+s; \mathbf{p}) \geq 1\}$. Then,

$$\begin{aligned} \Pr[\tau_{1-t, \mathbf{p}} > s] &= \Pr[D(1-t, 1-t+s; \mathbf{p}) = 0] \\ &= \exp\{-B_T(1-t, 1-t+s)(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon})\}, \end{aligned}$$

and the density of $\tau_{1-t, \mathbf{p}}$ is

$$f_{\tau_{1-t, \mathbf{p}}}(s) = (\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) \partial_2 B_T(1-t, 1-t+s) e^{-B_T(1-t, 1-t+s)(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon})}. \quad (\text{C.20})$$

Then, the Bellman equation is

$$\begin{aligned} V_k(t, \mathbf{p}) &= \mathbb{E} \left[\mathbf{1}_{\tau_{1-t, \mathbf{p}} < t} \left(\frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} + V_{k-1}^*(t - \tau_{1-t, \mathbf{p}}) \right) \right] \\ &= \int_0^t f_{\tau_{1-t, \mathbf{p}}}(s) \left(\frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} + V_{k-1}^*(t-s) \right) ds \\ &= \int_0^t (\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) \partial_2 B_T(1-t, 1-t+s) e^{-B_T(1-t, 1-t+s)(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon})} \\ &\quad \times \left(\frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} + V_{k-1}^*(t-s) \right) ds. \end{aligned} \quad (\text{C.21})$$

Given $\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}$, $V_k(t, \mathbf{p})$ is maximized if and only if $\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}$ is maximized. Again, this implies that $p_a = p_b$. Consequently, $\max_{p_a, p_b > 0} V_k(t, \mathbf{p}) = \max_{p > 0} V_k(t, p, p)$. Let $V_k^*(t) = \max_{p > 0} V_k(t, p, p)$. We now show by induction that

$$V_k^*(t) = \alpha_{k,s}^c [(\xi_a + \xi_b)B_T(1-t, 1)]^{\frac{1}{\varepsilon}}, \quad (\text{C.22})$$

where $\alpha_s^c(0) = 0$ and

$$\alpha_{k,s}^c = \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}}(1 - e^{-q}) + \alpha_{k-1,s}^c \int_0^1 qe^{-sq}(1-s)^{\frac{1}{\varepsilon}} ds \right\}.$$

The result holds for $k = 0$ since $V_0^*(1-t) = 0$. Now, suppose that (C.22) is true for $k-1 \geq 0$. First, by (C.21) and the fact that $p_a = p_b = p$ at optimum, we have

$$V_k(t, p, p) = \int_0^t (\xi_a + \xi_b) p^{-\varepsilon} \partial_2 B_T(1-t, 1-t+s) e^{-B_T(1-t, 1-t+s)(\xi_a + \xi_b) p^{-\varepsilon}} [p + V_{k-1}^*(t-s)] ds \quad (\text{C.23})$$

By using the change of variable $z = B_T(1-t, 1-t+s)/B_T(1-t, 1)$ and applying (C.22) for $V_{k-1}^*(t)$ in Equation (C.23), we get

$$\begin{aligned} V_k(t, p, p) &= \int_0^1 (\xi_a + \xi_b) B_T(1-t, 1) p^{-\varepsilon} e^{-B_T(1-t, 1)(\xi_a + \xi_b) p^{-\varepsilon} z} \\ &\quad \left(p + [(\xi_a + \xi_b) B_T(1-t, 1)(1-z)]^{\frac{1}{\varepsilon}} \alpha_{k-1,s}^c \right) dz \\ &= [(\xi_a + \xi_b) B_T(1-t, 1)]^{\frac{1}{\varepsilon}} \left(q^{-\frac{1}{\varepsilon}}(1 - e^{-q}) + \alpha_{k-1,s}^c \int_0^1 qe^{-qz}(1-z)^{\frac{1}{\varepsilon}} dz \right), \end{aligned}$$

where $q = (\xi_a + \xi_b) B_T(1-t, 1) p^{-\varepsilon}$. As a consequence,

$$\begin{aligned} V_k^*(t) &= \max_{p>0} V_k(t, p, p) \\ &= [(\xi_a + \xi_b) B_T(1-t, 1)]^{\frac{1}{\varepsilon}} \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}}(1 - e^{-q}) + \alpha_{k-1,s}^c \int_0^1 qe^{-qz}(1-z)^{\frac{1}{\varepsilon}} dz \right\} \\ &= \alpha_{k,s}^c [(\xi_a + \xi_b) B_T(1-t, 1)]^{\frac{1}{\varepsilon}}, \end{aligned}$$

and (C.22) is true for k . Thus, (C.22) holds for all $k \in \{0, \dots, C\}$. Finally, by taking $t = 1, k = C$ and the expectation with respect to $B_T|X_T$, we obtain the expression in Appendix C.3.

Stopping-time pricing with M fares. As before, one can prove that the optimal prices should be the same for intermediate and final destinations for a given train. As a result, we have $p = p_a = p_b$. Then, let us denote by $V_k(0; t, p, m)$ (resp. $V_k(1; t, p, m)$) the expected revenue of the firm at time $1-t$, with a current price p , a remaining capacity k and a remaining number of fares m , if it decides to keep the same price p (resp. to choose a new price). Then, we have the following Bellman equations:

$$\begin{cases} V_k(1; t, p, m) = \max_{p'>0} \int_0^t f_{\tau_{1-t}, p', p'}(s) [p' + V_{k-1}^*(t-s, p', m-1)] ds, \\ V_k(0; t, p, m) = \int_0^t f_{\tau_{1-t}, p, p}(s) [p + V_{k-1}^*(t-s, p, m)] ds, \\ V_k^*(t, p, m) = \max_{d \in \{0,1\}} V_k(d; t, p, m), \end{cases} \quad (\text{C.24})$$

with initial conditions $V_0^*(t, p, m) = 0$. We show by induction on k that for all $(k, m) \in \{0, \dots, C\} \times \mathbb{N}$,

$$V_k^*(t, p, m) = \alpha_{k,m}(q(t, p)) [(\xi_a + \xi_b)B_T(1-t, 1)]^{\frac{1}{\varepsilon}}, \quad (\text{C.25})$$

where $q(t, p) = p^{-\varepsilon}(\xi_a + \xi_b)B_T(1-t, 1)$, $\alpha_{k,0}(q) = q^{-\frac{1}{\varepsilon}}\mathbb{E}[D(q) \wedge k]$ and for $m \geq 1$,

$$\alpha_{k,m}(q) = \max \left\{ q \int_0^1 e^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m \wedge (k-1)}(q(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz, \right. \\ \left. \max_{q' > 0} q' \int_0^1 e^{-q'z} \left[q'^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m-1}(q'(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz \right\}.$$

Because for any $m \geq k$ and $d \in \{0, 1\}$, we have $V_k(d; t, p, m) = V_k(d; t, p, k)$, it suffices to prove the result for $m \leq k$. The result holds for $k = m = 0$ since $V_0^*(t, p, m) = 0$. Now, suppose that (C.25) holds for $k-1 \geq 0$ and all $m \leq k-1$. If $m = 0$, the price cannot be changed anymore, so $V_k^*(t, p, m)$ is simply the revenue with price p from $1-t$ to 1 , and (C.25) holds.

If $m \geq 1$, we have, by Equations (C.20), (C.24), the change of variable $z = B_T(1-t, 1-t+s)/B_T(1-t, 1)$ and the induction hypothesis,

$$\begin{aligned} & V_k(0; t, p, m) \\ &= \int_0^t f_{\tau_{1-t, p, p}}(s) [p + V_{k-1}^*(t-s, p, m \wedge (k-1))] ds \\ &= \int_0^t (\xi_a + \xi_b) p^{-\varepsilon} \partial_2 B_T(1-t, 1-t+s) e^{-(\xi_a + \xi_b) p^{-\varepsilon} B_T(1-t, 1-t+s)} \\ & \quad \left[p + \alpha_{k-1, m \wedge (k-1)} ((\xi_a + \xi_b) B_T(1-t+s, 1) p^{-\varepsilon}) [(\xi_a + \xi_b) B_T(1-t+s, 1)]^{\frac{1}{\varepsilon}} \right] ds \\ &= \int_0^1 (\xi_a + \xi_b) p^{-\varepsilon} B_T(1-t, 1) e^{-(\xi_a + \xi_b) p^{-\varepsilon} B_T(1-t, 1) z} \\ & \quad \left[p + \alpha_{k-1, m \wedge (k-1)} ((\xi_a + \xi_b) B_T(1-t, 1) p^{-\varepsilon} (1-z)) [(\xi_a + \xi_b) B_T(1-t, 1)]^{\frac{1}{\varepsilon}} (1-z)^{\frac{1}{\varepsilon}} \right] dz \\ &= [(\xi_a + \xi_b) B_T(1-t, 1)]^{\frac{1}{\varepsilon}} \\ & \quad \times \int_0^1 q(t, p) e^{-q(t, p) z} \left[q(t, p)^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m \wedge (k-1)}(q(t, p)(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz, \end{aligned} \quad (\text{C.26})$$

