Université Joseph Fourier

THESE

pour obtenir le grade de

DOCTEUR DE L'UJF

Spécialité : "mathématiques appliquées"

préparée au laboratoire de modélisation et calcul (LMC)
dans le cadre de **l'Ecole Doctorale "mathématiques, sciences et technologie de l'information, informatique"**

présentée et soutenue publiquement

par

Vincent GUIGUES

le 30 juin 2005

Titre :

Inférence statistique pour l'optimisation stochastique Applications en finance et en gestion de production

---

Directeur de thèse : Anatoli IOUDITSKI

JURY

M.	Guy Cohen	Cermics ENPC Paris	Président
M.	Vladimir Spokoiny	WIAS Berlin (Allemagne)	Rapporteur
M.	Georg Pflug	Université de Vienne (Autriche)	Rapporteur
M.	Anatoli Iouditski	Université Joseph Fourier de Grenoble	Directeur de thèse
M.	François Oustry	Raise Partner SAS Grenoble	Co-directeur de thèse
M.	René Aïd	EDF R&D Clamart	Examinateur
M.	Ollivier Taramasco	ENSIMAG Grenoble	Examinateur

À mes parents,

Remerciements

Je tiens tout d'abord à remercier sincèrement Anatoli Iouditski, mon directeur de thèse, pour sa disponibilité, sa sympathie et son discernement dans le choix des pistes de recherche passionnantes qu'il m'a proposées. Certaines problématiques auxquelles ce travail de thèse s'attaque (en particulier dans les chapitres 3 et 4) ont aussi été proposées par François Oustry, mon co-directeur de thèse. Je remercie donc aussi François pour avoir su poser des problèmes pratiques intéressants ayant servi de base à une partie de mes investigations dans cette thèse.

Je tiens à exprimer ma reconnaissance envers Vladimir Spokoiny et Georg Pflug pour avoir accepté d'être rapporteurs de mon manuscrit de thèse. Mes remerciements s'adressent aussi à Guy Cohen, René Aïd et Ollivier Taramasco qui m'ont fait l'honneur de faire partie de mon jury de soutenance.

Je tiens également à exprimer ma gratitude envers Pape NDiaye avec qui j'ai noué une collaboration étroite dans le cadre d'un projet de recherche effectué pour EDF. Nous avons eu l'occasion d'échanger à de nombreuses reprises et de manière constructive sur des problèmes tant théoriques que pratiques. Les discussions menées avec Pape, que ce soit chez Raise Partner, au LMC ou moins formellement au restaurant étaient des plaisirs toujours renouvelés.

Pour sa disponibilité, sa gentillesse et les compétences qu'il ma transmises, je tiens à exprimer ma gratitude envers Claude Lemaréchal, tout d'abord enseignant en optimisation à l'ENSIMAG puis dépanneur incollable en optimisation. Je dois aussi beaucoup à mes anciens enseignants de statistique, qui m'ont donné mon goût pour cette discipline, en particulier Olivier Gaudoin dont la clarté des cours à l'ENSIMAG était exemplaire.
J'ai par ailleurs trouvé, tout au long des enseignements dispensés pendant la thèse, une aide chaleureuse auprès de Carole Desprez et Serge Dégerine.

Je salue le dynamisme de l'équipe administrative, en particulier celui de Claudine Meyrieux qui s'est investi, par humanisme, au delà de son rôle de secrétaire en me donnant d'utiles conseils lors de mes démarches administratives.

Je tiens aussi à remercier mes collègues de bureaux pour leur bonne humeur : Sophie, Alain, Robin et Claire. Plus généralement, j'adresse mes amitiés à tous les thésards du laboratoire et amis grenoblois avec qui j'ai eu l'occasion de partager des discussions sympathiques ou des loisirs : Basile, Julie, Laurent (Doyen), Laurent (Tournier), Etienne, Antoine, Elimane, Claire (Chauvin), Fabien, Samir, Eva, Kartono, Christian, Guillaume,

Martin, Pietro, Rosa, Rachel, Aline, Christophe et Fred. Pendant ces années passées à Grenoble, j'ai pu bénéficier de la richesse des activités culturelles, sportives et associatives riches de cette région. J'adresse donc aussi un remerciement et un clin d'oeil à tous ceux, handballeurs, magiciens, oenologues en herbe, danseurs de rock ou de salsa, skieurs, véliplanchistes, jongleurs, joggers, judokas ou nageurs avec qui j'ai pu continuer à partager mes centres d'intérêt et qui ont contribué à faire de ces années, des années riches d'expérience, intenses de projets et d'activités et favorisant un climat serein pendant ces années de thèse.

Enfin et surtout je n'oublierai pas de saluer le soutien de mes parents tout au long de cette thèse. Plus généralement, je les remercie pour la liberté qu'ils m'ont laissé dans les différents choix que j'ai eu à faire jusqu'à présent.

Table des matières

Introduction **9**

Notations et abbréviations **13**

1 Modélisations en optimisation stochastique **15**
1.1 Introduction . . . 15
1.2 Modèle bi-étapes avec recours . . . 17
1.3 Modèle multi-étapes avec recours . . . 18
1.3.1 Formulation du modèle . . . 19
1.3.2 Méthodes de résolution . . . 19
1.3.3 Génération d'arbres de scénarios . . . 20
1.4 Contraintes probabilistes . . . 21
1.4.1 Différents types de contraintes probabilistes et de fonction coût . . . 21
1.4.2 Equivalents déterministes . . . 22
1.5 Contreparties robustes . . . 24
1.5.1 Présentation . . . 24
1.5.2 Différents ensembles d'incertitude . . . 27
1.5.3 Contreparties robustes de problèmes convexes incertains . . . 28
1.5.4 L'optimisation robuste à l'épreuve . . . 31
1.6 Contreparties robustes ajustées . . . 32
1.6.1 Présentation . . . 32
1.6.2 Problèmes linéaires . . . 34
1.6.3 Contreparties robustes ajustées affinement . . . 35
1.6.4 Problèmes quadratiques et coniques . . . 37
1.7 Spécificité des méthodes développées dans cette thèse . . . 38
1.8 Conclusion . . . 40

2 Inférence statistique pour la gestion portefeuilles **43**
2.1 Introduction . . . 43
2.1.1 Modélisation du processus des rendements . . . 44
2.1.2 Problèmes d'allocation d'actifs . . . 45
2.1.2.1 Modèle Value-at-Risk (VaR) . . . 46
2.1.2.2 Modèle moyenne variance de Markowitz . . . 47
2.1.3 Problème d'estimation des paramètres . . . 47
2.2 Contrôle de la précision des problèmes approchés . . . 48
2.2.1 Hypothèses sur la moyenne, les variances et covariances . . . 48
2.2.2 Précision des problèmes approchés . . . 49

2.2.2.1 Définition des problèmes approchés 50
2.2.2.2 Le cas stationnaire 51
2.2.2.3 Paramètres variant lentement 52
2.3 Méthode adaptative 54
2.3.1 Description de l'algorithme 54
2.3.2 Choix des ensembles $\mathcal{I}$, $\mathcal{I}(I)$ et du paramètre λ 56
2.3.3 Qualité de l'estimation 58
2.3.4 Modèle à sauts 60
2.3.5 L'algorithme adaptatif utilisant les estimations empiriques de la moyenne et de la matrice de covariance 62
2.3.5.1 Précision des problèmes VaR et de Markowitz approchés . . 64
2.3.5.2 Erreurs de première et deuxième espèce 65
2.4 Détection de rupture en utilisant une fenêtre glissante 65
2.4.1 Description de l'algorithme 66
2.4.2 Résultats théoriques 67
2.4.2.1 Cas où la matrice de covariance est constante 67
2.4.2.2 Cas où la moyenne est constante 68
2.4.3 Illustration numérique 69
2.4.4 Discussion 72
2.5 Simulations numériques 73
2.5.1 Simulations de Monte Carlo dans le cas Gaussien 74
2.5.1.1 Erreur de première espèce 74
2.5.1.2 Erreur de deuxième espèce 75
2.5.2 Simulations sur des données réelles 80
2.6 Conclusion sur l'inférence statistique pour la gestion de portefeuilles 85
2.7 Preuves 86

3 Analyse de sensibilité et calibration stable et robuste de matrices de covariances 105
3.1 Introduction 105
3.2 Modèle de Markowitz et sources d'instabilité 106
3.2.1 Modèle moyenne variance de Markowitz 106
3.2.2 Quelques propriétés du modèle de Markowitz 107
3.2.3 Sources d'instabilité 108
3.3 Analyse de sensibilité du modèle de Markowitz 109
3.3.1 Quelques résultats connus 109
3.3.1.1 Sur la sensibilité des changements de ρ pour P 109
3.3.1.2 Sur la sensibilité liée à des perturbations générales 110
3.3.2 Sensibilité de la valeur optimale du problème P 110
3.3.3 Analyse de sensibilité des solutions de $P(k, \rho, Q)$ 110
3.3.4 Analyse de sensibilité d'un problème quadratique et de P' 112
3.4 Contreparties robustes et stables 114
3.4.1 Pourquoi corriger les estimations adaptatives 114
3.4.2 Contreparties robustes du problème de Markowitz 115
3.4.2.1 Formulation générale du problème 115
3.4.2.2 Contraintes de boîte sur la moyenne et la matrice de covariance 115

3.4.2.3 Incertitude polytopique . . . 116
3.4.3 Calibration stable de la matrice de covariance . . . 116
3.4.3.1 Motivations . . . 117
3.4.3.2 Maximisation de la plus petite valeur propre . . . 117
3.4.3.3 Optimisation du conditionnement . . . 118
3.5 Autres calibrations de la matrice de covariance . . . 118
3.6 Résultats numériques . . . 120
3.6.1 Tests de stabilité . . . 120
3.6.1.1 Réduction du conditionnement . . . 121
3.6.1.2 Évolution de la composition du portefeuille dans le temps . 121
3.6.1.3 Influence de la perturbation des moyennes sur la composition du portefeuille optimal . . . 123
3.6.2 Comparaison des précisions . . . 123
3.6.3 Comparaison de la diversification des portefeuilles . . . 126
3.6.4 Comparaison des calibrations de la matrice de covariance sur des données réelles . . . 128
3.7 Conclusion sur la calibration stable et robuste de la matrice de covariance . 129
3.8 Preuves . . . 130

4 Analyse de problèmes de gestion de production 135
4.1 Modélisation du problème de gestion de production électrique . . . 137
4.1.1 Modélisation du fonctionnement des centrales thermiques . . . 137
4.1.2 Modélisation du fonctionnement des centrales hydrauliques . . . 138
4.1.3 Modélisation du fonctionnement des contrats EJP . . . 140
4.2 Formulation déterministe du problème . . . 141
4.3 Modélisation des paramètres incertains . . . 143
4.3.1 Modélisation des taux de disponibilité . . . 143
4.3.2 Modélisation par arbre de scénarios . . . 145
4.3.3 Modèle statistique pour la demande et les apports . . . 147
4.3.4 Ensembles d'incertitude . . . 147
4.4 Contreparties robustes et stables . . . 147
4.4.1 Régularisation de la fonction duale . . . 148
4.4.2 Contreparties robustes ajustées . . . 149
4.4.3 Contreparties robustes ajustées affinement . . . 151
4.4.3.1 Ensemble d'incertitude polytopique . . . 152
4.4.3.2 Contraintes de boîte sur les paramètres incertains . . . 152
4.4.3.3 Ensemble d'incertitude ellipsoïdal sur les consommations . 153
4.5 Méthodes de résolution . . . 154
4.5.1 Méthodologie . . . 154
4.5.2 Résolution directe des différents sous-problèmes . . . 155
4.5.3 Résolution des sous-problèmes hydrauliques et EJP par programmation dynamique . . . 156
4.5.3.1 Préliminaires . . . 156
4.5.3.2 Application pour résoudre les sous-problèmes hydrauliques et EJP . . . 162
4.5.4 Stratégies de gestion . . . 165
4.6 Calibration des ensembles d'incertitude . . . 166

4.6.1 Méthode de régression . . . 168
4.6.2 Modèle SARIMA . . . 170
4.6.3 Modèle saisonnier simple . . . 170
4.6.4 Comparaison des méthodes . . . 173
4.6.5 D'autres méthodes de calibration . . . 174
4.6.6 Conclusion sur la calibration des ensembles d'incertitude . . . 174
4.7 Comparaison des méthodes de gestion . . . 175
4.7.1 Simulations de scénarios . . . 176
4.7.1.1 Optimisation sur un arbre de scénarios . . . 176
4.7.1.2 Régularisation de la fonction duale . . . 180
4.7.1.3 Contrepartie robuste ajustée . . . 180
4.7.1.4 Contrepartie robuste ajustée affinement . . . 182
4.7.1.5 Tableau comparatif des méthodes de gestion . . . 184
4.7.2 Simulations sur des données réelles . . . 184
4.8 Conclusion sur les problèmes de gestion de production électrique . . . 186
4.9 Le problème de gestion du transport et de la distribution du gaz . . . 187
4.9.1 Présentation du problème . . . 187
4.9.2 Modélisation . . . 188
4.9.3 Contrepartie robuste ajustée . . . 190
4.10 CRA et CRAA du problème de gestion de production . . . 193

Conclusion générale et perspectives **197**

Contributions logicielles **201**

Bibliographie **202**

Introduction

Les problèmes d'optimisation déterministes sont utilisés pour modéliser et analyser un grand nombre de systèmes pour lesquels on recherche une stratégie optimale (vis à vis d'un certain critère appelé encore fonction objectif) satisfaisant un certain nombre de contraintes. Dans ces problèmes, les paramètres définissant les contraintes et la fonction objectif sont supposés connus. Cependant, un grand nombre de problèmes d'optimisation sont des problèmes d'optimisation stochastique dont les paramètres ne sont pas tous connus avec certitude lorsque la décision doit être prise. Par exemple, le problème de gestion de portefeuilles financiers consiste à choisir dans quels actifs investir pour un certain nombre de dates fixées à l'avance, de façon à maximiser sa richesse. Lorsqu'un rebalancement de portefeuille est opéré, le choix des actifs financiers doit se faire sans connaître les rendements exacts des différents actifs sur la période d'investissement à venir. Le problème de gestion d'un parc de production d'électricité consiste à prévoir les quantités d'électricité à produire par chacune des usines du parc de production pour chaque pas de temps sur une période donnée, sans connaître à l'avance les consommations électriques effectives, les apports en eau des réserves des centrales hydrauliques ainsi que les taux d'indisponibilité des usines. L'objet de cette thèse est de modéliser et analyser des problèmes d'optimisation stochastique et de proposer des méthodes de résolution pour ces problèmes. En particulier, nous étudierons en détail les deux problèmes que nous venons d'évoquer. Un intérêt tout particulier sera porté sur l'interfaçage entre les statistiques et l'optimisation. Autrement dit, les objets statistiques intervenant dans les problèmes d'optimisation que l'on considère seront étudiés en tenant compte de leur utilisation ultérieure pour résoudre ces problèmes d'optimisation. Le plan de la thèse est le suivant.

Le premier chapitre est un chapitre introductif dans lequel nous présentons différentes manières de prendre en compte l'incertitude des paramètres d'un problème d'optimisation stochastique. On présente successivement les problèmes d'optimisation stochastique multi-étapes, l'introduction de contraintes probabilistes dans le modèle et les modélisations robustes. La partie consacrée aux méthodologies robustes se termine par la présentation des contreparties robustes ajustées étudiées dans [11] dans le cas de problèmes linéaires. Nous étendons certains de leurs résultats pour des problèmes quadratiques et coniques quadratiques. Nous terminons ce chapitre en précisant quels sont les apports et les spécificités des méthodes développées dans cette thèse. La thèse se scinde ensuite en deux parties.

La première partie, regroupée dans les chapitres 2 et 3, est consacrée à l'étude de problèmes de gestion de portefeuilles.

Dans le chapitre 2, l'objectif principal est d'analyser le modèle de gestion de portefeuilles

Value-at-Risk (VaR) et le modèle de Markowitz. Une étude statistique est faite afin de déterminer des estimations des paramètres de ces modèles : les rendements moyens et la matrice de covariance des rendements sur la période d'investissement. On montre ensuite comment se servir de ces estimations statistiques afin de mettre en oeuvre les modèles VaR et de Markowitz. Plus précisément, le chapitre 2 aborde les problèmes suivants.

Après avoir précisé les hypothèses faites sur le processus des rendements et présenté le modèle Value-at-Risk et le modèle de Markowitz, on propose des mises en oeuvre de ces modèles dont on contrôle la qualité d'un point de vue théorique. Les mises en oeuvre proposées reviennent à se servir d'estimations adaptatives des paramètres de ces modèles. La qualité de ces mises en oeuvre est aussi mesurée d'un point de vue expérimental, sur des données réelles (en utilisant les actifs du Dow Jones) et simulées.

On présente alors un algorithme adaptatif dans la section 2.3 permettant de déterminer ces estimations adaptatives. Il s'agit de généraliser l'étude menée dans [70] (faite dans un cadre Gaussien unidimensionnel) à un contexte non paramétrique et multidimensionnel. On suppose qu'il existe un intervalle dit d'homogénéité temporelle locale tel que les moyennes, variances et covariances varient peu sur cet intervalle. On présente un algorithme adaptatif permettant de déterminer une estimation de cet intervalle d'homogénéité temporelle locale. Une fois cette estimation obtenue, les données de cet intervalle sont alors utilisées pour construire des estimateurs simples de la moyenne et de la matrice de covariance (par exemple les estimateurs empiriques). On étudie ensuite la qualité des estimations adaptatives d'un point de vue théorique. Cette étude est l'occasion d'obtenir des résultats sur la qualité de divers estimateurs de la moyenne et de la matrice de covariance utilisant un horizon d'estimation fixe dans le cas stationnaire. Les performances de l'algorithme adaptatif sont aussi mesurées d'un point de vue pratique à travers des simulations numériques.

Nous utilisons alors cet algorithme adaptatif afin de déterminer l'instant de la rupture dans un modèle à saut (rupture dans la moyenne et/ou la matrice de covariance). On fournit ensuite dans la section 2.3.4 des bornes sur les erreurs de première et deuxième espèce.

Dans la section 2.4, nous présentons une autre méthode afin de déterminer l'instant de la rupture dans un modèle à saut. La qualité de l'estimateur de l'instant de la rupture alors obtenu est examinée d'un point de vue théorique et numérique.

Dans le chapitre 3, on s'attache à étudier plus en profondeur le modèle de Markowitz en essayant de rechercher une calibration de la matrice de covariance qui stabiliserait les solutions du problème par rapport à des perturbations de ses paramètres sans pour autant trop affecter les performances des portefeuilles. On étudie enfin un moyen simple de robustifier ce modèle. Plus précisément, le plan du chapitre 3 est le suivant.

Dans la section 3.2.2 quelques propriétés du modèle sont tout d'abord données.

Dans la section 3.3, on s'intéresse ensuite à une analyse de sensibilité du modèle de Markowitz qui est l'occasion de proposer une analyse de sensibilité pour une classe de problèmes quadratiques particuliers.

À partir de cette analyse de sensibilité et en utilisant les résultats du chapitre précédent, on propose dans la section 3.4, deux calibrations originales de la matrice de covariance devant permettre de rendre les solutions du problème plus stables par rapport à des perturbations de ses paramètres. L'efficacité de l'une de ces calibrations est illustrée en effectuant des tests sur des données réelles. Dans cette même section, nous nous servons aussi des estimations adaptatives des paramètres du modèle de Markowitz et des erreurs sur ces

estimations (quantifiées au chapitre 2) afin de proposer des contreparties robustes simples du problème de Markowitz.

On présente enfin, dans la section 3.5, d'autres calibrations de la matrice de covariance. Le chapitre se termine par des simulations numériques pour comparer les comportements des différentes calibrations présentées dans le chapitre 3.

La deuxième partie correspond au chapitre 4 et est consacrée à l'étude de problèmes de gestion de production. Le problème de gestion de production électrique à l'horizon annuel est étudié en détail dans les sections 4.1 à 4.8. La deuxième partie se termine par la section 4.9 dans laquelle on présente le problème de transport et de distribution du gaz. Nous donnons une propriété des solutions de ce problème et nous montrons comment robustifier ce problème en utilisant des techniques présentées dans le chapitre 1. L'étude du problème de gestion de production électrique est découpée de la façon suivante.

La section 4.1 est consacrée à la présentation du problème et à la modélisation des unités de production. Lorsque tous les paramètres du problème sont connus, le problème revient à résoudre un problème d'optimisation linéaire. On donne des conditions assurant la faisabilité de ce problème. En pratique, les paramètres sont incertains sur la période d'étude et on a donc à résoudre un problème d'optimisation stochastique.

Dans la section 4.3.2, on montre comment ce problème d'optimisation stochastique peut se résoudre par programmation stochastique multi-étapes. Contrairement aux approches employées en général pour traiter le problème de gestion de production électrique, l'approche que l'on présente dans la section 4.3.2 permet de fournir des stratégies de gestion et non des plans de production. Un certain nombre de limitations de cette approche est cependant mentionné. Les stratégies de gestion sont testées sur un certain nombre de scénarios pour les paramètres incertains. On constate qu'avec cette approche, on obtient d'une part des coûts très importants sur certains scénarios et d'autre part, un écart type des coûts de gestion simulés assez élevé. Notre objectif est alors de proposer des méthodes permettant de réduire cet écart type tout en fournissant des méthodes dont le coût moyen sur l'ensemble des scénarios est raisonnable.

Dans la section 4.4.1, on montre comment une stabilisation de la fonction duale permet de satisfaire ces contraintes. On montre ensuite dans les sections 4.4.2 et 4.4.3, comment appliquer la méthodologie de l'optimisation robuste (présentée dans le chapitre 1) au problème de gestion de production électrique. Un des atouts de ces méthodes est de ne pas reposer sur un arbre de scénarios, ce qui permet de résoudre le problème plus rapidement et ne nécessite pas, pour la méthode de stabilisation de la fonction duale, d'étude statistique approfondie.

Les contreparties robustes proposées sont des contreparties robustes ajustées (CRA) et des contreparties robustes ajustées affinement (CRAA) du problème de gestion de production électrique. Les commandes de la CRA ne peuvent pas être utilisées directement d'un point de vue pratique car la dépendance de ces commandes en fonction des paramètres incertains passés n'est pas connue explicitement. Pour utiliser la méthodologie de la CRA, nous proposons alors, pour déterminer les commandes au pas de temps t, de résoudre la CRA du problème de gestion de production électrique correspondant aux $T-t+1$ derniers pas de temps (si T est le nombre de pas de temps sur l'année). Nous appellerons cette méthode CRA glissante. Les commandes correspondant au premier pas de temps de ce

problème sont alors les commandes appliquées au pas de temps t. Dans ce contexte, on reste bien préoccupé par la recherche d'un plan de production optimal sur l'année. D'autre part, l'intérêt de cette approche est d'intégrer les informations accumulées jusqu'au pas de temps t lorsqu'on cherche à déterminer les commandes pour ce pas de temps. En particulier, lorsqu'on cherche à déterminer les commandes au pas de temps t, les commandes et états du système sont connus de façon exacte pour tous les pas de temps précédents. On connaît notamment les niveaux des réserves. Cette approche sera testée sur des données simulées.

La section 4.5 porte sur les aspects algorithmiques liés à la résolution des problèmes introduits dans les sections 4.2, 4.3.2 et 4.4. On explique comment la version déterministe du problème et le problème introduit en section 4.3.2 sont en général résolus en utilisant une méthode de décomposition par les prix revenant à dualiser les contraintes couplantes du problème. On présente ensuite une méthode efficace de résolution par programmation dynamique de la version déterministe du problème de gestion de production électrique lorsqu'on ne considère que des centrales thermiques et une centrale hydraulique. On montre alors comment en déduire une méthode de résolution efficace du sous-problème hydraulique et du sous-problème EJP pour calculer la fonction duale introduite en section 4.5.1.

La section 4.6 est dédiée à une étude statistique de la série chronologique des consommations électriques en France de 1996 à 2003. Ces données sont utilisées dans la section 4.7.2 pour comparer les différentes méthodes de gestion de production électrique présentées dans cette thèse sur des données réelles. Enfin dans la section 4.7.1, ces méthodes sont comparées sur des données simulées.

Notations et abbréviations

$\mathcal{M}_n$	ensemble des matrices carrées réelles d'ordre n,
$\mathcal{S}_n$	ensemble des matrice symétriques de $\mathcal{M}_n$,
$L_1(\mathbb{R})$	ensemble des fonctions réelles intégrables,
$L_2(\mathbb{R})$	ensemble des fonctions réelles de carré intégrables,
I_n	matrice identité d'ordre n,
$< A, B >=\mathrm{Tr}(A^*B)$	produit scalaire usuel des deux matrices A et B de $\mathcal{S}_n$,
$\|A\|$	norme de Frobeniüs de la matrice $A \in \mathcal{S}_n$ associée au produit scalaire usuel défini ci-dessus,
$\|A\|_\infty$	si $A \in \mathcal{M}_n$, $\|A\|_\infty = \max A(i,j),\ \ i = 1, \ldots, n,\ j = 1, \ldots, n$,
$A \succeq 0$	signifie que la matrice $A \in \mathcal{S}_n$ est semidéfinie positive,
$A \succ 0$	signifie que la matrice $A \in \mathcal{S}_n$ est définie positive,
$\lambda_{\min}(A)$	plus petite valeur propre de A,
$\lambda_{\max}(A)$	plus grande valeur propre de A,
$C_i(A)$	i-ème colonne de la matrice A,
A^T	transposée de la matrice A,
$a \geq_K b$	signifie que $a - b$ appartient à l'ensemble K, si $K = \mathbb{R}^n$, on notera simplement $a \geq b$,
$\|x\|_1$	si $x \in \mathbb{R}^n$, $\|x\|_1 = \sum_{i=1}^{n} x_i$,
e (en gras)	vecteur de 1,
e	$exp(0)$,
$(e_1, \ldots, e_n)$	base canonique de $\mathbb{R}^n$,,
a_+	si $a \in \mathbb{R}^n$, $a_+(i) = \max(a(i), 0),\ i = 1, \ldots, n$,
a_-	si $a \in \mathbb{R}^n$, $a_-(i) = \min(a(i), 0),\ i = 1, \ldots, n$,
$\sum_{i=a}^{b} x_i$	si $a < b$, on pose que cette somme est nulle,
$\Phi(.)$	fonction de répartition de la loi normale centrée réduite,
$E(X)$	si X est une variable aléatoire désigne l'espérance de X,
$Var(X)$	si X est une variable aléatoire désigne la variance de X,
$\sigma(X)$	si X est une variable aléatoire désigne l'écart type de X,
$E[x]$	si x est un réel désigne la partie entière de x,
Δ_n	simplexe unité de $\mathbb{R}^n$: $\Delta_n = \{x \in \mathbb{R}^n,\ \ x \geq 0,\ \ x^T\mathbf{e} \leq 1\}$,
VaR	Value-at-Risk,
CR	contrepartie robuste,
CRA	contrepartie robuste ajustée,
CRAA	contrepartie robuste ajustée affinement.

Chapitre 1

Modélisations en optimisation stochastique

1.1 Introduction

Dans un problème d'optimisation stochastique, un certain nombre de paramètres peuvent être considérés comme inconnus et sont modélisés par des variables aléatoires. Plus généralement, l'incertitude est représentée par des expériences aléatoires dont l'issue est notée ω. L'ensemble de toutes les issues possibles est l'espace fondamental Ω. On choisit une collection d'évènements $\mathcal{A}$ et une mesure de probabilité P telle que le triplet $(\Omega, \mathcal{A}, P)$ soit un espace probabilisé. Par exemple, si les éléments de Ω sont les conditions climatiques, ils permettent de décrire des variables aléatoires telles que la consommation électrique.
Dans un problème d'optimisation stochastique, on cherche souvent à prendre une décision avant que les valeurs des réalisations des variables aléatoires ne soient observées. La formulation générale de tels problèmes d'optimisation stochastique s'écrit alors :

$$\begin{cases} \min \quad Ef_0(x,\omega) \\ Ef_i(x,\omega) \leq 0, \quad i = 1,\ldots,k \\ Ef_i(x,\omega) = 0, \quad i = k+1,\ldots,m, \\ x \in X, \end{cases} \tag{1.1}$$

avec $\omega \in \Omega$, $E = E_P$ l'espérance associée à la mesure de probabilité P, X un ensemble fermé non vide et, $f_0 : \mathbb{R}^n \times \Omega \to \mathbb{R} \cup \{+\infty\}$, $f_i : \mathbb{R}^n \times \Omega \to \mathbb{R}$, $1 \leq i \leq m$, des fonctions données.
La formulation ci-dessus couvre une large variété de problèmes d'optimisation stochastique y compris les problèmes incluant des contraintes probabilistes (chance-constraints en anglais) et les modèles avec recours, introduits dans le cas linéaire par Beale and Dantzig ([3],[33]). Pour s'assurer que (1.1) est bien défini, on suppose que (i) à la fois l'espérance dans la fonction objectif et les contraintes sont finies pour tout $x \in X$, (ii) l'ensemble de faisabilité de (1.1) est non vide et (iii) un minimum de (1.1) est atteint sur son ensemble de faisabilité.
La formulation (1.1) d'un problème d'optimisation stochastique s'écrit sous la forme d'un problème d'optimisation déterministe dont la fonction objectif et les contraintes se présentent sous la forme d'espérances. Le calcul de ces espérances, pour un point $x \in X$, nécessite de calculer une intégrale multidimensionnelle des fonctions mises en jeu dans la

fonction objectif et les contraintes, ces fonctions elles-mêmes n'étant pas forcément définies explicitement. Pour surmonter ce problème, différentes solutions approchées à la fois stochastiques et déterministes ont été proposées. On peut remplacer les fonctions f_i par des fonctions plus simples ou approcher la distribution de probabilité P. Le but étant d'obtenir un problème d'optimisation traitable dont la solution soit une solution approchée du problème posé. Si P est une distribution de probabilité unidimensionnelle, on peut l'approcher par des distributions uniformes par morceaux, utiliser un estimateur à noyau de la densité ([50], [51]) ou choisir d'approcher P par une mesure de probabilité discrète. Dans le cas multidimensionnel, on approche le plus souvent la distribution de probabilité P par une distribution de probabilité discrète concentrée en un certain nombre de points $\omega_1, \ldots, \omega_s$ appelés scénarios. On peut aussi utiliser des méthodes d'échantillonnage de Monte Carlo pour approcher la fonction objectif qui est une intégrale multidimensionnelle.
L'objet de ce chapitre est de présenter différents choix de modélisation et de traitement des paramètres incertains d'un problème d'optimisation. Ces choix induisent des méthodes de résolution propres à chaque modélisation qui seront brièvement évoquées. En particulier, nous insisterons sur les aspects de modélisation que nous privilégions dans cette thèse en soulignant leurs originalités.
On commencera par présenter les problèmes d'optimisation stochastique bi-étapes avec recours puis leur généralisation à travers les problèmes d'optimisation stochastique multi-étapes. Pour la mise en oeuvre de ces méthodes, on soulignera l'importance de la construction d'un arbre de scénarios pour les paramètres incertains. Différentes méthodes de construction de ces arbres seront présentées. Dans le chapitre 4, nous étudions le problème de gestion de production électrique. L'une des méthodes proposées pour résoudre ce problème sera d'effectuer une optimisation multi-étapes sur un arbre de scénarios modélisant l'évolution des paramètres incertains sur la période de gestion (la consommation électrique, les apports en eau des réserves hydrauliques ainsi que les taux d'indisponibilité des usines). Un autre choix de modélisation consiste à inclure des contraintes probabilistes dans le problème. Les principaux résultats dans ce domaine ont trait à la forme des équivalents déterministes des problèmes d'optimisation avec contraintes probabilistes étudiés. Ces équivalents déterministes pourront être exploités si des méthodes exactes et efficaces de résolution existent pour ces équivalents. On présentera différentes situations (reprises dans le chapitre 2) pour lesquelles les équivalents déterministes sont des problèmes convexes et donc solubles, en théorie, de façon exacte en temps polynomial. On présente ensuite une modélisation alternative (de l'incertitude des paramètres d'un problème d'optimisation stochastique) basée sur l'optimisation robuste. On trouve les premières traces de cette modélisation dans [86], [87], [88], [43] et elle a été récemment activement étudiée par Ben-Tal et Nemirovski ([6], [7], [8], [9], [10], [5], [38], [11]). Dans ce contexte, on suppose que les paramètres incertains appartiennent à un certain ensemble d'incertitude $\mathcal{Z}$. Une optimisation pire cas est alors envisagée, visant à prendre une décision qui sera optimale si le pire scénario se produit et qui satisfera les contraintes du problème quelles que soient les réalisations des paramètres incertains dans l'ensemble $\mathcal{Z}$. Pour des problèmes d'optimisation linéaires, coniques et semidéfinis positifs, on peut obtenir des contreparties robustes traitables ou des approximations de ces contreparties robustes traitables pour des ensembles d'incertitude $\mathcal{Z}$ que nous rappellerons. On s'intéresse ensuite aux contreparties robustes ajustées et aux contreparties robustes ajustées affinement qui peuvent être vue comme un prolongement des contreparties robustes classiques présentées dans la section 1.5. Les méthodes d'optimisation robustes présentées dans ce chapitre seront utilisées pour deux problèmes de gestion

de production et notamment le problème de gestion de production électrique dans le chapitre 4. Enfin, on termine ce chapitre en spécifiant les caractéristiques des modélisations développées dans cette thèse.

1.2 Modèle bi-étapes avec recours

Dans cette section et dans la suivante, lorsqu'interviendront des réalisations de variables aléatoires représentées par des lettres grecques, les variables aléatoires correspondantes seront notées avec la même lettre surmontée d'une tilde.
Dans un problème d'optimisation stochastique bi-étapes avec recours, une première décision doit être prise avant la réalisation des paramètres incertains (modélisés par un vecteur aléatoire). Lorsqu'une réalisation de ce vecteur aléatoire est connue, une décision de recours doit être mise en place ('second stage decision' en anglais). Les deux décisions sont soumises à des contraintes et un coût est associé à chacune de ces décisions. Il s'agit alors de prendre une décision à la premiére étape (avant que ne soient connus les paramètres incertains du problème) telle que la valeur moyenne du coût global (coût dû aux deux prises de décision) soit optimisée. Nous donnons un exemple classique ([17]) d'optimisation stochastique bi-étapes avec recours afin d'illustrer ce processus. Dans cet exemple, un vendeur de journaux achète chaque jour à l'éditeur x journaux au prix unitaire c. La quantité maximale de journaux qu'il peut acheter est limitée par son pouvoir d'achat et par les quantités maximales que l'éditeur s'autorise à vendre à chaque client. On désignera par u un majorant du nombre x de journaux pouvant être achetés. Autant de journaux que possible sont vendus dans la journée au prix unitaire q. Les journaux non vendus en fin de journée peuvent être revendus à l'éditeur au prix $r < c$. Dans ce problème, c'est le nombre de clients qui achèteront effectivement un journal au vendeur qui est inconnu. Ce nombre est modélisé par une variable aléatoire $\tilde{\xi}$. Soit alors y le nombre de journaux effectivement vendus dans la journée et w le nombre de journaux retournés à l'éditeur en fin de journée. Le nombre x de journaux à acheter à l'éditeur en début de journée est une décision qui doit être prise avant que ne soit connu le nombre exact de clients qui seront intéressés pour acheter un journal. Lorsque le nombre effectif de clients intéressés est connu, une décision de recours peut être prise dans une deuxième étape. Si la demande en journaux ξ était connue, alors pour x journaux achetés, $y^*(\xi) = \min(\xi, x)$ journaux seront vendus et $w^*(\xi) = \max(x - \xi, 0)$ seront retournés à l'éditeur induisant ainsi un coût $\mathcal{Q}(x, \xi) = -q \min(\xi, x) - r \max(x - \xi, 0)$ (ou un profit $-\mathcal{Q}(x, \xi)$). Le nombre de journaux x acheté devra permettre de minimiser le coût moyen global qui est la somme du coût d'achat cx et du coût moyen dû aux ventes de journaux $\mathcal{Q}(x) = E_{\tilde{\xi}}\mathcal{Q}(x, \tilde{\xi})$. Le problème revient donc à trouver x solution du problème :

$$\begin{cases} \min \ \ cx + \mathcal{Q}(x) \\ 0 \leq x \leq u. \end{cases} \tag{1.2}$$

$\mathcal{Q}$ est une fonction convexe, continue et différentiable lorsque $\tilde{\xi}$ est une variable aléatoire continue. Soit F la fonction de répartition de $\tilde{\xi}$. Notons que $F(0) = 0$. La solution du problème (1.2) est alors donnée par $x^* = u$ si $\frac{q-c}{q-r} > F(u)$ et par $x^* = F^{-1}(\frac{q-c}{q-r})$ sinon. On voit donc qu'une information d'ordre statistique (l'évaluation de la fonction de répartition de $\tilde{\xi}$) est nécessaire pour résoudre le problème. Nous insisterons dans cette thèse sur le développement d'outils statistiques tant pour estimer des paramètres incertains d'un

problème d'optimisation stochastique que pour calibrer des ensembles d'incertitude d'une contrepartie robuste du problème étudié.
D'une manière plus générale, un problème d'optimisation stochastique bi-étapes avec recours consiste à minimiser $f_0(x) + E_{\tilde{\xi}}\mathcal{Q}(x,\tilde{\xi})$ sur un ensemble convexe fermé $X \subseteq \mathbb{R}^{n_1}$; le coût de la décision de recours pour un x fixé et une réalisation ξ des paramètres incertains étant donné par :

$$\mathcal{Q}(x,\xi) = \begin{cases} \min & f_1(x,y,\xi) \\ & y \in X'(x,\xi), \quad y \in \mathbb{R}^{n_2}_+; \end{cases} \tag{1.3}$$

avec $f_0 : \mathbb{R}^{n_1} \to \mathbb{R}$, $f_1 : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}$, des fonctions données et X' un ensemble dépendant de x et de la réalisation ξ des paramètres incertains. Initialement, ce sont les problèmes d'optimisation stochastique bi-étapes linéaires avec recours qui ont été étudiés (par Beale [3] et Dantzig [33]). Il permettent, en pratique, de modéliser une grande variété de problèmes d'optimisation stochastique bi-étapes réels et consistent à minimiser :

$$E_{\tilde{\xi}}(c^T x + \mathcal{Q}(x,\tilde{\xi})), \quad \text{sur l'ensemble} \quad \{x \in \mathbb{R}^{n_1}_+ \mid Ax = b\}; \tag{1.4}$$

la fonction de recours $\mathcal{Q}(x,\xi)$ étant définie pour x et ξ fixés par la valeur optimale du problème d'optimisation auxiliaire de la deuxième étape :

$$\mathcal{Q}(x,\xi) = \begin{cases} \min & q(\xi)^T y \\ & W(\xi)y + T(\xi)x = h(\xi), \quad y \in \mathbb{R}^{n_2}_+, \end{cases} \tag{1.5}$$

où W est la matrice de recours. En pratique, on considère que les paramètres incertains prennent un ensemble de valeurs décrit par les scénarios : $\xi_s = [q_s, T_s, W_s, h_s]$, $1 \le s \le S$, où le scénario s est de probabilité $0 \le p_s \le 1$ ($\sum_{s=1}^{S} p_s = 1$). On cherchera alors à résoudre :

$$\min \quad c^T x + \sum_{s=1}^{S} p_s \, q_s^T y_s \tag{1.6}$$

$$\text{sous les contraintes} \quad \begin{array}{lllllllll} Ax & & & & & & & = & b \\ T_1 x & + & W_1 y_1 & & & & & = & h_1 \\ T_2 x & + & & & W_2 y_2 & & & = & h_2 \\ \vdots & & & & & \cdots & & \vdots & \\ T_S x & + & & \cdots & & + & W_S y_S & = & h_S \end{array} \tag{1.7}$$

$$x \ge 0, \quad y_s \ge 0, \quad s = 1, \ldots, S.$$

La résolution de ce problème d'optimisation fournit la décision x à prendre avant la réalisation des paramètres incertains et donne ensuite, lorsque se réalise le scénario s, $(1 \le s \le S)$, une décision de recours y_s.

1.3 Modèle multi-étapes avec recours

Le problème d'optimisation stochastique bi-étapes avec recours se généralise naturellement en un problème d'optimisation multi-étapes pour prendre en compte la dynamique temporelle d'un processus de prises de décisions complexes dans un environnement incertain. Dans ce cas, l'incertitude est modélisée par une suite de vecteurs aléatoires formant un processus stochastique.

1.3.1 Formulation du modèle

Considérons à présent le cas d'un problème d'optimisation stochastique à T étapes pour lequel une suite de décisions $(x_t)_{t=1}^T$ doit être prise dans un environnement incertain représenté par un processus stochastique $(\tilde{\xi}_t)_{t=1}^T$. A l'étape t, la décision $x_t \in \mathbb{R}^{n_t}$ est prise en ayant seulement connaissance des décisions passées $x_1, \ldots, x_{t-1}$ et des réalisations $(\xi_1, \ldots, \xi_t)$ des vecteurs aléatoires. Il s'agit alors, pour déterminer x_t, de minimiser l'espérance conditionnelle d'une fonction objectif $f_t(x_1, \ldots, x_t, \tilde{\xi}_1, \ldots, \tilde{\xi}_{t+1})$ au vu des réalisations $(\xi_1, \ldots, \xi_t)$ de $(\tilde{\xi}_1, \ldots, \tilde{\xi}_t)$. La décision x_t est soumise à des contraintes dépendant des décisions passées $(x_1, \ldots, x_{t-1})$ et des réalisations $(\xi_1, \ldots, \xi_t)$. On suppose que $\tilde{\xi}_1$ est un vecteur aléatoire dégénéré prenant la valeur ξ_1 avec la probabilité 1 et que la loi de probabilité gouvernant l'évolution de $(\tilde{\xi}_t)_{t=1}^T$ est connue et ne dépend pas de $(x_t)_{t=1}^T$. Reprenant notamment les notations de [44], soit $\xi^t = (\xi_1, \ldots, \xi_t)$ l'historique des paramètres incertains jusqu'au pas de temps t, et $x^t = (x_1, \ldots, x_t)$ l'historique des décisions jusqu'au pas de temps t. Un problème d'optimisation stochastique à T étapes se formule alors ainsi :

$$\begin{cases} \min E[f_1(x_1, \tilde{\xi}^2)|\xi^1] \\ x_1 \in X_1(\xi^1), \end{cases} \tag{1.8}$$

où pour $t = 2, \ldots, T-1$:

$$f_{t-1}(x^{t-1}, \xi^t) = \begin{cases} \min\limits_{x_t} E[f_t(x^{t-1}, x_t, \tilde{\xi}^{t+1})|\tilde{\xi}^t = \xi^t] \\ x_t \in X_t(x^{t-1}, \xi^t), \end{cases} \tag{1.9}$$

et,

$$f_{T-1}(x^{T-1}, \xi^T) = \begin{cases} \min\limits_{x_T} f_T(x^{T-1}, x_T, \xi^T)] \\ x_T \in X_T(x^{T-1}, \xi^T). \end{cases} \tag{1.10}$$

Les fonctions f_t, $t = 1, \ldots, T$ sont des fonctions à valeurs réelles. L'ensemble de contraintes X_t dépend à la fois de l'histoire de x^{t-1} et de $\tilde{\xi}^t$. On trouvera dans [30] des conditions assurant que (1.8)-(1.10) est bien défini.

1.3.2 Méthodes de résolution

En général, l'évolution des paramètres incertains est représenté par un arbre de scénarios. Comme nous l'avons fait pour les problèmes d'optimisation stochastique bi-étapes linéaires, nous donnons alors l'équivalent déterministe d'un problème d'optimisation stochastique multi-étapes linéaire lorsque les réalisations de $(\tilde{\xi}_t)_{t=1}^T$ forment un arbre de scénarios représentant toutes les évolutions possibles de $(\tilde{\xi}_t)_{t=1}^T$. On se limite à des arbres de scénarios finis i.e que pour tout $t = 1, \ldots, T$, les réalisations de $\tilde{\xi}^t$ prennent un nombre fini de valeurs $(\xi^{t,1}, \ldots . \xi^{t,k_t})$. Au pas de temps t, l'arbre de scénarios comprend alors k_t noeuds et pour ce pas de temps, le scénario père du scénario $\xi^{t,s}$ (pour $1 \leq s \leq k_t$) est noté $\xi^{t-1,P_t(s)}$. Sous ces hypothèses, pour $t = 1, \ldots, T$, en notant $p_t^s = P(\tilde{\xi}^t = \xi^{t,s})$, $1 \leq s \leq k_t$, on peut alors exprimer un problème d'optimisation stochastique multi-étapes linéaire sous la forme du problème déterministe équivalent suivant :

$$\begin{cases} \min \sum\limits_{t=1}^{T} \sum\limits_{s=1}^{k_t} p_t^s \, q_t^{sT} \, x_t^s \\ T_t^s x_{t-1}^{P_t(s)} + W_t^s x_t^s = h_t^s, \ s = 1, \ldots, k_t, \ t = 1, \ldots, T \\ x_t^s \geq 0, \ s = 1, \ldots, k_t, \ t = 1, \ldots, T, \end{cases} \tag{1.11}$$

avec $T_1^1 x_0^{P_1(1)} = 0$ et $x_{t-1}^{P_t(s)}$ est le vecteur de décision au pas de temps $t-1$ pour le noeud père de la feuille associée au scénario $\xi^{t,s}$. Si le nombre d'étapes et de scénarios intervenant dans le problème (1.11) n'est pas trop grand, ce problème pourra être résolu de manière exacte en utilisant des méthodes de points intérieurs efficaces par exemple.
Pour des problèmes d'optimisation stochastique multi-étapes de petites tailles, il peut aussi exister une solution analytique (comme pour le problème présenté dans la section précédente pour illustrer le fonctionnement d'un problème d'optimisation stochastique bi-étapes). D'autres problèmes de petite taille contenant peu de variables et des contraintes simples peuvent être résolus par programmation dynamique ([4],[12]). Dans le chapitre 4, nous présentons une méthode de résolution d'un problème d'optimisation stochastique multi-étapes de grande taille par programmation dynamique. Lorsque le nombre d'étapes et de scénarios conduit à un équivalent déterministe de très grande taille, des méthodes prenant efficacement en compte la structure particulière du problème étudié ont été développées afin de le résoudre. Il s'agit essentiellement de méthodes de décomposition. Les premières de ces méthodes décomposent le problème par étape et un ensemble de sous problèmes d'optimisation est ensuite à résoudre en chaque noeud de l'arbre de scénarios. La méthode L-shaped ([91]) visant à approcher la partie non linéaire de la fonction objectif d'un problème d'optimisation stochastique bi-étapes ou multi-étapes, entre dans cette catégorie. D'autres méthodes décomposent le problème par scénarios en utilisant, par exemple, une relaxation lagrangienne des contraintes non anticipatives, ce qui revient à résoudre un problème déterministe pour chaque scénario de l'arbre. Le problème de gestion de production électrique étudié dans le chapitre 4, sera en particulier résolu par relaxation lagrangienne.
La génération d'arbre de scénarios est donc une étape importante dans le processus de résolution d'un problème d'optimisation stochastique multi-étapes. Nous abordons ce problème dans la section suivante.

1.3.3 Génération d'arbres de scénarios

Nous indiquons ici les grands axes utilisés pour la génération d'arbres de scénarios (pour un développement de ces notions voir par exemple [36], [78]). Dans le cas où la distribution de probabilité P du problème (1.1) est connue, les scénarios peuvent être obtenus par échantillonnage suivant la distribution de probabilité P (voir [23] ou [92] dans le cas normal, [74] dans le cas uniforme).
On peut aussi choisir entre différentes distributions de probabilité, différents modèles paramétriques, celui qui paraît le plus vraissemblable. Pour chaque modèle testé, on recherche la distribution de probabilité parmi une famille de lois $\mathcal{P}$ paramétrées par $\theta : \mathcal{P} = \{\mathcal{P}_\theta, \theta \in \Theta\}$. Une étude statistique doit permettre alors d'estimer dans un premier temps le paramètre θ du modèle par une valeur $\hat{\theta}$ en utilisant les historiques disponibles donnant l' évolution passée des paramètres incertains. Typiquement ce paramètre est obtenu par maximisation de la vraissemblance. On pourra ensuite choisir le modèle permettant d'assurer le meilleur compromis entre complexité et explication de l'évolution des paramètres passés. Dans cet objectif différents critères tels que les critères AIC ou BIC pourront être utilisés afin de choisir un des modèles. Une fois le modèle choisi, les noeuds fils d'un noeud donné sont obtenus par échantillonnage suivant la distribution de probabilité correspondant au modèle retenu (voir [85] par exemple). Typiquement, le modèle reliera la variable aléatoire à l'instant t, X_t à des observations passées $X_t^{k_1} = (X_{t-1}, \ldots, X_{t-k_1})$ et à des réalisations

$\varepsilon_t^{k_2} = (\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-k_2})$ d'un bruit ε_t de loi connue :

$$X_t = f(X_t^{k_1}) + g(\varepsilon_t^{k_2}). \tag{1.12}$$

En un noeud donné, les valeurs X_t de X pour les noeuds fils seront obtenues en déterminant un échantillon de valeurs pour la valeur ε_t du bruit à l'instant t, et en appliquant le modèle paramétrique (1.12) pour lequel les fonctions f et g sont remplacées par les estimations obtenues lors de phase d'estimation des paramètres du modèle. On assigne dans ces conditions la même probabilité de transition vers chacun des fils. L'utilisation de quantiles de la distribution estimée pour X_t permet d'obtenir éventuellement des probabilités de transition différentes.
On peut aussi, en chaque noeud de l'arbre de scénarios, choisir les descendants en déterminant, pour les noeuds fils, un échantillon de valeurs ayant même moments jusqu'à un certain ordre (disons 4) que les moments empiriques estimés à l'aide des historiques disponibles. Une mise en oeuvre de cette démarche est proposée dans [60]. Plus généralement, les historiques disponibles, peuvent être utilisés afin de déterminer une distribution de probabilité empirique.
Enfin dans le cas où peu ou pas d'historiques sont disponibles, on peut recourir à des prévisions données par des experts ayant une connaissance fine des données susceptibles d'expliquer les valeurs futures des paramètres que l'on cherche à estimer. On trouve une utilisation d'informations fournies par des experts en finance pour construire des scénarios pour un modèle financier dans [83]. Des experts économiques peuvent être utilisés pour fournir des prévisions sur la demande pour des problèmes de gestion de production ([68], [42]).
Si l'arbre de scénarios, obtenu à l'aide d'une première analyse, est de trop grande taille, on peut utiliser des techniques de réduction du nombre de scénarios ([37],[58]) en agrégeant les scénarios proches. L'idée est de comparer la distance entre les distributions de probabilité de l'arbre initial et de l'arbre réduit, et de supprimer les scénarios si l'arbre réduit est encore assez proche de l'arbre initial. Actuellement, en optimisation stochastique, un intérêt croissant est porté sur la construction de modèles appropriés pour la distribution des consommations électriques. Nous apporterons une contribution dans ce domaine en étudiant la série chronologique des consommations hebdomadaires d'électricité en France de 1996 à 2003 inclus.

1.4 Contraintes probabilistes

Nous nous intéressons dans cette section aux problèmes d'optimisation pour lesquels les contraintes doivent être satisfaites avec une probabilité fixée à l'avance. Ceci suppose que les contraintes peuvent être violées pour certains évènements élémentaires ω. La méthode de résolution de tels problèmes consiste en général à déterminer un équivalent déterministe du problème initial. Avant de donner les équivalents déterministes de différents types de contraintes probabilistes, nous présentons la forme des fonctions objectifs et contraintes que l'on rencontre typiquement dans ce genre de situations.

1.4.1 Différents types de contraintes probabilistes et de fonction coût

Un problème d'optimisation avec contraintes probabilistes consiste à minimiser une fonction f sous les contraintes $P(f_i(x,\omega) \leq 0) \geq \alpha_i,\ i = 1, \ldots, m$, avec pour $i =$

$1, \ldots, m$, $0 \leq \alpha_i \leq 1$ et $f_i(x, \omega) : \mathbb{R}^n \times \Omega \to \mathbb{R}$ des fonctions données. La fonction objectif à minimiser appartiendra typiquement à l'une des trois classes suivantes :

- la classe des E-modèles (la fonction objectif est une espérance) : $f(x) = E(c(\omega)^T x)$;
- la classe des V-modèles (la fonction objectif est une caractéristique de dispersion comme la variance) : $f(x) = E((c(\omega) - Ec(\omega))^T x)^2$ ou $f(x) = E(c(\omega)^T x - c_0^T x_0)^2$ avec en général $c(\omega)^T x \neq c_0^T x_0$;
- la classe des P-modèles (la fonction objectif est une probabilité) : $f(x) = P(c(\omega)^T x \geq c_0)$.

Une variante du P-modèle (présentée dans [62]) consiste à minimiser γ de telle sorte que la contrainte $P(c(\omega)^T x \leq \gamma) = 1 - \varepsilon$ ou $P(c(\omega)^T x \leq \gamma) \geq 1 - \varepsilon$ ($\varepsilon \leq \frac{1}{2}$) soit satisfaite et soit donc rajoutée aux autres contraintes que doit satisfaire la solution. Dans les problèmes d'optimisation avec contraintes probabilistes les paramètres ne sont donc pas connus avec exactitude. Cette approche nécessite d'émettre des hypothèses sur la distribution de probabilité des variables aléatoires associées aux paramètres.

1.4.2 Equivalents déterministes

Nous allons préciser différents ensembles de contraintes probabilistes conduisant à des problèmes d'optimisation convexes. Un problème d'optimisation convexe est un cas particulier de problème d'optimisation soluble en temps polynomial. La classe des problèmes d'optimisation convexes a été largement étudiée et une large gamme d'algorithmes efficaces existe pour la résolution de tels problèmes, différentiables ou non. Certains, comme la méthode des faisceaux ([65]) ou l'ACCPM, peuvent résoudre n'importe quel type de problème convexe non différentiable (pourvu que la valeur de la fonction objectif ainsi qu'un sous-gradient soient connus en chaque point). Enfin d'autres méthodes sont dédiées à la résolution de problèmes convexes bien spécifiques : le simplexe pour la programmation linéaire ou les méthodes de points intérieurs utilisant des fonctions barrières adaptées au problème qu'elles permettent de résoudre par exemple. Par conséquent, d'un point de vue pratique, un équivalent déterministe ou une approximation de cet équivalent sera exploitable s'il définit un problème convexe (ou quasi-convexe). Notons que si $F_{i,x}(.)$ est la fonction de répartition de $f_i(x, \omega)$ et si pour $p \in [0, 1]$,

$$F_{i,x}^{-1}(p) = \inf\{y \mid F_{i,x}(y) \geq p\},$$

alors la contrainte $P(f_i(x, \omega) \leq 0) \geq \alpha_i$ peut se réécrire $F_{i,x}^{-1}(\alpha_i) \leq 0$. Quand bien même la fonction $F_{i,x}$ est connue, l'ensemble de faisabilité ainsi défini n'est pas forcément convexe. Nous donnons quelques cas où l'ensemble de contraintes est convexe. Commençons par considérer un ensemble de contraintes définies par :

$$\{x \mid P(\sum_{j=1}^{n} a_{ij}(\omega) x_j \geq b_i(\omega)) \geq \alpha_i,\ 0 \leq \alpha_i \leq 1,\ 1 \leq i \leq m\}. \tag{1.13}$$

L'ensemble défini par ces contraintes n'est pas nécessairement convexe (voir [17] p.104). Dans le cas où la matrice $A(\omega) = (a_{ij}(\omega))_{ij} = (a_{ij})_{ij}$ est fixe, et si F_i est la fonction de répartition de b_i alors les contraintes énoncées en (1.13) se réécrivent sous la forme compacte $Ax \geq b^*$ (et définissent donc un ensemble convexe) avec $b^*(i) = F_i^{-1}(\alpha_i)$. Dans le cas où les contraintes (1.13) doivent être satisfaites conjointement et si la matrice $A(\omega) = A$ est fixe, nous pouvons les écrire sous la forme $K(\alpha) = P(Ax \geq b(\omega)) \geq \alpha$ où $0 \leq \alpha \leq 1$. Dans

ce cas, il existe un vaste ensemble de mesures de probabilité sur $b(\omega)$ qui définissent un ensemble $K(\alpha)$ convexe.

Définition 1.4.1 *Nous dirons qu'une mesure de probabilité P est une mesure quasi-concave si pour tous ensembles mesurables U et V et pour tout $0 \leq \lambda \leq 1$,*

$$P((1-\lambda)U + \lambda V) \geq \min\{P(U), P(V)\}. \tag{1.14}$$

On trouve alors dans [17], le théorème suivant :

Théorème 1.4.1 *[17] Si A est fixe et si la mesure de probabilité associée à b est quasi-concave alors $K(\alpha)$ est un ensemble fermé et convexe pour $0 \leq \alpha \leq 1$.*

Prékopa a étudié une classe de mesure de probabilité plus restreinte conduisant néanmoins à un ensemble $K(\alpha)$ convexe. C'est la classe des mesures de probabilité logarithmiquement concaves.

Définition 1.4.2 *Une mesure de probabilité P est une mesure logarithmiquement concave si pour tous boréliens U et V et pour tout $0 \leq \lambda \leq 1$,*

$$P((1-\lambda)U + \lambda V) \geq P(U)^{\lambda} P(V)^{1-\lambda}. \tag{1.15}$$

Les distributions de Weibull, Normale multivariée, beta et de Dirichlet, trés utilisées en pratique, sont logarithmiquement concaves.

Théorème 1.4.2 *[17] p.105. Si A est fixe, si les composantes b_i de b sont des variables aléatoires indépendantes de fonction de répartition $F_i(x)$ et dont les mesures de probabilité associées P_i sont logarithmiquement concaves alors $K(\alpha) = \{x \mid \sum_{i=1}^{m} \ln(F_i(A_i^T x)) \geq \ln \alpha\}$ est convexe.*

Dans [71], un cas particulier de ce résultat a été présenté. Il s'agissait de rajouter au système de contraintes $Ax \leq b, x \geq 0$, la contrainte $P(b > b^*) \geq \alpha$ qui définit un ensemble convexe d'après ce qui précède.
Considérons les contraintes (1.13) dans lesquelles les coefficients a_{ij} de la matrice des contraintes et du second membre b_i sont gaussiens :

$$a_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2), \quad \text{et} \quad b_i \sim \mathcal{N}(\eta_i, \theta_i^2).$$

On suppose que $0.5 \leq \alpha_i \leq 1$ et on définit les corrélations $v_{ij} = E(b_i(\omega) - \eta_i)(a_{ij}(\omega) - \mu_{ij})$ et $v_{ijk} = E(a_{ij}(\omega) - \mu_{ij})(a_{ik}(\omega) - \mu_{ik})$. Notons que la variance σ_{ij}^2 de $a_{ij}(\omega)$ est alors $v_{ijj} = \sigma_{ij}^2$. Dans ce contexte, [61] montre alors que les contraintes (1.13) s'écrivent

$$\Phi^{-1}(\alpha_i) \left(\sum_{j,k} v_{ijk} x_j x_k + 2 \sum_{j} v_{ij} x_j + \theta_i^2 \right)^{\frac{1}{2}} \leq -\eta_i + \sum_{j=1}^{n} \mu_{ij} x_j, \quad i = 1, \ldots, m. \tag{1.16}$$

Puisque la forme quadratique est définie positive, l'ensemble de contraintes (1.16) est bien convexe. Au jeu de contraintes qui vient d'être étudié, Charnes et Cooper dans [27], [28] et [29] ajoutent le jeu de contraintes $Db(\omega) = x, \ x \geq 0$ (où D est une matrice connue de taille

(n, m)) et fournissent des équivalents déterministes convexes dans le cas d'un E-modèle, d'un V-modèle et d'un P-modèle.
Nous terminons ces développements sur les problèmes incluant des contraintes probabilistes par un jeu de contraintes qui sera réutilisé dans le chapitre 2 pour une application en finance. Dans la situation qui nous préoccupe ici, le coût induit par le choix d'un vecteur x est linéaire et aléatoire et vaut : $c(\omega)^T x$, lorsque la réalisation du vecteur des coûts est $c(\omega)$. On se fixe alors un niveau de confiance ε qui sera typiquement petit et que nous supposons compris entre 0 et 0.5 au sens large. Nous cherchons alors à déterminer un vecteur $x \in X$, X étant un ensemble convexe fermé, et une valeur γ^* de sorte que γ^* soit la plus petite valeur telle que l'on s'assure avec une probabilité $1 - \varepsilon$ (ou au moins $1 - \varepsilon$) que le coût obtenu en choisissant le vecteur x est inférieur à γ^*. Ainsi, on a de grandes chances que le coût réel soit effectivement inférieur à γ^*. Mathématiquement nous cherchons donc ici à résoudre le problème suivant :

$$\begin{cases} \min\ \gamma \\ P(c(\omega)^T x \leq \gamma) \geq 1 - \varepsilon, \\ x \in X. \end{cases} \tag{1.17}$$

Dans le cas où $c(\omega)$ est un vecteur Gaussien de moyenne $\bar{c}$ et de matrice de covariance Q, Kataoka ([62]) a alors montré que l'équivalent déterministe de ce problème est convexe.

Théorème 1.4.3 *[62] Si* $0 \leq \varepsilon \leq 0.5$, *et* $c(\omega) \sim \mathcal{N}(\bar{c}, Q)$, *alors l'équivalent déterministe du problème* (1.17) *est :*

$$\begin{cases} \min\ \bar{c}^T x + \Phi^{-1}(1 - \varepsilon)\sqrt{x^T Q x} \\ x \in X. \end{cases} \tag{1.18}$$

Le problème ci-dessus est bien convexe puisque pour $0 \leq \varepsilon \leq 0.5$, on a $\Phi^{-1}(1 - \varepsilon) \geq 0$ et Q étant une matrice de variance covariance est semidéfinie positive. Notons que dans la modélisation ci-dessus, l'objectif est de minimiser le coût lié au choix d'un vecteur x. On peut reprendre la modélisation ci-dessus en supposant que notre objectif est plutôt de déterminer un vecteur x maximisant un certain profit. Dans ce cas le vecteur $c(\omega)$ est un vecteur de gain. On cherche alors naturellement à trouver la valeur maximale de gain γ telle que la probabilité que le gain $c(\omega)^T x$ soit supérieure à γ est supérieure à $1 - \varepsilon$: $\max \gamma,\ |\ P(c(\omega)^T x \geq \gamma) \geq 1 - \varepsilon$. Mathématiquement, on passe aisément d'une formulation à l'autre par changement de signe, puisque maximiser un gain $c(\omega)^T x$ revient à minimiser les pertes $-c(\omega)^T x$. Bien qu'ancienne, cette modélisation a connu un regain d'intérêt dans les années 1990 dans le domaine financier. Nous reviendrons sur cette modélisation à la fois pour l'analyser, préciser son implémentation et la tester sur des problèmes de gestion de portefeuilles dans le chapitre 2.

1.5 Contreparties robustes

1.5.1 Présentation

L'optimisation robuste est un choix de modélisation (voir notamment [87], [6], [7], [8], [40]) doté d'outils pour le mettre en oeuvre, afin de résoudre des problèmes d'optimisation dont une partie des données est incertaine. On suppose que ces données peuvent prendre

n'importe quelle valeur dans un ensemble d'incertitude $\mathcal{Z}$. Plus précisément, un problème d'optimisation incertain peut être défini comme une famille

$$\mathcal{P}_{\mathcal{Z}} = \left\{ \min_{x \in \mathbb{R}^n} f_0(x,\xi) \quad : \quad F(x,\xi) \in K, \;\; x \in X \right\}_{\xi \in \mathcal{Z}}, \tag{1.19}$$

de problèmes d'optimisation usuels (appelés instances) paramétrés par le vecteur $\xi \subset \mathbb{R}^M$ variant dans un ensemble d'incertitude donné $\mathcal{Z} \subset \mathbb{R}^M$ (les M paramètres sont inconnus à l'avance mais prennent des valeurs finies). Par la suite, nous supposons que $\mathcal{Z}$ est un ensemble non vide, convexe et compact et que pour tout $\xi \in \mathcal{Z}$, $f_0(x,\xi)$ est convexe en x. Nous faisons également les hypothèses suivantes :

Hypothèse A :

- $K \subset \mathbb{R}^m$ est un cône convexe fermé, d'intérieur non vide ;
- $X \subset \mathbb{R}^n$ est un ensemble convexe fermé, d'intérieur non vide ;
- $F(x,\xi) : X \times \mathcal{Z} \to \mathbb{R}^m$ est une application continuement différentiable sur $X \times \mathcal{Z}$, K concave en x, i.e :

  $$\forall (x_1, x_2 \in X, \; \xi \in \mathcal{Z}) \;\; \forall (\lambda \in [0,1]) : \quad F(\lambda x_1 + (1-\lambda) x_2, \xi) \geq_K \lambda F(x_1,\xi) + (1-\lambda) F(x_2,\xi);$$

  et K-concave en ξ i.e :

  $$\forall (x \; \in X, \; \xi_1, \xi_2 \in \mathcal{Z}) \;\; \forall (\lambda \in [0,1]) : \quad F(x, \lambda \xi_1 + (1-\lambda)\xi_2) \geq_K \lambda F(x,\xi_1) + (1-\lambda) F(x,\xi_2).$$

Ce contexte couvre le cas (traité par la programmation multi-étapes par arbre de scénarios) où les paramètres incertains sont, par nature, la réalisation de variables aléatoires comme la demande future, les prix de vente futurs. Il couvre aussi le cas où les paramètres, bien qu'incertains, ne sont pas aléatoires par nature. Citons l'exemple d'un paramètre correspondant à une mesure physique (température, pression, etc,....) obtenue par un appareil fournissant la mesure de façon approchée. Dans le cadre de la programmation robuste, afin de s'immuniser contre les variations des paramètres dans l'ensemble $\mathcal{Z}$, on recherche une solution x satisfaisant les contraintes :

$$x \in X, \;\; F(x,\xi) \in K, \quad \forall \xi \in \mathcal{Z}. \tag{1.20}$$

Ceci sous entend que l'on considère que les contraintes du problème sont dures et qu'on ne peut pas tolérer leurs violations. On dira qu'un vecteur $x \in \mathbb{R}^n$ est robuste faisable s'il satisfait les contraintes (1.20) ci-dessus. Si x est robuste faisable, le critère ne pourra pas dépasser une valeur dite garantie définie par :

$$\sup_{\xi \in \mathcal{Z}} f_0(x,\xi).$$

On définira alors une solution robuste du problème $\mathcal{P}_{\mathcal{Z}}$ comme un vecteur robuste faisable fournissant la plus petite valeur garantie et qui est donc une solution du problème suivant :

$$(CR) \;\; \min \left\{ \sup_{\xi \in \mathcal{Z}} f_0(x,\xi) \quad : \quad x \in X, \quad F(x,\xi) \in K, \quad \forall \xi \in \mathcal{Z} \right\}. \tag{1.21}$$

Ce problème est la contrepartie robuste du problème d'optimisation $\mathcal{P}_{\mathcal{Z}}$. On est donc certain, en choisissant une solution robuste, que la valeur du critère qui sera effectivement

obtenue pour la valeur exacte de ξ, sera inférieure à la valeur optimale de la contrepartie robuste (1.21). Remarquons, que le modèle (1.21) correspond au cas où le vecteur décisionnel x doit être complètement spécifié avant que les vraies valeurs des paramètres ne soient connues. D'autre part, cette modélisation est particulièrement adaptée aux situations où l'on ne peut pas tolérer de violations des contraintes. C'est le cas par exemple de contraintes de satisfaction de la demande, ou de contraintes de charge pour la construction d'un pont.