With the same reasoning, we also obtain

$$\begin{aligned} & V_k(1; t, p, m) \\ &= \max_{p' > 0} \int_0^t f_{\tau_{1-t, p', p'}}(s) [p' + V_{k-1}^*(t-s, p', m-1)] ds \\ &= [(\xi_a + \xi_b) B_T(1-t, 1)]^{\frac{1}{\varepsilon}} \max_{q > 0} \int_0^1 q e^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m-1}(q(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz. \end{aligned}$$

Then,

$$\begin{aligned} V_k^*(t, p, m) &= \max_{d \in \{0, 1\}} V_k(d; t, p, m) \\ &= \alpha_{k,m}(q(t, p)) [(\xi_a + \xi_b) B_T(1-t, 1)]^{\frac{1}{\varepsilon}}, \end{aligned}$$

Thus, (C.25) holds for k , and hence for all $k \in \{0, \dots, C\}$. By setting $t = 0$ and optimizing $V_k^*(t, p, m)$

over p (or equivalently over $q(t, p)$) and taking the expectation with respect to $B_T|X_T$, we obtain the desired expression in Appendix C.3.

Stopping-time pricing with M increasing prices. The reasoning is very similar to the previous case. The only change in (C.24) is in the formula of $V_k(1; t, p, m)$: the maximization is now over $p' \geq p$ rather than $p' \geq 0$, since the new price has to be higher than the current one. Then, following a similar strategy by induction, we get

$$V_k^*(t, p, m) = \alpha_{k,m}^+(q(t, p)) [(\xi_a + \xi_b)B_T(1 - t, 1)]^{\frac{1}{\varepsilon}},$$

where $\alpha_{k,0}^+(q) = \alpha_{k,0}(q)$ and

$$\alpha_{k,m}^+(q) = \max \left\{ q \int_0^1 e^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m \wedge (k-1)}^+(q(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz, \right. \\ \left. \max_{q' \in (0, q]} q' \int_0^1 e^{-q'z} \left[q'^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m-1}^+(q'(1-z))(1-z)^{\frac{1}{\varepsilon}} \right] dz. \right\}$$

We obtain the result by taking $t = 0$, $k = C$ and defining $\alpha_{C, sM+}^c = \max_{q>0} \alpha_{C, M}^+(q)$.

C.7.2 Incomplete Information

Uniform pricing. As in the case of complete information, the revenue $R_u^i(\mathbf{p}; B_T)$ with a price vector \mathbf{p} and conditional on B_T satisfies

$$\begin{aligned} R_u^i(\mathbf{p}; B_T) &= \mathbb{E}[p_a D_{aT}(0, \tau_C \wedge 1; p_a) + p_b D_{bT}(0, \tau_C \wedge 1; p_b) | B_T] \\ &= \frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} \mathbb{E}[D([\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}]B_T) \wedge C | B_T], \end{aligned}$$

Then, the expected revenue (taken with respect to B_T) with prices \mathbf{p} is

$$\begin{aligned} R_u^i(\mathbf{p}) &= \frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} E_{B_T|X_T}[\mathbb{E}[D([\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}]B_T) \wedge C]] \\ &= \frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} \int_{z>0} \mathbb{E}[D([\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}]z) \wedge C] f(z) dz. \end{aligned} \quad (\text{C.27})$$

As above, the function $\mathbf{p} \mapsto R_u^i(\mathbf{p})$ reaches its maximum when $p_a = p_b$. Thus,

$$\begin{aligned} R_u^i &= \max_{p>0} R_u^i(p, p; \xi_a, \xi_b, \varepsilon, f) \\ &= \max_{p>0} p \int_{z>0} \mathbb{E}[D([\xi_a + \xi_b]p^{-\varepsilon}z) \wedge C] f(z) dz. \end{aligned}$$

By the change of variable $q = [\xi_a + \xi_b]p^{-\varepsilon}$, we obtain the desired formula. If Assumption 2 also holds, the result follows using the change of variable $q = [\xi_a + \xi_b]p^{-\varepsilon} e^{X_T \beta_0}$.

Full dynamic pricing. Define $V_k(t, \mathbf{p}, f)$ as the expected revenue at time $1 - t$ when there remains k vacant seats before the departure, the current seat is priced at \mathbf{p} and the density of η_T , given the

current information, is f . Let also $V_k^*(t, f) = \max_{p_a, p_b > 0} V_k(t, \mathbf{p}, f)$. When $\eta_T \sim \Gamma(\lambda, \mu)$, we use respectively $V_k(t, \mathbf{p}, \lambda, \mu)$ and $V_k^*(t, \lambda, \mu)$ instead of $V_k(t, \mathbf{p}, g_{\lambda, \mu})$ and $V_k^*(t, g_{\lambda, \mu})$.

Between $1-t$ and $1-t+\Delta t$, if one seat is sold, which occurs with probability $(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) \partial_1 B_T(1-t, 1) \Delta t + o(\Delta t)$, the posterior cumulative distribution function (cdf) of η_T , $F_1(\cdot; \Delta t)$ satisfies

$$F_1(\eta; \Delta t) \propto [(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) e^{X_T \beta_0} \partial_1 g(1-t, 1) \eta \Delta t + o(\Delta t)] \eta^{\lambda-1} e^{-\mu \eta},$$

and the corresponding density is

$$f_1(\eta; \Delta t) = \eta^\lambda e^{-\mu \eta} \frac{\mu^{\lambda+1}}{\Gamma(\lambda+1)} + o(\Delta t).$$

As $\Delta t \rightarrow 0$, the posterior density converges to $g_{\lambda+1, \mu}$. If the seat is not sold between $1-t$ and $1-t+\Delta t$, then the posterior cdf of η_T is

$$F_0(\eta; \Delta t) \propto \eta^{\lambda-1} \exp(-\mu(t, \Delta t, \mathbf{p}) \eta),$$

where $\mu(t, \Delta t, \mathbf{p}) = \mu + (\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) e^{X_T \beta_0} g_T(1-t, 1-t+\Delta t)$. Therefore, the posterior density is $g_{\lambda, \mu(t, \Delta t, \mathbf{p})}$. Then, the Bellman equation can be written as:

$$\begin{aligned} V_k(t, \mathbf{p}, \lambda, \mu) = & \int \left\{ [(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) \eta e^{X_T \beta_0} \partial_1 g(1-t, 1) \Delta t + o(\Delta t)] \right. \\ & \times \left[\frac{(\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon})}{(\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon})} + V_{k-1}^*(t-\Delta t, f_1(\cdot; \Delta t)) \right] \\ & + \left[1 - (\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) \eta e^{X_T \beta_0} \partial_1 g(1-t, 1) \Delta t - o(\Delta t) \right] \\ & \left. \times V_k^*(t-\Delta t, \lambda, \mu(\mathbf{p}, t, \Delta)) \right\} g_{\lambda, \mu}(\eta) d\eta. \end{aligned}$$

The maximum of $V(t, \mathbf{p}, f)$ is reached when $p_a = p_b$, and $V_k(t, p, \lambda, \mu) = V_k(t, p, p, \lambda, \mu)$ satisfies

$$\begin{aligned} V_k(t, p, \lambda, \mu) = & V_k^*(t-\Delta t, \lambda, \mu) + \int \left\{ [(\xi_a + \xi_b) p^{-\varepsilon} e^{X_T \beta_0} \eta \partial_1 g(1-t, 1) \Delta t + o(\Delta t)] \right. \\ & \times [p + V_{k-1}^*(t-\Delta t, f_1(\cdot; \Delta t))] + [V_k^*(t-\Delta t, \lambda, \mu(\mathbf{p}, t, \Delta)) - V_k^*(t-\Delta t, \lambda, \mu)] \\ & \left. - V_k^*(t-\Delta t, \lambda, \mu(\mathbf{p}, t, \Delta)) [(\xi_a + \xi_b) p^{-\varepsilon} e^{X_T \beta_0} \eta \partial_1 g(1-t, 1) \Delta t + o(\Delta t)] \right\} g_{\lambda, \mu}(\eta) d\eta. \end{aligned}$$

Then, using $V_k^*(t, \lambda, \mu) = \max_{p>0} V_k(t, p, \lambda, \mu)$ and letting $\Delta t \rightarrow 0$, we obtain:

$$\begin{aligned}
& \partial_1 V_k^*(t, \lambda, \mu) \\
&= \max_{p>0} \int \left\{ (\xi_a + \xi_b) p^{-\varepsilon} e^{X_T \beta_0} \eta \partial_1 g(1-t, 1) [p + V_{k-1}^*(t, \lambda + 1, \mu) - V_k^*(t, \lambda, \mu)] \right. \\
&\quad \left. + \lim_{\Delta t \rightarrow 0} \frac{V_k^*(t - \Delta t, \lambda, \mu(t, \Delta t, \mathbf{p})) - V_k^*(t - \Delta t, \lambda, \mu)}{\Delta t} \right\} g_{\lambda, \mu}(\eta) d\eta \\
&= \partial_1 g(1-t, 1) (\xi_a + \xi_b) e^{X_T \beta_0} \max_{p>0} \int \left\{ p^{-\varepsilon} \eta [p + V_{k-1}^*(t, \lambda + 1, \mu) - V_k^*(t, \lambda, \mu)] \right. \\
&\quad \left. + \partial_3 V_k^*(t, \lambda, \mu) p^{-\varepsilon} \right\} g_{\lambda, \mu}(\eta) d\eta \\
&= \partial_1 g(1-t, 1) (\xi_a + \xi_b) e^{X_T \beta_0} \max_{p>0} \left\{ p^{-\varepsilon} \frac{\lambda}{\mu} [p + V_{k-1}^*(t, \lambda + 1, \mu) - V_k^*(t, \lambda, \mu)] \right. \\
&\quad \left. + \partial_3 V_k^*(t, \lambda, \mu) p^{-\varepsilon} \right\}.
\end{aligned}$$

Solving for the optimal price, we then obtain:

$$\begin{aligned}
\partial_1 V_k^*(t, \lambda, \mu) &= \left[\frac{\varepsilon}{\varepsilon - 1} \right]^{-\varepsilon} \frac{\lambda}{\mu(\varepsilon - 1)} \partial_1 g(1-t, 1) (\xi_a + \xi_b) e^{X_T \beta_0} \\
&\quad \times \left[-V_{k-1}^*(t, \lambda + 1, \mu) + V_k^*(t, \lambda, \mu) - \frac{\mu}{\lambda} \partial_3 V_k^*(t, \lambda, \mu) \right]^{1-\varepsilon}.
\end{aligned}$$

Letting $z(t) = g(1-t, 1) (\xi_a + \xi_b) e^{X_T \beta_0}$ and $\bar{V}^*(z(t), \lambda, \mu) = V^*(t, \lambda, \mu)$, we obtain:

$$\begin{aligned}
\partial_1 \bar{V}_k^*(z, \lambda, \mu) &= \left[\frac{\varepsilon}{\varepsilon - 1} \right]^{-\varepsilon} \frac{\lambda}{\mu(\varepsilon - 1)} \left[-V_{k-1}^*(z, \lambda + 1, \mu) + \bar{V}_k^*(z, \lambda, \mu) \right. \\
&\quad \left. - \frac{\mu}{\lambda} \partial_3 \bar{V}_k^*(z, \lambda, \mu) \right]^{1-\varepsilon}. \tag{C.28}
\end{aligned}$$

We prove by induction on k that for all $k \in \{0, \dots, C\}$.

$$\bar{V}_k^*(z, \lambda, \mu) = \left(\frac{z}{\mu} \right)^{\frac{1}{\varepsilon}} \alpha_{k,f}^i(\lambda), \tag{C.29}$$

where $\alpha_f^i(0, \lambda) = 0$ and for $k \geq 1$,

$$\alpha_{k,f}^i(\lambda) = \lambda \left(1 - \frac{1}{\varepsilon} \right)^{\varepsilon-1} \left[-\alpha_{k-1,f}^i(\lambda + 1) + \left(1 + \frac{1}{\lambda \varepsilon} \right) \alpha_{k,f}^i(\lambda) \right]^{1-\varepsilon}.$$

The result holds for $k = 0$ since $V_0^*(z, \lambda, \mu) = 0$. Suppose that (C.29) holds for $k - 1$. Then, (C.28) and the induction hypothesis yield

$$\begin{aligned}
\partial_1 \bar{V}_k^*(z, \lambda, \mu) &= \left[\frac{\varepsilon}{\varepsilon - 1} \right]^{1-\varepsilon} \frac{\lambda}{\mu(\varepsilon - 1)} \left[- \left(\frac{z}{\mu} \right)^{\frac{1}{\varepsilon}} \alpha_{k-1,f}^i(\lambda + 1) + \bar{V}_k^*(z, \lambda, \mu) \right. \\
&\quad \left. - \frac{\mu}{\lambda} \partial_3 \bar{V}_k^*(z, \lambda, \mu) \right]^{1-\varepsilon}. \tag{C.30}
\end{aligned}$$

The function $(z, \lambda, \mu) \mapsto \alpha_{k,f}^i(\lambda) (z/\mu)^{1/\varepsilon}$ is a solution to (C.30). We now show that $\bar{V}_k^*(z, \lambda, \mu)$ is equal to this solution. First, note that $V_k^*(t, \lambda, \mu)$ remains unchanged if the distribution of $B_T(t, t')$ remains unchanged. Now,

$$B_T(t, t') = g_T(t, t') e^{X_T \beta_0} \eta_T = (g_T(t, t') e^{X_T \beta_0} / \delta) \times (\delta \eta_T),$$

with $\delta \eta_T \sim \Gamma(\lambda, \mu/\delta)$. Hence, $V_k^*(t, \lambda, \mu)$ remains unchanged if we replace μ by μ/δ and $z(t) = g(1-t, 1)(\xi_a + \xi_b) e^{X_T \beta_0}$ by $z(t)/\delta$. Given the definition of $\bar{V}_k^*(z, \lambda, \mu)$, this implies $\bar{V}_k^*(z/\delta, \lambda, \mu/\delta) = \bar{V}_k^*(z, \lambda, \mu)$ for all $\delta > 0$. Then, to prove the induction step, we only need to show that $V(x) := V_k^*(x, \lambda, 1)$ satisfies $V(x) = \alpha_{k,f}^i(\lambda) x^{1/\varepsilon}$. By Equation (C.30),

$$V'(x) = \left[\frac{\varepsilon}{\varepsilon - 1} \right]^{1-\varepsilon} \frac{\lambda}{\varepsilon - 1} \left[-x^{\frac{1}{\varepsilon}} \alpha_{k-1,f}^i(\lambda + 1) + V(x) + \frac{x}{\lambda} V'(x) \right]^{1-\varepsilon}, \quad (\text{C.31})$$

with initial condition $V(0) = 0$. Suppose that (C.31) has two distinct solutions V_1, V_2 and let x_0 be such that $V_1(x_0) \neq V_2(x_0)$, say $V_1(x_0) > V_2(x_0)$. Define $x_m = \sup\{x \leq x_0 : V_1(x) \leq V_2(x)\}$. Because $V_1(0) = V_2(0)$ and $V_1(x_0) > V_2(x_0)$, we have $0 \leq x_m < x_0$ and $V_1(x) > V_2(x)$ for $x \in (x_m, x_0]$. Moreover, because both solutions are continuous, $V_1(x_m) = V_2(x_m)$. According to (C.31), because $\varepsilon > 1$, as long as $V_1(x) > V_2(x)$, we have $V_1'(x) < V_2'(x)$. Then,

$$V_1(x_0) - V_2(x_0) = \int_{x_m}^{x_0} [V_1'(x) - V_2'(x)] dx < 0,$$

which contradicts $V_1(x_0) > V_2(x_0)$. Hence, $V(x) = \alpha_{k,f}^i(\lambda) x^{1/\varepsilon}$, and the induction step holds. Thus, (C.29) is satisfied for $k \in \{0, \dots, C\}$. Finally, we obtain the result in Appendix C.3 by taking $t = 0$ and $k = C$.

Stopping-time pricing The difference from the stopping-time pricing under complete information is that the firm updates in a Bayesian way its belief on the distribution of B_T . Even if the firm continuously updates its belief, only moments where a sale occurs matter, since this is the time where it can decide to change its prices. Thus, starting at time $1-t$, we can focus on time $1-t + \tau_{t,\mathbf{p}}$. The next lemma characterizes the corresponding posterior distribution of B_T .