Les principaux problèmes posés en optimisation robuste consistent d'une part à déterminer des ensembles d'incertitude raisonnables et adaptés aux problèmes étudiés et d'autre part à reformuler la contrepartie robuste (1.21) (ou à défaut une approximation de cette contrepartie robuste) par un problème d'optimisation traitable. Les premiers travaux portant sur ces méthodes d'optimisation robuste sont ceux de Soyster ([86], [87], [88]) et Falk ([43]). Dans ces articles, Soyster étudie notamment un problème du type $\sup_{x\in\mathbb{R}^n} \inf_{c\in C} c^T x$, sous les contraintes $x \geq 0$ et

$$\sum_{i=1}^{n} x_i a_i \leq b, \quad a_i \in K_i, \tag{1.22}$$

où $b \in \mathbb{R}^m$ et les ensembles C et $K_i \subset \mathbb{R}^m$ sont convexes. En notant $\bar{a}_j = (\bar{a}_j(1), \ldots, \bar{a}_j(m))^T$ avec $\bar{a}_j(i) = \sup_{a_j \in K_j} a_j(i)$, $i = 1, \ldots, m$, et si $\bar{A}$ est la matrice $\bar{A} = (\bar{a}_1, \ldots, \bar{a}_n)$ alors les contraintes (1.22) se réécrivent sous la forme $\bar{A}x \leq b$ et définissent donc un ensemble convexe. Dans [87], Soyster montre comment une solution optimale du problème dual du problème robuste ci-dessus peut être utilisée pour construire un programme linéaire dont une solution est la solution optimale du problème robuste. [43] a prolongé cette étude pour montrer comment déterminer, d'un point de vue algorithmique, la solution robuste optimale. Notons que si $\bar{a}_j(i) = +\infty$, on doit avoir $x_j = 0$ (remarquons que cette situation ne se produit pas si K_j est un convexe, fermé). Néanmoins, en général, ce sont plutôt les paramètres d'une même contrainte et non d'une même colonne dans la matrice des contraintes qui sont liés entre eux. La robustification d'un problème d'optimisation linéaire telle qu'elle vient d'être présentée est très conservative dans le sens où elle revient à remplacer chaque paramètre par la pire valeur qu'il puisse prendre. Dans le cas où les paramètres sont de nature stochastique, il est très improbable qu'ils prennent tous simultanément leur pire valeur.

Depuis ces premiers travaux dans le domaine de l'optimisation robuste, des contreparties robustes traitables ou des approximations traitables de contreparties robustes de problèmes d'optimisation incertains ont été obtenues pour des problèmes d'optimisation linéaires, coniques et semidéfinis positifs et en utilisant différents ensembles d'incertitude. La méthodologie robuste a été par ailleurs appliquée avec succès pour un bon nombre de problèmes pratiques, allant de la construction d'antenne ([5]), la gestion de portefeuilles ([10], [39]), l'étude de systèmes dynamiques incertains ([5]) ou la construction mécanique ([9], [81]).

Dans la suite de cette section, nous développons quelques résultats importants concernant la traitabilité de contreparties robustes de problèmes d'optimisation. Les contraintes (1.20) conduisent à un problème d'optimisation semi infini. De tels problèmes, même convexes, sont rarement solubles efficacement d'un point de vue numérique, et en particulier ne permettent pas l'utilisation d'outils d'optimisation récents comme les méthodes de points intérieurs. Par conséquent, comme dans le cas de problèmes contenant des contraintes probabilistes, un objectif de premier plan en matière d'optimisation robuste, est de trans-

former (si possible) les contreparties robustes de problèmes convexes incertains (ou à défaut des approximations de ces contreparties robustes) en problèmes d'optimisation convexes formulés explicitement. La difficulté de cette tâche dépendra d'une part de la structure analytique du problème incertain étudié, et d'autre part de la géométrie de l'ensemble d'incertitude $\mathcal{Z}$.

1.5.2 Différents ensembles d'incertitude

Ayant supposé (Hypothèse A) que $F(x,\xi)$ est K-concave en ξ, il n'est pas restrictif de supposer que l'ensemble d'incertitude est un ensemble convexe fermé. En effet, si $F(x,\xi)$ est K-concave en ξ alors la contrepartie robuste du problème d'optimisation incertain (1.19) (où $\mathcal{Z}$ est un ensemble quelconque) est identique à la contrepartie robuste du problème obtenu à partir de (1.19) en remplaçant l'ensemble $\mathcal{Z}$ par son enveloppe convexe ([7]). Si on suppose que les fonctions $f_0(x,\xi)$ et $F(x,\xi)$ sont des fonctions affines des paramètres ξ (ce qui est le cas dans de nombreuses applications), et si $\mathcal{Z}$ est l'enveloppe convexe d'un ensemble de scénarios $(\xi_1,\ldots,\xi_S)$, alors la contrepartie robuste de $\mathcal{P}_{\mathcal{Z}}$ est le problème :

$$\min\{\tau \ : \ \tau \geq f_0(x,\xi_i), \ : \ x \in X, \ F(x,\xi_i) \leq 0, \ i=1,\ldots,S\}\,.$$

La structure analytique de ce problème est exactement la même que celle des instances et ce problème sera donc traitable si les instances le sont. Ce sera en particulier le cas si les instances sont des problèmes d'optimisation linéaires, coniques ou semidéfinis positifs. Lorsque les paramètres incertains sont de nature stochastique, l'enveloppe convexe des réalisations disponibles des paramètres incertains peut être utilisée comme une estimation du support de la distribution de ces paramètres (Moore [72], Ripley et Rasson [80]). Afin de rendre l'approche un peu moins conservative, on peut utiliser comme ensemble d'incertitude $\mathcal{Z}$ un polytope (une enveloppe convexe de scénarios $(\xi'_1,\ldots,\xi'_S)$) obtenu par contraction de l'enveloppe convexe des scénarios $(\xi_1,\ldots,\xi_S)$. Plus précisément, si g est l'isobarycentre des points $\xi_1,\ldots,\xi_S$ alors les scénarios ξ'_i sont définis par $\xi'_i = g+\alpha(\xi_i - g)$, où $0<\alpha<1$. La contrepartie robuste utilisant ce polytope a ainsi la même forme analytique et est moins conservative que la contrepartie robuste utilisant l'enveloppe convexe des historiques des paramètres. Néanmoins les ensembles d'incertitude les plus couramment utilisés en optimisation robuste sont les ellipsoïdes et les intersections d'un nombre fini d'ellipsoïdes. Comme le souligne [7] l'utilisation de tels ensembles d'incertitude est motivée par un certain nombre de facteurs :

- Les ellipsoïdes ont une représentation paramétrique simple.
- Des intersections d'ellipsoïdes bien choisis permettent d'approximer des ensembles d'incertitude plus complexes.
- Dans certains cas où les paramètres incertains sont de nature stochastique, des arguments d'ordre probabiliste permettent d'interpréter l'utilisation d'un ensemble d'incertitude ellipsoïdal. Prenons l'exemple d'un problème de programmation linéaire dont les coefficients de la matrice des contraintes sont aléatoires. Remarquons, qu'en utilisant une variable slack, ce cas peut facilement se généraliser au cas où l'incertitude porte aussi sur la fonction objectif et le second membre des contraintes. Soit alors $a_j^T x \leq b_j$ la j-ème contrainte du problème. Si $\bar{a}_j = Ea_j$ est la moyenne du vecteur a_j et $Q_j = E(a_j-\bar{a}_j)(a_j-\bar{a}_j)^T$ sa matrice de variance covariance alors $Ea_j^T x = \bar{a}_j^T x$ et $Var(a_j^T x) = x^T Q_j x$. Si κ est un paramètre positif, dans le cas où le vecteur a_j est

gaussien, la probabilité que $a_j^T x$ soit supérieur ou égal à $\bar{a}_j^T x + \kappa\sqrt{x^T Q_j x}$ est majorée par $e^{-\kappa^2/2}$. Autrement dit, en prenant κ suffisamment grand (disons supérieur ou égal à 2) et si a_j est gaussien, alors si x satisfait la contrainte :

$$\bar{a}_j^T x + \kappa\sqrt{x^T Q_j x} \leq b_j, \tag{1.23}$$

alors il satisfait la contrainte $a_j^T x \leq b_j$ avec une grande probabilité (supérieure à $1 - e^{-\kappa^2/2}$). Dans le cas où $a_j^T x$ a de faibles queues de distribution, l'approche ci-dessus reste valide et l'on peut considérer que la contrainte (1.23) est une version sûre, robuste, de la contrainte $a_j^T x \leq b_j$. De plus, remarquons que la contrainte (1.23) est exactement la contrepartie robuste

$$a_j^T x \leq b_j, \quad \forall a_j \in \mathcal{Z}_j,$$

de la contrainte initiale $a_j^T x \leq b_j$, si l'on choisit comme ensemble d'incertitude $\mathcal{Z}_j$ l'ellipsoïde

$$\mathcal{Z}_j = \{a \mid (a - \bar{a}_j)^T Q_j^{-1}(a - \bar{a}_j) \leq \kappa^2\}.$$

Un ensemble d'incertitude ellipsoïdal permet donc de prendre en compte les corrélations possibles entre les paramètres et fournit des contreparties robustes moins conservatives que celles initialement étudiées par Soyster dans [86]. En contrepartie, la mise en oeuvre de ce procédé nécessite de calibrer la moyenne et la matrice de variance covariance du vecteur a_j. Ce problème sera étudié dans les chapitres 2 et 3.

Notons enfin que les ensembles définis par des contraintes de boîte, sont des ellipsoïdes particuliers couramment utilisés comme ensemble d'incertitude en optimisation robuste. En effet, des procédures statistiques permettent d'obtenir facilement, sous certaines hypothèses sur les paramètres incertains, de telles régions de confiance avec un niveau de confiance donné (cf section 4.6).

1.5.3 Contreparties robustes de problèmes convexes incertains

L'ensemble de faisabilité de la contrepartie robuste (CR) est un ensemble convexe fermé, donc la contrepartie robuste (CR) du problème d'optimisation incertain convexe $P_{\mathcal{Z}}$ est un problème convexe. Si la contrepartie robuste (CR) est faisable, alors toutes les instances du problème d'optimisation incertain $P_{\mathcal{Z}}$ sont faisables, et la valeur optimale de la contrepartie robuste est supérieure ou égale aux valeurs optimales de toutes les instances. Cependant, à l'inverse, il se peut très bien que toutes les instances du problème $P_{\mathcal{Z}}$ soient faisables sans que pour autant la contrepartie robuste le soit ou que la valeur optimale de la contrepartie robuste soit "pire" que celles de toutes les instances. Afin que, notamment, la valeur optimale de la pire des instances et la valeur optimale de la contrepartie robuste soient les mêmes nous introduisons l'hypothèse supplémentaire suivante.

Hypothèse B. $K = \mathbb{R}^n_+$ et pour tout $i = 1, \ldots, m$, la i-ème composante de F, ne dépend que de x et d'une portion ξ_i du vecteur des paramètres incertains ξ. L'ensemble d'incertitude $\mathcal{Z}$ est alors le produit cartésien d'ensembles d'incertitude convexes $\mathcal{Z}_i$ dans l'espace des ξ_i, $i = 1, \ldots, m$.

Sous l'hypothèse B, le vecteur ξ peut être partitionné en $(\xi_1, \ldots, \xi_m)$, où ξ_i regroupe

les données incertaines de la i-ème contrainte et la contrepartie robuste (CR) est le programme :

$$(CR) \quad \min \left\{ c^T x \ : \ x \in X,\ f_i(x, \xi_i) \geq 0,\ \ \forall i = 1, \ldots, m,\ \ \forall \xi_i \in \mathcal{Z}_i \right\}, \tag{1.24}$$

où c est un vecteur fixé. Dans [7] on montre que si les hypothèses A et B explicitées ci-dessus sont satisfaites et si X est compact, alors la contrepartie robuste (CR) du problème $\mathcal{P}_{\mathcal{Z}}$ (1.19) est faisable si et seulement si toutes les instances de ce problème sont faisables. De plus, la valeur optimale de la contrepartie robuste est le supremum des valeurs optimales des instances. Nous faisons cette hypothèse dans la suite de cette section où nous présentons des contreparties robustes de problèmes d'optimisation linéaires, coniques et semidéfini positifs (SDP). Les ensembles d'incertitude retenus sont en général des ellipsoïdes ou des intersections d'ellipsoïdes. Puisqu'on peut toujours se ramener au cas où la fonction objectif est connue, les résultats seront présentés dans ce cadre.

Programmation linéaire robuste

La contrepartie robuste d'un problème d'optimisation linéaire incertain :

$$\left\{ \min_x \ \left\{ c^T x \ : \ Ax \geq b \right\} \ | \ (A, b) \in \mathcal{Z} \subset \mathbb{R}^{m \times n} \times \mathbb{R}^m \right\}, \tag{1.25}$$

est équivalente à un problème traitable défini explicitement en fonction des paramètres du problème, à condition que l'ensemble d'incertitude $\mathcal{Z}$ soit lui même traitable.

Théorème 1.5.1 *[6] Supposons que l'ensemble d'incertitude $\mathcal{Z}$ dans* (1.25) *soit l'image affine d'un ensemble borné $\mathcal{Z}' \subset \mathbb{R}^N$, et que $\mathcal{Z}'$ soit donné par : (i) un système de contraintes d'inégalités $P\xi \leq p$ ou (ii) un système d'inégalités coniques quadratiques $\|P_i\xi - p_i\|_2 \leq q_i^T\xi - r_i, i = 1, \ldots, M$ ou (iii) un système d'inégalités matricielles linéaires $P_0 + \sum_{i=1}^{dim\,\xi} \xi_i P_i \succeq 0$. Dans les cas (ii), (iii), supposons que le système de contraintes définissant $\mathcal{Z}$ soit strictement faisable. Alors la contrepartie robuste du programme linéaire* (1.25) *est équivalente à un problème linéaire dans le cas (i), un problème conique quadratique dans le cas (ii) et un problème semidéfini positif dans le cas (iii).*

Le problème de gestion de production étudié dans le chapitre 4 est un problème linéaire pour lequel la méthodologie de l'optimisation robuste sera appliquée.

Programmation conique robuste

Considérons un problème d'optimisation convexe avec contraintes quadratiques :

$$\left\{ \min_x \ \left\{ c^T x \ : \ x^T A_i x \leq 2 b_i^T x + c_i,\ i = 1, \ldots, m \right\} \ | \ \{(A_i, b_i, c_i)\}_{i=1}^m \in \mathcal{Z} \right\}. \tag{1.26}$$

Si l'ensemble d'incertitude $\mathcal{Z}$ est une intersection d'ellipsoïdes :

$$\mathcal{Z} = \left\{ \left\{ (A_i, b_i, c_i) = (A_i^0, b_i^0, c_i^0) + \sum_{\ell=1}^{k} \xi_\ell \, (A_i^\ell, b_i^\ell, c_i^\ell) \right\}_{i=1}^{m} \ | \ \xi^T Q_j \xi \leq 1,\ j = 1, \ldots, k \right\} \tag{1.27}$$

où $Q_j \succeq 0,\ j = 1, \ldots, k$, et $\sum_{j=1}^{k} Q_j \succ 0$; alors [5] fournit un problème d'optimisation SDP dépendant explicitement des paramètres du problème qui, si $k = 1$, (un seul ellipsoïde) est la contrepartie robuste du problème (1.26),(1.27) et qui dans le cas où $k > 1$ est une approximation de cette contrepartie robuste. Un problème d'optimisation quadratique sera robustifié dans le chapitre 3 en utilisant un ensemble d'incertitude simple défini par des contraintes de boîte.
Considérons à présent le cas plus général d'un problème d'optimisation conique quadratique qui s'exprime sous la forme :

$$\left\{\min_x \ \left\{c^T x \ : \ \|A_i x + b_i\|_2 \leq \alpha_i^T x + \beta_i,\ i = 1, \ldots, m\right\} \ | \ \{(A_i, b_i, \alpha_i, \beta_i)\}_{i=1}^m \in \mathcal{Z}\right\}. \tag{1.28}$$

On se place ici aussi dans le contexte où l'ensemble

$$\mathcal{Z} = \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_m \tag{1.29}$$

est le produit cartésien d'ensembles $\mathcal{Z}_i$; l'ensemble $\mathcal{Z}_i = \mathcal{V}_i \times \mathcal{W}_i$, associé à la i-ème contrainte, étant lui même le produit cartésien de deux ellipsoïdes $\mathcal{V}_i$ et $\mathcal{W}_i$:

$$\mathcal{V}_i = \left\{ [A_i; b_i] = [A_i^0; b_i^0] + \sum_{j=1}^{k_i} u_j [A_i^j; b_i^j] + \sum_{j=1}^{q_i} v_j [E_i^j; f_i^j] \mid u^T u \leq 1 \right\};$$

$$\mathcal{W}_i = \left\{ [\alpha_i; \beta_i] = [\alpha_i^0; \beta_i^0] + \sum_{j=1}^{k_i'} u_j [\alpha_i^j; \beta_i^j] + \sum_{j=1}^{q_i'} v_j [g_i^j; h_i^j] \mid u^T u \leq 1 \right\}.$$

La contrepartie robuste du problème conique quadratique (1.28) correspondant à l'ensemble d'incertitude (1.29) défini ci-dessus, est alors équivalente à un problème SDP ([7]).

Programmation SDP robuste

La programmation SDP robuste a été essentiellement étudiée dans [38] et [40] ainsi que dans les articles [7] et [5] synthétisant les résultants importants en optimisation robuste. La contrepartie robuste d'un problème d'optimisation semidéfini positif incertain

$$\left\{\min_x \ \left\{c^T x \ : \ \mathcal{A}(x) = A_0 + \sum_{j=1}^{n} x_j A_j \succeq 0\right\} \ | \ (A_0, \ldots, A_n) \in \mathcal{Z}\right\}. \tag{1.30}$$

(dans lequel les matrices A_j sont des matrices $M \times M$ symétriques) est déjà NP-difficile pour un ensemble d'incertitude $\mathcal{Z}$ ellipsoïdal. Dans le cas où l'ensemble d'incertitude est une enveloppe convexe :

$$\mathcal{Z} = \mathrm{Conv}\{(A_0^1, A_1^1, \ldots, A_n^1), \ldots, (A_0^L, A_1^L, \ldots, A_n^L)\};$$

la contrepartie robuste du problème (1.30) est le problème semidéfini positif :

$$\min_x \left\{ c^T x \ : \ A_0^\ell + \sum_{j=1}^{n} x_j A_j^\ell, \ell = 1, \ldots, L \right\}.$$

On peut donner un autre ensemble d'incertitude conduisant à une contrepartie robuste équivalente à un problème SDP. Soit d un vecteur fixe non nul de dimension M. On appelle perturbation de rang 2 de $\mathcal{A}(.)$ associée à d, une perturbation de la forme $\mathcal{A}(x) + b(x)d^T + db(x)^T$, où $b(x)$ est une fonction affine de x à valeur dans $\mathbb{R}^M$. On considère alors le problème d'optimisation semidéfini positif incertain obtenu à partir de (1.30) en envisageant toutes les perturbations de rang 2 associées à un vecteur d non nul fixe et $b(.)$ variant dans un ellipsoïde :

$$\left\{ \min_x \left\{ c^T x \ : \ \mathcal{A}(x) + [\sum_{\ell=1}^{L} u_\ell b^\ell(x)]d^T + d[\sum_{\ell=1}^{L} u_\ell b^\ell(x)]^T \succeq 0 \right\} \mid u^T u \leq 1 \right\}, \qquad (1.31)$$

où $b^\ell(x)$ sont des fonctions affines de x à valeur dans $\mathbb{R}^M$. La contrepartie robuste du problème d'optimisation semidéfini positif (1.31) est alors un problème d'optimisation semidéfini positif explicite (voir [5]).
Dans le cas où l'ensemble d'incertitude $\mathcal{Z}$ intervenant dans (1.30) est défini par :

$$\mathcal{Z} = \left\{ (A_0, \ldots, A_n) = (A_0, \ldots, A_n) + \sum_{\ell=1}^{L} \xi_\ell (A_0^\ell, \ldots, A_n^\ell) \mid \xi \in \mathcal{Z}' \right\}, \qquad (1.32)$$

avec $\mathcal{Z}' = \{\xi \in \mathbb{R}^L \mid \|\xi\|_\infty \leq 1\}$ ou $\mathcal{Z}' = \{\xi \in \mathbb{R}^L \mid \|\xi\|_2 \leq 1\}$, alors on peut définir un problème d'optimisation semidéfini positif ([5]) dont la solution fournit une solution approchée de la contrepartie robuste du problème (1.30) associé à l'ensemble d'incertitude $\mathcal{Z}$ (1.32) défini ci-dessus.
Citons un dernier cas, où l'on connaît une approximation traitable de la contrepartie robuste d'un problème d'optimisation SDP de la forme :

$$\left\{ \min \{c^T x \ : \ \mathcal{A}_i(x) \succeq 0, \ i = 1, \ldots, m\} \ \ \mathcal{A}_i(.) \in \mathcal{A}_i^0(.) + \nu_i, \ i = 1, \ldots, m \right\}, \qquad (1.33)$$

où $\mathcal{A}_i(.)$ et $\mathcal{A}_i^0(.)$ sont des applications de $\mathbb{R}^n$ dans $\mathcal{S}_{\ell_i}$ et où ν_i est un ensemble convexe dans l'espace des applications de ce type.

Hypothèse C. [7] Pour tout $i \leq m$, l'ensemble ν_i peut être approchée à un facteur γ_i près par un ellipsoïde, c'est à dire que l'on peut trouver k_i applications affines $\mathcal{A}_i^j(.) : \mathbb{R}^n \to \mathcal{S}_{\ell_i}$, $j = 1, \ldots, k_i$, de sorte que $\nu_i^- \subseteq \nu_i \subseteq \gamma_i \nu_i^-$, où :

$$\nu_i^- = \left\{ \sum_{j=1}^{k_i} u_j \mathcal{A}_i^j(.) \mid u^T u \leq 1 \right\}.$$

Sous l'hypothèse C, on peut définir un programme SDP ([7]) qui est une approximation de la contrepartie robuste du problème (1.33). La qualité des approximations des contreparties robustes mentionnées dans cette section est définie et précisée dans [5] et [7].

1.5.4 L'optimisation robuste à l'épreuve

L'optimisation robuste a été utilisée dans de nombreuses applications pratiques ([5], [10], [39], [9], [81]). Dans [8], Ben-Tal et Nemirovski montrent sur des exemples tirés de la librairie NETLIB, en quoi, dans le cadre de la programmation linéaire, l'optimisation

robuste est particulièrement intéressante pour obtenir une solution stable, i.e une solution qui soit pratiquement optimale pour toute une gamme d'instances obtenues en perturbant légèrement les paramètres d'un problème nominal. L'analyse de sensibilité permet de quantifier l'écart entre la solution d'un problème nominal et la solution d'un problème perturbé. D'autre part une analyse de sensibilité permet de quantifier localement (en général) la stabilité de la solution nominale pour de petites perturbations des paramètres. Cette étude renseigne rarement sur les méthodes à utiliser pour améliorer la stabilité des solutions si nécessaire. Dans [8], on montre qu'une solution nominale d'un problème d'optimisation linéaire peut fortement violer les contraintes de problèmes d'optimisation linéaire obtenus à partir du problème nominal par de petites perturbations des paramètres. Pour l'un des problèmes étudiés par exemple, les coefficients d'une seule contrainte sont perturbés aléatoirement et uniformément, le niveau de perturbation étant de 0.01%. La plus importante violation de la contrainte en utilisant la solution nominale sur les problèmes perturbés a représenté 450% du second membre. D'autre part, en général, sur les exemples étudiés, une petite correction de la solution nominale ne permet pas d'obtenir une solution robuste conduisant à une valeur optimale proche de celle du problème nominal. En revanche, en envisageant des perturbations uniformes d'ordre 0.1%, pour 23 problèmes d'optimisation de la librairie NETLIB, et en appliquant la méthodologie de l'optimisation robuste à ces problèmes, la valeur optimale robuste était au plus à 1% de la valeur optimale du problème perturbé. D'autre part, par définition de la contrepartie robuste, la solution robuste était évidemment faisable pour chacun des problèmes perturbés.

1.6 Contreparties robustes ajustées

1.6.1 Présentation

Dans le problème robuste (CR) défini par (1.24) dans la section 1.5, on suppose que toutes les variables de décision, regroupées dans le vecteur x, doivent être déterminées avant que les réalisations des paramètres incertains soient connues. Si c'est le cas, une telle modélisation paraît réellement adéquate. Cependant, dans la majorité des problèmes d'optimisation incertains, seule une partie des variables de décision doit être déterminée avant la réalisation des paramètres incertains. Les autres variables, quant à elles, pourront *s'ajuster* aux données incertaines, lorsque celles ci seront connues. Ces variables ajustables sont de deux types. D'une part, elles regroupent les variables auxiliaires, telles que les variables de surplus ou celles introduites pour présenter le modèle sous une forme de programme linéaire (en éliminant des fonctions linéaires par morceaux comme $|x_i|$ ou $\max(x_i, 0)$). D'autre part, elles regroupent les variables qui peuvent être déterminées lorsqu'une partie des données incertaines (incertaines au moment de la prise de décision initiale) devient connue (variables dénommées "wait and see" en anglais). Pour illustrer ce partitionnement des variables, reprenons l'exemple du problème du vendeur de journaux détaillé dans la section 1.2. Dans ce problème, trois variables de décision interviennent. D'une part le nombre x de journaux que le vendeur de journaux achète à l'éditeur chaque matin, d'autre part le nombre y de journaux vendus et enfin le nombre w de journaux revendus à l'éditeur en fin de journée. Le vendeur de journaux doit décider du montant x de journaux à acheter le matin, avant de savoir le nombre de clients qu'il aura. Par conséquent, cette variable x n'est pas ajustable aux données incertaines et doit être déterminée avant de connaître le nombre de journaux qu'il vendra effectivement dans la journée. En revanche, en fin de journée, le

nombre de journaux vendus ξ est connu (c'était le paramètre incertain du problème) et les variables y et w peuvent *s'ajuster* à cette valeur ξ et être déterminées à partir de cette valeur ξ puisque $y = \min(\xi, x)$ et $w = \max(x - \xi, 0)$. Nous venons donc de partitionner les variables en deux groupes : les variables qui doivent être déterminées avant la réalisation des paramètres incertains (variables de décisions "here and now" en anglais) et celles qui peuvent s'ajuster à tout ou partie des données incertaines (variables "wait and see" en anglais). Nous appellerons ces dernières variables des variables ajustables. En général, chaque variable ajustable peut avoir sa propre *base d'information*, c'est à dire dépendre d'une partie des données incertaines qui lui est propre. Afin de simplifier les notations et la présentation des résultats qui suivent, nous supposons qu'une partie des variables, regroupée dans un vecteur u, n'est pas ajustable, et l'autre partie, regroupée dans un vecteur v, peut s'ajuster à toutes les données incertaines. Les résultats que nous énoncerons peuvent être facilement généralisés au cas où chaque variable ajustable a sa propre base d'information. Par conséquent le vecteur de décision x est partitionné en $x = (u^T, v^T)^T$. On peut, alors, sous les hypothèses A et B (introduites dans la section 1.5) et sans perdre de généralité, (quitte à introduire une variable de "surplus" si nécessaire), écrire un problème d'optimisation générique incertain sous la forme :

$$\mathcal{P}_{\mathcal{Z}} = \left\{ \min_{(u,v)} \{c^T u \ : \ (u,v) \in X, \ f_i(u,v,\xi_i) \geq 0, \ i = 1, \ldots, m\} \right\}_{(\xi_1,\ldots,\xi_m) \in (\mathcal{Z}_1 \times \ldots \times \mathcal{Z}_m)} . \tag{1.34}$$

Sa contrepartie robuste peut alors s'écrire :

$$(CR) \ \min_{u,v} \left\{ c^T u \ : \ (u,v) \in X, \ \inf_{\xi_i \in \mathcal{Z}_i} f_i(u,v,\xi_i) \geq 0, \ \ i = 1, \ldots, m \right\};$$

où ce qui revient au même :

$$(CR) \quad \min_{u} \left\{ c^T u \ : \exists\, v \mid (u,v) \in X, \ \ \forall i = 1, \ldots, m, \, \forall (\xi_i \in \mathcal{Z}_i) \ f_i(u,v,\xi_i) \geq 0 \right\}. \tag{1.35}$$

Si $\mathcal{Z} = \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_m$, on définit alors la contrepartie robuste ajustée du problème d'optimisation incertain (1.34) par :

$$(CRA) \ \min_{u} \left\{ c^T u \ : \forall \xi = (\xi_1, \ldots, \xi_m) \in \mathcal{Z}, \, \exists\, v \mid (u,v) \in X, \ \ \forall i = 1, \ldots, m, \ f_i(u,v,\xi_i) \geq 0 \right\}. \tag{1.36}$$

La contrepartie robuste ajustée (1.36), introduite récemment dans [11], fournit donc une politique robuste plus flexible que la contrepartie robuste définie par (1.35). En effet, l'ensemble de faisabilité de la contrepartie robuste ajustée est plus vaste, ce qui permet d'obtenir une meilleure valeur optimale tout en obtenant une solution satisfaisant les contraintes quelle que soit la réalisation des paramètres aléatoires. Prenons l'exemple d'une contrainte linéaire d'égalité incertaine $u + v = a$, l'ensemble d'incertitude étant $\mathcal{Z} = \{a \mid a \in [0,1]\}$. L'ensemble de faisabilité de la contrepartie robuste est vide alors que l'ensemble de faisabilité de la contrepartie robuste ajustée est $\mathbb{R}$ (pour tout $u \in \mathbb{R}$, pour tout $0 \leq a \leq 1$, $v = a - u$ satisfait la contrainte). Par contre, les contreparties robustes ajustées conduisent en général à des problèmes beaucoup plus durs à résoudre que les contreparties robustes pour lesquelles il existe des traitements numériques efficaces pour une large gamme de problèmes et d'ensembles d'incertitude (cf section 1.5). Dans la section suivante, nous nous intéressons aux contreparties robustes ajustées de problèmes d'optimisation linéaires incertains.

1.6.2 Problèmes linéaires

Adoptant donc un partitionnement du vecteur décisionnnel en $x = (u^T, v^T)^T$ comme il a été précisé ci-dessus, nous écrivons un problème d'optimisation linéaire incertain sous la forme :

$$LP_{\mathcal{Z}} \quad \left\{ \min_{u,v} \{c^T u \ : \ Uu + Vv \leq b\} \right\}_{\xi=[U,V,b]\in\mathcal{Z}}, \tag{1.37}$$

où $\mathcal{Z}$ est un ensemble convexe compact non vide. Reprenant la terminologie de la programmation stochastique muti-étapes, nous appellerons la matrice V, la matrice de recours. La contrepartie robuste ajustée (notée CRA par la suite) d'un tel problème est alors donnée par :

$$(CRA) \ \min_{u}\{c^T u \ : \ \forall\, (\xi = [U,V,b] \in \mathcal{Z}) \ \exists\, v \ : Uu + Vv \leq b\}, \tag{1.38}$$

et sa contrepartie robuste (notée CR par la suite) est :

$$(CR) \ \min_{u}\{c^T u \ : \ \exists v \ \forall\, (\xi = [U,V,b] \in \mathcal{Z}) \ : Uu + Vv \leq b\}. \tag{1.39}$$

Sous certaines hypothèses assez restrictives, la CR et la CRA d'un problème d'optimisation linéaire incertain sont équivalentes [11]. L'une de ces hypothèses est l'hypothèse B donnée en section 1.5, à savoir que l'incertitude affectant une contrainte est indépendante de l'incertitude affectant les autres contraintes. On peut néanmoins facilement construire des exemples montrant que la CRA est plus flexible que la CR. Au delà du cas de contraintes d'égalités incertaines déjà évoqué, prenons un exemple avec deux contraintes d'inégalités incertaines pour illustrer ce phénomène. Considérons le problème d'optimisation incertain :

$$\min_{u,v} u \ : \ \ u \geq -1, \quad v \leq (\xi - 1)u + 1, \quad v \geq \xi u + 1,$$

où $\xi \in [-1, 1]$ est un paramètre incertain. Le seul point faisable pour la CR est $(u, v) = (0, 1)$ et la valeur optimale de la CR est donc 0. La valeur optimale de la CRA est meilleure puisqu'elle vaut -1. En effet, l'ensemble de faisabilité de la CRA est [-1,0] : pour tout $u \in [-1, 0]$ et tout $\xi \in [-1, 1]$, le point $v = \xi u = 1$ (par exemple) est faisable. On remarque d'ailleurs que l'hypothèse B n'est pas satisfaite. Dans le cas d'un problème d'optimisation linéaire incertain, la contrepartie robuste est un problème traitable dès que l'ensemble d'incertitude est traitable. Un cas simple où la CRA (1.38) du problème (1.37) est traitable (et est en fait un problème linéaire) est le cas où la matrice de recours V est fixe et l'ensemble d'incertitude $\mathcal{Z}$ est l'enveloppe convexe d'un ensemble de scénarios :

$$\mathcal{Z} = \mathrm{Conv}\{[U_1, V, b_1], \ldots, [U_S, V, b_S]\}. \tag{1.40}$$

Dans ce cas, la CRA est donnée par le programme linéaire :

$$\min_{u,v_1,\ldots,v_S} \ c^T u \ : \ \ U_\ell u + V v_\ell \leq b_\ell, \ \ell = 1, \ldots, S.$$

Dans le cas où la matrice de recours n'est pas fixe et où $\mathcal{Z}$ est aussi donné par une enveloppe convexe de scénarios : $\mathcal{Z} = \mathrm{Conv}\{[U_1, V_1, b_1], \ldots, [U_S, V_S, b_S]\}$, la CRA peut être NP difficile. De même, dans le cas où la matrice de recours est fixe, si $\mathcal{Z}$ est un polytope défini par une liste d'inégalités linéaires, la CRA peut être NP difficile. On doit alors chercher des approximations de la CRA traitables pour une plus large gamme d'ensembles d'incertitude. C'est ce qui motive l'introduction de contreparties robustes ajustées affinement décrites dans la section suivante.