Lemma 13. *Suppose that the density function of $B_T|X_T$ at time $1-t$ is f and the firm prices the next seat at \mathbf{p} . Then, the posterior distribution of $B_T|\tau_{1-t,\mathbf{p}} = s$ is $T(f; q(t, s, \mathbf{p}))$, with $q(t, s, \mathbf{p}) = (\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) g_T(1-t, 1-t+s)$ and*

$$T(f; u)(z) = \frac{z e^{-uz} f(z)}{\int z e^{-uz} f(z) dz}.$$

Proof. As Equation (C.20) shows, given $B_T = z$, the density function of $\tau_{1-t,\mathbf{p}}$ is

$$f_{\tau_{1-t,\mathbf{p}}|B_T}(s|z) = (\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) z \partial_2 g_T(1-t, 1-t+s) e^{-zq(t,s,\mathbf{p})}. \quad (\text{C.32})$$

Then, the joint distribution of $(\tau_{1-t,\mathbf{p}}, B_T)$ is

$$f_{\tau_{1-t,\mathbf{p}}, B_T}(s, z) = (\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) z \partial_2 g_T(1-t, 1-t+s) e^{-q(t,s,\mathbf{p})z} f(z)$$

The result follows. \square

Now, using the same notation as in the full dynamic pricing case above and the same arguments as in proof of (C.21), we have

$$V_k(t, \mathbf{p}, f) = \int_0^t f_{\tau_{1-t, \mathbf{p}}}(s) \left[\frac{\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}}{\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}} + V_{k-1}^*(t-s; T(f; q(t, s, \mathbf{p}))) \right] ds.$$

Given $\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}$, $V_k(t, \mathbf{p}, f)$ is maximized if and only if $\xi_a p_a^{1-\varepsilon} + \xi_b p_b^{1-\varepsilon}$ is maximized. Thus, $V_k(t, \mathbf{p}, f)$ is maximized at $p_a = p_b$ and

$$V_k^*(t, f) = \max_{p>0} \int_0^t f_{\tau_{1-t, p, p}}(s) [p + V_{k-1}^*(t-s; T(f; q(t, s, p)))] ds. \quad (\text{C.33})$$

where $q(t, s, p) = q(t, s, p, p)$. We now prove by induction on k that for all $k \in \{0, \dots, C\}$,

$$V_k^*(t; f) = [(\xi_a + \xi_b)g_T(1-t, 1)]^{\frac{1}{\varepsilon}} \alpha_{k, s}^i(f). \quad (\text{C.34})$$

where $\alpha_s^i(0, f) = 0$ and for all $k \in \{1, \dots, C\}$,

$$\alpha_{k, s}^i(f) = \max_{q>0} q \int_0^1 \left[q^{-1/\varepsilon} + (1-u)^{\frac{1}{\varepsilon}} \alpha_{k-1, s}^i(T(f; qu)) \right] \int_0^\infty z e^{-quz} f(z) dz du.$$

The result holds for $k = 0$ since $V_0^*(t; f) = 0$. Suppose that it holds for $k-1 \geq 0$. First, by (C.32), we have

$$f_{\tau_{1-t, \mathbf{p}, B_T}}(s, z) = \int_0^\infty (\xi_a p_a^{-\varepsilon} + \xi_b p_b^{-\varepsilon}) z \partial_2 g_T(1-t, 1-t+s) e^{-q(t, s, \mathbf{p})z} f(z) dz. \quad (\text{C.35})$$

Then, letting $q(t, p) = q(t, t, p)$ and using (C.33), we obtain

$$\begin{aligned} V_k^*(t, f) &= \max_{p>0} \int_0^t f_{\tau_{1-t, p, p}}(s) \left\{ p + [(\xi_a + \xi_b)g_T(1-t+s, 1)]^{\frac{1}{\varepsilon}} \right. \\ &\quad \left. \times \alpha_{k-1, s}^i(T(f; q(t, s, p))) \right\} ds \\ &= \max_{p>0} \int_0^1 \left[\int_0^\infty q(t, p) z e^{-q(t, p)uz} f(z) dz \right] \\ &\quad \times \left\{ p + [(\xi_a + \xi_b)g_T(1-t, 1)(1-u)]^{\frac{1}{\varepsilon}} \alpha_{k-1, s}^i(T(f; q(t, p)u)) \right\} du \\ &= [(\xi_a + \xi_b)g_T(1-t, 1)]^{\frac{1}{\varepsilon}} \max_{q>0} q \int_0^1 \left[\int_0^\infty z e^{-quz} f(z) dz \right] \\ &\quad \times \left[q^{-1/\varepsilon} + (1-u)^{\frac{1}{\varepsilon}} \alpha_{k-1, s}^i(T(f; qu)) \right] du. \end{aligned}$$

The second equality follows using the change of variable $u = g_T(1-t, 1-t+s)/g_T(1-t, 1)$ and the third by the change of variable $q = q(t, p)$. Hence, the induction step holds, and (C.34) is satisfied for all $k \in \{0, \dots, C\}$. We obtain the desired expression by taking $t = 0$ and $k = C$.

If Assumption 2 also holds, we obtain by Lemma 13 that if $f = g_{\lambda, \mu}$, then $T(f; u) = g_{\lambda+1, \mu+\exp(X'_T \beta_0)u}$. Let $V_k(t, \mathbf{p}; \lambda, \mu)$ and $V_k^*(t; \lambda, \mu)$ be defined as in the full dynamic pricing case. Then, by the same

induction as above, we have, for all $k \in \{0, \dots, C\}$,

$$V_k^*(t; \lambda, \mu) = \left[\frac{(\xi_a + \xi_b) e^{X_T' \beta_0} g_T(1-t, 1)}{\mu} \right]^{\frac{1}{\varepsilon}} \alpha_{k,s}^i(\lambda), \quad (\text{C.36})$$

where $\alpha_s^i(0, \lambda) = 0$ for $\lambda > 0$, and

$$\alpha_{k,s}^i(\lambda) = \max_{q>0} q \int_0^1 \frac{\lambda}{(1+qs)^{\lambda+1}} \left[q^{-\frac{1}{\varepsilon}} + \left(\frac{1-s}{1+qs} \right)^{\frac{1}{\varepsilon}} \alpha_{k-1,s}^i(\lambda+1) \right] ds.$$

The result follows by taking $t = 0$ and $k = C$, we obtain the desired expression.

Stopping-time pricing with M fares. Like previously, one can show that the optimal prices should satisfy $p_a = p_b = p$. As in the complete information case, let $V_k(0; t, p, m)$ (resp. $V_k(1; t, p, m, f)$) denote the optimal revenue at time $1-t$, with a current price p , a remaining capacity k , a remaining number of fares m and a density of f for η_T (conditional on the current information) if the firm decides to keep the same price (resp. to change it). Then, as (C.24), we have:

$$\begin{aligned} V_k(0; t, p, m, f) &= \int_0^t f_{\tau_{1-t,p,p}}(s) [p + V_{k-1}^*(t-s, p, m, T(f; q(t, s, p)))] ds, \\ V_k(1; t, p, m, f) &= \max_{p'>0} \int_0^t f_{\tau_{1-t,p',p'}}(s) [p' + V_{k-1}^*(t-s, p', m-1, T(f; q(t, s, p)))] ds, \\ V_k^*(t, p, m, f) &= \max_{d \in \{0,1\}} V_k(d; t, p, m, f), \end{aligned} \quad (\text{C.37})$$

with the initial conditions $V_0^*(t, p, m, f) = 0$. We prove by induction on k that for all $(k, m) \in \{0, \dots, C\} \times \mathbb{N}$,

$$V_k^*(t, p, m, f) = c_{k,m}(q(t, p), f) [(\xi_a + \xi_b) g_T(1-t, 1)]^{\frac{1}{\varepsilon}}, \quad (\text{C.38})$$

where $c_{k,0}(q, f) = q^{-\frac{1}{\varepsilon}} \int \mathbb{E}[D(qz) \wedge k] f(z) dz$ and

$$\begin{aligned} c_{k,m}(q, f) &= \max \left\{ q \int_0^1 \int z e^{-qzu} f(z) dz \left[q^{-1/\varepsilon} + c_{k-1, m \wedge (k-1)}(q(1-u), T(f; qu)) \right. \right. \\ &\quad \left. \left. (1-u)^{\frac{1}{\varepsilon}} \right] du, \max_{q'>0} q' \int_0^1 \int z e^{-q'zu} f(z) dz \left[q'^{-1/\varepsilon} \right. \right. \\ &\quad \left. \left. + c_{k-1, m-1}(q'(1-u), T(f; q'u))(1-u)^{\frac{1}{\varepsilon}} \right] du \right\}. \end{aligned}$$

The result holds for $k = 0$ since $c_{0,m} = V_0^*(t, p, m, f) = 0$. Suppose that it holds for $k-1 \geq 0$ and all $m \leq k-1$ (recall that $V_k^*(t, p, m, f) = V_k^*(t, p, m \wedge k, f)$). If $m = 0$, the price cannot be changed anymore, so $V_k^*(t, p, m)$ is simply the revenue with price p from $1-t$ to 1, and (C.25) holds.