1.6.3 Contreparties robustes ajustées affinement

Dans de nombreuses applications de la théorie de la commande, on cherche des contrôles optimaux qui sont des fonctions linéaires des observations de l'état du système. Cette restriction sur la liaison entre les variables de décision et l'état du système permet souvent le calcul efficace de la commande. Dans la même veine, on peut imposer, pour u fixé, que la variable ajustable v du problème $LP_{\mathcal{Z}}$ soit une fonction affine des données :

$$v = w + W\xi, \quad \text{où} \quad \xi = [U, V, b].$$

Dans ce contexte, un vecteur non ajustable u peut être complété en une solution (u, w, W) robuste si et seulement si $Uu + V(w + W\xi) \leq b \quad \forall (\xi = [U, V, b] \in \mathcal{Z})$. On définit alors la contrepartie robuste ajustée affinement (CRAA) comme le programme :

$$(CRAA) \min_{u,w,W} \{c^T u \ : \ Uu + V(w + W\xi) \leq b \quad \forall (\xi = [U, V, b] \in \mathcal{Z})\}. \tag{1.41}$$

Notons que l'ensemble de faisabilité de la CR est contenu dans l'ensemble de faisabilité de la CRAA, lui même contenu dans l'ensemble de faisabilité de la CRA. Par conséquent, la CRAA est un peu plus conservative que la CRA mais moins conservative que la CR.
Si la matrice de recours V est fixe et si l'ensemble d'incertitude $\mathcal{Z}$ est traitable alors la CRAA (1.41) du problème $LP_{\mathcal{Z}}$ est traitable ([11]). Afin de fournir, dans ce cas, des équivalents explicites de la CRAA pour divers ensembles d'incertitude, nous paramétrons affinement l'ensemble $\mathcal{Z}$ par un vecteur de perturbations ζ variant dans un ensemble convexe compact non vide $\chi \subset \mathbb{R}^p$:

$$\mathcal{Z} = \left\{ [U, V, b] = [U^0, V^0, b^0] + \sum_{\ell=1}^{p} \zeta_\ell, [U^\ell, V^\ell, b^\ell], \ \zeta \in \chi \right\}.$$

On peut alors exprimer les variables ajustables v sous la forme :

$$v = v^0 + \sum_{\ell=1}^{p} \zeta_\ell \, v^\ell.$$

La CRAA du problème $LP_{\mathcal{Z}}$ est alors :

$$(CRAA) \left\{ \begin{array}{l} \min \ c^T u \\ \forall \, \zeta \ \in \chi \quad \alpha_i(x) + 2\zeta^T \beta_i(x) + \zeta^T \Gamma_i(x) \zeta \geq 0, \ i = 1, \ldots, m, \end{array} \right. \tag{1.42}$$

où $\alpha_i(x), \beta_i(x)$ et $\Gamma_i(x)$ sont des fonctions affines de $x = [u, v^0, \ldots, v^p]$ définies dans [11]. Dans le cas où la matrice de recours V est fixe, on a pour tout $1 \leq i \leq m$, $\Gamma_i(x) = 0$, $\alpha_i(x) = -[U_i^0 u + V_i^0 v^0 - b_i^0]$, et $\beta_i^\ell(x) = -\frac{1}{2}(U_i^\ell u + V_i^0 v^\ell - b_i^\ell), \ \ell = 1, \ldots, p$.

Incertitude polytopique. On suppose que χ est l'enveloppe convexe de $(\bar{\zeta}_1, \ldots, \bar{\zeta}_k)$. Dans ce cas, puisque $\min_{\zeta \in \chi} \zeta^T \beta_i(x) = \min_{1 \leq j \leq k} \beta_i(x)^T \bar{\zeta}_j$, la CRAA de $LP_{\mathcal{Z}}$ avec matrice de recours fixe est :

$$\left\{ \begin{array}{l} \min \ \ c^T u \\ \alpha_i(x) + \gamma_i \geq 0, \quad i = 1, \ldots, m, \\ \gamma_i \leq 2\beta_i(x)^T \bar{\zeta}_j, \ \ i = 1, \ldots, m, \quad j = 1, \ldots, k. \end{array} \right.$$

Contraintes de boîte sur ζ. Soit χ défini par des contraintes de boîte :

$$\chi = \{\zeta \mid \zeta_{\text{inf}} \leq \zeta \leq \zeta_{\text{sup}}\}, \tag{1.43}$$

et soient

$$\theta_1 = \zeta_{\text{inf}} + \zeta_{\text{sup}}, \quad \theta_2 = \zeta_{\text{inf}} - \zeta_{\text{sup}}.$$

Comme $\min_{\zeta \in \chi} 2\zeta^T \beta_i(x) = \theta_1^T \beta_i(x) + \theta_2^T |\beta_i(x)|$, la CRAA de $LP_{\mathcal{Z}}$ avec matrice de recours fixe et ensemble d'incertitude (1.43) pour χ est alors :

$$\begin{cases} \min \quad c^T u \\ \alpha_i(x) + \theta_1^T \beta_i(x) + \theta_2^T \gamma_i \geq 0, \quad i = 1, \ldots, m, \\ -\gamma_i \leq \beta_i(x) \leq \gamma_i, \quad i = 1, \ldots, m. \end{cases}$$

Incertitude ellipsoïdale. Soit χ l'ellipsoïde suivant :

$$\chi = \{\zeta \mid (\zeta - \bar{\zeta})^T Q^{-1} (\zeta - \bar{\zeta}) \leq \kappa^2\}. \tag{1.44}$$

Comme $\min_{\zeta \in \chi} \zeta^T \beta_i(x) = \bar{\zeta}^T \beta_i(x) - \kappa \sqrt{\beta_i(x)^T Q \beta_i(x)}$, la CRAA de $LP_{\mathcal{Z}}$ avec matrice de recours fixe et ensemble d'incertitude (1.44) pour χ s'écrit :

$$\begin{cases} \min \quad c^T u \\ \alpha_i(x) + 2\bar{\xi}^T \beta_i(x) - 2\kappa \sqrt{\beta_i(x)^T Q \beta_i(x)} \geq 0, \quad i = 1, \ldots, m. \end{cases} \tag{1.45}$$

Dans le cas où la matrice de recours V est fixe, en supposant que les variables ajustables sont des fonctions affines des paramètres incertains, on obtient donc des contreparties robustes ajustées affinement qui sont traitables si l'ensemble d'incertitude $\mathcal{Z}$ est traitable. Cette hypothèse de linéarité peut donner matière à discussion. La dépendance des variables ajustables en fonction des données incertaines peut très bien ne pas être affine. Cependant, dans certains cas, on peut penser que cette dépendance peut être approchée par une fonction affine. Prenons le cas d'un problème dans lequel on doit produire chaque jour une entité afin de satisfaire une certaine demande. Si la demande, aléatoire, suit un modèle AR (autorégressif), alors une estimation de la demande moyenne un jour donné sera fonction des demandes passées. La production devant suivre la demande, envisager une dépendance linéaire de la production d'un jour en fonction des productions des jours passées fait donc sens dans ce contexte. Par ailleurs, cette démarche sera justifiée si elle fournit des solutions robustes satisfaisantes (en particulier bien meilleures que celles obtenues par application de la contrepartie robuste (1.39)) sur le problème étudié.
Si la matrice de recours n'est pas fixe, et si l'ensemble d'incertitude χ est défini par

$$\chi = \{\zeta \mid \zeta^T \, S_\ell \, \zeta \leq \rho^2, \; \ell = 1, \ldots, L\},$$

avec $\rho > 0, S_\ell \succeq 0, \sum S_\ell \succ 0$, i.e si l'ensemble d'incertitude est défini comme une intersection d'un nombre fini d'ellipsoïdes concentriques, alors si $L = 1$ la CRAA est équivalente à un problème SDP et si $L > 1$, on peut concevoir un problème SDP qui est une bonne approximation de la CRAA ([11]).

1.6.4 Problèmes quadratiques et coniques

Considérons à présent un problème d'optimisation incertain avec objectif et contraintes quadratiques. Comme il a été précisé ci-dessus, ce problème est équivalent à un problème dont les contraintes sont quadratiques et dont l'objectif est à la fois linéaire et indépendant des données incertaines et des variables ajustables v. On s'intéresse plus précisément à un problème du type :

$$QP_{\mathcal{Z}} \quad \left\{ \min_{u,v} \{c^T u \ : \ U(i,:)u + V(i,:)v + u^T A_i u + v^T B_i v + u^T C_i v \leq b_i\}, \ i = 1, \ldots, m \right\}_{\zeta \in \mathcal{Z}}; \tag{1.46}$$

où $\zeta = [U, V, A_i, B_i, C_i, b]$, $\mathcal{Z}$ est un ensemble non vide convexe compact et $U(i,:)$ et $V(i,:)$ désignent les i-èmes lignes des matrices U et V. Supposons les matrices B_i connues et reparamétrons comme précédemment le problème en écrivant les autres paramètres incertains sous la forme :

$$[U, V, A_i, C_i, b] = [U^0, V^0, A_i^0, C_i^0, b^0] + \sum_{\ell=1}^{p} \zeta_\ell, [U^\ell, V^\ell, A_i^\ell, C_i^\ell, b^\ell],$$

le paramètre ζ appartenant à un ensemble convexe compact non vide χ. On peut alors écrire les variables ajustables sous la forme $v = v^0 + \sum_{\ell=1}^{p} \xi_\ell \, v^\ell$. Dans ce contexte, en notant $x = [u, v^0, \ldots, v^p]$, les contraintes de la CRAA du problème $QP_{\mathcal{Z}}$ peuvent s'écrire :

$$\forall \zeta \in \chi \ \ \alpha_i(x) + 2\zeta^T \beta_i(x) + \zeta^T \Gamma_i(x) \zeta \geq 0, \ i = 1, \ldots, m,$$

où maintenant, pour tout $i = 1, \ldots, m$:

- $\alpha_i(x) = -U_i^0 u - V_i^0 v^0 - u^T A_i^0 u - {v^0}^T B_i v^0 - u^T C_i^0 v^0 + b_i^0$,
- $\beta_i^\ell(x) = \frac{1}{2}(-U_i^\ell u - V_i^\ell v^0 - V_i^0 v^\ell - u^T A_i^\ell u - 2{v^0}^T B_i v^\ell - u^T C_i^0 v^\ell - u^T C_i^\ell v^0 + b_i^\ell), \ \ \ell = 1, \ldots, p,$
- $\Gamma_i^{\ell,k}(x) = -\frac{1}{2}(V_i^k v^\ell + V_i^\ell v^k + u^T C_i^k v^\ell + u^T C_i^\ell v^k + 2{v^\ell}^T B_i v^k) \ \ \ell, k = 1, \ldots, p.$

Dans les notations ci-dessus, $U^\ell(i,:)$ et $V^\ell(i,:)$ ont été abrégés en U_i^ℓ et V_i^ℓ. Le schéma de la preuve du théorème 4.2 [11] peut alors être appliqué pour prouver le résultat suivant.

Lemme 1.6.1 *Soient $Q \succ 0$ et $\rho > 0$. La CRAA formulée par le problème quadratique incertain $QP_{\mathcal{Z}}$ avec ensemble d'incertitude χ :*

$$\chi = \{\zeta \mid \zeta^T Q \, \zeta \leq \rho^2\}$$

est le problème semidéfini positif :

$$\left\{ \begin{array}{l} \min \ \ c^T u \\ \begin{pmatrix} \Gamma_i(x) + \frac{\lambda}{\rho^2} Q & \beta_i(x) \\ \beta_i^T(x) & \alpha_i(x) - \lambda \end{pmatrix} \succeq 0, \quad i = 1, \ldots, m, \\ \lambda \geq 0. \end{array} \right. \tag{1.47}$$

Lemme 1.6.2 *En partitionnant toujours le vecteur x en $x = (u^T, v^T)^T$, envisageons à présent un problème conique quadratique du type :*

$$\min c^T u \quad : \quad \|A_i u + b_i\|_2 \leq p_i^T u + q_i^T v + d_i, \ \ i = 1, \ldots, m, \tag{1.48}$$

pour lequel, pour tout $i = 1, \ldots, m$, les vecteurs $(p_i; d_i)$ sont incertains, et appartiennent à l'enveloppe convexe de S scénarios $((p_i^1; d_i^1), \ldots, (p_i^S, d_i^S))$. Dans ces conditions, la robustification du problème (1.48) *est le problème :*

$$\min_{u, v_1, \ldots, v_S} c^T u \quad : \quad \|A_i u + b_i\|_2 \leq p_i^{\ell T} u + q_i^T v_\ell + d_i^\ell, \;\; i = 1, \ldots, m, \;\; \ell = 1, \ldots, S. \tag{1.49}$$

Preuve. Soit u appartenant à l'ensemble de faisabilité de la CRA du problème (1.48). Pour tout α tel que $\alpha \geq 0, \;\; \alpha^T \mathbf{e} = 1, \exists\, v(\alpha)$ |

$$\|A_i u + b_i\|_2 \leq \sum_{j=1}^{S} \alpha_j u^T p_i^j + q_i^T v(\alpha) + \sum_{j=1}^{S} \alpha_j \, d_i^j, \;\; i = 1, \ldots, m. \tag{1.50}$$

Pour $\ell = 1, \ldots, S$, en choisissant $\alpha = e_\ell$, on voit que u est faisable pour le problème (1.49) (en prenant $v_\ell = v(e_\ell)$). Inversement, si u est faisable pour (1.49), alors pour tout α tel que $\alpha \geq 0, \;\; \alpha^T \mathbf{e} = 1$, et pour tout $1 \leq i \leq m$, en multipliant membre à membre $\|A_i u + b_i\|_2 \leq p_i^{jT} u + q_i^T v_j + d_i^j$ par α_j (pour $1 \leq j \leq S$) et en additionnant membre à membre ces S inégalités on voit qu'il existe $v(\alpha) = \sum_{j=1}^{S} \alpha_j \, v_j$ tel que $(u, v(\alpha))$ satisfait les contraintes (1.50) et u est donc faisable pour la CRA de (1.48).

□

1.7 Spécificité des méthodes développées dans cette thèse

Nous venons de passer en revue diverses manières de traiter un problème d'optimisation stochastique. Nous étudions dans cette thèse des problèmes d'optimisation stochastique dans le domaine de la gestion de portefeuilles et dans le domaine de la gestion de production.

Aspects de modélisation.

Du point de vue de la modélisation, les différentes approches présentées précédemment ont été utilisées dans le domaine de la gestion de portefeuilles. Le problème de gestion de portefeuilles peut être modélisé sous la forme d'un problème d'optimisation stochastique multi-étapes ou d'un problème d'optimisation robuste ([10]). Notre contribution dans ce domaine est de proposer des méthodes de calibration des paramètres de deux problèmes de gestion de portefeuilles.

Concernant les problèmes de gestion de production, la modélisation qui prévaut (dans le cas des problèmes de gestion de production électrique en particulier). est celle de la programmation stochastique bi-étapes ou multi-étapes. En utilisant ces approches, on obtient une solution indexée par les noeuds de l'arbre de scénarios. Lorsqu'il y a un saut de dualité (ce qui est le cas lorsqu'on s'intéresse au problème de mise en route/arrêt des centrales) se pose le problème supplémentaire de reconstruction d'une solution primale faisable. Cette méthode ne fournit donc pas de stratégie de gestion quelle que soit la réalisation de la demande. D'autre part, les variables de recours de la deuxième étape (lorsque le problème est modélisé sous la forme d'un problème d'optimisation stochastique bi-étapes) ne peuvent

être mises en place qu'une fois les consommations électriques connues ! Enfin, la construction des arbres de scénarios est une étape délicate. Le nombre de scénarios doît être choisi de telle sorte que l'on obtienne une approximation de la distribution de la consommation électrique relativement bonne mais sans que la vitesse de la phase d'optimisation en soit affectée. En effet, le temps de résolution du problème croît rapidement avec le nombre de scénarios.
Nous proposons différentes méthodes pour pallier certains de ces défauts. Dans un premier temps, bien que modélisant le problème sous la forme d'un problème d'optimisation sur un arbre de scénarios, nous présentons une modélisation du problème permettant de proposer des stratégies de gestion (quelles que soient les consommations) et non un plan de production fixé à l'avance. Diverses variantes de ce modèle sont ensuite détaillées dans le chapitre 4 afin de diminuer la variance des coûts de gestion simulés sur un ensemble de scénarios tests. Enfin, nous montrons comment s'affranchir de l'utilisation d'arbres de scénarios. On propose dans cet optique :

- d'une part, une méthode de stabilisation de la fonction duale dont le but est de stabiliser les multiplicateurs de Lagrange et réduire la variance des coûts de gestion sur un ensemble de scénarios tests ;
- d'autre part, diverses variantes robustes et en particulier les CRA et CRAA introduites dans ce chapitre.

Afin d'élargir la discussion autour des choix de modélisations d'un problème d'optimisation stochastique, mentionnons une dernière méthode utilisée en pratique. Il s'agit de résoudre une version simplifiée du problème d'optimisation stochastique posé. Cette méthode (que nous testerons dans le chapitre 4) consiste à résoudre un problème déterministe obtenu en remplaçant toutes les variables aléatoires par une estimation de leurs espérances. La solution obtenue de cette façon et appliquée en situation réelle peut violer certaines contraintes du problème (ce qui n'est pas le cas d'une solution robuste si l'ensemble d'incertitude est choisi suffisamment grand). De plus, dans le cas d'un problème linéaire, on connaît l'instabilité des solutions face à de petites perturbations des paramètres. Cette caractéristique, rend pratiquement inutilisable une telle solution dans le cadre d'un problème d'optimisation stochastique linéaire.

Mesures de la qualité des modèles.

Afin de mesurer la qualité d'une solution obtenue avec une méthode donnée, et de comparer différentes méthodes de résolution entre elles, divers critères ont été introduits pour étudier les problèmes d'optimisation stochastique. Le premier, appelé EVPI (initiales de Expected Value of Perfect Information en anglais) mesure l'écart entre la solution optimale et la solution du problème (1.8). Le second appelé VSS (pour Value of the Stochastic Solution en anglais) mesure l'écart entre la solution optimale et la solution obtenue en remplaçant toutes les variables aléatoires par leurs espérances. Afin de donner quelques bornes sur ces critères, nous notons RP la valeur optimale d'un problème d'optimisation avec recours bi-étapes sous la forme : $RP = \min_{x \in X} E_{\tilde{\xi}} f_1(x, \tilde{\xi})$ et WS (Wait and See en anglais) la valeur $WS = E_{\tilde{\xi}} f_1(\bar{x}(\tilde{\xi}), \tilde{\xi})$, où $\bar{x}(\xi)$, est la solution optimale du problème $\min f_1(x, \xi)$ sous contraintes $x \in X$. On peut alors définir le critère $EVPI$ par $EVPI = RP - WS$. La valeur optimale obtenue en remplaçant toutes les variables aléatoires par leurs espérances est notée $EV = \min_{x \in X} f_1(x, E\tilde{\xi})$. On définit alors la valeur moyenne obtenue en utilisant

la solution $\bar{x}(E\tilde{\xi})$ par $EEV = E_{\tilde{\xi}}\ f_1(\bar{x}(E\tilde{\xi}), \tilde{\xi})$. Le critère VSS peut alors être défini par $VSS = EEV - RP$. Madansky a montré que :

$$WS \leq RP \leq EEV.$$

D'autre part, pour tout problème d'optimisation stochastique ([17]), $EVPI \geq 0$ et $VSS \geq 0$. Enfin, pour tout problème d'optimisation stochastique dont la matrice de recours est fixe et dont les coefficients de la fonction objectif sont fixes, nous avons :

$$EVPI \leq EEV - EV, \quad VSS \leq EEV - EV.$$

L'EVPI représente donc la somme maximale qu'un gestionnaire est prêt à dépenser afin qu'on lui fournisse à l'avance les valeurs des paramètres incertains du problème. VSS mesure la valeur pouvant être attribuée à la connaissance et à l'utilisation de la distribution des paramètres. Il existe des cas où EVPI est nul alors que VSS est strictement positif (et vice versa, voir [17]). Ces résultats sont théoriques et en pratique, on ne connaît ni la distribution de probabilité des paramètres incertains, ni leurs espérances, ni la solution optimale.
Nous mesurerons alors la qualité des méthodes que nous développons dans cette thèse de la façon suivante :

- Concernant les problèmes d'allocations d'actifs on commence par mesurer la pertinence de la méthode de calibration des paramètres proposée sur des données simulées. La qualité des implémentations des modèles d'allocation d'actifs est ensuite étudiée à la fois sur des données réelles et simulées. On s'attachera ici à observer les rendements réels des portefeuilles et leur volatilité sur la période d'investissement.
- Concernant les problèmes de gestion de production électrique, nous commençons par utiliser un ensemble de 456 scénarios. Ces scénarios sont vus comme des historiques des paramètres incertains sur 456 années différentes. Chacune des méthodes de gestion proposées tire un certain nombre d'informations de ces scénarios (moyenne et écart type des paramètres chaque jour,...) et les utilise pour mettre en place un plan de production ou une stratégie de gestion. On applique alors ces différents plans de production ou stratégies de gestion sur l'ensemble des 456 scénarios et l'on compare les coûts moyens et l'écart type de ces coûts sur l'ensemble des scénarios pour les différentes méthodes de gestion. On compare ensuite les différentes méthodes de gestion sur des données réelles en observant le coût de ces méthodes de gestion sur différentes années successives (en utilisant donc à chaque fois une année d'historique supplémentaire).

1.8 Conclusion

Les différents choix de modélisation qui viennent d'être présentés ne s'excluent pas mutuellement. On peut par exemple adapter au cadre robuste la problématique des problèmes avec recours. Pour un problème nominal avec recours de la forme

$$\min_{x,y} f(x, y, \omega),$$

le problème robuste correspondant sera

$$\min_{x} \sup_{\omega \in \Omega} \inf_{y} f(x, y, \omega).$$

D'un point de vue pratique, un gestionnaire optera ou non pour solution robuste en fonction, essentiellement, de trois facteurs.

- D'une part son aversion au risque : dans le cas d'un gestionnaire de portefeuilles financiers par exemple, un gestionnaire très conservateur décidera de placer l'essentiel de ses liquidités dans des opérations dont le rendement est connu à l'avance (emprunts d'état par exemple).
- D'autre part, l'estimation du volume de l'ensemble d'incertitude $\mathcal{Z}$. En effet, si l'ensemble d'incertitude est de volume important, un grand nombre de scénarios (et en particulier des scénarios peu favorables) pourront se produire. Il sera donc plus important, dans une telle situation de s'immuniser contre les paramètres incertains.
- Enfin, les conséquences des variations envisagées pour les paramètres incertains du problème sont à prendre en compte. En particulier, si aucun paramètre dans l'ensemble $\mathcal{Z}$ ne conduit à de scénarios difficiles, une solution robuste n'est pas justifiée. Cependant, même dans de telles situations, une solution robuste peut être rechercheée pour avoir au moins une idée du pire coût auquel elle conduirait. Dans de nombreux cas pratiques (voir [8],[9],[10] par exemple) la mise en oeuvre d'une solution robuste se fait à peu de frais. Par contre, si certains scénarios dans l'ensemble $\mathcal{Z}$ ont de graves conséquences telles qu'une faillite, une solution robuste semble indispensable. Néanmoins, le choix d'un ensemble d'incertitude $\mathcal{Z}$ de volume trop important conduira alors à une solution très conservative. Par conséquent, une fois la décision prise de rechercher une solution robuste, un compromis devra être trouvé afin de s'immuniser contre les pires scénarios sans que la solution soit trop conservative.

L'utilisation de l'optimisation stochastique multi-étapes sur un arbre de scénarios se limite aux cas où l'incertitude est de nature stochastique (ce qui n'est pas toujours le cas) et où l'on est capable d'estimer les distributions de probabilité sous-jacentes, (ce qui est encore plus problématique, tout particulièrement en grande dimension). D'autre part, la plupart des modèles de programmation stochastique autorise qu'une solution viole les contraintes dont les paramètres sont incertains, ce qui n'est pas le cas de l'optimisation robuste. Cependant, le grief majeur qui est en général porté envers l'utilisation de méthodes robustes pour résoudre des problèmes réels est leur conservatisme.
Nous utiliserons certains modèles présentés dans ce chapitre et nous en développerons de nouveaux afin de résoudre des problèmes d'allocation d'actifs dans le chapitre 2 et de gestion de production dans le chapitre 4. Nous verrons notamment en quoi la calibration des ensembles d'incertitude est une étape déterminante pour mettre en place une contrepartie robuste.

Chapitre 2

Inférence statistique pour la gestion portefeuilles

Nota Bene : Deux articles utilisant et généralisant les résultats de ce chapitre sont en cours ([56] et [57]).

La contribution majeure de ce chapitre est de montrer comment définir des problèmes de Markowitz et VaR approchés dont on peut contrôler la qualité. On détermine pour cela, de bonnes estimations des paramètres des modèles de Markowitz et Value-at-Risk à savoir de la moyenne et de la matrice de covariance entre les rendements d'un certain nombre d'actifs risqués sur une période d'investissement donnée. L'étude est conduite dans un contexte non paramétrique pour le processus des rendements. On suppose simplement que la moyenne et la matrice de covariance varient lentement sur un certain intervalle passé. Notre principale tâche est alors de déterminer cet intervalle d'homogénéité temporelle locale. On utilise pour cela un algorithme adaptatif dont le comportement théorique est étudié. Les estimations de la moyenne et de la matrice de covariance peuvent ensuite être données en utilisant plusieurs techniques, par moyennage local par exemple. On montre alors comment utiliser ces estimations pour déterminer des problèmes de Markowitz et Value-at-Risk approchés (en tenant compte des frais de transactions). Des simulations numériques mettent les procédures développées dans ce chapitre à l'épreuve sur des données réelles et simulées.

2.1 Introduction

La plan de ce chapitre est le suivant. Dans cette section, nous présentons la modélisation du processus des rendements et nous rappelons brièvement en quoi consistent les modèles de Markowitz et Value-at-Risk qui tiennent compte des frais de transactions. Nous soulignons le problème d'estimation des paramètres de ces modèles. Dans la section suivante, on définit des problèmes de Markowitz et Value-at-Risk approchés qui consistent à utiliser des estimations spécifiques des paramètres des problèmes. Nous définissons également la précision de ces problèmes approchés et nous montrons que la précision des problèmes approchés est contrôlée. On considère ensuite deux cas particuliers pour le processus des rendements et deux procédures d'estimations associées. Le premier cas est le cas stationnaire où des

méthodes non adaptatives peuvent être utilisées afin d'estimer la moyenne, les variances et les covariances. Nous montrons comment choisir la taille de l'échantillon de manière à contrôler la précision des problèmes approchés. Le second cas est le cas où la moyenne, les variances et covariances varient lentement dans le temps. Après avoir expliqué ce que l'on entend par variation lente, nous supposons qu'on dispose d'un oracle capable de donner avec une bonne précision le plus grand intervalle passé tel que la moyenne, les variances et les covariances varient lentement sur cet intervalle. Nous donnons ensuite la précision de la procédure d'optimisation utilisant cet oracle. La troisième section montre comment construire cet oracle à partir d'un algorithme d'estimation adaptative. Cet algorithme a été utilisé dans un contexte similaire dans [70] (pour un processus unidimensionnel Gaussien). Contrairement à [70], notre étude est conduite dans un contexte multidimensionnel et non paramétrique et nous nous intéressons à deux objets : la moyenne et la matrice de covariance. Ainsi, au lieu d'avoir seulement un seul intervalle d'homogénéité temporelle locale comme dans [70] nous avons deux intervalles d'homogénéité : un pour la moyenne et l'autre pour les variances et covariances. Nous donnons ensuite quelques propriétés théoriques de l'algorithme. Nous envisageons aussi l'utilisation de cet algorithme afin de détecter la dernière rupture dans un modèle à sauts. La détection d'une rupture dans un modèle à sauts ayant une seule rupture, peut aussi être faite en utilisant une autre méthode non paramétrique utilisant une fenêtre glissante. Cette méthode est aussi étudiée. Nous terminons par quelques simulations numériques. Les preuves sont consignées à la fin du chapitre.

2.1.1 Modélisation du processus des rendements

Soit M l'horizon d'investissement, τ la date de l'investissement et soit $s_t = (s_t(1), \ldots, s_t(n))$ le processus à temps discret du prix des actifs, $t = 1, \ldots, \tau$, où n est le nombre d'actifs risqués, dont les cours au temps t sont regroupés dans le vecteur s_t de taille n. Nous définissons le rendement r_t^M à M pas de temps au pas de temps t par $r_t^M = \frac{s_{t+M}}{s_t}$, $t = 1, \ldots, \tau - M$, où la division doit être comprise composante par composante : $r_t^M(i) = \frac{s_{t+M}(i)}{s_t(i)}$, $i = 1, \ldots, n$. Remarquons qu'au pas de temps τ, les données $(r_t^M)_{t=\tau-M+1,\ldots,\tau}$ ne sont pas disponibles. Cependant, les rendements qui nous intéressent sont les rendements $(r_\tau^M(i))_{1\leq i\leq n}$ sur la période d'investissement. Ainsi, afin d'obtenir une estimation raisonnable de ces rendements, nous supposons qu'il y a un intervalle passé d'extrémité droite τ tel que les rendements varient lentement sur cet intervalle. Nous supposons aussi que cet intervalle est de longueur supérieure ou égale à $M + k$ où k est le nombre minimal de données demandées pour mettre en oeuvre une estimation. Sans perdre de généralité, et afin d'alléger les notations, nous supposerons que l'horizon d'investissement $M = 1$ et l'on notera r_t les rendements au temps t. On suppose ensuite que le processus r_t satisfait l'hypothèse 1 ci-dessous.

Hypothèse 1. Le processus r_t est généré par le modèle :

$$r_t = \rho_t + \zeta_t, \text{ avec } Er_t = \rho_t \text{ and } E\zeta_t\zeta_t^T = Q_t \succeq 0, \ t = 1, \ldots, \tau - 1, \tag{2.1}$$

où les ζ_t sont des vecteurs aléatoires indépendants de moyennes nulles. Également, pour un certain $\sigma > 0$, on suppose que $E\|r_t\|_\infty^4 \leq \sigma^4$.

Nous considérerons deux cas particuliers : le cas stationnaire où ρ_t et Q_t ne varient pas dans le temps et le cas où ils varient lentement. Nous présentons maintenant brièvement le

problème d'allocation d'actifs et l'on rappelle en particulier en quoi consistent les modèles de Markowitz et Value-at-Risk. Nous supposons par la suite, sauf mention contraire, que les rendements des actifs risqués suivent le modèle (2.1).

2.1.2 Problèmes d'allocation d'actifs

Soit τ la date de l'investissement. Nous considérons n actifs risqués ($n \geq 1$), un actif sans risque, et l'on tient compte des frais de transactions. On s'intéresse à une seule période d'investissement. Les quantités se rapportant à l'actif sans risque seront indexés par $n+1$. Dans ce contexte, un problème simplifié de gestion de portefeuilles ([34]) se présente de la façon suivante. Soit x_i le montant de l'actif i dans le portefeuille au début de la période d'investissement. Les équations de conservation reliant les montants x_i sont données par :

$$\begin{cases} x_i & = & x_i^- - y_i + z_i, \quad \text{pour les actifs risqués, } 1 \leq i \leq n, \\ x_{n+1} & = & x_{n+1}^- + \displaystyle\sum_{i=1}^{n} (1-\mu_i)\, y_i - \sum_{i=1}^{n} (1+\nu_i)\, z_i, \quad \text{pour l'actif sans risque,} \end{cases}$$

après avoir défini :

- x_i^- : le montant initial détenu en actif i avant le rebalancement du portefeuille ;
- y_i : le montant vendu en actif i au début de la période d'investissement, μ_i étant le coût de transaction unitaire correspondant ;
- z_i : le montant acheté en actif i au début de la période d'investissement, avec le coût de transaction unitaire correspondant ν_i.

Le but est de maximiser la valeur finale totale du portefeuille :

$$\sum_{i=1}^{n+1} r_\tau(i)\, x_i,$$

où $r_\tau(i)$ est le rendement du i-ème actif sur la période d'investissement fournissant indirectement la valeur du cours du i-ème actif à la fin de la période d'investissement. Dans ce qui suit, un portefeuille est donc un vecteur $x = (x_1, \ldots, x_n)^T$ de n montants correspondant aux sommes investies dans les différents actifs risqués auquel s'ajoute la somme x_{n+1} investie dans l'actif sans risque et $r_\tau = (r_\tau(1), \ldots, r_\tau(n))^T$ est le vecteur des rendements des actifs risqués sur la période d'investissement. Remarquons que le rendement de l'actif sans risque $r_\tau(n+1) = r(n+1)$ est indépendant du temps.

Information complète. Si l'on connaissait les rendements r_τ et les frais de transactions, on pourrait résoudre le problème d'optimisation linéaire suivant :

$$\mathcal{ALLOC} \begin{cases} \max & w & & \\ w & \leq & \displaystyle\sum_{i=1}^{n+1} r_\tau(i)\, x_i, & \\ x_i & = & x_i^- - y_i + z_i, & i = 1, ..., n, \\ x_{n+1} & = & x_{n+1}^- + \displaystyle\sum_{i=1}^{n} (1-\mu_i)\, y_i - \sum_{i=1}^{n} (1+\nu_i)\, z_i, & \\ x \geq 0, & x_{n+1} \geq 0, & y \geq 0, & z \geq 0. \end{cases}$$

On notera $X(\mu, \nu, x^-)$ l'ensemble des portefeuilles admissibles défini par les équations de conservation des montants et les contraintes de positivité sur x, x_{n+1}, y et z.

Le problème de l'incertitude des paramètres. En fait, les paramètres qui sont connus au moment du choix du portefeuille, sont les vecteurs des frais de transactions μ et ν et le rendement $r(n+1)$ de l'actif sans risque. Comme r_τ est un vecteur aléatoire, on doit résoudre un problème d'optimisation stochastique. Afin de résoudre ce problème d'allocation d'actifs, on pourrait estimer les rendements sur la période d'investissement (en utilisant des estimations de leurs moyennes par exemple) et reporter ces valeurs dans le problème $\mathcal{ALLOC}$. La réelle efficacité du portefeuille optimal obtenu de cette façon dépendra grandement de notre procédure d'estimation. Cette approche ne permet en particulier pas d'assurer un rendement cible fixé avec une grande probabilité. Nous n'avons aucun contrôle sur nos gains étant donné que nous pouvons, par exemple, estimer le rendement d'un actif par une grande valeur qui ne sera que rarement atteinte. Cela signifie que l'on doit prendre en compte le risque de l'investissement. Les deux paragraphes suivants rappellent brièvement les modèles d'allocation d'actifs Value-at-Risk et de Markowitz. Ces modèles proposent deux différentes définitions du risque d'un investissement. Ce chapitre est consacré à l'étude de ces modèles et en particulier à la calibration des paramètres de ces modèles (la moyenne et la matrice de covariance entre les rendements des actifs risqués sur la période d'investissement). Nous montrerons comment utiliser ces estimations pour résoudre les problèmes Value-at-Risk et de Markowitz.

2.1.2.1 Modèle Value-at-Risk (VaR)

Dans le modèle Value-at-Risk, on considérera qu'un investissement n'est pas risqué si son rendement a de grandes chances de dépasser une certaine valeur raisonnable fixé à l'avance. Plus précisément, étant donné la distribution des rendements, si l'on se fixe un niveau de confiance $0 < \varepsilon < \frac{1}{2}$, le rendement maximum que l'on peut obtenir avec une probabilité supérieure ou égale à $1-\varepsilon$, pour un portefeuille admissible $(x_1, \ldots, x_{n+1})$, est la valeur à risque de niveau ε, $V_\varepsilon(x_1, \ldots, x_{n+1})$, de ce portefeuille :

$$V_\varepsilon(x, x_{n+1}) = \max\ \gamma \ \text{ sous contraintes } \ P(r_\tau^T\, x + r(n+1)x_{n+1} \geq \gamma) \geq 1-\varepsilon.$$

Un portefeuille (x, x_{n+1}) sera d'autant plus intéressant que sa Value-at-Risk, $V_\varepsilon(x, x_{n+1})$, sera élevée. Un problème d'allocation d'actif VaR revient ainsi à résoudre :

$$\max\ \ V_\varepsilon(x, x_{n+1}) \ \text{ sous contraintes } (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-). \tag{2.2}$$

Si la distribution des rendements des actifs risqués r_τ est Gaussienne de moyenne $E[r_\tau] = \rho_\tau$ et de matrice de covariance $E[(r_\tau - \rho_\tau)(r_\tau - \rho_\tau)^T] = Q_\tau$ alors $V_\varepsilon(x, x_{n+1})$ est donnée par ${\rho_\tau}^T\, x + r(n+1)\, x_{n+1} - \Phi^{-1}(1-\varepsilon)\, \sqrt{x^T Q_\tau x}$. En utilisant la version exacte de la borne de Chebyschev (voir [13],[84]) on peut montrer que si les rendements sont Gaussiens, une solution prudente de (2.2) (en fait une borne supérieure sur la valeur optimale) est donnée en résolvant :

$$(\mathcal{V}) \begin{cases} \max & E(r_\tau^T x + r(n+1)x_{n+1}) - \kappa(\varepsilon)\, \sigma(r_\tau^T x + r(n+1)x_{n+1}) \\ & (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-), \end{cases} \tag{2.3}$$

où maintenant $\kappa(\varepsilon) = \sqrt{\frac{1-\varepsilon}{\varepsilon}}$. Ainsi, dans tous les cas, le problème de maximisation de la Value-at-Risk sur l'ensemble des portefeuilles admissibles revient à résoudre (2.3) où $\kappa(\varepsilon)$ est un facteur de risque dépendant des hypothèses sur la distribution des rendements :

– $\kappa(\varepsilon) = \Phi^{-1}(1-\varepsilon) > 0$ pour des rendements Gaussiens,

– $\kappa(\varepsilon) = \sqrt{\frac{1-\varepsilon}{\varepsilon}}$ pour des rendements aléatoires dans $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$.

Comme le remarque [6], le problème d'optimisation $(\mathcal{V})$ ci-dessus est une contrepartie robuste du problème d'allocation d'actifs $\mathcal{ALLOC}$ en utilisant l'ellipsoïde suivant comme ensemble d'incertitude auquel les rendements des actifs risqués appartiennent avec une grande probabilité :

$$\mathcal{E}(\kappa(\varepsilon)) = \{x \in \mathbb{R}^n \mid (x - \rho_\tau)^T Q_\tau^{-1} (x - \rho_\tau) \leq \kappa(\varepsilon)^2\}. \tag{2.4}$$

Par exemple, $\kappa(\varepsilon)$ sera tel que $r_\tau \in \mathcal{E}(\kappa(\varepsilon))$ avec une probabilité supérieure ou égale à $1-\varepsilon$. Dans [6], on remarque aussi que $(\mathcal{V})$ revient à proposer une "version ingénieur sûre" de la contrainte $w \leq r_\tau^T x + r(n+1)x_{n+1}$ du problème $\mathcal{ALLOC}$. En fait, pour un ingénieur, la valeur de la variable aléatoire $r_\tau^T\, x$ ne descendra pas en dessous de sa moyenne $\rho_\tau^T\, x$ moins θ (disons entre 1 et 2) fois son écart type $\sqrt{x^T\, Q_\tau\, x}$. Bien sûr, le facteur θ dépend de la distribution des rendements. Ainsi, une approche ingénieur orientée pire cas, consisterait à maximiser la pire valeur possible $\rho_\tau^T\, x + r(n+1)\, x_{n+1} - \theta\sqrt{x^T Q_\tau x}$ d'un portefeuille sur l'ensemble des portefeuilles admissibles, ce qui est en fait une approche VaR. Ainsi, le problème d'allocation d'actifs Value-at-Risk s'exprime sous la forme :

$$(\mathcal{V}) \begin{cases} \max \;\; \rho_\tau^T\, x + r(n+1)x_{n+1} - \kappa(\varepsilon)\sqrt{x^T Q_\tau x} \\ (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-), \end{cases} \tag{2.5}$$

où $\kappa(\varepsilon)$ a été défini ci-dessus.

2.1.2.2 Modèle moyenne variance de Markowitz

Dans le modèle moyenne variance de Markowitz ([69]), le risque d'un investissement est la variance (ou l'écart type) du rendement du portefeuille $r_\tau^T\, x + r(n+1)x_{n+1}$. Nous imposons par ailleurs que le rendement du portefeuille soit supérieur à une certaine valeur cible ℓ (disons $\ell = 1.1$). Nous minimisons ensuite la variance du portefeuille sous une contrainte de rendement :

$$(\mathcal{M}) \begin{cases} \min \;\; \frac{1}{2}\, x^T Q_\tau x \\ \rho_\tau^T\, x + r(n+1)x_{n+1} \geq \ell\, \mathbf{e}^T\, x^-, \\ (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-). \end{cases} \tag{2.6}$$

2.1.3 Problème d'estimation des paramètres

Si l'on investit au temps τ, afin d'implémenter le modèle $\mathcal{ALLOC}$, on a besoin des rendements $(r_\tau(i))_{1 \leq i \leq n}$ des actifs risqués. Cependant, on voit que l'implémentation des modèles VaR et de Markowitz nécessitent seulement l'estimation des rendements moyens ρ_τ ainsi que la matrice de covariance Q_τ entre ces rendements sur la période d'investissement. En pratique, ces paramètres doivent être obtenus à partir des données disponibles. Ce problème a été abondamment étudié. Notre contribution est de montrer que si l'on utilise des estimations $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ de respectivement ρ_τ et Q_τ telles que $\|\hat{\rho}_\tau - \rho_\tau\|_\infty$ et $\|\hat{Q}_\tau - Q_\tau\|_\infty$ sont petites avec une grande probabilité, alors on peut définir des problèmes approchés de bonne qualité à partir de ces estimations. Nous détaillons ensuite la calibration d'un tel vecteur $\hat{\rho}_\tau$ et d'une telle matrice $\hat{Q}_\tau$. Dans le cas particulier où la moyenne $\rho_t = \rho$ et la matrice de covariance $Q_t = Q$ ne varient pas dans le temps, les estimations $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ utilisent autant de données r_t que possible. Étant données des observations $(r_t)_{1 \leq t \leq \tau - 1}$, si

$N = \tau - 1$ est le nombre de données disponibles, on choisit pour $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ la moyenne et matrice de covariance d'un processus α_t obtenu à partir du processus r_t par seuillage :

$$\hat{\rho}_\tau = \frac{1}{N}\sum_{t=1}^{N} \alpha_t, \quad \text{et} \quad \hat{Q}_\tau = \frac{1}{N}\sum_{t=1}^{N} (\alpha_t - \hat{\rho}_\tau)(\alpha_t - \hat{\rho}_\tau)^T, \tag{2.7}$$

où pour $i = 1, \ldots, n$,

$$\alpha_t(i) = r_t(i)\, 1(|r_t(i)| \leq K_0) + K_0 \frac{r_t(i)}{|r_t(i)|}\, 1(|r_t(i)| > K_0), \quad K_0 = \sigma\left(\frac{N}{\ln n(n+1) + \lambda \ln N}\right)^{\frac{1}{4}}, \tag{2.8}$$

λ est un paramètre positif à choisir et σ est la constante intervenant dans l'hypothèse 1. Dans le cas plus général où les paramètres ρ_t et Q_t varient lentement (sous l'hypothèse d'homogénéité temporelle locale), la mise en place d'une procédure adaptative se trouve justifiée. Ceci signifie que l'on doit déterminer une estimation $\hat{I}$ du meilleur intervalle à utiliser pour l'estimation des paramètres. Cette question est traitée dans la troisième section. Une fois le meilleur intervalle $\hat{I}$ trouvé, les estimations $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ de ρ_τ et Q_τ sont données par $\hat{\rho}_{\hat{I}}$ et $\hat{Q}_{\hat{I}}$ où pour tout intervalle I, on définit $\hat{\rho}_I$ et $\hat{Q}_I$ par :

$$\hat{\rho}_I = \frac{1}{|I|}\sum_{t\in I} \alpha_t^I, \quad \text{et} \quad \hat{Q}_I = \frac{1}{|I|}\sum_{t\in I} (\alpha_t^I - \hat{\rho}_I)(\alpha_t^I - \hat{\rho}_I)^T, \tag{2.9}$$

avec pour $i = 1, \ldots, n$,

$$\alpha_t^I(i) = r_t(i)\, 1(|r_t(i)| \leq K_0^I) + K_0^I \frac{r_t(i)}{|r_t(i)|}\, 1(|r_t(i)| > K_0^I), \quad K_0^I = \sigma\left(\frac{|I|}{\ln n(n+1) + \lambda \ln |I|}\right)^{\frac{1}{4}}. \tag{2.10}$$

Remarquons que cela signifie que l'on accorde plus ou moins de crédit aux données passées pour expliquer le futur. En ayant ce même souci en tête, [67] affecte des poids aux données. Si w_t est le poids appliqué aux données au temps t, la moyenne, sur un intervalle donné I, est alors donnée par $\hat{\rho}_I(w) = \frac{1}{|I|}\sum_{t\in I} w_t r_t$ et la matrice de covariance par $\frac{1}{|I|}\sum_{t\in I} w_t(r_t - \hat{\rho}_I(w))(r_t - \hat{\rho}_I(w))^T$. Plus une donnée est éloignée dans le passé, plus son poids est faible. On accorde ainsi plus de crédit aux données les plus récentes. Les poids constituent une suite géométrique et sont déterminés par maximisation de la vraisemblance en choisissant une loi mixture pour les rendements. D'un point de vue théorique, si les corrélations des rendements des actifs financiers varient lentement, les variances des rendements sont plus instables (les rendements étant souvent modélisés par un processus ARCH). Ainsi, le processus des rendements ne peut pas être considéré comme stationnaire. Notre approche tient compte de cette constatation, mais contrairement aux modèles que l'on vient de mentionner, nous utilisons une approche non paramétrique qui repose sur un modèle simple pour les rendements.

2.2 Contrôle de la précision des problèmes approchés

2.2.1 Hypothèses sur la moyenne, les variances et covariances

Sous l'hypothèse d'homogénéité, on suppose qu'il existe un intervalle passé d'extrémité droite τ tel que la moyenne, les variances et covariances varient lentement ou sont

presque constantes sur cet intervalle. La procédure adaptative que l'on décrit dans la section suivante, vise à déterminer le plus grand intervalle satisfaisant cette hypothèse. Pour un intervalle passé donné d'extrémité droite τ, nous devons donc être capable de décider, d'un point de vue théorique, si l'on peut considérer que la moyenne, les variances et covariances varient lentement sur cet intervalle ou non. Si ρ_t et Q_t varient lentement sur l'intervalle $I = [\tau - m, \tau]$ où $m \in \mathbb{N}^*$, alors les quantités :

$$\Delta_I^2(\rho) = \frac{1}{|I|}\sum_{t\in I} \|\rho_t - \rho_\tau\|_\infty^2, \quad \text{et} \quad \Delta_I^2(Q) = \frac{1}{|I|}\sum_{t\in I} \|Q_t - Q_\tau\|_\infty^2,$$

doivent être petites. De façon similaire, on s'attend à ce que $\Delta_J(\rho)$ et $\Delta_J(Q)$ soient petits pour tout sous-intervalle J de l'intervalle I. En particulier, si $\mathcal{I}(I)$ est un ensemble fini de sous-intervalles tests de l'intervalle I, (le choix de $\mathcal{I}(I)$ est discuté plus loin) alors $\Delta_J(\rho)$ et $\Delta_J(Q)$ doivent être petits pour tout intervalle $J \in \mathcal{I}(I)$. Mais afin de tenir compte des variances des composantes, on définit pour tout intervalle I :

$$V_I(\rho) = E\|\hat{\rho}_I - E\hat{\rho}_I\|_\infty, \quad V_I(Q) = E\|\hat{Q}_I - E\hat{Q}_I\|_\infty.$$

En fin de compte, nous dirons que ρ_t et Q_t varient lentement sur l'intervalle I si :

$$\Delta_J(\rho) \leq DV_J(\rho), \;\; \Delta_J(Q) \leq DV_J(Q), \text{ pour } J = I \text{ et } J \in \mathcal{I}(I), \tag{2.11}$$

où D est une constante fixe. Remarquons que si l'on considère un modèle à sauts où la moyenne et la matrice de covariance sont constantes par morceaux alors $D = 0$. L'intervalle idéal $\mathbb{I}$ d'homogénéité temporelle locale que l'oracle que l'on construit en section 2.3 vise à estimer est ensuite le plus grand intervalle (parmi une famille $\mathcal{I}$ d'intervalles candidats) tel que (2.11) ait lieu :

$$\mathbb{I} = \text{argmax}\,\{|I| \;\mid\; I \in \mathcal{I}, \;\; \Delta_J(\rho) \leq DV_J(\rho), \;\; \Delta_J(Q) \leq DV_J(Q), \text{ pour } J = I \text{ et } J \in \mathcal{I}(I)\}. \tag{2.12}$$

Remarquons que notre définition d'intervalle idéal d'homogénéité temporelle locale diffère de celle de [70]. Dans [70], la condition (2.11) doit être satisfaite seulement pour $J = I$. Cependant, supposons qu'un intervalle suffisamment grand I est tel que $Q_t = Q$ pour tout t dans I et que ρ_t varie beaucoup seulement sur la gauche de l'intervalle. En utilisant la définition d'un intervalle d'homogénéité temporelle locale donnée dans [70], nous accepterions probablement I comme un intervalle d'homogénéité temporelle locale alors que notre définition conclurait probablement le contraire, ce qui est plus raisonnable dans ce contexte.

2.2.2 Précision des problèmes approchés

Notre objectif est de construire des problèmes approchés, $(\widehat{\mathcal{V}})$ et $(\widehat{\mathcal{M}})$ pour $(\mathcal{V})$ et $(\mathcal{M})$ respectivement qui consisteront à utiliser des estimations spécifiques $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ de la moyenne et de la matrice de covariance sur la période d'investissement afin de résoudre $(\mathcal{V})$ et $(\mathcal{M})$. Nous commençons par définir la notion de précision des problèmes approchés. Pour alléger les notations, la fonction objectif des problèmes (2.5) et (2.6) sera notée f, le contexte permettant de lever l'ambiguïté. Les précisions de nos problèmes d'allocation d'actifs sont définies comme suit :

– pour le problème VaR, soient $(\widehat{x}, \widehat{y}, \widehat{z})$ une solution optimale du problème approché $(\widehat{\mathcal{V}})$ et (x^*, y^*, z^*) une solution de $(\mathcal{V})$. La précision du problème approché est :

$$\epsilon(\widehat{\mathcal{V}}) \equiv f(x^*) - f(\widehat{x}).$$

– pour le problème de Markowitz, soient $(\widehat{x}, \widehat{y}, \widehat{z})$ une solution optimale du problème approché $(\widehat{\mathcal{M}})$ et (x^*, y^*, z^*) une solution optimale de $(\mathcal{M})$. La précision du problème approché est :

$$\epsilon(\widehat{\mathcal{M}}) \equiv \max\left[f(\widehat{x}) - f(x^*),\ \left(\ell\, \mathbf{e}^T x^- - \rho_\tau^T \widehat{x} - r(n+1)\hat{x}_{n+1}\right)_+\right],$$

Remarquons que si $f(\hat{x}) \leq f(x^*)$, le risque du portefeuille $\hat{x}$ est inférieur au risque du portefeuille optimal et l'erreur correspond seulement à la violation de la contrainte de rendement.

2.2.2.1 Définition des problèmes approchés

L'idée la plus naturelle pour construire des problèmes Value-at-Risk et de Markowitz approchés est alors probablement d'utiliser les estimations $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ mentionnées ci-dessus pour les paramètres des problèmes (voir section 2.1.3). En substituant ces estimations dans le problème $(\mathcal{V})$ (resp. le problème $(\mathcal{M})$), on peut s'attendre à obtenir une approximation "assez bonne" de la solution du problème $(\mathcal{V})$ (resp. du problème $(\mathcal{M})$). C'est en effet vrai pour le modèle de gestion de portefeuilles VaR. Cependant, ce n'est pas vrai pour le modèle de Markowitz. Remarquons, par exemple, que l'utilisation directe de l'estimation de la moyenne dans la contrainte de performance du problème (2.6) peut même changer le statut de faisabilité du problème. Pour le modèle Value-at-Risk $(\mathcal{V})$, on définit le problème approché $(\widehat{\mathcal{V}})$ suivant :

$$(\widehat{\mathcal{V}}) \left\{ \begin{array}{l} \max \quad x^T \widehat{\rho}_\tau + r(n+1)x_{n+1} - \kappa(\varepsilon)\sqrt{x^T \widehat{Q}_\tau x} \\ (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-), \end{array} \right. , \tag{2.13}$$

où les définitions de $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ sont données en section 2.1.3 et leur calcul est détaillé en section 2.3. Le problème $(\widehat{\mathcal{M}})$, approximation du problème $(\mathcal{M})$, est alors défini comme le problème quadratique suivant :

$$(\widehat{\mathcal{M}}) \left\{ \begin{array}{l} \min \ \ \frac{1}{2} x^T \widehat{Q}_\tau x \\ (\widehat{\rho}_\tau + \eta_\rho \mathbf{e})^T x + r(n+1)x_{n+1} \geq \ell\, \mathbf{e}^T x^-, \\ (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-), \end{array} \right. \tag{2.14}$$

où η_ρ est tel que $\|\hat{\rho}_\tau - \rho_\tau\|_\infty \leq \eta_\rho$ avec une grande probabilité. Dans le cas où la moyenne, les variances et covariances sont constantes, on choisit :

$$\eta_\rho = \sigma\left(\frac{7}{3}\left(\frac{\ln n(n+1) + \lambda \ln N}{N}\right)^{\frac{3}{4}} + \sqrt{2}\sqrt{\frac{\ln 2n + \lambda \ln N}{N}}\right),$$

où $\lambda > 0$ est le paramètre utilisé pour calculer les estimations des paramètres (voir (2.8)), $N = \tau - 1$ est le nombre d'observations et dans le cas de paramètres variant lentement

$$\eta_\rho = (7 + 3\sqrt{2} + 6\sqrt{2e}D)\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}.$$

Remarquons que cela revient à utiliser une estimation optimiste ("meilleur cas") de la moyenne ρ_τ. D'autre part, avec une grande probabilité, l'ensemble de faisabilité de $(\widehat{\mathcal{M}})$ contient l'ensemble de faisabilité de $(\mathcal{M})$.
Dans ce qui suit, nous supposons que l'hypothèse de stationnarité ou d'homogénéité temporelle locale est satisfaite pour le processus des rendements. Dans chacun de ces cas, nous donnons la précision des problèmes de Markowitz et VaR approchés $(\widehat{\mathcal{M}})$ et $(\widehat{\mathcal{V}})$ définis ci-dessus. Dans le cas où les paramètres varient lentement, la définition donnée ci-dessus du problème $(\widehat{\mathcal{M}})$ fait intervenir l'intervalle $\mathbb{I}$ qui est inconnu. D'un point de vue pratique, cet intervalle sera remplacé par son estimation $\hat{I}$ obtenue en utilisant l'algorithme adaptatif décrit en section 2.3.

2.2.2.2 Le cas stationnaire

Nous considérons tout d'abord le cas où les moyennes, variances et covariances sont constantes. On suppose que N rendements $(r_i)_{1\leq i\leq N}$ satisfaisant l'hypothèse 1 sont disponibles afin de calculer les estimations empiriques $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ de la moyenne et de la matrice de covariance du processus α_t défini dans la section 2.1.3.

Définition 2.2.1 *Pour toute matrice symétrique Q, soit $\beta(Q)$ tel que la fonction quadratique x^TQx soit $\beta(Q)$-fortement convexe par rapport à $\|.\|_1$, i.e :*

$$\forall (x,y)\ \ (x-y)^TQ(x-y)\geq \beta(Q)\,\|x-y\|_1^2.$$

Remarquons que $\beta(Q)\geq \frac{\lambda_{\min}(Q)}{n}$ et pour $Q=I_n$, $\beta(Q)=\frac{\lambda_{\min}(Q)}{n}=\frac{1}{n}$. Le théorème suivant donne la précision du problème $(\widehat{\mathcal{V}})$ lorsque les estimations empiriques $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ de ρ_τ et Q_τ sont utilisées dans le cas stationnaire.

Théorème 2.2.1 *Supposons les moyennes $\rho_t=\rho$ et matrices de covariances $Q_t=Q$ constantes et soient $(\hat{\rho}_\tau,\hat{Q}_\tau)$ les estimations de (ρ_τ,Q_τ) données dans (2.7). Soit $\lambda>0$ le paramètre intervenant dans (2.8). Si $\ln n(n+1)+\lambda\ln N\leq N$, alors il existe des constantes k_1 et k_2 et un ensemble $S\subseteq\Omega$ de probabilité au moins $1-\frac{3}{N^\lambda}$ tel que pour tout $\omega\in S$, la précision du problème $(\widehat{\mathcal{V}})$ est majorée de la façon suivante :*

$$\epsilon(\widehat{\mathcal{V}})\leq (k_1+2\kappa(\varepsilon)\sqrt{k_2})\sigma\|x^-\|_1\left(\frac{\ln n(n+1)+\lambda\ln N}{N}\right)^{\frac{1}{4}}. \tag{2.15}$$

De plus, si la matrice Q est non dégénérée, i.e $\min_{\|x\|_1=1}x^TQx=\beta(Q)>0$, alors

$$\epsilon(\widehat{\mathcal{V}})\leq\left(k_1+\frac{2\kappa(\varepsilon)k_2\sigma}{\sqrt{\beta(Q)}}\right)\sigma\|x^-\|_1\sqrt{\frac{\ln n(n+1)+\lambda\ln N}{N}}. \tag{2.16}$$

Remarquons la décroissance rapide des majorants en fonction de N. D'autre part, puisque $\ln n(n+1)+\lambda\ln N\leq N$, si $\beta(Q)$ est suffisamment grand alors la majoration (2.16) est effectivement meilleure que la majoration (2.15). La précision du problème approché $(\widehat{\mathcal{V}})$ peut aussi être donnée sans la condition $\ln n(n+1)+\lambda\ln N\leq N$. Dans le cas général, on obtient des seconds membres compliqués dans les inégalités ci-dessus (2.15) et (2.16) (voir la preuve de ce théorème).
Il se trouve que la précision du problème $(\widehat{\mathcal{M}})$ peut aussi être contrôlée grâce au théorème suivant.

Théorème 2.2.2 *Supposons que les moyennes $\rho_t = \rho$ et les matrices de covariances $Q_t = Q$ soient constantes et soient $(\hat{\rho}_\tau, \hat{Q}_\tau)$ les estimations de (ρ_τ, Q_τ) données dans (2.7). Soit $\lambda > 0$ le paramètre intervenant dans (2.8) et tel que $\ln n(n+1) + \lambda \ln N \leq N$. Il existe alors des constantes k_1 et k_2 et un ensemble $S \subseteq \Omega$ de probabilité au moins $1 - \frac{3}{N^\lambda}$ tel que pour tout $\omega \in S$, $(\widehat{\mathcal{M}})$ est faisable et*

$$\epsilon(\widehat{\mathcal{M}}) \leq \max\left(k_1, k_2\sigma\|x^-\|_1\right)\sigma\|x^-\|_1\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}; \tag{2.17}$$

et si Q est définie positive,

$$\epsilon(\widehat{\mathcal{M}}) \leq \max\left(k_1, k_2\sigma\|x^-\|_1\tilde{\varepsilon}\right)\sigma\|x^-\|_1\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}; \tag{2.18}$$

où

$$\tilde{\varepsilon} = \min(1, \frac{2k_2\sigma^2}{\beta(Q)}\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}).$$

On peut faire les mêmes remarques que pour le théorème précédent. Les constantes k_1 et k_2 ci-dessus (dans les théorèmes 2.2.1 et 2.2.2) sont données par $k_1 = 2(\sqrt{2} + \frac{7}{3})$ et $k_2 = \frac{26}{3}\sqrt{2} + \frac{160}{9}$. Remarquons que dans le cas où $\ln n(n+1) + \lambda \ln N \leq N$, le théorème 2.2.2 est valide avec la version simplifiée suivante de η_ρ utilisé pour définir le problème approché $(\widehat{\mathcal{M}})$:

$$\eta_\rho = \frac{k_1}{2}\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}.$$

Remarquons que l'on a plus précisément, concernant la satisfaction de la contrainte de rendement :

$$P\left(\rho^T\widehat{x} + r(n+1)\hat{x}_{n+1} - \ell\,\mathbf{e}^T x^- > -k_1\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\|x^-\|_1\right) \geq 1 - \frac{1}{N^\lambda};$$

avec $k_1 = 2(\sqrt{2} + \frac{7}{3})$.

2.2.2.3 Paramètres variant lentement

Dans le cas où les paramètres varient lentement, la précision des problèmes approchés est obtenue grossièrement en remplaçant dans les résultats donnés dans le cas stationnaire, le nombre d'observations N par la longueur$|\mathbb{I}|$ de l'intervalle idéal d'homogénéité temporelle locale. Ceci tient au fait que les estimations $\hat{\rho}_{\hat{I}}$ et $\hat{Q}_{\hat{I}}$ obtenues dans le cas où les paramètres varient lentement, utilisent les données de l'intervalle $\hat{I}$ et on peut montrer que $\hat{\rho}_{\hat{I}}$ et $\hat{Q}_{\hat{I}}$ sont proches de $\hat{\rho}_{\mathbb{I}}$ et $\hat{Q}_{\mathbb{I}}$ (voir les théorèmes 2.3.2 et 2.3.3). Nous donnons maintenant la précision des problèmes $(\widehat{\mathcal{V}})$ et $(\widehat{\mathcal{M}})$ sous l'hypothèse d'homogénéité temporelle locale.
Pour mettre en oeuvre l'algorithme adaptatif, différents choix de sous intervalles $\mathcal{I}(I)$ peuvent être envisagés. Si l'on utilise les sous intervalles définis dans [70], on définit :

$$\mathcal{I}_+(I) = \left\{\{I' \mid I' \in \mathcal{I} \cap I' \subseteq I\} \bigcup \{J \mid J \in \mathcal{I}(I') \cap I' \in \mathcal{I} \cap I' \subseteq I\}\right\}. \tag{2.19}$$

Lorsque l'on choisit d'autres stratégies pour le choix des sous intervalles tests (voir section 2.3.2), les résultats donnés dans cette section sur la précision des problèmes $(\widehat{\mathcal{V}})$ et $(\widehat{\mathcal{M}})$

sont valables en redéfinissant $\mathcal{I}_+(I)$ (voir section 2.3.2).

Modèle VaR

Théorème 2.2.3 *Soient $(\hat{\rho}_\tau, \hat{Q}_\tau)$ les estimations de (ρ_τ, Q_τ) données dans la section 2.1.3 dans le cas où les paramètres varient lentement. Soit $\lambda > 0$ le paramètre intervenant dans* (2.10) *et tel que $\ln n(n+1) + \lambda \ln |\mathbb{I}| \leq |\mathbb{I}|$. Il existe alors des constantes k_1', k_2' et un ensemble $S \subseteq \Omega$ de probabilité au moins $1 - \sum_{J \in \mathcal{I}_+(\mathbb{I})} \frac{3}{|J|^\lambda}$ tel que pour tout $\omega \in S$,*

$$\epsilon(\widehat{\mathcal{V}}) \leq (k_1' + 2\kappa(\varepsilon)\sqrt{k_2'})\sigma\|x^-\|_1 \left(\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}\right)^{\frac{1}{4}}. \tag{2.20}$$

De plus, si la matrice Q_τ est non dégénérée, i.e $\min_{\|x\|_1=1} x^T Q_\tau x = \beta(Q_\tau) > 0$, alors

$$\epsilon(\widehat{\mathcal{V}}) \leq \left(k_1' + \frac{2\kappa(\varepsilon)k_2'\sigma}{\sqrt{\beta(Q_\tau)}}\right)\sigma\|x^-\|_1 \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}. \tag{2.21}$$

Modèle de Markowitz. La précision du problème $(\widehat{\mathcal{M}})$ est donnée par le théorème suivant.

Théorème 2.2.4 *Soient $(\hat{\rho}_\tau, \hat{Q}_\tau)$ les estimations de (ρ_τ, Q_τ) données en section 2.1.3 dans le cas où les paramètres varient lentement. Soit $\lambda > 0$ le paramètre intervenant dans* (2.10) *et tel que $\ln n(n+1) + \lambda \ln |\mathbb{I}| \leq |\mathbb{I}|$. Il existe alors des constantes k_1', k_2' et un ensemble $S \subseteq \Omega$ de probabilité au moins $1 - \sum_{J \in \mathcal{I}_+(\mathbb{I})} \frac{3}{|J|^\lambda}$ tel que pour tout $\omega \in S$, le problème $(\widehat{\mathcal{M}})$ est faisable et*

$$\epsilon(\widehat{\mathcal{M}}) \leq \max\left(k_1', k_2'\sigma\|x^-\|_1\right)\sigma\|x^-\|_1 \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}; \tag{2.22}$$

et si Q_τ est définie positive alors

$$\epsilon(\widehat{\mathcal{M}}) \leq \max\left(k_1', k_2'\sigma\|x^-\|_1\tilde{\varepsilon}\right)\sigma\|x^-\|_1 \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}; \tag{2.23}$$

où

$$\tilde{\varepsilon} = \min(1, \frac{2k_2'\sigma^2}{\beta(Q_\tau)} \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}).$$

On constate dans les deux théorèmes 2.2.3 et 2.2.4, la décroissance rapide des majorants en fonction de $|\mathbb{I}|$. D'autre part, puisque $\ln n(n+1) + \lambda \ln |\mathbb{I}| \leq |\mathbb{I}|$, si $\beta(Q_\tau)$ est suffisamment grand alors la majoration (2.23) est effectivement meilleure que la majoration (2.22). Enfin, il est intéressant de remarquer que les majorations obtenues augmentent faiblement avec le nombre de composantes n du processus r_t. Ainsi, si $\mathbb{I}$ est suffisamment long, on peut construire un problème approché $(\widehat{\mathcal{V}})$ ou $(\widehat{\mathcal{M}})$ de bonne qualité même lorsque le nombre d'actifs n est très élevé.