If $m \geq 1$, we have, using (C.35) and (C.37) and the same change of variables as above, we obtain

$$\begin{aligned}
& V_k(0; t, p, m, f) \\
&= \int_0^t \left[\int_0^\infty (\xi_a + \xi_b) p^{-\varepsilon} \partial_2 g_T(1-t, 1-t+s) z e^{-q(t,s,p)z} f(z) dz \right] \\
&\quad \left[p + c_{k-1, m \wedge (k-1)}(q(t-s, p), T(f; q(t, s, p))) [(\xi_a + \xi_b) g_T(1-t+s, 1)]^{\frac{1}{\varepsilon}} \right] ds \\
&= \int_0^1 \left[\int_0^\infty q(t, p) z e^{-q(t,p)uz} f(z) dz \right] \left[p + c_{k-1, m \wedge (k-1)}(q(t, p)(1-u), \right. \\
&\quad \left. T(f; q(t, p)u)) [(\xi_a + \xi_b) g_T(1-t, 1)]^{\frac{1}{\varepsilon}} (1-u)^{\frac{1}{\varepsilon}} \right] du \\
&= [(\xi_a + \xi_b) g_T(1-t, 1)]^{\frac{1}{\varepsilon}} q(t, p) \int_0^1 \left[\int_0^\infty z e^{-q(t,p)uz} f(z) dz \right] \\
&\quad \times \left[q(t, p)^{-1/\varepsilon} + c_{k-1, m \wedge (k-1)}(q(t, p)(1-u), T(f; q(t, p)u)) (1-u)^{\frac{1}{\varepsilon}} \right] du.
\end{aligned}$$

By the same reasoning and the change of variable $q = q(t, p)$,

$$\begin{aligned}
V_k(1; t, p, m, f) &= [(\xi_a + \xi_b) g_T(1-t, 1)]^{\frac{1}{\varepsilon}} \max_{q>0} q \int_0^1 \left[\int_0^\infty z e^{-qzu} f(z) dz \right] \\
&\quad \left[q^{-1/\varepsilon} + c_{k-1, m-1}(q(1-u), T(f; qu)) (1-u)^{\frac{1}{\varepsilon}} \right] du.
\end{aligned}$$

Then,

$$\begin{aligned}
V_k^*(t, p, m, f) &= \max_{d \in \{0,1\}} V_k(d; t, p, m, f) \\
&= c_{k,m}(q(p), f) [(\xi_a + \xi_b) g_T(1-t, 1)]^{\frac{1}{\varepsilon}}.
\end{aligned}$$

This concludes the induction step, proving that (C.38) holds for all $k \in \{0, \dots, C\}$.

Stopping-time pricing with M increasing fares The proof follows by making the same changes of the previous case as those made in the complete information set-up.

C.8 Intermediate- K stopping time pricing

We extend Theorem 10 to pricing strategies called intermediate- K stopping-time pricing where the firm can only adjust prices for the first $K \leq C$ seats. Then a uniform pricing applies from the $(K+1)^{th}$ seat till departure. These pricing strategies connect uniform pricing and stopping-time pricing, since the former corresponds to $K=0$ and the latter to $K=C$. More importantly, we can infer from an increase in K how uncertainty is swept out under complete and incomplete information, and quantify its impact on revenues. They are simulated with the estimated parameters and the results are used to depict the blue curve of Figure 3.1.

Complete information As above, let $V_k^*(t; K, C)$ denote the optimal revenue at time $1-t$ with $k \geq C-K$ remaining seats. For such k , we have

$$V_k^*(t, K, C) = c'_{k,K} [(\xi_a + \xi_b) B_T(1-t, 1)]^{\frac{1}{\varepsilon}},$$

where $c'_{C-K,K} = \max_{q>0} q^{-\frac{1}{\varepsilon}} \mathbb{E}[D(q) \wedge (C - K)]$ and, for any $k > C - K$,

$$c'_{k,K} = \max_{q>0} q^{-\frac{1}{\varepsilon}} (1 - e^{-q}) + c'_{k-1,K} \int_0^1 q e^{-qs} (1 - s)^{\frac{1}{\varepsilon}} ds.$$

The proof is the same as that for (C.22) except for the initial value because the firm must apply uniform pricing whenever there remain $C - K$ seats. Thus, the Bellman equation and the updating of the constants $c'_{k,K}$ take the same form as under the stopping-time pricing strategy in (C.22) for $k \geq C - K$ and the initial value becomes $c'_{C-K,K}$, which comes from the optimal uniform pricing with $C - K$ seats.

Incomplete information Let $V_k^*(t; K, C, \lambda, \mu)$ denote the optimal expected revenue at time $1 - t$ with $k \geq C - K$ remaining seats if $\eta_T \sim \Gamma(\lambda, \mu)$. Then, for any $k \in \{C - K + 1, \dots, C\}$,

$$V_k^*(t; K, C, \lambda, \mu) = \left[\frac{(\xi_a + \xi_b) e^{X'_T \beta_0} g_T(t, 1)}{\mu} \right]^{\frac{1}{\varepsilon}} c'_{k,K}(\lambda),$$

where $c'_{C-K,K}(\lambda) = \max_{q \geq 0} q^{-\frac{1}{\varepsilon}} \int_{\mathbb{R}_+} \mathbb{E}[D(qz) \wedge (C - K)] g_{\lambda,1}(z) dz$ and for $k > C - K$,

$$c'_{k,K}(\lambda) = \max_{q>0} \int_0^1 \frac{\lambda}{(1 + qs)^{\lambda+1}} q \left[q^{-\frac{1}{\varepsilon}} + \left(\frac{1 - s}{1 + qs} \right)^{\frac{1}{\varepsilon}} c'_{k-1,K}(\lambda + 1) \right] ds.$$

As with complete information, the proof is the same as for (C.36), with the same change due to a different initial value based on uniform pricing.

Bibliography

- Adams, W. J. and Yellen, J. L. (1976). Commodity bundling and the burden of monopoly. *The quarterly journal of economics*, pages 475–498. (Cited on page 2.)
- Allen, R. and Rehbeck, J. (2019a). Identification with additively separable heterogeneity. *Econometrica*, 87(3):1021–1054. (Cited on pages 1, 4, 9, 37, 42, 52 and 68.)
- Allen, R. and Rehbeck, J. (2019b). Latent complementarity in bundles models. *Available at SSRN 3257028*. (Cited on pages 42 and 53.)
- Amemiya, T. (1977). The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica (pre-1986)*, 45(4):955. (Cited on page 55.)
- Amemiya, T. (1985). *Advanced econometrics*. Harvard university press. (Cited on page 13.)
- Andrews, D. W. (2017). Examples of l2-complete and boundedly-complete distributions. *Journal of econometrics*, 199(2):213–220. (Cited on pages 40 and 126.)
- Andrews, D. W. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157. (Cited on page 81.)
- Andrews, D. W. K. (2002). Generalized method of moments estimation when a parameter is on a boundary. *Journal of Business & Economic Statistics*, 20(4):530–544. (Cited on pages 54 and 68.)
- Armstrong, M. (2013). A more general theory of commodity bundling. *Journal of Economic Theory*, 148(2):448–472. (Cited on page 2.)
- Armstrong, M. (2016a). Nonlinear pricing. *Annual Review of Economics*, 8:583–614. (Cited on page 16.)
- Armstrong, M. and Vickers, J. (2010). Competitive non-linear pricing and bundling. *The Review of Economic Studies*, 77(1):30–60. (Cited on page 2.)
- Armstrong, T. B. (2016b). Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica*, 84(5):1961–1980. (Cited on pages 5 and 138.)
- Augereau, A., Greenstein, S., and Rysman, M. (2006). Coordination versus differentiation in a standards war: 56k modems. *The RAND Journal of Economics*, 37(4):887–909. (Cited on pages 3 and 41.)
- Aviv, Y. and Pazgal, A. (2002). Pricing of short life-cycle products through active learning. Working paper. (Cited on page 71.)
- Berry, S., Gandhi, A., and Haile, P. (2013). Connected substitutes and invertibility of demand. *Econometrica*, 81(5):2087–2111. (Cited on pages 3, 5, 9, 10, 18, 19, 40, 42, 50, 94, 95 and 96.)
- Berry, S., Khwaja, A., Kumar, V., Musalem, A., Wilbur, K. C., Allenby, G., Anand, B., Chintagunta, P., Hanemann, W. M., Jeziorski, P., et al. (2014). Structural models of complementary choices. *Marketing Letters*, 25(3):245–256. (Cited on pages 1, 2, 18, 37 and 41.)

- Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890. (Cited on pages 5, 37, 39, 41, 50, 55 and 138.)
- Berry, S., Levinsohn, J., and Pakes, A. (2004a). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*, 112(1):68–105. (Cited on page 37.)
- Berry, S., Linton, O. B., and Pakes, A. (2004b). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, 71(3):613–654. (Cited on page 5.)
- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262. (Cited on pages 3, 5, 8, 18, 19 and 39.)
- Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797. (Cited on pages 4, 9, 15, 16, 40, 43 and 51.)
- Bitran, G. R. and Mondschein, S. V. (1997). Periodic pricing of seasonal products in retailing. *Management science*, 43(1):64–79. (Cited on page 71.)
- Boothby, W. M. (1986). *An introduction to differentiable manifolds and Riemannian geometry*, volume 120. Academic press. (Cited on page 98.)
- Bowden, R. (1973). The theory of parametric identification. *Econometrica: Journal of the Econometric Society*, pages 1069–1074. (Cited on pages 4 and 13.)
- Bronnenberg, B. J., Kruger, M. W., and Mela, C. F. (2008). Database paper—the iri marketing data set. *Marketing science*, 27(4):745–748. (Cited on pages 21 and 58.)
- Brumelle, S. and McGill, J. (1993). Airline seat allocation with multiple nested fare classes. *Operations Research*, 41(1):127–137. (Cited on page 71.)
- Bugni, F. A., Canay, I. A., and Shi, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, 8(1):1–38. (Cited on page 81.)
- Canay, I. A., Santos, A., and Shaikh, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, 81(6):2535–2559. (Cited on page 40.)
- Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334. (Cited on page 55.)
- Chan, T. Y. (2006). Estimating a continuous hedonic-choice model with an application to demand for soft drinks. *The Rand journal of economics*, 37(2):466–482. (Cited on page 3.)
- Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica*, 75(5):1243–1284. (Cited on pages 2, 13 and 100.)
- Chesher, A. and Rosen, A. M. (2017). Generalized instrumental variable models. *Econometrica*, 85(3):959–989. (Cited on page 4.)

- Cho, S., Lee, G., Rust, J., and Yu, M. (2018). Optimal dynamic hotel pricing. (Cited on pages 72 and 73.)
- Chu, C. S., Leslie, P., and Sorensen, A. (2011). Bundle-size pricing as an approximation to mixed bundling. *The American Economic Review*, 101(1):263–303. (Cited on page 2.)
- Ciarlet, P. G. (2013). *Linear and nonlinear functional analysis with applications*, volume 130. Siam. (Cited on page 111.)
- Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828. (Cited on page 81.)
- Compiani, G. (2019). Market counterfactuals and the specification of multi-product demand: A non-parametric approach. *Working Paper*. (Cited on page 5.)
- Conlon, C. and Gortmaker, J. (2019). Best practices for differentiated products demand estimation with pyblp. Technical report. (Cited on pages 20, 50 and 55.)
- Cournot, A. (1838). *Researches into the Mathematical Principles of the Theory of Wealth*. Macmillan Co. (Cited on pages 3, 35, 36, 41 and 66.)
- Crawford, G. S., Lee, R. S., Whinston, M. D., and Yurukoglu, A. (2018). The welfare effects of vertical integration in multichannel television markets. *Econometrica*, 86(3):891–954. (Cited on pages 3, 37 and 41.)
- Crawford, G. S. and Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television markets. *American Economic Review*, 102(2):643–85. (Cited on pages 3, 37 and 41.)
- Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*. Number 17. Yale University Press. (Cited on page 1.)
- den Boer, A. V. (2015). Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18. (Cited on page 71.)
- den Boer, A. V. and Zwart, B. (2015). Dynamic pricing and learning with finite inventories. *Operations research*, 63(4):965–978. (Cited on page 71.)
- D’Haultfoeuille, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory*, 27(3):460–471. (Cited on pages 40, 52, 123, 126 and 143.)
- Dubé, J.-P. (2004). Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science*, 23(1):66–81. (Cited on pages iv, 1, 3, 7, 21, 22, 40, 41, 46, 93 and 94.)
- Dubé, J.-P. H. (2018). Microeconomic models of consumer demand. Technical report, National Bureau of Economic Research. (Cited on page 41.)
- Dubé, J.-P. H. (2019). Microeconomic models of consumer demand. *Handbook on the Economics of Marketing*. (Cited on pages 3 and 7.)
- Dunker, F., Hoderlein, S., and Kaido, H. (2017). Nonparametric identification of random coefficients in endogenous and heterogeneous aggregated demand models. Technical report, cemap working paper. (Cited on page 42.)

- Ershov, D., Laliberté, W. P., and Orr, S. (2018). Mergers in a model with complementarity. *Working Paper*, pages 1–50. (Cited on pages 2, 4, 18, 37, 41 and 45.)
- Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review*, 103(5):1598–1628. (Cited on pages 5, 41, 42 and 53.)
- Feng, Y. and Gallego, G. (1995). Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Science*, 41(8):1371–1391. (Cited on page 71.)
- Feng, Y. and Xiao, B. (2000). Optimal policies of yield management with multiple predetermined prices. *Operations Research*, 48(2):332–343. (Cited on pages 71 and 76.)
- Fosgerau, M., Monardo, J., and de Palma, A. (2019). The inverse product differentiation logit model. *Working Paper*. (Cited on pages 3, 4, 21 and 42.)
- Fox, J. T., Kim, K., Ryan, S. P., and Bajari, P. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, 166(2):204–212. (Cited on pages 4 and 14.)
- Fox, J. T. and Lazzati, N. (2017). A note on identification of discrete choice models for bundles and binary games. *Quantitative Economics*, 8(3):1021–1036. (Cited on pages 1, 4 and 53.)
- Freyberger, J. (2015). Asymptotic theory for differentiated products demand models with many markets. *Journal of Econometrics*, 185(1):162–181. (Cited on pages 5 and 48.)
- Freyberger, J. (2017). On completeness and consistency in nonparametric instrumental variable models. *Econometrica*, 85(5):1629–1644. (Cited on page 40.)
- Gale, D. and Nikaido, H. (1965). The jacobian matrix and global univalence of mappings. *Mathematische Annalen*, 159(2):81–93. (Cited on pages 3, 40, 50 and 113.)
- Gallego, G. and Van Ryzin, G. (1994). Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management science*, 40(8):999–1020. (Cited on pages 71 and 158.)
- Gallego, G. and van Ryzin, G. (1994). Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8):999–1020. (Cited on page 75.)
- Gandhi, A. and Houde, J.-F. (2016). Measuring substitution patterns in differentiated products industries. (Cited on page 55.)
- Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *The American Economic Review*, 97(3):713–744. (Cited on pages 1, 2, 3, 4, 6, 7, 8, 10, 17, 18, 21, 25, 26, 29, 37, 40, 41, 45, 47, 48 and 53.)
- Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2014). Competition and ideological diversity: Historical evidence from us newspapers. *American Economic Review*, 104(10):3073–3114. (Cited on page 3.)
- Ghosh, J., Singh, R., et al. (1966). Unbiased estimation of location and scale parameters. *The Annals of Mathematical Statistics*, 37(6):1671–1675. (Cited on page 143.)