2.3 Méthode adaptative

Nous supposons que la moyenne ρ_t du processus r_t et la matrice de covariance Q_t sont dépendantes du temps. Sur la base des observations $(r_t)_{1\leq t\leq \tau-1}$ d'un processus (r_t) satisfaisant l'hypothèse 1, le but est d'estimer ρ_τ et Q_τ sous l'hypothèse d'homogénéité temporelle locale. Sous cette hypothèse d'homogénéité temporelle locale, nous supposons qu'il existe des intervalles $\mathbb{I}_\rho = [\tau - m_\rho, \tau]$ et $\mathbb{I}_Q = [\tau - m_Q, \tau]$, tels que la moyenne ρ_t ne varie pas beaucoup sur $\mathbb{I}_\rho$ et Q_t ne varie pas beaucoup sur $\mathbb{I}_Q$. D'un point de vue théorique, il peut y avoir un changement dans la moyenne, les variances et covariances restant constantes. Cependant, dans ce cas, nous devrions tenir compte du changement dans la moyenne pour estimer la matrice de covariance et le choix d'un bon intervalle adaptatif $\mathbb{I}_Q$ serait plus difficile. Nous rechercherons donc le plus grand intervalle $\hat{I}$ (une estimation de $\mathbb{I}$) tel qu'à la fois les moyennes, les variances et les covariances varient lentement sur $\hat{I}$; ce qui revient à déterminer une estimation de l'intersection de $\mathbb{I}_\rho$ et $\mathbb{I}_Q$. Une méthode alternative consiste à déterminer tout d'abord une estimation $\hat{I}_\rho$ de $\mathbb{I}_\rho$ et de déterminer ensuite une estimation $\hat{I}_Q$ du plus grand intervalle contenu dans $\hat{I}_\rho$ tel que Q_t varie lentement sur cet intervalle. Nous donnons la précision théorique des estimateurs obtenus en utilisant un intervalle d'homogénéité. Les preuves peuvent s'adapter directement au cas où l'on recherche deux intervalles d'homogénéité afin de montrer la précision des estimations $\hat{\rho}_{\hat{I}_\rho}$ et $\hat{Q}_{\hat{I}_Q}$ lorsque deux intervalles d'homogénéité $\hat{I}_\rho$ et $\hat{I}_Q$ sont déterminés. Pour illustration, nous donnons la précision de l'estimation de la moyenne par $\hat{\rho}_{\hat{I}_\rho}$ quand des intervalles d'homogénéité séparés sont déterminés pour estimer, d'une part, la moyenne et d'autre part, les variances et covariances. Les deux méthodes seront testées numériquement. Dans la section suivante, nous décrivons un algorithme adaptatif qui permet de déterminer une estimation $\hat{I}$ de $\mathbb{I}$. La question centrale sera de décider, à la fois d'un point de vue théorique et pratique, si sur un intervalle donné, on peut considérer que les fonctions ρ_t et Q_t varient lentement ou non.

2.3.1 Description de l'algorithme

Le choix de l'intervalle adaptatif d'homogénéité temporelle locale est effectué de la façon suivante. Soit $\mathcal{I}$ une famille d'intervalles candidats ordonnés de la forme $[\tau - m, \tau[, m \in \mathbb{N}^*$, et pour tout $I \in \mathcal{I}$, soit $\mathcal{I}(I)$ une famille de sous-intervalles (de l'intervalle I) tests J. On suppose que l'on dispose d'une règle permettant de savoir si on peut considérer que sur un intervalle candidat I donné, la moyenne, les variances et covariances varient lentement sur cet intervalle I ou non (ce problème est abordé plus loin). L'intervalle d'homogénéité temporelle sélectionné $\hat{I}$ est alors tel que pour tout $I \in \mathcal{I}$ satisfaisant $I \subseteq \hat{I}$, I est accepté et le plus petit intervalle $I \in \mathcal{I}$ tel que $\hat{I} \subsetneq I$ est rejeté. Une fois l'intervalle $\hat{I}$ déterminé, les estimations adaptatives ρ_τ et Q_τ sont données par $\hat{\rho}_{\hat{I}}$ et $\hat{Q}_{\hat{I}}$. Le choix adaptatif de l'intervalle d'homogénéité est ainsi effectué en utilisant la procédure représentée par la figure 2.1 ci-après.

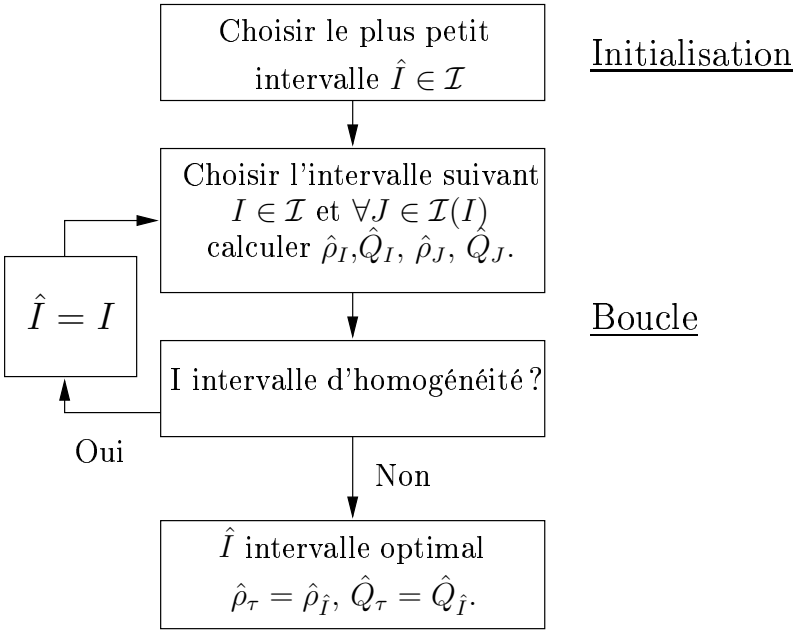


FIG. 2.1 – Représentation schématique de l'algorithme adaptatif

Nous devons alors être capable de décider, d'un point de vue théorique et pratique, si sur un intervalle I donné , la moyenne ρ_t et la matrice de covariance Q_t varient lentement ou non. Si sur $I = [\tau - m, \tau[$, ρ_t ne varie "pas beaucoup", alors la valeur moyenne de ρ_t sur I est proche de la valeur moyenne de ρ_t sur tout sous-intervalle $J \in \mathcal{I}(I)$. Mais la valeur moyenne de ρ_t sur I est proche de la valeur moyenne de r_t sur I qui est proche de la valeur moyenne de α_t sur I. Ainsi, si ρ_t ne varie "pas beaucoup" sur I, alors pour tout sous-intervalle $J \in \mathcal{I}(I)$, $\hat{\rho}_J$ est proche de $\hat{\rho}_I$ où $\hat{\rho}_I$ est défini en (2.9). De manière analogue, si Q_t est pratiquement constante sur I, alors pour tout sous-intervalle $J \in \mathcal{I}(I)$, $\hat{Q}_J$ est proche de $\hat{Q}_I$. En fait, décider si un intervalle donné I est d'homogénéité temporelle ou non, revient à faire un test. Par exemple, dans le cas particulier de fonctions constantes par morceaux ρ_t et Q_t, on doit faire le test :

$$H_0^I : \forall (t,t') \in I^2,\ \rho_t = \rho_{t'}, Q_t = Q_{t'} \qquad H_1^I :\ \exists (t,t') \in I^2,\ \mid\ \rho_t \neq \rho_{t'} \text{ ou } Q_t \neq Q_{t'}.$$

Nous avons alors besoin du théorème suivant qui quantifie la proximité entre deux estimations de la moyenne $\hat{\rho}_I$ et $\hat{\rho}_J$ ou de la matrice de covariance $\hat{Q}_I$ et $\hat{Q}_J$ faites sur des intervalles d'homogénéité temporelle locale I et J.

Théorème 2.3.1 *Supposons que r_t satisfasse l'hypothèse 1. Soient I et J deux intervalles d'homogénéité temporelle locale. Soient $\hat{\rho}_I$ et $\hat{Q}_I$ les estimateurs définis par* (2.9)*. Alors pour tout $\lambda > 0$ tel que $\ln n(n+1) + \lambda \ln |I| \leq |I|$ et $\ln n(n+1) + \lambda \ln |J| \leq |J|$, nous avons :*

$$P\left(\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \geq \gamma_\rho(|I|,|J|,\lambda)\right) \leq \frac{1}{|I|^\lambda} + \frac{1}{|J|^\lambda}, \tag{2.24}$$

$$P\left(\|\hat{Q}_I - \hat{Q}_J\|_\infty \geq \gamma_Q(|I|,|J|,\lambda)\right) \leq 2(\frac{1}{|I|^\lambda} + \frac{1}{|J|^\lambda}), \tag{2.25}$$

avec $\gamma_\rho(|I|,|J|,\lambda) = 2\sqrt{2e}D\sigma\left(\sqrt{\frac{\ln n}{|I|}}+\sqrt{\frac{\ln n}{|J|}}\right)+(\frac{7}{3}+\sqrt{2})\sigma\left(f(|I|,\lambda)+f(|J|,\lambda)\right),$

$$\gamma_Q(|I|,|J|,\lambda) = (k_Q D + k'_Q)\sigma^2(f(|I|,\lambda)+f(|J|,\lambda)),$$

où $f(|I|,\lambda) = \sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$, *et* k_Q *et* k'_Q *sont des constantes fournies dans la preuve du théorème.*

Remarque 2.3.1 *Les conditions* $\ln n(n+1)+\lambda\ln|I| \leq |I|$ *et* $\ln n(n+1)+\lambda\ln|J| \leq |J|$ *dans le théorème ci-dessus 2.3.1 peuvent être supprimées. Dans ce cas, on obtient des expressions plus compliquées de* γ_ρ *et* γ_Q. *Ce cas plus général est traité dans la preuve de ce théorème.*

D'après le théorème 2.3.1, nous accepterons I en tant qu'intervalle d'homogénéité si pour tout sous-intervalle $J \in \mathcal{I}(I)$:

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq \gamma_\rho(|I|,|J|,\lambda) \text{ et } \|\hat{Q}_I - \hat{Q}_J\|_\infty \leq \gamma_Q(|I|,|J|,\lambda), \tag{2.26}$$

où λ est le paramètre positif intervenant dans (2.10), et rejetterons I sinon. Ainsi, si I est effectivement un intervalle d'homogénéité temporelle locale, alors la probabilité que (2.26) ait lieu peut être contrôlée par un choix approprié de λ. Notons que lorsque les longueurs de I et J croissent, alors les membres droits de (2.26) tendent rapidement vers 0, comme on s'y attendait. Nous allons maintenant discuter du choix des ensembles $\mathcal{I}$, $\mathcal{I}(I)$ et du paramètre λ.

2.3.2 Choix des ensembles $\mathcal{I}$, $\mathcal{I}(I)$ et du paramètre λ

Le moyen le plus simple de choisir les ensembles $\mathcal{I}$ et $\mathcal{I}(I)$ pour un tel algorithme adaptatif est décrit dans [70]. Nous rappelons brièvement ce choix. On se fixe une grille $\mathcal{G} = \{t_k = m_0\,k,\ \ k \in \mathbb{N}^*\}$ où $m_0 \in \mathbb{N}^*$ est le pas de la grille et on pose $t_0 = 1$. Soit $\tau \geq 2m_0$, un point pour lequel on veut mettre en oeuvre l'algorithme adaptatif pour déterminer des estimations de ρ_τ et Q_τ. Si $k^* = E[\frac{\tau}{m_0}]$, on peut choisir pour $\mathcal{I}$ l'ensemble des intervalles :

$$\mathcal{I} = \{[t_k, t_{k^*}[,\ \ 0 \leq k < k^*\}.$$

Pour chaque $I_k = [t_k, t_{k^*}[\in \mathcal{I}$, l'ensemble $\mathcal{I}(I_k)$ des sous-intervalles de tests de l'intervalle I_k est alors l'ensemble de tous les intervalles plus petits dont les extrémités appartiennent à $\mathcal{G}$ et avec soit la même extrémité gauche soit la même extrémité droite que l'intervalle I_k :

$$\mathcal{I}(I_k) = \{[t_k, t_{k'}[\cup [t_{k'}, t_{k^*}[,\ \text{ avec }\ k < k' < k^*\}.$$

On pourrait aussi utiliser une grille dynamique adaptée au temps τ de l'estimation. Dans ce cas, les intervalles de l'ensemble $\mathcal{I}$ sont du type $[\tau - m_0 k, \tau[$ où $k \in \mathbb{N}^*$.
Quelques modifications de ces choix peuvent améliorer la performance ou la rapidité de l'algorithme.

- Pour commencer, on peut introduire plus de flexibilité dans le choix de l'ensemble $\mathcal{I}$. En effet, la différence de longueur de deux intervalles successifs de $\mathcal{I}$ n'est pas nécessairement m_0, où m_0 est la longueur du plus petit sous-intervalle de test. D'autre part, si on décide de déterminer deux intervalles d'homogénéité $\hat{I}_\rho$ et $\hat{I}_Q$, afin d'estimer respectivement la moyenne et la matrice de covariance, la longueur $m_0(\rho)$ du plus

petit sous-intervalle de test choisi pour déterminer $\hat{I}_\rho$ et la longueur $m_0(Q)$ du plus petit sous-intervalle de test choisi pour déterminer $\hat{I}_Q$ ne sont pas nécessairement les mêmes. Cependant, les choix de $m_0(\rho)$ et $m_0(Q)$ doivent être tels que les estimateurs de la moyenne et de la matrice de covariance utilisant respectivement $m_0(\rho)$ et $m_0(Q)$ données aient de petites variances. Ces valeurs, bien sûr, dépendent de la variance des composantes sur $\mathbb{I}_\rho$ et $\mathbb{I}_Q$. Nous avons besoin de moins de données pour obtenir une bonne estimation de la moyenne. Dans un contexte financier, au vu de la volatilité des rendements, l'algorithme adaptatif peut être implémenté avec disons $m_0(\rho) = 15$ pour déterminer une estimation de $\mathbb{I}_\rho$. Cependant, pour déterminer une estimation de $\mathbb{I}_Q$, on doit prendre des valeurs de $m_0(Q)$ plus grande, disons au moins n où n est le nombre de composantes (si $m_0(Q)$ est inférieur au nombre de composantes alors l'estimation adaptative de la matrice de covariance n'est pas inversible). Cela signifie que l'on aura une estimation très grossière de $\mathbb{I}_Q$ si le pas $m_0(Q)$ est grand. Le même phénomène se produit pour $\mathbb{I}_\rho$ si le pas $m_0(\rho)$ est grand. On trouvera, au mieux, l'intervalle optimal $\mathbb{I}_Q$ avec une probabilité $\frac{1}{m_0(Q)}$. Dans ce qui suit, on ômet la dépendance de m_0 en fonction de ρ et Q (toutes les remarques qui suivent sur m_0 seront alors valides pour $m_0(\rho)$ et $m_0(Q)$). Une modification simple de l'ensemble $\mathcal{I}$ permet de déterminer les intervalles d'homogénéité optimaux avec une plus grande probabilité si l'amplitude des changements est suffisamment importante. Pour un choix fixé de m_0, supposons que le plus petit intervalle de l'ensemble $\mathcal{I}$ soit constitué des m_0 dernières observations. Nous sommes alors seulement capables d'identifier des ruptures s'étant produites il y a plus de m_0 pas de temps. Si $I = [t_\ell, \tau[$ est un intervalle de l'ensemble $\mathcal{I}$, le successeur de I dans $\mathcal{I}$ peut être obtenu en ajoutant k données au lieu de m_0 conduisant ainsi à l'intervalle $I = [t_\ell - k, \tau[$ où $k \in \mathbb{N}^*$. Même des valeurs de k aussi petites que 1 (le successeur d'un intervalle de l'ensemble $\mathcal{I}$ a la longueur de son père plus un) peut donner des améliorations spectaculaires (voir la section 2.5).

– On peut aussi modifier les sous-intervalles de tests sans changer beaucoup la performance de l'algorithme tout en améliorant sa rapidité. Pour une itération donnée de l'algorithme, si on a accepté un intervalle $I = [t_\ell, \tau[$ comme un intervalle d'homogénéité, les données de l'intervalle successeur $I' = [t_\ell - k, \tau[$ les plus susceptibles d'être avant la rupture (s'il y en a une) dans I' sont les k données les plus à gauche. C'est pourquoi une modification simple des sous-intervalles de tests consiste à choisir un seul sous-intervalle $J = [t_\ell - k, t_\ell - k + m_0[$ de l'intervalle I'. Si un intervalle est d'homogénéité, alors cette procédure l'acceptera avec une plus grande probabilité que l'algorithme adaptatif tel qu'il est présenté dans [70] puisque le test effectué pour savoir si cet intervalle est accepté est un des tests effectués par l'algorithme adaptatif utilisé dans [70]. De plus, s'il y a une rupture dans I', puisque l'intervalle précédent I était accepté, la rupture se situe certainement sur la gauche de l'intervalle. Les m_0 données se trouvant sur la gauche de l'intervalle I' sont alors les m_0 données les plus susceptibles de fournir des estimations éloignées de celles obtenues en utilisant toutes les données de l'intervalle I'. Ce choix d'implémentation sera noté (a) par la suite. Par ailleurs, au lieu de s'intéresser à la différence des estimations sur $I' = [t_\ell - k, \tau[$ et $J = [t_\ell - k, t_\ell - k + m_0[$, on peut tester la différence des estimations sur J et $J' = [\tau - m_0, \tau[$, qui est donc un intervalle fixe (choix d'implémentation (b)). De cette façon, la moyenne et la matrice de covariance varient lentement sur l'ensemble d'au moins un des deux intervalles (l'intervalle J'). Ayant le même objectif en tête,

on peut aussi tester la différence entre les estimations sur J et $J' = [\tau - m_0, \tau[$ si la longueur de I' est inférieure ou égale à $2m_0$ et sur J et $J' = [t_\ell - k + m_0, \tau[$ sinon (choix d'implémentation (c)). Ainsi, à chaque itération de l'algorithme, nous devons mettre en place un seul test, dans lequel on aura confiance si $m_0(\rho)$ et $m_0(Q)$ ne sont pas trop petits.

Nous montrerons, à la fois d'un point de vue théorique et pratique, les comportements proches de ces variantes de l'algorithme adaptatif. Par ailleurs, nous calibrerons λ en utilisant différentes techniques. On peut, par exemple, déterminer λ de telle sorte que l'erreur de type I soit contrôlée dans un modèle à sauts. λ peut aussi être choisi en utilisant des simulations de Monte Carlo.

Enfin, précisons que si les choix d'implémentation (b) et (c) sont retenus alors $\mathcal{I}_+(I)$ n'est plus défini par (2.19) mais regroupe simplement l'ensemble des sous-intervalles de tests de tous les intervalles candidats $I' \subseteq I$.

2.3.3 Qualité de l'estimation

La question centrale que nous traitons maintenant est de déterminer la qualité de l'approximation de ρ_τ et Q_τ par les estimations adaptatives $\hat{\rho}_{\hat{I}}$ et $\hat{Q}_{\hat{I}}$. Rappelons que l'intervalle idéal $\mathbb{I}$ est l'intervalle qui satisfait :

$$\mathbb{I} = \operatorname{argmax}\{|I| \mid I \in \mathcal{I},\ \Delta_J(\rho) \leq DV_J(\rho),\ \Delta_J(Q) \leq DV_J(Q), \text{ pour } J = I \text{ et } J \in \mathcal{I}(I)\}, \tag{2.27}$$

où D est une constante fixée. On donne maintenant les qualités des estimations $\hat{\rho}_{\mathbb{I}}$ et $\hat{Q}_{\mathbb{I}}$ qui seraient utilisées si l'intervalle idéal d'homogénéité temporelle locale était connu.

Théorème 2.3.2 *Soit $\lambda > 0$ tel que $\ln n(n+1) + \lambda \ln |\mathbb{I}| \leq |\mathbb{I}|$, alors*

$$P\left(\|\hat{\rho}_{\mathbb{I}} - \rho_\tau\|_\infty \geq (\frac{7}{3} + \sqrt{2} + 2\sqrt{2e}D)\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}\right) \leq \frac{1}{|\mathbb{I}|^\lambda}, \tag{2.28}$$

$$P\left(\|\hat{Q}_{\mathbb{I}} - Q_\tau\|_\infty \geq (k_Q D + k'_Q)\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}\right) \leq \frac{2}{|\mathbb{I}|^\lambda}, \tag{2.29}$$

où $k_Q = 2\left(1 + 4\sqrt{e}(\sqrt{2} + 2\sqrt{e} + 2)\right)$ et $k'_Q = \frac{160}{9} + \frac{26}{3}\sqrt{2} + 4\sqrt{2e}D$.

Le théorème suivant donne ensuite la précision des estimations adaptatives.

Théorème 2.3.3 *Soit $\hat{I}$ l'intervalle sélectionné par l'algorithme adaptatif et λ le paramètre intervenant dans (2.10). On suppose que $\ln n(n+1) + \lambda \ln m_0 \leq m_0$, alors*

$$P\Bigg((\|\hat{\rho}_{\hat{I}} - \rho_\tau\|_\infty > (7 + 3\sqrt{2} + 6\sqrt{2e}D)\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}) \bigcup$$
$$(\|\hat{Q}_{\hat{I}} - Q_\tau\|_\infty > 3(k_Q D + k'_Q)\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}})\Bigg) \leq \sum_{J \in \mathcal{I}_+(\mathbb{I})} \frac{3}{|J|^\lambda}, \tag{2.30}$$

où k_Q et k'_Q sont définis dans le théorème 2.3.2.

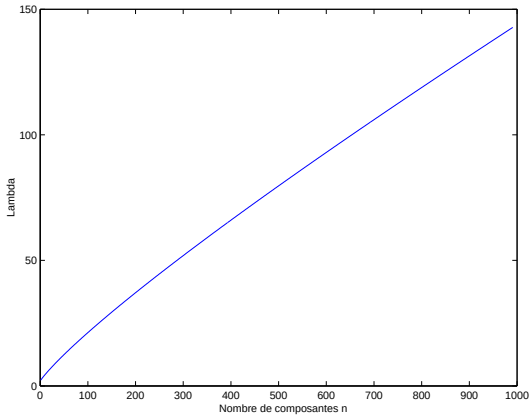


FIG. 2.2 – Évolution de la valeur maximale admissible pour λ en fonction du nombre de composantes n si $m_0 = n$.

Le théorème 2.3.3 montre en fait que la qualité des estimations adaptatives est proche de la qualité des estimations $\hat{\rho}_{\mathbb{I}}$ et $\hat{Q}_{\mathbb{I}}$ qui seraient utilisées si l'intervalle idéal $\mathbb{I}$ était connu. L'algorithme adaptatif peut être vu comme une oracle qui, à partir d'une série d'observations $(r_t)_{t=1,\ldots,\tau-1}$ d'un processus satisfaisant l'hypothèse 1, donne des estimations de ρ_τ et Q_τ qui sont proches des vraies valeurs inconnues. Si on détermine deux intervalles d'homogénéité $\hat{I}_\rho$ et $\hat{I}_Q$ pour respectivement la moyenne ρ et la matrice de covariance Q, alors les définitions des intervalles optimaux d'homogénéité se déduisent facilement de la définition (2.27) de l'intervalle intersection de $\mathbb{I}_\rho$ et $\mathbb{I}_Q$. Par exemple, l'intervalle optimal d'homogénéité temporelle locale $\mathbb{I}_\rho$ pour ρ satisfait :

$$\mathbb{I}_\rho = \operatorname{argmax}\{|I| \mid I \in \mathcal{I},\ \Delta_J(\rho) \leq DV_J(\rho), \text{ pour } J = I \text{ et } J \in \mathcal{I}(I)\}. \tag{2.31}$$

Une estimation adaptative $\hat{I}_\rho$ de $\mathbb{I}_\rho$ peut ensuite être déterminée en utilisant la même règle d'acceptation concernant la moyenne que pour la détermination de $\hat{I}$. En suivant la preuve du théorème ci-dessus, on peut ensuite obtenir la précision de $\hat{\rho}_{\hat{I}_\rho}$ suivante :

$$P\left(\|\hat{\rho}_{\hat{I}_\rho} - \rho_\tau\|_\infty > k\,\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}_\rho|}{|\mathbb{I}_\rho|}}\right) \leq \sum_{J \in \mathcal{I}_+(\mathbb{I}_\rho)} \frac{1}{|J|^\lambda},$$

où la constante $k = 7 + 3\sqrt{2} + 6\sqrt{2e}D$. Un résultat similaire peut être obtenu pour la précision de l'estimation de Q_τ par $\hat{Q}_{\hat{I}_Q}$ en utilisant les données de l'intervalle $\hat{I}_Q$. La condition $\ln n(n+1) + \lambda \ln m_0 \leq m_0$ dans le théorème ci-dessus peut être supprimée mais cela conduit à des seconds membres plus compliqués. Cependant, cette condition n'est pas trop restrictive (voir figure 2.2). Par exemple, pour $n = 40$, si on choisit $m_0 = n$, on peut alors prendre des valeurs de λ aussi grande que 8.84. Si on choisit $\lambda = 1$ et $m_0 = n$, alors il suffit que (r_t) ait plus de 6 composantes ($n \geq 6$) pour avoir $\ln n(n+1) + \lambda \ln m_0 \leq m_0$. Remarquons qu'en suivant la preuve du théorème 2.3.3, on peut borner la précision des estimations adaptatives obtenues en utilisant les différentes implémentations de l'algorithme adaptatif décrites en section 2.3.2.

Théorème 2.3.4 *Soit $\mathbb{I}_\rho = [\tau - m_0 - k_{\mathbb{I}}, \tau[$ ($k_{\mathbb{I}} \in \mathbb{N}$) l'intervalle d'homogénéité temporelle pour ρ, soit $\lambda > 0$ tel que $\ln n(n+1) + \lambda \ln m_0 \leq m_0$ et soit*

$$P_\rho = P\left(\|\hat{\rho}_{\hat{I}_\rho} - \rho_\tau\|_\infty > k\,\sigma \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}_\rho|}{|\mathbb{I}_\rho|}} \right),$$

avec $k = 7 + 3\sqrt{2} + 6\sqrt{2e}D$. Différents choix de sous-intervalles de tests ont été envisagés en section 2.3.2. Si on veut tester l'homogénéité sur $I = [\tau - k, \tau[$, on peut contrôler la différence des estimations sur I et $J = [\tau - k, \tau - k + m_0 - 1]$ (choix (a)), sur J et $J' = [\tau - m_0, \tau[$ (choix (b)), ou sur J et ($J' = [\tau - m_0, \tau[$ si $k \leq 2m_0$ et $J' = [\tau - k + m_0, \tau[$ sinon)(choix (c)). Pour ces différents choix d'implémentation, nous avons :

$$(a)\ \ P_\rho \leq \tfrac{k_{\mathbb{I}}}{m_0^\lambda} + \sum_{k=1}^{k_{\mathbb{I}}} \frac{1}{(m_0+k)^\lambda}, \quad (b)\ \ P_\rho \leq \tfrac{k_{\mathbb{I}}+1}{m_0^\lambda}, \quad (c)\ \ P_\rho \leq \tfrac{k_{\mathbb{I}}+1}{m_0^\lambda} + \sum_{k=1}^{k_{\mathbb{I}}-m_0} \frac{1}{(m_0+k)^\lambda}.$$

Ce théorème se prouve de manière analogue au théorème 2.3.3. Un résultat similaire peut être donné sur la précision de $\hat{Q}_{\hat{I}_Q}$.
Nous appliquons maintenant cet algorithme adaptatif pour détecter les ruptures dans un modèle à sauts.

2.3.4 Modèle à sauts

Un cas particulier intéressant est celui où ρ_t et Q_t sont des fonctions constantes par morceaux. Pour une date fixée τ, on cherche ainsi à déterminer la dernière rupture c'est à dire le plus grand intervalle d'extrémité droite τ sans rupture dans la moyenne ni dans les variances et covariances. Cela revient à résoudre un problème de tests multiples. Afin d'obtenir une faible erreur de première espèce (faite lorsqu'un intervalle d'homogénéité est rejeté), chaque test devrait être conduit avec une faible erreur de première espèce. Le théorème suivant fournit une borne supérieure pour l'erreur de première espèce si on utilise notre algorithme adaptatif pour détecter la dernière rupture d'un modèle à sauts.

Théorème 2.3.5 *Soit I un intervalle d'homogénéité temporelle locale tel que pour tout $(t,t') \in I^2$, $\rho_t = \rho'_t$ et $Q_t = Q'_t$. Si $\lambda > 0$ est le paramètre intervenant dans la relation* (2.10) *alors*

$$P(I \ \ \textit{est rejeté}\ ) \leq \frac{6\,Card(\mathcal{I}(I))}{m_0^\lambda}. \tag{2.32}$$

Si $\ln n(n+1) + \lambda \ln m_0 \leq m_0$, la plus petite erreur de première espèce est obtenue en remplaçant λ par $\frac{m_0 - \ln n(n+1)}{\ln m_0}$; la plus grande de ses valeurs admissibles. La précision de l'algorithme adaptatif peut être obtenue sans cette restriction (en utilisant d'autres règles d'acceptation que (2.26), comme on peut le voir dans les preuves fournies à la fin de ce chapitre). Par ailleurs, même si la borne supérieure fournie par le théorème 2.3.5 est très grossière, elle donne, au moins d'un point de vue théorique, une manière de choisir λ, à savoir

$$\lambda \geq \frac{\ln(\frac{6Card(\mathcal{I}(I))}{\alpha})}{\ln m_0},$$

telle que l'erreur de première espèce soit au plus α. Nous aimerions aussi que notre algorithme détecte des ruptures dans la moyenne, les variances et les covariances, si ces changements sont suffisamment importants. Le théorème suivant précise que si sur un intervalle donné I, il y a une importante rupture dans la moyenne, les variances ou les covariances au temps $T_{cp} \in I$, alors il y a peu de chances que notre algorithme adaptatif accepte cet intervalle en tant qu'intervalle d'homogénéité temporelle locale (i.e que l'algorithme fasse une erreur de deuxième espèce).

Théorème 2.3.6 *Soit $m \in \mathbb{N}^*$ multiple de m_0 et $I = [T_{cp} - m, \tau[$ un intervalle où une rupture intervient au temps T_{cp} à la fois dans les moyennes, les variances et les covariances. Sur $J = [T_{cp} - m, T_{cp}[$, $\rho_t = \rho_2$, $Q_t = Q_2$ et sur $[T_{cp}, \tau[$, $\rho_t = \rho_1$, $Q_t = Q_1$. Soit λ le paramètre de lissage de l'algorithme adaptatif. On suppose que $\ln n(n+1) + \lambda \ln m_0 \leq m_0$. Si les changements dans les moyennes, les variances et les covariances sont suffisamment importants i.e si*

$$\|\rho_2 - \rho_1\|_\infty \geq \frac{m+m'}{m'} \left(2\eta_1(m+m', \lambda) + 2\eta_1(m, \lambda) + 2\sqrt{2e} D\sigma \left(\sqrt{\frac{\ln n}{m+m'}} + \sqrt{\frac{\ln n}{m}} \right) \right) \tag{2.33}$$

et

$$\|Q_2 - Q_1\|_\infty \geq \frac{m+m'}{m'} (2k'_Q + k_Q D)\sigma^2 \left(\sqrt{\frac{\ln n(n+1) + \lambda \ln m}{m}} + \sqrt{\frac{\ln n(n+1) + \lambda \ln(m+m')}{m+m'}} \right), \tag{2.34}$$

où $m' = \tau - T_{cp}$ et $\eta_1(|I|, \lambda) = (\frac{7}{3} + \sqrt{2})\, \sigma \sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}}$ et k_Q et k'_Q sont définis dans le théorème 2.3.2, alors

$$P(I \quad accepté\ ) \leq 3 \left(\frac{1}{m^\lambda} + \frac{1}{(m+m')^\lambda} \right).$$

On a supposé m multiple de m_0 afin de simplifier la présentation du résultat mais l'erreur de deuxième espèce peut être contrôlée sans cette condition. La borne supérieure dans le théorème ci-dessus peut par ailleurs être améliorée (comme on peut le voir en suivant la preuve de ce théorème) si les moyennes sont constantes et seules les variances et covariances varient ou, de manière symétrique, si les variances et covariances sont constantes et que les moyennes changent au temps T_{cp}. Dans le théorème ci-dessus, la condition $\ln n(n+1) + \lambda \ln m_0 \leq m_0$ a été introduite afin d'obtenir des contraintes plus simples sur l'amplitude minimale des changements permettant une bonne détection de rupture. On peut enfin, de façon analogue, contrôler le pouvoir de détection lorsque la règle d'acceptation d'un intervalle $I = [\tau - k, \tau[$, consiste à mesurer la différence des estimations sur $J = [\tau - k, \tau - k + m_0 - 1]$ et ($J' = [\tau - m_0, \tau[$ si $k \leq 2m_0$ et $J' = [\tau - k + m_0, \tau[$ sinon) (règle d'acceptation (c)).

Théorème 2.3.7 *Soit (r_t) un processus satisfaisant l'hypothèse 1 et soit $I = [T_{cp} - m, \tau - 1]$ un intervalle sur lequel une rupture dans les moyennes se produit au temps T_{cp} : pour $t = T_{cp} - m, \ldots, T_{cp} - 1$ ($m \geq m_0$), $\rho_t = \rho_2$ et pour $t = T_{cp}, \ldots, \tau - 1$, $\rho_t = \rho_1$ ($m' = \tau - T_{cp} \geq m_0$). Soit $\lambda > 0$ tel que $\ln n(n+1) + \lambda \ln m_0 \leq m_0$. Si les changements dans les moyennes sont suffisamment importants, i.e si*

$$\|\rho_2 - \rho_1\|_\infty \geq \frac{m+m'-m_0}{m'} \left(2\eta_1(m_0, \lambda) + 2\eta_1(m+m'-m_0, \lambda) + 2\sqrt{2e} D\sigma \left(\sqrt{\frac{\ln n}{m_0}} + \sqrt{\frac{\ln n}{m+m'-m_0}} \right) \right),$$

où $\eta_1(|I|,\lambda) = (\frac{7}{3}+\sqrt{2})\,\sigma\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$, *alors la probabilité que l'intervalle* I *soit considéré comme un intervalle d'homogénéité en utilisant la règle d'acceptation (c) rappelée ci-dessus est majorée de la façon suivante :*

$$P(I \quad accepté\ ) \leq \frac{1}{m_0^\lambda} + \frac{1}{(m+m'-m_0)^\lambda}.$$

2.3.5 L'algorithme adaptatif utilisant les estimations empiriques de la moyenne et de la matrice de covariance

Au lieu d'utiliser les estimations de la moyenne et de la matrice de covariance données par (2.9), nous aurions pu utiliser les estimations empiriques plus naturelles. Une des raisons de notre choix est qu'il permet de fournir un majorant simple (en tant que fonction du paramètre λ) sur les erreurs de première et deuxième espèce (théorèmes 2.3.5, 2.3.6 et 2.3.7), fournissant ainsi une estimation grossière de λ permettant de contrôler les erreurs de première et deuxième espèce. D'autre part, dans ce cadre, nous contrôlons plus finement et sous des hypothèses moins restrictives sur le processus r_t, la précision des problèmes $(\widehat{\mathcal{V}})$ et $(\widehat{\mathcal{M}})$. Cependant, si les estimations empiriques sont utilisées à la place des estimations données par (2.9), on peut encore donner la précision des estimations empiriques adaptatives $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ qui en résultent, majorer les erreurs de première et deuxième espèce, et contrôler la précision des problèmes Value-at-Risk et de Markowitz approchés qui utilisent ces estimations $\hat{\rho}_\tau$ et $\hat{Q}_\tau$. Pour obtenir ces résultats, nous faisons l'hypothèse suivante sur le processus des rendements.

Hypothèse 2. Les rendements r_i sont générés par le modèle

$$r_i = \rho_i + \zeta_i, \quad i = 1, ..., \tau - 1, \tag{2.35}$$

où (ζ_i) sont des vecteurs aléatoires indépendants de moyenne nulle, $E\,\zeta_i = 0$ et tels que $E\|\zeta_i\|_\infty^p \leq \sigma^p$ pour un certain $p \geq 4$ et pour $\sigma > 0$. Enfin on suppose que pour tout i, $\|\rho_i\|_\infty \leq \sigma'$, que $n > 1$, $\tau - 1 > 1$ et on pose $\gamma = \frac{\ln n}{\ln(\tau-1)}$.

Dans ce qui suit, les estimations de la moyenne et de la matrice de covariance obtenues en utilisant l'algorithme adaptatif et, à chaque itération de l'algorithme, les estimations empiriques de la moyenne et de la matrice de covariance au lieu de (2.9), seront appelées les estimations empiriques adaptatives. Rappelons que l'estimation empirique de la moyenne utilisant les données de l'intervalle non vide I est donnée par $\hat{\rho}_I = \dfrac{1}{|I|}\displaystyle\sum_{k\in I} r_k$ et l'estimation empirique de la matrice de covariance par $\hat{Q}_I = \dfrac{1}{|I|}\displaystyle\sum_{k\in I}(r_k - \hat{\rho}_I)(r_k - \hat{\rho}_I)^T$. L'utilisation des estimations empiriques introduit cependant un degré de liberté additionnel. En effet, l'algorithme adaptatif qui en résulte a alors deux paramètres λ et μ qui contrôlent respectivement la calibration de la moyenne (pour λ) et la calibration de la matrice de covariance (pour μ). Plus précisément, supposons que les paramètres de lissage $\lambda > 0$, $\mu > 0$ de l'algorithme adaptatif vérifient $10(\ln n + \lambda \ln m_0) \leq 3m_0$ et $10(\ln n(n+1) + \mu \ln m_0) \leq 3m_0$ et supposons que les estimations empiriques de la moyenne et de la matrice de covariance sont utilisées à la place de (2.9). Dans cette version de l'algorithme adaptatif, un intervalle I

sera alors accepté comme intervalle d'homogénéité pour la moyenne si pour tout intervalle $J \in \mathcal{I}(I)$

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq f_1(|I|,\lambda) + f_1(|J|,\lambda) + \sqrt{2e}D\sigma\left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}}\right), \tag{2.36}$$

où une forme simplifiée de f_1 (si $10(\ln n + \lambda \ln m_0) \leq 3m_0$) est donnée par

$$f_1(|I|,\lambda) = 2\sqrt{\frac{10}{3}}\sigma\sqrt{\frac{\ln n + \lambda \ln |I|}{|I|}}. \tag{2.37}$$

Cet intervalle I sera accepté comme intervalle d'homogénéité pour les variances et covariances si pour tout intervalle $J \in \mathcal{I}(I)$

$$\|\hat{Q}_I - \hat{Q}_J\|_\infty \leq f_2(|I|,\mu) + f_2(|J|,\mu) + (4e\sigma^2 + 2\sqrt{e}\sigma''')D\left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}}\right), \tag{2.38}$$

avec $\sigma''' = 2(2\sigma' + \sigma)^2$ et où une forme simplifiée pour f_2 (si $10(\ln n(n+1) + \mu \ln m_0) \leq 3m_0$) est donnée par

$$f_2(|I|,\mu) = 2\sqrt{\frac{10}{3}}\left(2\sqrt{\frac{10}{3}}\sigma^2 + \sigma''\right)\sqrt{\frac{\ln n(n+1) + \mu \ln |I|}{|I|}}; \tag{2.39}$$

avec $\sigma'' = \sigma^2 + (2\sigma' + \sigma)^2$. En suivant la preuve du théorème 2.3.3, on peut borner la précision des estimations empiriques adaptatives. On définit tout d'abord :

$$K_1(\lambda) = 2 + \left(\frac{10\,p(\gamma+\lambda)}{3e(p-2-2\lambda)}\right)^{\frac{p}{2}} \text{ et } K_2(\lambda) = 1 + \left(\frac{10\,E[\frac{p}{2}](2\gamma+\lambda)}{3\,e(E[\frac{p}{2}]-2-2\lambda)}\right)^{\frac{E[\frac{p}{2}]}{2}}. \tag{2.40}$$

Théorème 2.3.8 *Soient $\hat{\rho}_{\hat{I}_\rho}$ et $\hat{Q}_{\hat{I}_Q}$ les estimations empiriques adaptatives de la moyenne et de la matrice de covariance. Soient (λ,μ) les paramètres de l'algorithme adaptatif tels que $E[\frac{p}{2}] > 2(1+\lambda)$, $10(\ln n + \lambda \ln m_0) \leq 3m_0$ et $10(\ln n(n+1) + \mu \ln m_0) \leq 3m_0$. Alors*

$$P\left(\|\hat{\rho}_{\hat{I}_\rho} - \rho_\tau\|_\infty > (6\sqrt{\tfrac{10}{3}} + 3\sqrt{2e}D)\sigma\sqrt{\tfrac{\ln n + \lambda \ln |\mathbb{I}_\rho|}{|\mathbb{I}_\rho|}}\right) \leq \sum_{J \in \mathcal{I}_+(\mathbb{I}_\rho)} \frac{K_1(\lambda)}{|J|^\lambda},$$

$$P\left(\|\hat{Q}_{\hat{I}_Q} - Q_\tau\|_\infty > f(\sigma,\sigma')\sqrt{\tfrac{\ln n(n+1) + \mu \ln |\mathbb{I}_Q|}{|\mathbb{I}_Q|}}\right) \leq \sum_{J \in \mathcal{I}_+(\mathbb{I}_Q)} \frac{K_1(\mu) + K_2(\mu)}{|J|^\mu},$$

avec $f(\sigma,\sigma') = 6\sqrt{\frac{10}{3}}(2\sqrt{\frac{10}{3}}\sigma^2 + \sigma'') + 3D(4e\sigma^2 + 2\sqrt{e}\sigma''')$, $\sigma'' = \sigma^2 + (2\sigma' + \sigma)^2$ et $\sigma''' = 2(2\sigma' + \sigma)^2$.

Ce théorème 2.3.8 et les lemmes 2.7.1 et 2.7.2 (donnés en fin de ce chapitre) peuvent ensuite être utilisés afin de majorer la précision des problèmes VaR et de Markowitz approchés qui utilisent les estimations empiriques adaptatives de la moyenne et de la matrice de covariance.

2.3.5.1 Précision des problèmes VaR et de Markowitz approchés

En utilisant les lemmes 2.7.1 et 2.7.2 (donnés en fin de ce chapitre), on peut montrer que dans le cas stationnaire, la précision des problèmes VaR et de Markowitz approchés qui utilisent les estimations empiriques de ρ_τ et Q_τ est encore donnée par les théorèmes 2.2.1 et 2.2.2, les constantes k_1 et k_2 étant à présent données par $k_1 = 4\sqrt{\frac{10}{3}}$ et $k_2 = \frac{40}{3} + (2^{E[\frac{p}{2}]}+1)\sqrt{\frac{28}{3}}$. Le paramètre $\lambda > 0$ vérifie à présent $28(\ln n(n+1) + \lambda \ln N) \leq 3N$, l'ensemble S étant maintenant de probabilité au moins $1\text{-}\frac{2K_1(\lambda)+K_2'(\lambda)}{N^\lambda}$ avec $K_2'(\lambda) = 1 + 2^{E[\frac{p}{2}]}\Big(\frac{28\,E[\frac{p}{2}](2\gamma+\lambda)}{3\,e(E[\frac{p}{2}]-2-2\lambda)}\Big)^{\frac{E[\frac{p}{2}]}{2}}$.

Théorème 2.3.9 *Soient $(\hat{\rho}_\tau, \hat{Q}_\tau)$ les estimations empiriques adaptatives de (ρ_τ, Q_τ) obtenues en utilisant les données de l'intersection des intervalles d'homogénéité. Soient $\lambda > 0$, $\mu > 0$ les paramètres de l'algorithme adaptatif tels que $10(\ln n + \lambda \ln m_0) \leq 3m_0$ et $10(\ln n(n+1) + \mu \ln m_0) \leq 3m_0$. Si $\tilde{\gamma} = \max(\lambda, \mu)$, alors il existe des constantes k_1', $k_2'(\sigma, \sigma') = f(\sigma, \sigma')$ (défini dans le théorème 2.3.8) et un ensemble $S \subseteq \Omega$ de probabilité au moins $1 - \sum_{J \in \mathcal{I}_+(\mathbb{I})} \frac{K_1(\lambda)}{|J|^\lambda} + \frac{K_1(\mu)+K_2(\mu)}{|J|^\mu}$ tel que pour tout $\omega \in S$,*

$$\epsilon(\widehat{\mathcal{V}}) \leq (k_1'\sigma + 2\kappa(\varepsilon)\sqrt{k_2'(\sigma,\sigma')})\|x^-\|_1 \left(\frac{\ln n(n+1) + \tilde{\gamma}\ln|\mathbb{I}|}{|\mathbb{I}|}\right)^{\frac{1}{4}}. \tag{2.41}$$

De plus, si la matrice Q_τ est non dégénérée, i.e $\min_{\|x\|_1=1} x^T Q_\tau x = \beta(Q_\tau) > 0$, alors

$$\epsilon(\widehat{\mathcal{V}}) \leq \left(k_1'\sigma + \frac{2\kappa(\varepsilon)k_2'(\sigma,\sigma')}{\sqrt{\beta(Q_\tau)}}\right)\|x^-\|_1\sqrt{\frac{\ln n(n+1) + \tilde{\gamma}\ln|\mathbb{I}|}{|\mathbb{I}|}}. \tag{2.42}$$

Théorème 2.3.10 *Soient $(\hat{\rho}_\tau, \hat{Q}_\tau)$ les estimations empiriques adaptatives de (ρ_τ, Q_τ) obtenues en utilisant les données de l'intersection des intervalles d'homogénéité. Soient $\lambda > 0$, $\mu > 0$ les paramètres de l'algorithme adaptatif tels que $10(\ln n + \lambda \ln m_0) \leq 3m_0$ et $10(\ln n(n+1) + \mu \ln m_0) \leq 3m_0$. Si $\tilde{\gamma} = \max(\lambda, \mu)$, alors il existe des constantes k_1', $k_2' = f(\sigma, \sigma')$ (défini dans le théorème 2.3.8) et un ensemble $S \subseteq \Omega$ de probabilité au moins $1 - \sum_{J \in \mathcal{I}_+(\mathbb{I})} \frac{K_1(\lambda)}{|J|^\lambda} + \frac{K_1(\mu)+K_2(\mu)}{|J|^\mu}$ tel que pour tout $\omega \in S$, le problème $(\widehat{\mathcal{M}})$ est faisable et*

$$\epsilon(\widehat{\mathcal{M}}) \leq \max\left(k_1'\sigma, k_2'(\sigma,\sigma')\|x^-\|_1\right)\|x^-\|_1\sqrt{\frac{\ln n(n+1) + \tilde{\gamma}\ln|\mathbb{I}|}{|\mathbb{I}|}}; \tag{2.43}$$

et si Q_τ est définie positive alors

$$\epsilon(\widehat{\mathcal{M}}) \leq \max\left(k_1'\,\sigma, k_2'(\sigma,\sigma')\|x^-\|_1\tilde{\varepsilon}\right)\|x^-\|_1\sqrt{\frac{\ln n(n+1) + \tilde{\gamma}\ln|\mathbb{I}|}{|\mathbb{I}|}}; \tag{2.44}$$

où

$$\tilde{\varepsilon} = \min(1, \frac{2k_2'(\sigma,\sigma')}{\beta(Q_\tau)}\sqrt{\frac{\ln n(n+1) + \tilde{\gamma}\ln|\mathbb{I}|}{|\mathbb{I}|}}).$$

La constante k_1' dans les théorèmes 2.3.9 et 2.3.10 ci-dessus est donnée par $k_1' = 2(6\sqrt{\frac{10}{3}} + 3\sqrt{2e}D)$. Remarquons que le problème de Markowitz approché défini dans ce contexte (lorsque l'on cherche à déterminer les estimations empiriques adaptatives) est aussi différent. On constate néanmoins que les majorants sur la précision des problèmes approchés sont des fonctions de n et $|\mathbb{I}|$ du même type que celles rencontrées en section 2.2.2.

2.3.5.2 Erreurs de première et deuxième espèce

Les erreurs de première et deuxième espèce sont contrôlées de la façon suivante.

Théorème 2.3.11 *Soit I un intervalle d'homogénéité temporelle tel que pour tout $(t,t') \in I^2$, $\rho_t = \rho_t'$ et $Q_t = Q_t'$. Alors si (λ, μ) sont les paramètres de lissage de l'algorithme adaptatif,*

$$P(I \text{ est rejeté }) \leq 2\,Card(\mathcal{I}(I)) \left(\frac{K_1(\lambda)}{m_0^\lambda} + \frac{K_1(\mu) + K_2(\mu)}{m_0^\mu} \right). \tag{2.45}$$

Théorème 2.3.12 *Soit (r_t) un processus satisfaisant l'hypothèse 2 et soit $I = [T_{cp} - m, \tau - 1]$ un intervalle où une rupture dans les moyennes intervient au temps T_{cp} : pour $t = T_{cp} - m, \ldots, T_{cp} - 1$ ($m \geq m_0$ est multiple de m_0), $\rho_t = \rho_2$ et pour $t = T_{cp}, \ldots, \tau - 1$, $\rho_t = \rho_1$ ($m' = \tau - T_{cp} \geq m_0$). Soit $\lambda > 0$ tel que $10(\ln n + \lambda \ln m_0) \leq 3m_0$. Si les changements dans les moyennes sont suffisamment importants, c'est à dire si*

$$\|\rho_2 - \rho_1\|_\infty \geq \frac{m+m'}{m'} \left(2f_1(m+m', \lambda) + 2f_1(m, \lambda) + \sqrt{2e} D\sigma(\sqrt{\frac{\ln n}{m+m'}} + \sqrt{\frac{\ln n}{m}}) \right),$$

avec f_1 défini en (2.37), alors la probabilité que l'intervalle I soit accepté comme un intervalle d'homogénéité est majorée comme suit :

$$P(I \text{ accepté }) \leq K_1(\lambda) \left(\frac{1}{m^\lambda} + \frac{1}{(m+m')^\lambda} \right).$$

Un résultat similaire peut être énoncé si l'on s'intéresse à la détection de ruptures dans les variances et les covariances.

2.4 Détection de rupture en utilisant une fenêtre glissante

L'algorithme adaptatif décrit dans la section précédente peut être utilisé pour détecter la dernière rupture dans un processus à sauts si les variations de moyennes, de variances ou de covariances sont suffisamment importantes. En utilisant le même cadre statistique, plus précisément pour un processus satisfaisant l'hypothèse 1 ou l'hypothèse 2 (introduites dans les sections 2.1.1 et 2.3.5), nous pourrions utiliser une autre méthode non paramétrique afin de détecter la rupture d'un processus à sauts ayant une rupture (ou d'une manière plus générale la rupture la plus importante s'il y en a plusieurs). Le but de cette section est de décrire cette méthode et de donner des résultats théoriques sur cette méthode. Nous terminons cette section en présentant des résultats numériques illustrant les performances de cette méthode sur des données simulées.
Soient $(r_t)_{t=1,\ldots,N}$, N observations indépendantes d'un processus n dimensionnel vérifiant

l'hypothèse 1 ou l'hypothèse 2. Nous allons nous intéresser à deux processus à sauts particuliers. Dans le premier de ces modèles, on suppose que la matrice de covariance $Q_t = Q$ est constante et que la moyenne change brutalement entre les instants $t = T_{cp}$ et $t = T_{cp} + 1$. Plus précisément, pour $t = 1, \ldots, T_{cp}$, $\rho_t = \rho_1$ et pour $t = T_{cp}+1, \ldots, N$, $\rho_t = \rho_2$. Le second modèle étudie la cas symétrique où la moyenne ρ_t est constante et la matrice de covariance Q_t change brutalement entre les instants $T_{cp} \in [1, \ldots, N]$ et $T_{cp} + 1$ (pour $t = 1, \ldots, T_{cp}$, $Q_t = Q_1$ et pour $t = T_{cp} + 1, \ldots, N$, $Q_t = Q_2$). Le cas général (rupture à la fois dans la moyenne et les covariances/variances) peut être étudié de la même façon.

2.4.1 Description de l'algorithme

Pour détecter la rupture, nous procédons de la façon suivante. Soit $h \in \mathbb{N}^*$ la largeur de la fenêtre. Pour $h \leq t \leq N - h + 1$, s'il y a une rupture dans la matrice de covariance au temps t, la moyenne restant constante, alors l'estimation de la matrice de covariance sur une fenêtre de largeur h dont l'extrémité droite est t (utilisant ainsi les observations $(r_k)_{t-h+1 \leq k \leq t}$) et l'estimation de la matrice de covariance sur une fenêtre de largeur h dont l'extrémité gauche est t (utilisant ainsi les observations $(r_k)_{t \leq k \leq t+h-1}$) devraient être assez différentes (si la rupture est suffisamment importante). D'un autre coté, si pour $k = t - h + 1, \ldots, t + h - 1$, $Q_k = Q$ est constante alors les estimations gauches et droites au temps t devraient être proches. Ainsi, pour tout $h \leq t \leq N - h + 1$, nous calculons l'estimation gauche $Q_g^t(h)$ et l'estimation droite $Q_d^t(h)$ de la matrice de covariance au temps t par :

$$Q_g^t(h) = \frac{1}{h} \sum_{k=t-h+1}^{t} (r_k - \rho_g^t(h))(r_k - \rho_g^t(h))^T, \quad Q_d^t(h) = \frac{1}{h} \sum_{k=t}^{t+h-1} (r_k - \rho_d^t(h))(r_k - \rho_d^t(h))^T, \tag{2.46}$$

avec $\rho_g^t(h)$ et $\rho_d^t(h)$ les estimations gauches et droites de la moyenne :

$$\rho_g^t(h) = \frac{1}{h} \sum_{k=t-h+1}^{t} r_k, \quad \rho_d^t(h) = \frac{1}{h} \sum_{k=t}^{t+h-1} r_k. \tag{2.47}$$

L'estimation $\hat{T}_{cp}(h)$ de l'instant de la rupture lorsqu'on utilise une fenêtre de largeur h est alors l'instant où les estimations de la matrice de covariance sur les fenêtres gauche et droite de largeur h sont les plus éloignées :

$$\hat{T}_{cp}(h) = \underset{h \leq t \leq N-h+1}{\text{Argmax}} \|Q_g^t(h) - Q_d^t(h)\|_\infty. \tag{2.48}$$

Dans les cas où les variances et covariances sont constantes, afin de détecter la plus importante rupture dans la moyenne, on utilise l'estimateur suivant :

$$\hat{T}_{cp}(h) = \underset{h \leq t \leq N-h+1}{\text{Argmax}} \|\rho_g^t(h) - \rho_d^t(h)\|_\infty. \tag{2.49}$$

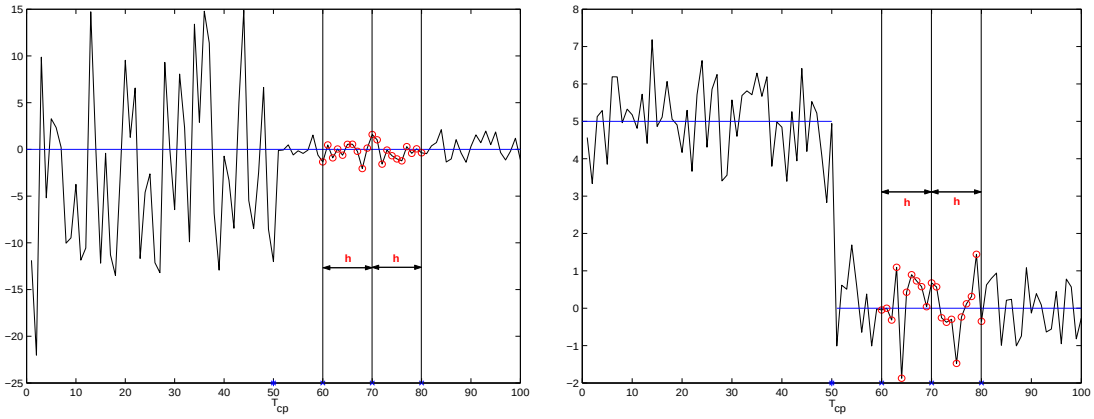


FIG. 2.3 – Illustration de la méthode. Pour la figure de gauche, les 50 premières observations sont issues d'une loi normale $\mathcal{N}(0, 100)$ et les 50 dernières observations d'une loi normale $\mathcal{N}(0, 1)$. On utilise une fenêtre glissante de longueur $h = 11$. Pour $t = 70$, la fenêtre de gauche utilise les observations $(r_t)_{t=60,\ldots,70}$ et la fenêtre de droite les observations $(r_t)_{t=70,\ldots,80}$. La figure de droite illustre la méthode lorsque seule une rupture dans la moyenne se produit.

2.4.2 Résultats théoriques

Dans cette section, $K_1(\lambda)$ et $K_2(\lambda)$ désignent les quantités définies par (2.40) dans la section précédente.

2.4.2.1 Cas où la matrice de covariance est constante

Si le changement dans les moyennes est suffisamment important, $\hat{T}_{cp}(h)$ a de grandes chances d'appartenir à $[T_{cp} - h + 1, T_{cp} + h - 1]$. Plus précisément, nous pouvons énoncer le résultat suivant.

Théorème 2.4.1 *Supposons l'hypothèse 2 vérifiée. Soient $\lambda > 0$ et $h \in \mathbb{N}^* \backslash \{1\}$ tels que $10(\ln n + \lambda \ln h) \leq 3h$. Si $\|\rho_2 - \rho_1\|_\infty > \frac{4h}{h-1} f_1(h, \lambda)$ où f_1 est défini par (2.37) alors*

$$P(\hat{T}_{cp}(h) \in [T_{cp} - h + 1, T_{cp} + h - 1]) \geq 1 - \frac{2(N - 4h + 5)K_1(\lambda)}{h^\lambda}. \tag{2.50}$$

Remarquons que f_1 est une fonction croissante de n (le nombre de composantes), λ et σ et une fonction décroissante de h comme on s'y attendait. Également, lorsque h tend vers l'infini, alors $f_1(h, \lambda)$ tend vers 0. Si h n'est pas trop grand et si λ suffisamment grand, alors ce théorème montre que l'estimateur $\hat{T}_{cp}(h)$ est de bonne qualité. Mais lorsque λ croît, f_1 aussi. Ceci signifie que l'estimateur sera bien sûr d'autant plus précis que la rupture dans la moyenne est importante. Par exemple, si le nombre de composantes $n = 30$, et si on choisit $\lambda = 0.5$, alors la condition ci-dessus implique que $h \geq 16$. Cette condition n'est pas trop restrictive puisque les simulations numériques faites dans cette section, tendent à montrer que les "bonnes" valeurs de h sont plutôt élevées. Par la suite, pour $h \leq t \leq N - h + 1$, on notera $I_g^t = [t - h + 1, t]$ l'intervalle de longueur h d'extrémité droite t et $I_d^t = [t, t + h - 1]$ l'intervalle de longueur h d'extrémité gauche t.

Remarque 2.4.1 *Sous l'hypothèse 1, supposons que l'on utilise les estimations gauche (resp. droite) de la matrice de covariance et de la moyenne obtenues en remplaçant r_k par $\alpha_k^{I_g^t}$ (resp. $\alpha_k^{I_d^t}$) dans les expressions de $Q_g^t(h)$ et $\rho_g^t(h)$ (resp. $Q_d^t(h)$ et $\rho_d^t(h)$) données dans* (2.46) *et* (2.47). *Soient $\lambda > 0$ et $h \in \mathbb{N}^* \backslash \{1\}$. Si $\|\rho_2 - \rho_1\|_\infty > \frac{4h}{h-1}\,\eta_1(h,\lambda)$ avec*

$$\eta_1(h,\lambda) = \sigma\left(\frac{7}{3}\left(\frac{\ln n(n+1) + \lambda \ln h}{h}\right)^{\frac{3}{4}} + \sqrt{2}\sqrt{\frac{\ln n(n+1) + \lambda \ln h}{h}}\right), \tag{2.51}$$

alors

$$P(\hat{T}_{cp}(h) \in [T_{cp} - h + 1, T_{cp} + h - 1]) \geq 1 - \frac{2(N - 4h + 5)}{h^\lambda}; \tag{2.52}$$

une expression simplifiée de $\eta_1(h,\lambda)$ (lorsque $\ln n(n+1) + \lambda \ln h \leq h$), étant $\eta_1(h,\lambda) = (\frac{7}{3} + \sqrt{2})\sigma\sqrt{\frac{\ln n(n+1)+\lambda \ln h}{h}}$.

En utilisant une condition plus restrictive sur l'amplitude des changements dans la moyenne, on peut aussi borner inférieurement la probabilité que cet algorithme fournisse effectivement la vraie valeur T_{cp} ou $T_{cp}+1$ (rappelons que la rupture intervient entre les observations aux temps T_{cp} et $T_{cp}+1$).

Théorème 2.4.2 *Supposons l'hypothèse 2 vérifiée. Soient $\lambda > 0$ et $h \in \mathbb{N}^*$ tels que $10(\ln n + \lambda \ln h) \leq 3h$. Si $\|\rho_2 - \rho_1\|_\infty > 4h\, f_1(h,\lambda)$ avec f_1 défini par* (2.37) *alors*

$$P(\{\hat{T}_{cp}(h) = T_{cp}\} \cup \{\hat{T}_{cp}(h) = T_{cp} + 1\}) \geq 1 - \frac{2(N - 2h + 2)K_1(\lambda)}{h^\lambda}. \tag{2.53}$$

Remarque 2.4.2 *Sous l'hypothèse 1, supposons que l'on utilise les estimations gauche (resp. droite) de la matrice de covariance et de la moyenne obtenues en remplaçant r_k par $\alpha_k^{I_g^t}$ (resp. $\alpha_k^{I_d^t}$) dans les expressions de $Q_g^t(h)$ et $\rho_g^t(h)$ (resp. $Q_d^t(h)$ et $\rho_d^t(h)$) données dans* (2.46) *et* (2.47). *Soient $\lambda > 0$ et $h \in \mathbb{N}^* \backslash \{1\}$. Si $\|\rho_2 - \rho_1\|_\infty > 4h\,\eta_1(h,\lambda)$, avec η_1 donné en* (2.51) *alors :*

$$P(\{\hat{T}_{cp}(h) = T_{cp}\} \cup \{\hat{T}_{cp}(h) = T_{cp} + 1\}) \geq 1 - \frac{2(N - 2h + 2)}{h^\lambda}. \tag{2.54}$$

2.4.2.2 Cas où la moyenne est constante

Théorème 2.4.3 *Supposons l'hypothèse 2 vérifiée. Soient $\lambda > 0$ et $h \in \mathbb{N}^* \backslash \{1\}$ tels que $10(\ln n(n+1) + \lambda \ln h) \leq h$. Si $\|Q_2 - Q_1\|_\infty > \frac{4h}{h-1}\, f_2(h,\lambda)$ où f_2 est défini par* (2.39) *alors*

$$P(\hat{T}_{cp}(h) \in [T_{cp} - h + 1, T_{cp} + h - 1]) \geq 1 - \frac{2(N - 4h + 5)(K_1(\lambda) + K_2(\lambda))}{h^\lambda}. \tag{2.55}$$

Théorème 2.4.4 *Sous les hypothèses du théorème 2.4.3, si $\|Q_2 - Q_1\|_\infty > 4\,h\, f_2(h,\lambda)$ où f_2 est défini par* (2.39) *alors*

$$P(\{\hat{T}_{cp}(h) = T_{cp}\} \cup \{\hat{T}_{cp}(h) = T_{cp} + 1\}) \geq 1 - \frac{2(N - 2h + 2)(K_1(\lambda) + K_2(\lambda))}{h^\lambda}. \tag{2.56}$$

Remarque 2.4.3 *Sous l'hypothèse 1, supposons que l'on utilise les estimations gauche (resp. droite) de la matrice de covariance et de la moyenne obtenues en remplaçant r_k par $\alpha_k^{I_g^t}$ (resp. $\alpha_k^{I_d^t}$) dans les expressions de $Q_g^t(h)$ et $\rho_g^t(h)$ (resp. $Q_d^t(h)$ et $\rho_d^t(h)$) données dans (2.46) et (2.47). Soient $\lambda > 0$ et $h \in \mathbb{N}^* \backslash \{1\}$ tels que $\ln n(n+1) + \lambda \ln h \leq h$. Alors si*

$$\|Q_2 - Q_1\|_\infty > \frac{4k'_Q}{h-1}\,\sigma^2\,\sqrt{h\,(\ln n(n+1) + \lambda \ln h)}$$

(où k'_Q est défini dans le théorème 2.3.2) la probabilité que l'estimation $\hat{T}_{cp}(h)$ soit dans $[T_{cp} - h + 1, T_{cp} + h - 1]$ est supérieure à $1 - \frac{2(N-4h+5)}{h^\lambda}$ et si

$$\|Q_2 - Q_1\|_\infty > 4k'_Q\,\sigma^2\,\sqrt{h\,(\ln n(n+1) + \lambda \ln h)},$$

alors la probabilité que l'estimation $\hat{T}_{cp}(h)$ soit égale à T_{cp} ou $T_{cp}+1$ est supérieure à $1 - \frac{2(N-2h+2)}{h^\lambda}$.

2.4.3 Illustration numérique

Cas unidimensionnel. Pour observer le comportement de l'algorithme ainsi que l'influence du choix de la largeur de fenêtre h, nous commençons par étudier un processus Gaussien S_t où pour $1 \leq t \leq T_{cp} = 250$, $S_t \sim \mathcal{N}(m_1, 1)$, $251 \leq t \leq T_{cp} = 750$, $S_t \sim \mathcal{N}(m_2, 1)$. Pour ce processus, la moyenne change brutalement après l'instant $T_{cp} = 250$ alors que la variance reste constante. La méthode du maximum de vraisemblance est traditionnellement utilisée pour déterminer, dans un contexte paramétrique, un certain nombre (fixé) de ruptures. On se propose ici d'utiliser l'approche non paramétrique présentée dans la section précédente. On s'intéresse à trois cas : $m_1 = 0, m_2 = 1$(cas 1), $m_1 = 0, m_2 = 5$ (cas 2) et $m_1 = 0, m_2 = 10$ (cas 3). Dans chacun de ces cas, on génère 400 réalisations du processus et pour des valeurs de h allant de 2 à $T_{cp} = 250$, on calcule la moyenne et l'écart type empiriques de l'estimateur $\hat{T}_{cp}(h)$ de l'instant de la rupture, ainsi que la probabilité de détecter correctement la rupture. Puisque cette rupture a lieu entre les instants T_{cp} et $T_{cp}+1$, cette probabilité est la probabilité que $\hat{T}_{cp}(h) = T_{cp}$, ou $T_{cp}+1$. Les résultats sont représentés sur la figure 2.4 ci-après.