- Gillen, B. J., Montero, S., Moon, H. R., and Shum, M. (2019). Blp-2lasso for aggregate discrete choice models with rich covariates. *The Econometrics Journal*, 22(3):262–281. (Cited on page 39.)
- Gordon, B. R. and Hartmann, W. R. (2013). Advertising effects in presidential elections. *Marketing Science*, 32(1):19–35. (Cited on page 39.)
- Grieco, P. L. (2014). Discrete games with flexible information structures: An application to local grocery markets. *The RAND Journal of Economics*, 45(2):303–340. (Cited on page 81.)
- Grzybowski, L. and Verboven, F. (2016). Substitution between fixed-line and mobile access: the role of complementarities. *Journal of Regulatory Economics*, 49(2):113–151. (Cited on pages 2, 3, 17, 37, 41, 45 and 53.)
- Hendel, I. (1999). Estimating multiple-discrete choice models: An application to computerization returns. *The Review of Economic Studies*, 66(2):423–446. (Cited on pages iv, 1, 3, 7, 21, 22, 40, 41, 46 and 93.)
- Ho, K., Ho, J., and Mortimer, J. H. (2012). The use of full-line forcing contracts in the video rental industry. *The American Economic Review*, 102(2):686–719. (Cited on page 4.)
- Hu, Y. and Shiu, J.-L. (2018). Nonparametric identification using instrumental variables: sufficient conditions for completeness. *Econometric Theory*, 34(3):659–693. (Cited on page 40.)
- Iaria, A. and Wang, A. (2019a). Identification and estimation of demand for bundles. *Working Paper*, pages 1–92. (Cited on pages 41, 42, 45, 46, 48, 50, 53, 121, 125 and 129.)
- Iaria, A. and Wang, A. (2019b). Inferring complementarity from correlations rather than structural estimation. *Working Paper*, pages 1–17. (Cited on pages 3 and 7.)
- il Kim, K. (2014). Identification of the distribution of random coefficients in static and dynamic discrete choice models. *The Korean Economic Review*, 30(2):191–216. (Cited on pages 4 and 14.)
- Jevons, D., Meaney, A., Robins, N., Dargay, J., Preston, J., Goodwin, P., and Wardman, M. (2005). How do passengers respond to change? Working paper, Transport Studies Unit and Oxford University Centre for the Environment. (Cited on pages 70 and 81.)
- Kaido, H., Molinari, F., and Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, 87(4):1397–1432. (Cited on page 81.)
- Ketz, P. (2019). On asymptotic size distortions in the random coefficients logit model. *Journal of Econometrics*. (Cited on page 54.)
- Koijen, R. S. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515. (Cited on page 39.)
- Komunjer, I. (2012). Global identification in nonlinear models with moment restrictions. *Econometric Theory*, 28(4):719–729. (Cited on pages 4 and 13.)
- Krantz, S. G. and Parks, H. R. (2002). *A primer of real analytic functions*. Springer Science & Business Media. (Cited on page 108.)

- Kretschmer, T., Miravete, E. J., and Pernías, J. C. (2012). Competitive pressure and the adoption of complementary innovations. *American Economic Review*, 102(4):1540–70. (Cited on pages 4 and 41.)
- Kwak, K., Duvvuri, S. D., and Russell, G. J. (2015). An analysis of assortment choice in grocery retailing. *Journal of Retailing*, 91(1):19–33. (Cited on pages 2, 17, 37, 41 and 53.)
- Lazarev, J. (2013). The welfare effects of intertemporal price discrimination: An empirical analysis of airline pricing in U.S. monopoly markets. Working paper. (Cited on pages 71 and 72.)
- Lee, A. (1990). *Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process*. PhD thesis, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA. (Cited on pages 70 and 78.)
- Lee, S., Kim, J., and Allenby, G. M. (2013). A direct utility model for asymmetric complements. *Marketing Science*, 32(3):454–470. (Cited on pages 4 and 41.)
- Lewbel, A. (1985). Bundling of substitutes or complements. *International Journal of Industrial Organization*, 3(1):101–107. (Cited on page 2.)
- Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, 30(1):67–80. (Cited on page 4.)
- Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903. (Cited on page 4.)
- Lewis, A. D. (2009). Semicontinuity of rank and nullity and some consequences. *Author's notes on semicontinuity of rank and nullity*. (Cited on page 96.)
- Li, J., Granados, N., and Netessine, S. (2014). Are consumers strategic? structural estimation from the air-travel industry. *Management Science*, 60(9):2114–2137. (Cited on page 147.)
- Lin, K. Y. (2006). Dynamic pricing with real-time demand learning. *European Journal of Operational Research*, 174(1):522–538. (Cited on page 71.)
- Littlewood, K. (1972). Forecasting and control of passenger bookings. *Airline Group International Federation of Operational Research Societies Proceedings, 1972*, 12:95–117. (Cited on page 71.)
- Liu, H., Chintagunta, P. K., and Zhu, T. (2010a). Complementarities and the demand for home broadband internet services. *Marketing Science*, 29(4):701–720. (Cited on page 3.)
- Liu, H., Chintagunta, P. K., and Zhu, T. (2010b). Complementarities and the demand for home broadband internet services. *Marketing Science*, 29(4):701–720. (Cited on page 41.)
- Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, pages 343–360. (Cited on page 4.)
- Manski, C. F. (2003). *Partial identification of probability distributions*. Springer Science & Business Media. (Cited on page 4.)
- Manski, C. F. and Sherman, L. (1980). An empirical analysis of household choice among motor vehicles. *Transportation Research Part A: General*, 14(5-6):349–366. (Cited on pages 1, 3, 37, 39 and 41.)

- Manzini, P., Mariotti, M., and Ülkü, L. (2018). Stochastic complementarity. *The Economic Journal*. (Cited on pages 3 and 7.)
- Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). *Microeconomic theory*, volume 1. Oxford university press New York. (Cited on page 1.)
- Mattner, L. (1992). Completeness of location families, translated moments, and uniqueness of charges. *Probability Theory and Related Fields*, 92(2):137–149. (Cited on pages 40, 52, 122 and 126.)
- Mattner, L. (1993). Some incomplete but boundedly complete location families. *The Annals of Statistics*, 21(4):2158–2162. (Cited on pages 40, 126 and 143.)
- McAfee, R. and te Velde, V. (2008a). Dynamic pricing with constant demand elasticity. *Production and Operations Management*, 17(4):432–438. (Cited on pages 71, 74, 75 and 76.)
- McAfee, R. P., McMillan, J., and Whinston, M. D. (1989). Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, 104(2):371–383. (Cited on page 2.)
- McAfee, R. P. and te Velde, V. (2008b). Dynamic pricing with constant demand elasticity. *Production and Operations Management*, 17(4):432–438. (Cited on page 78.)
- McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, pages 447–470. (Cited on page 6.)
- Merlo, A. and Paula, Á. d. (2017). Identification and estimation of preference distributions when voters are ideological. *The Review of Economic Studies*, 84(3):1238–1263. (Cited on page 39.)
- Milligan, K. and Rekkas, M. (2008). Campaign spending limits, incumbent spending, and election outcomes. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(4):1351–1374. (Cited on page 39.)
- Mityagin, B. (2015). The zero set of a real analytic function. *arXiv preprint arXiv:1512.07276*. (Cited on pages 108, 110 and 132.)
- Nevo, A. (2000). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics*, pages 395–421. (Cited on pages 3, 20, 21, 24, 25, 32, 33, 34, 35, 36 and 58.)
- Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342. (Cited on pages 3, 20, 21, 24, 25, 32, 33, 34, 35, 36 and 58.)
- Nevo, A., Rubinfeld, D. L., and McCabe, M. (2005). Academic journal pricing and the demand of libraries. *American Economic Review*, 95(2):447–452. (Cited on pages 3 and 41.)
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245. (Cited on pages 4, 54, 114, 116 and 119.)
- Pozzi, A. (2012). Shopping cost and brand exploration in online grocery. *American Economic Journal: Microeconomics*, 4(3):96–120. (Cited on pages 3, 7, 21, 22 and 40.)

- Rekkas, M. (2007). The impact of campaign spending on votes in multiparty elections. *The Review of Economics and Statistics*, 89(3):573–585. (Cited on page 39.)
- Reynaert, M. and Verboven, F. (2014). Improving the performance of random coefficients demand models: the role of optimal instruments. *Journal of Econometrics*, 179(1):83–98. (Cited on pages 55 and 68.)
- Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, 40(6):2798–2822. (Cited on pages 2, 13, 102 and 103.)
- Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica: Journal of the Econometric Society*, pages 577–591. (Cited on pages 2, 4, 12 and 13.)
- Ruiz, F. J., Athey, S., and Blei, D. M. (2017). Shopper: A probabilistic model of consumer choice with substitutes and complements. *working paper*. (Cited on pages 2, 4, 18 and 37.)
- Samuelson, P. A. (1974). Complementarity: An essay on the 40th anniversary of the hicks-allen revolution in demand theory. *Journal of Economic literature*, 12(4):1255–1289. (Cited on pages 3, 7 and 39.)
- Sargan, J. D. (1983). Identification and lack of identification. *Econometrica: Journal of the Econometric Society*, pages 1605–1633. (Cited on page 4.)
- Sauvant, J. (2002). Le transport ferroviaire de voyageurs en france : enfin un bien normal? *Note de synthèse du SES*. (Cited on page 152.)
- Sher, I. and Kim, K. i. (2014). Identifying combinatorial valuations from aggregate demand. *Journal of Economic Theory*, 153:428–458. (Cited on pages 9, 37 and 42.)
- Song, I. and Chintagunta, P. K. (2006). Measuring cross-category price effects with aggregate store data. *Management Science*, 52(10):1594–1609. (Cited on page 42.)
- Song, M., Nicholson, S., and Lucarelli, C. (2017). Mergers with interfirm bundling: a case of pharmaceutical cocktails. *The RAND Journal of Economics*, 48(3):810–834. (Cited on page 4.)
- Stefanescu, C. (2012). Multivariate demand: Modeling and estimation from censored sales. Working paper. (Cited on pages 70 and 78.)
- Swan, W. (1990). Revenue management forecasting biases. Working paper, Boeing Commercial Aircraft, Seattle. (Cited on pages 70 and 78.)
- Talluri, K. and van Ryzin, G. (2005). *The theory and practice of revenue management*. Springer Verlag. (Cited on page 70.)
- Thomassen, Ø., Smith, H., Seiler, S., and Schiraldi, P. (2017). Multi-category competition and market power: a model of supermarket pricing. *American Economic Review*, 107(8):2308–51. (Cited on pages 1, 3, 4, 7, 21, 22, 40, 41 and 46.)
- Varian, H. R. (1992). Microeconomic analysis. 3rd. ed. ed. New York, London: WW Norton & Company. (Cited on page 1.)

- Wang, A. (2019). A blp demand model of product-level market shares with complementarity. *Working Paper*. (Cited on pages 9 and 37.)
- Wang, A. (2020). Identifying the distribution of random coefficients in blp demand models using one single variation in product characteristics. *Available at SSRN 3569077*. (Cited on page 53.)
- Wardman, M. (1997). Inter-urban rail demand elasticities and competition in Great Britain: Evidence from direct demand models. *Transportation Research*, 33E(1):15–28. (Cited on page 81.)
- Wardman, M. (2006). Demand for rail travel and the effects of external factors. *Transportation Research*, 42E:129–148. (Cited on page 81.)
- Wardman, M., Lythgoe, W., and Whelan, G. (2007). Rail passenger demand forecasting: cross-sectional models revisited. *Research in Transportation Economics*, 20(3):119–152. (Cited on page 81.)
- Williams, K. R. (2017). Dynamic airline pricing and seat availability. (Cited on pages 72 and 87.)
- Zhao, W. and Zheng, Y.-S. (2000). Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Management science*, 46(3):375–388. (Cited on pages 71 and 87.)
- Zhou, J. (2017). Competitive bundling. *Econometrica*, 85(1):145–172. (Cited on page 2.)

Titre : Trois essais sur les modèles microéconométriques de la demande et leurs applications dans l'organisation industrielle empirique

Mots clés : organisation industrielle empirique, économétrie structurelle, identification, demande de bundles, revenue management

Résumé : La thèse se compose de trois chapitres qui étudient les modèles microéconométriques de la demande et leurs applications dans l'organisation industrielle empirique.

Les deux premiers articles se concentrent sur les modèles de demande de bundles et étudient l'identification et l'estimation sous différentes disponibilités de données. Le premier article est un travail conjoint avec Alessandro Iaria (Université de Bristol) et se concentre sur les situations où les données d'achat au niveau du bundle sont disponibles. Nous présentons de nouveaux résultats d'identification et d'estimation pour un modèle logit mixte de demande de faisceaux. En particulier, nous proposons un nouvel inverse de la demande en présence de complémentarité qui permet de concentrer hors de la fonction de vraisemblance les effets fixes (potentiellement nombreux) spécifiques aux produits de marché, atténuant sensiblement le défi de dimensionnalité inhérent à l'estimation. Pour illustrer l'utilisation de nos méthodes, nous estimons la demande et l'offre dans l'industrie américaine des céréales prêtes à consommer, où le MLE proposé réduit la recherche numérique d'environ 12 000 à 130 paramètres. Nos estimations suggèrent que le fait d'ignorer la complémentarité hicksienne entre différents produits souvent achetés en lots peut entraîner des estimations de la demande et des contrefactuels trompeurs.

Le deuxième article se concentre sur les situations où seules des données d'achat agrégées au niveau du produit sont disponibles. Il propose un modèle de demande de Berry, Levinsohn et Pakes (BLP, 1995). Comparé aux modèles BLP de demande de produits uniques, ce modèle ne restreint pas les produits à être des substituts et, notamment, permet des complémentarités hicksiennes entre les produits qui peuvent être choisis conjointement dans un bundle. En s'appuyant sur l'inverse de la demande du premier article, il propose des arguments d'identification constructifs du modèle et un estimateur de

la méthode généralisée des moments (GMM) pratiquement utile. En particulier, cet estimateur peut gérer des ensembles de choix potentiellement importants et sa mise en œuvre est simple, essentiellement comme un estimateur BLP standard. Enfin, j'illustre la mise en œuvre pratique des méthodes et j'évalue la demande de céréales et de lait prêts-à-manger (PAM) aux États-Unis. Les estimations de la demande suggèrent que les céréales et le lait PAM sont globalement complémentaires Hicksian et ces complémentarités sont hétérogènes entre les paquets. Ignorer ces complémentarités entraîne des contrefactuels trompeurs.

Le troisième article est un travail conjoint avec Xavier d'Haultfoeuille, Philippe Février et Lionel Wilner et porte sur la gestion des revenus. Bien que cette gestion ait considérablement accru la flexibilité dans la façon dont les entreprises fixent les prix, les entreprises imposent toujours des contraintes à leur stratégie de prix. Il existe encore peu de preuves des gains ou des pertes de telles stratégies par rapport à des prix uniformes ou à des stratégies totalement flexibles. Dans cet article, nous quantifions ces gains et pertes et identifions leurs sources sous-jacentes dans le contexte du transport ferroviaire français. Cela est compliqué par la censure à la demande et l'absence de variations de prix exogènes. Nous développons une stratégie d'identification originale sur la demande qui combine les variations temporelles des prix relatifs et les inégalités de moment résultant de la rationalité de base du côté des consommateurs et des faibles conditions d'optimalité de la stratégie de tarification de l'entreprise. Nos résultats suggèrent des gains importants de la gestion des revenus réels par rapport à une tarification uniforme, mais également des pertes substantielles par rapport à la stratégie de tarification optimale. Enfin, nous soulignons le rôle clé de la gestion des revenus pour l'acquisition d'informations lorsque la demande est incertaine.

Title : Three essays on microeconomic models of demand and their applications in empirical industrial organisation

Keywords : empirical industrial organisation, structural econometrics, identification, demand for bundles, revenue management

Abstract : The thesis consists of three chapters that study microeconomic models of demand and their applications in empirical industrial organisation.

The first two papers focus on models of demand for bundles and study the identification and estimation under different data availabilities. The first paper is a joint work with Alessandro Iaria (University of Bristol) and focuses on the situations where purchase data at bundle-level is available. We present novel identification and estimation results for a mixed logit model of demand for bundles. In particular, we propose a new demand inverse in the presence of complementarity that enables to concentrate out of the likelihood function the (potentially numerous) market-product specific fixed effects, substantially alleviating the challenge of dimensionality inherent in estimation. To illustrate the use of our methods, we estimate demand and supply in the US ready-to-eat cereal industry, where the proposed MLE reduces the numerical search from approximately 12000 to 130 parameters. Our estimates suggest that ignoring Hicksian complementarity among different products often purchased in bundles may result in misleading demand estimates and counterfactuals.

The second paper focuses on the situations where only aggregate purchase data at product-level is available. It proposes a Berry, Levinsohn and Pakes (BLP, 1995) model of demand for bundles. Compared to BLP models of demand for single products, this model does not restrict products to be substitutes and, notably, allows for Hicksian complementarities among products that can be jointly chosen in a bundle. Leveraging the demand inverse of the first paper, it proposes constructive identification arguments of the mo-

del and a practically useful Generalized Method of Moments (GMM) estimator. In particular, this estimator can handle potentially large choice sets and its implementation is straightforward, essentially as a standard BLP estimator. Finally, I illustrate the practical implementation of the methods and estimate the demand for Ready-To-Eat (RTE) cereals and milk in the US. The demand estimates suggest that RTE cereals and milk are overall Hicksian complementary and these complementarities are heterogeneous across bundles. Ignoring such complementarities results in misleading counterfactuals.

The third paper is a joint work with Xavier d'Haultfoeuille, Philippe Février and Lionel Wilner and focuses on revenue management. Despite that this management has greatly increased flexibility in the way firms set prices, firms usually still impose constraints on their pricing strategy. There is yet scarce evidence on the gains or losses of such strategies compared to uniform pricing or fully flexible strategies. In this paper, we quantify these gains and losses and identify their underlying sources in the context of French railway transportation. This is complicated by the censoring on demand and the absence of exogenous price variations. We develop an original identification strategy on the demand that combines temporal variations in relative prices and moment inequalities stemming from basic rationality on consumers' side and weak optimality conditions on the firm's pricing strategy. Our results suggest significant gains of the actual revenue management compared to uniform pricing, but also substantial losses compared to the optimal pricing strategy. Finally, we highlight the key role of revenue management for acquiring information when demand is uncertain.