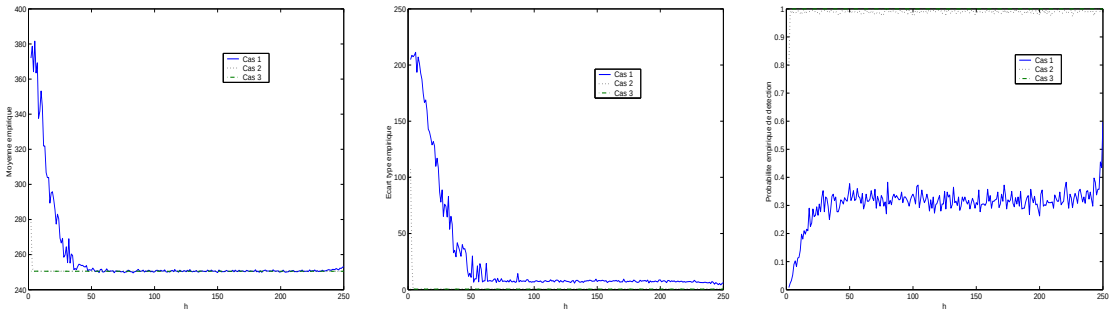


FIG. 2.4 – De gauche à droite, la moyenne et l'écart type empiriques de l'estimateur de l'instant de la rupture ainsi que la probabilité empirique de détecter correctement la rupture sont représentés pour des valeurs de h allant de 2 à T_{cp} et en utilisant 400 réalisations d'un processus Gaussien ayant un changement dans la moyenne entre les instants $T_{cp} = 250$ et $T_{cp} + 1 = 251$.

Dans les deuxièmes et troisièmes cas, (où les amplitudes des changements sont les plus importantes) la probabilité de détecter le véritable instant de la rupture est proche de 1 pour toutes les valeurs de $h \geq 3$. Dans le deuxième cas, la probabilité empirique de détection de l'instant de la rupture est 0.8 pour $h = 2$ et est supérieure à 0.97 pour toutes les autres valeurs de h testées. Dans le troisième cas, toutes les valeurs de h permettent de toujours trouver l'instant de la rupture. Dans le premier cas, nous devons prendre $h \geq 50$ pour avoir de bons résultats en moyenne et un faible écart type pour l'estimateur de l'instant de la rupture.

En utilisant cette méthode, nous pouvons aussi détecter un changement dans la variance d'un processus si la moyenne est constante. Considérons alors à présent un processus Gaussien S_t où pour $1 \leq t \leq T_{cp} = 250$, $S_t \sim \mathcal{N}(0, \sigma_1^2)$ et pour $251 \leq t \leq 750$, $S_t \sim \mathcal{N}(0, \sigma_2^2)$. Pour ce processus, la variance change brutalement après l'instant $T_{cp} = 250$, alors que la moyenne reste constante. On considère trois cas comme précédemment correspondant à des ruptures de plus en plus importantes : $\sigma_1^2 = 1, \sigma_2^2 = 2.25$ (cas 1), $\sigma_1^2 = 1, \sigma_2^2 = 5$ (cas 2), et $\sigma_1^2 = 1, \sigma_2^2 = 10$ (cas 3). On s'intéresse aux mêmes quantités que précédemment. Les résultats sont représentés ci-après sur la figure 2.5.

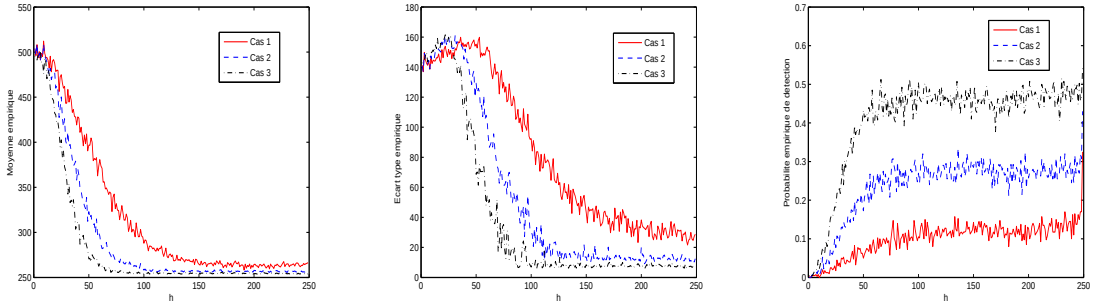


FIG. 2.5 – De gauche à droite, la moyenne et l'écart type empiriques de l'estimateur de l'instant de la rupture ainsi que la probabilité empirique de détecter correctement la rupture sont représentés pour des valeurs de h allant de 2 à T_{cp} et en utilisant 400 réalisations d'un processus Gaussien ayant un changement dans la variance entre les instants $T_{cp} = 250$ et $T_{cp} + 1 = 251$.

Dans les deuxièmes et troisièmes cas, si la valeur de h est suffisamment grande (disons $h \geq 100$) on obtient de bons résultats en moyenne et un écart type faible. Cependant, la probabilité de détecter l'instant de la rupture n'est pas grande (au mieux 0.5) même dans la troisième cas. Au vu de ces simulations, il semble plus difficile de détecter un changement dans la variance que dans la moyenne.

Pour chacune des situations considérées, on peut définir un choix optimal de la largeur de fenêtre h, qui serait la valeur $\hat{h}$ minimisant $|\hat{E}(\hat{T}_{cp}(h)) - T_{cp}| + \hat{\sigma}(\hat{T}_{cp}(h))$ où $\hat{E}(\hat{T}_{cp}(h))$ et $\hat{\sigma}(\hat{T}_{cp}(h))$ sont les estimations empiriques de la moyenne et de l'écart type de l'estimateur de l'instant de la rupture pour une largeur de fenêtre h. On obtient les valeurs de $\hat{h}$ données dans le tableau 2.1 ci-après.

Sortie	$m_1 = 0$ $m_2 = 1$	$m_1 = 0$ $m_2 = 5$	$m_1 = 0$ $m_2 = 10$	$\sigma_1^2 = 1$ $\sigma_2^2 = 2.25$	$\sigma_1^2 = 1$ $\sigma_2^2 = 5$	$\sigma_1^2 = 1$ $\sigma_2^2 = 10$
$\hat{h}$	246	129	132	198	204	204
$\hat{E}(\hat{T}_{cp}(\hat{h}))$	251.25	250.43	250.43	262.72	255.10	253.90
$\hat{\sigma}(\hat{T}_{cp}(\hat{h}))$	4.49	0.51	0.50	32.67	9.11	5.78
$\hat{P}(\hat{T}_{cp}(\hat{h}))$	0.36	0.99	1	0.13	0.28	0.48

TAB. 2.1 – Meilleure valeur $\hat{h}$ de h et pour cette valeur de h, moyenne et écart type empiriques de l'estimateur de l'instant de la rupture ainsi que la probabilité empirique de détecter correctement l'instant de la rupture.

Cas multidimensionnel. Nous testons maintenant la méthode sur un vecteur Gaussien de taille $n = 30$. Les observations S_t, $1 \leq t \leq T_{cp} = 250$ sont issues d'une distribution normale $\mathcal{N}(m_i, Q_j)$ et les observations restantes S_t, $251 \leq t \leq 750$ sont issues d'une distribution normale $\mathcal{N}(m_{i'}, Q_{j'})$ pour différentes combinaisons du couple (m_i, Q_j) où m_i et Q_j sont des moyennes et des matrices de covariance définies de la manière suivante. Si $\delta = \frac{0.05}{n}$, les moyennes m_j sont donnés pour $1 \leq i \leq n$ par : $m_1(i) = 1.15 + i\delta$, $m_2(i) = 1.2 + i\delta$, $m_3(i) = 1.4 + i\delta$, et $m_4(i) = 1.7 + i\delta$. Les matrices de covariance Q_j sont des matrices diagonales telles que pour $1 \leq i \leq n$: $Q_1(i,i) = \frac{in(n+1)}{18}\delta^2$, $Q_2(i,i) = \frac{5in(n+1)}{36}\delta^2$, $Q_3(i,i) = \frac{2in(n+1)}{9}\delta^2$ et $Q_4(i,i) = \frac{in(n+1)}{2}\delta^2$. On obtient les matrices Q_2, Q_3 et Q_4 en multipliant les écarts types donnés dans la matrice Q_1 par respectivement 1.5, 2 et 3. Ces données peuvent être vues comme des moyennes et matrices de covariance de rendements indépendants d'actifs financiers risqués (et sont proches des données utilisées dans les simulations de [6]). En effet, plus la moyenne d'un rendement est élevée, plus sa variance est forte. Nous considérons alors dans un premier temps un processus de matrice de covariance Q_1 constante et dont la moyenne change. On mesure la qualité de la détection de la rupture lorsque l'amplitude de cette rupture augmente. Le premier cas (Cas 1 sur la figure 2.6 ci-après) correspond à un processus ayant une moyenne m_1 jusqu'à T_{cp} et m_2 ensuite. Dans le deuxième cas (noté Cas 2) la moyenne passe de m_1 à m_3 à l'instant $T_{cp} + 1$ et dans le dernier cas (noté Cas 3) de m_1 à m_4. Le cas où la matrice de covariance change et où la moyenne est fixée à la valeur m_1 est ensuite traitée. Le premier cas étudié (Cas 1) correspond à un processus ayant une matrice de covariance Q_1 jusqu'à T_{cp} et Q_2 après. Dans le deuxième cas (Cas 2) la matrice de covariance passe de Q_1 à Q_3 à l'instant $T_{cp} + 1$ et dans le dernier cas (Cas 3) de Q_1 à Q_4. Les résultats sont représentés sur la figure 2.6 ci-après.

Comme dans le cas unidimensionnel, on peut déterminer la meilleure valeur $\hat{h}$ de h dans chacun des cas considérés. Les résultats sont donnés dans le tableau 2.2 ci-après. La première ligne de ce tableau se lit de la façon suivante. Par exemple, la première colonne correspond à un processus ayant une matrice de covariance Q_1 fixe et une moyenne changeante : la moyenne est m_1 jusqu'à T_{cp} et m_2 après.

Cette méthode détecte donc très bien les ruptures dans la moyenne (même dans le cas multidimensionnel). Les ruptures dans les variances/covariances doivent être importantes afin d'obtenir un bon estimateur de l'instant de la rupture, tout particulièrement dans le

cas multidimensionnel.

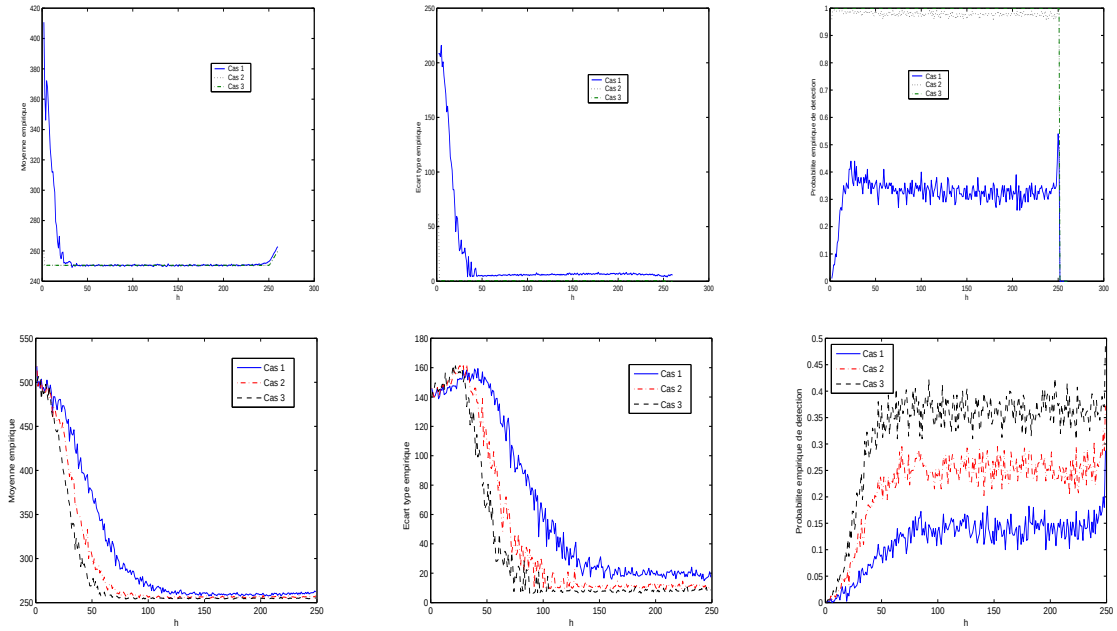


FIG. 2.6 – Cadres supérieurs, de gauche à droite, la moyenne et l'écart type empiriques de l'estimateur de l'instant de la rupture ainsi que la probabilité empirique de détecter correctement la rupture sont représentés pour des valeurs de h allant de 2 à 260 en utilisant 400 réalisations de trois processus Gaussiens ayant trois différentes ruptures dans la moyenne. Les trois figures du cadre inférieur correspondent au cas où la matrice de covariance change.

Sortie	m_1, m_2, Q_1	m_1, m_3, Q_1	m_1, m_4, Q_1	m_1, Q_1, Q_2	m_1, Q_1, Q_3	m_1, Q_1, Q_4
$\hat{h}$	37	3	150	162	202	134
$\hat{E}(\hat{T}_{cp}(\hat{h}))$	250.17	250.47	250.43	257.86	254.5	253.7
$\hat{\sigma}(\hat{T}_{cp}(\hat{h}))$	4.27	0.50	0.49	15.9	9.4	5.8
$\hat{P}(\hat{T}_{cp}(\hat{h}))$	0.38	0.9975	1	0.14	0.3	0.38

TAB. 2.2 – Meilleure valeur $\hat{h}$ de h et pour cette valeur de h, moyenne et écart type empiriques de l'estimateur de l'instant de la rupture ainsi que la probabilité empirique de détecter correctement la rupture.

2.4.4 Discussion

La méthode utilisant une fenêtre glissante permet de détecter la rupture la plus importante dans la moyenne ou dans les variances/covariances sur une période d'étude donnée. Pour que cette méthode soit efficace, elle devrait être utilisée dans le cas où il n'y a qu'une seule rupture et où au moins quelques données sont issues des deux distributions. L'algorithme adaptatif ne nécessite aucune hypothèse sur le nombre de ruptures sur la période d'étude. Il vise à déterminer l'instant de la dernière rupture importante. La performance des

deux algorithmes dépend du choix de leurs paramètres (le couple (λ, μ) pour l'algorithme adaptatif et la largeur de fenêtre h pour l'autre).

2.5 Simulations numériques

Le but de cette section est de tester l'efficacité de l'algorithme adaptatif présenté dans la section 2.3 à la fois sur des données simulées et sur des données réelles. Pour plus de flexibilité, nous utiliserons les estimations empiriques adaptatives de la moyenne et de la matrice de covariance. De la sorte, l'algorithme adaptatif dépendra de deux paramètres positifs λ et μ. À partir de l'étude théorique conduite en section 2.3, on voit que l'on peut choisir la règle d'acceptation suivante pour un intervalle I donné. Nous accepterons l'intervalle I comme intervalle d'homogénéité temporelle locale si pour tout $J \in \mathcal{I}(I)$, $\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq k_1\sigma(\sqrt{\frac{\ln n+\lambda \ln |I|}{|I|}} + \sqrt{\frac{\ln n+\lambda \ln |J|}{|J|}})$ et $\|\hat{Q}_I - \hat{Q}_J\|_\infty \leq k_2\sigma^2(\sqrt{\frac{\ln n(n+1)+\mu \ln |I|}{|I|}} + \sqrt{\frac{\ln n(n+1)+\mu \ln |J|}{|J|}})$ où (λ, μ, k_1, k_2) sont des constantes positives. Étant donné que l'on utilise les estimations empiriques adaptatives, σ (introduit dans l'hypothèse 2) est lié à l'amplitude du bruit. Puisque la moyenne ρ_t peut être arbitrairement petite, nous estimerons σ (ce qui est assez conservatif) par le plus grand des rendements $r_t(i)$ disponibles parmi tous les pas de temps t et tous les actifs i. Pour (I, J) fixés et pour des valeurs de λ et μ suffisamment grandes, l'étude théorique effectuée en section 2.3 fournit des constantes k_1 et k_2 (indépendantes de λ et μ) telles que les probabilités que les inégalités ci-dessus sont vérifiées sont arbitrairement petites. Bien sûr, les mêmes inégalités seront vérifiées avec au moins la même probabilité pour des valeurs plus grandes de k_1 et k_2. Cependant, le choix des constantes k_1 et k_2 issues de notre étude théorique pourrait être conservatif. Plus ces constantes sont élevées, plus on aura de chances de considérer qu'un intervalle où la moyenne, les variances/covariances sont constantes est un intervalle d'homogénéité mais moins on sera capable de détecter des ruptures. Nous essaierons donc de déterminer les plus petites valeurs des constantes k_1 et k_2 telles que les inégalités ci-dessus sont vérifiées avec un niveau de confiance donné pour un processus donné. Dans le cas particulier d'un processus unidimensionnel constant, pour tout I, J, $\hat{\rho}_I = \hat{\rho}_J$ et $\hat{Q}_I = \hat{Q}_J$ ce qui signifie que l'on peut prendre $k_1 = k_2 = 0$. Afin de ne pas choisir de trop grandes valeurs de k_1 et k_2 tout en détectant l'absence de ruptures avec une grande probabilité, nous devrions avoir une bonne estimation de σ. Nous allons ainsi introduire deux degrés de liberté supplémentaires et considérer k_1 et k_2 comme des paramètres de notre algorithme. Supposons que l'on cherche à calibrer les paramètres (λ, μ, k_1, k_2) de l'algorithme adaptatif dans un contexte paramétrique, pour un processus Gaussien par exemple. Choisissons une valeur $\rho_{\min}$ de telle sorte que l'on s'attende à ce que la moyenne $\rho_t > \rho_{\min}$ et une matrice $Q \succeq 0$ telle que les variances dans la matrice Q sont des bornes supérieures sur les variances attendues des composantes. On peut ensuite choisir les valeurs (λ, μ, k_1, k_2) des paramètres de telle sorte que l'on puisse détecter avec une grande probabilité l'absence de ruptures pour un vecteur Gaussien r' de loi $\mathcal{N}(\rho_{\min}, Q)$. Intuitivement, cette estimation conservative des paramètres devrait aussi permettre de détecter l'absence de ruptures pour des processus r ayant des moyennes suffisamment plus grandes que $\rho_{\min}$ et des variances $\sigma'_t(i)^2$ inférieures ou égales aux variances dans Q. En effet, pour un tel processus, les membres de gauche dans les inégalités testant s'il y a homogénéité ou non devraient être inférieurs et la plus grande réalisation des rendements sera certainement supérieure si pour tout temps t, il y a une composante i_t telle que $\rho_t(i_t) + \beta\sigma'_t(i_t) > \rho_{\min}(i_t) + \beta\sqrt{Q(i_t, i_t)}$; avec $\beta = 2$ par exemple

(en fait β est tel qu'avec une grande probabilité le rendement $r_t(i_t)$ de la i_t-ème composante au temps t est majoré par $\rho_t(i_t) + \beta\sigma'_t(i_t)$ et $r'_t(i_t)$ est majoré par $\rho_{\min}(i_t) + \beta\sqrt{Q(i_t, i_t)}$). D'autres calibrations des paramètres de l'algorithme seront présentées par la suite. Elles sont basées sur l'utilisation de l'erreur quadratique moyenne (Mean Squared Error ou MSE en anglais) et l'erreur absolue moyenne (Mean Absolute Error ou MAE en anglais) pour des données simulées et l'erreur de prédiction quadratique moyenne (Mean Squared Forecast Error ou MSFE en anglais) ou l'erreur de prédiction absolue moyenne (Mean Absolute Forecast Error ou MAFE en anglais) pour des données réelles. La section suivante s'intéresse aux simulations dans le cas Gaussien qui est fréquemment rencontré. La dernière section se focalisera sur l'utilisation de la méthode sur des données réelles, en utilisant les cours des actifs du Dow Jones de janvier 1992 à juin 2004. Pour cette application, nous présentons des méthodes supplémentaires de calibration des paramètres (λ, μ, k_1, k_2). Dans ce contexte, l'algorithme adaptatif est utilisé pour déterminer des estimations des paramètres des modèles VaR et de Markowitz dont on se sert pour gérer un portefeuille d'actifs de janvier 1995 à juin 2004.

2.5.1 Simulations de Monte Carlo dans le cas Gaussien

Dans cette partie, nous testons l'efficacité de l'algorithme sur des exemples simulés, ce qui nous donne une idée des valeurs à choisir pour les paramètres (λ, μ, k_1, k_2) dans le cas Gaussien. Nous commençons par étudier un processus Gaussien homogène. On envisage ensuite un modèle à sauts avec différentes amplitudes de la rupture dans la moyenne ou la matrice de covariance et l'on étudie la sensibilité de l'algorithme par rapport à l'amplitude de la rupture. On montre que, pour des valeurs bien choisies de ses paramètres, l'algorithme adaptatif peut détecter l'absence de changements dans un modèle où la moyenne et la matrice de covariance sont constantes et peut rapidement détecter une rupture dans les moyennes et/ou les variances/covariances si cette rupture est suffisamment importante.

2.5.1.1 Erreur de première espèce

Afin de choisir des valeurs des moyennes et des variances qui soient proches de celles d'actifs financiers, on calcule, le 2 janvier 2002, la moyenne empirique ρ et la matrice de covariance empirique Q des rendements à 3 mois, en utilisant 3 ans d'historiques pour les cours des actifs du Dow Jones. La valeur maximale des rendements moyens (valeur maximale de ρ) est 1.1 correspondant à une variance de 0.08. Nous considérons alors que les moyennes ne peuvent pas descendre en dessous de 0.5ρ et que les variances des rendements doivent être inférieures à 2.25 fois les variances dans Q (ou que les écarts types des rendements sont moins de 1.5 fois les écarts types dans Q). On considère ensuite le modèle Gaussien homogène $\mathcal{N}(0.5\rho, 2.25Q)$. On se fixe un niveau de risque $\alpha = 5\%$, et on s'intéresse aux valeurs des paramètres (λ, μ, k_1, k_2) telles que la probabilité de rejeter un intervalle d'homogénéité de longueur donnée est au plus α. Dans ce modèle, à la fois la moyenne ρ_t et la matrice de covariance Q_t sont donc indépendantes du temps. Le bruit ζ_t est Gaussien. On simule ainsi trois ensembles de séries chronologiques gaussiennes (i.i.d $\mathcal{N}(0.5\rho, 2.25Q)$, où $n = 30$ est la dimension de ζ_t) de longueurs $L = 120, 240, 360$. Chacun de ces ensembles contient 400 réalisations. Les ensembles $\mathcal{I}(I)$ sont définis en utilisant une grille dynamique (voir section 2.3.2), m_0 étant le pas de la grille. On choisit pour m_0 les valeurs 15, 30, 40 et 60. Remarquons que si $m_0 < n = 30$, on calcule des estimations dégénérées de la matrice de covariance. Les valeurs de λ et μ constituent une grille dont les

extrémités gauches et droites sont 0.1 et 2.9 et dont le pas est 0.4. Les valeurs de k_1 et k_2 testées sont 0.1, 0.2, 0.3, 0.5, 1 et 2. On retient ensuite les valeurs des paramètres donnant une erreur de première espèce d'au plus 5%. Parmi les 2304 valeurs des combinaisons des paramètres testées, le nombre de combinaison des paramètres (λ, μ, k_1, k_2) qui fournit une telle erreur de première espèce pour différentes valeurs du couple (L, m_0) est donné dans le tableau suivant :

(L, m_0)	(120,15)	(120,30)	(120,40)	(120,60)
Nombre	1282	1604	1675	1890
(L, m_0)	(240,15)	(240,30)	(240,40)	(240,60)
Nombre	1185	1443	1467	1658
(L, m_0)	(360,15)	(360,30)	(360,40)	(360,60)
Nombre	1185	1387	1444	1557

TAB. 2.3 – Nombre de quadruplets fournissant une erreur de première espèce d'au plus 5%.

Les observations faites dans ces simulations sont conformes à nos attentes. Bien sur, la probabilité de rejeter un intervalle d'homogénéité temporelle locale augmente avec la longueur L de cet intervalle. Cette probabilité est par ailleurs une fonction décroissante du pas de la grille m_0. En effet, pour de petites valeurs du pas de la grille, les tests d'homogénéité sont effectués très souvent. Finalement, pour L, m_0, k_1, k_2 et λ fixés (resp. μ), la probabilité de rejeter un intervalle homogène est une fonction décroissante de μ (resp. λ). Nous allons maintenant chercher, parmi les valeurs des paramètres qui fournissent une erreur de première espèce d'au plus 5%, ceux qui fournissent aussi une erreur de deuxième espèce raisonnable.

2.5.1.2 Erreur de deuxième espèce

Considérons une suite d'intervalles (I_t) de longueur T_1+t, où T_1 est fixé et $t = km_0$ avec $1 \leq k \leq E[\frac{T_2}{m_0}]$, $m_0 > 0$ et $T_2 \geq m_0$. Sur chaque intervalle I_t, on génère 400 réalisations du processus suivant. Les T_1 premières observations sont issues d'une loi normale n dimensionnelle ($n = 30$) $\mathcal{N}(\beta_1\rho, \beta_2 Q)$ où ρ et Q sont respectivement les moyennes et matrices de covariance empiriques des rendements d'actifs financiers utilisées dans le paragraphe précédent (calculées en utilisant des données réelles). Les t dernières observations sont issues d'une loi normale $N(\beta_1'\rho, \beta_2' Q)$. Pour chaque valeur $t = km_0$ avec $1 \leq k \leq E[\frac{T_2}{m_0}]$, on s'intéresse à l'erreur de deuxième espèce, i.e la probabilité d'accepter l'intervalle I_t qui ne devrait pas être considéré comme un intervalle d'homogénéité pour des valeurs suffisamment grandes de $|\beta_1 - \beta_1'|$ et $|\beta_2 - \beta_2'|$. Pour chaque intervalle I_t, et chaque réalisation du processus, on utilise donc l'algorithme adaptatif pour savoir, si, étant donné cette réalisation du processus, l'intervalle peut être considéré comme un intervalle d'homogénéité ou non. Le pas de la grille de l'algorithme est m_0. L'algorithme adaptatif est implémenté en utilisant un sous-ensemble des paramètres $(L, m_0, \lambda, \mu, k_1, k_2)$ determinés dans le paragraphe précédent et qui fournissent une erreur de première espèce d'au plus 5%. Afin de réduire le temps de calcul, on choisit les plus petites valeurs de ces paramètres qui devraient être les meilleures pour obtenir une erreur de deuxième espèce raisonnable. Plus précisément, on choisit $k_1 = 0.1$ et $k_2 \leq 0.3$. Pour chaque combinaison $(L, m_0, \lambda, \mu, k_1, k_2)$ des paramètres et chaque instant $t = km_0 \leq T_2$, on peut calculer une estimation empirique

$P_t(L, m_0, \lambda, \mu, k_1, k_2)$ de la puissance du test :

$$H_0^t : \ I_t \text{ est d'homogénéité temporelle locale}, \quad H_1^t : \ \bar{H}_0^t.$$

- Dans une première série d'expérience, on choisit $T_1 = T_2 = 360$, $\beta_1 = \beta_2 = \beta_2' = 1$ (on considère seulement un changement dans la moyenne). Pour $\beta_1' = 1.1, 1.2$ (et en théorie pour des valeurs de β_1' supérieures), on a pu trouver des valeurs des paramètres telles que l'intervalle I_t était toujours rejeté pour toutes les valeurs de t. Pour $\beta_1' = 1.05$, les meilleurs paramètres ont donné les puissances empiriques suivantes :

t	60	120	180	240	300	360
P_t	0.24	0.33	0.31	0.29	0.31	0.31

  en utilisant les paramètres $(L^*, m_0^*, \lambda^*, \mu^*, k_1^*, k_2^*) = (240, 60, 0.1, 1.7, 0.1, 0.2)$.
- On considère maintenant un changement dans la matrice de covariance seulement ($\beta_1 = \beta_1' = 1$). On pose $T_1 = 60$ et $T_2 = 300$. Si les intervalles I_t étaient d'homogénéité, nous aimerions que notre algorithme les retienne comme intervalle d'homogénéité. Pour obtenir une erreur de deuxième espèce raisonnable, si on considère des changements dans les variances/covariances, nous devons choisir des valeurs à tester supplémentaires pour les paramètres. En procédant par dichotomie, on essaie déterminer de petites valeurs des paramètres tels que l'erreur de première espèce (en simulant 400 réalisations d'un vecteur Gaussien $\mathcal{N}(0.5\rho, 2.25Q)$ de longueur 360) soit au plus 5%. Parmi ces paramètres, l'ensemble des valeurs $(L^*, m_0^*, \lambda^*, \mu^*, k_1^*, k_2^*) = (360, 60, 0.4, 0.8, 0.15, 0.016)$ est optimal pour l'erreur de deuxième espèce et donne les puissances suivantes :

t	60	120	180	240	300
P_t $(Q, 2Q)$	0.1	0.32	0.57	0.68	0.77
P_t $(0.5Q, 2Q)$	0.41	0.92	0.99	1	1
P_t $(0.2Q, 2Q)$	0.73	0.99	1	1	1

On considère maintenant un modèle à sauts Gaussien comme précédemment mais on s'intéresse aux erreurs faites sur les estimations de la moyenne et de la matrice de covariance ainsi qu'à l'évaluation de la vitesse de détection de la rupture. On fixe T_1 et T_2 tels que les T_1 premières observations sont issues d'une loi normale $\mathcal{N}(\beta_1\rho, \beta_2 Q)$ (ρ et Q sont définis au début de cette section) et les T_2 dernières observations d'une loi normale $\mathcal{N}(\beta_1'\rho, \beta_2' Q)$. On considère ensuite une grille de valeurs pour les paramètres $(m_0, \lambda, \mu, k_1, k_2)$ de l'algorithme adaptatif et pour chaque combinaison des paramètres, on effectue 100 simulations de ce processus. Pour chaque simulation, et pour chaque instant $t = t_0, \ldots, T_1 + T_2$ (où t_0 est supérieur où égal à la longueur du plus petit intervalle de l'ensemble $\mathcal{I}$), on calcule les estimations de l'intervalle adaptatif, de la moyenne et de la matrice de covariance des rendements. Si $T = T_1 + T_2 - t_0 + 1$, on définit les quatre critères d'erreur suivants :

$$MSE = E\frac{1}{T}\sum_{t=1}^{T} \|\hat{\rho}_t - \rho_t\|_\infty^2 + \frac{1}{T}\sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_\infty^2, \quad MSFE = E\frac{1}{T}\sum_{t=1}^{T} \|\hat{\rho}_t - r_t\|_\infty^2,$$

$$MAE = E\frac{1}{T}\sum_{t=1}^{T} \|\hat{\rho}_t - \rho_t\|_\infty + \frac{1}{T}\sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_\infty, \quad MAFE = E\frac{1}{T}\sum_{t=1}^{T} \|\hat{\rho}_t - r_t\|_\infty.$$

On choisit ensuite les valeurs $(m_0, \lambda, \mu, k_1, k_2)$ des paramètres qui fournissent les meilleures MSE, MSFE, MAE ou MAFE. Par la suite, nous considérons des changements dans la moyenne, dans la matrice de covariance ou dans les deux.
On compare tout d'abord les différents choix d'implémentation de l'algorithme adaptatif présentés dans le paragraphe 2.3.2, en considérant un changement dans la moyenne. Soit $\beta_1' = 1.5$, $\beta_1 = \beta_2 = \beta_2' = 1$, $T_1 = 120$, $T_2 = 240$ et $t_0 = 60$. On choisit tout d'abord pour l'ensemble $\mathcal{I}$, des intervalles de la forme $[\tau - km_0, \tau[$, où $m_0 = 15$ et $k \in \mathbb{N}^*$. On représente ci-après, sur la figure 2.7, l'erreur d'estimation de la moyenne (i.e la moyenne $E\|\hat{\rho}_t - \rho_t\|_\infty$ et le premier et le troisième quartile pour chaque instant t), l'erreur d'estimation de la matrice de covariance (i.e la moyenne $E\|\hat{Q}_t - Q_t\|_\infty$ et le premier et le troisième quartile pour chaque instant t) et la longueur moyenne de l'intervalle adaptatif, si on utilise les paramètres fournissant la meilleure valeur de la MAE. Dans les figures de cette section, le premier et le troisième quartile seront notés Q_1 et Q_3, la moyenne 'Moy' et la méthode adaptative 'Adapt'.

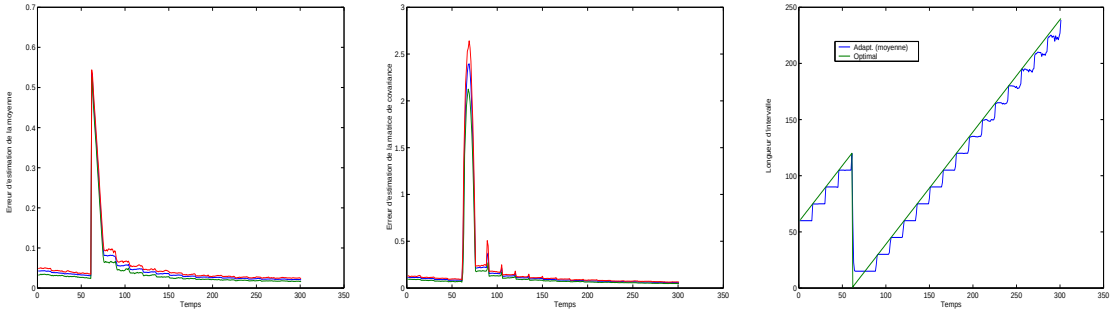
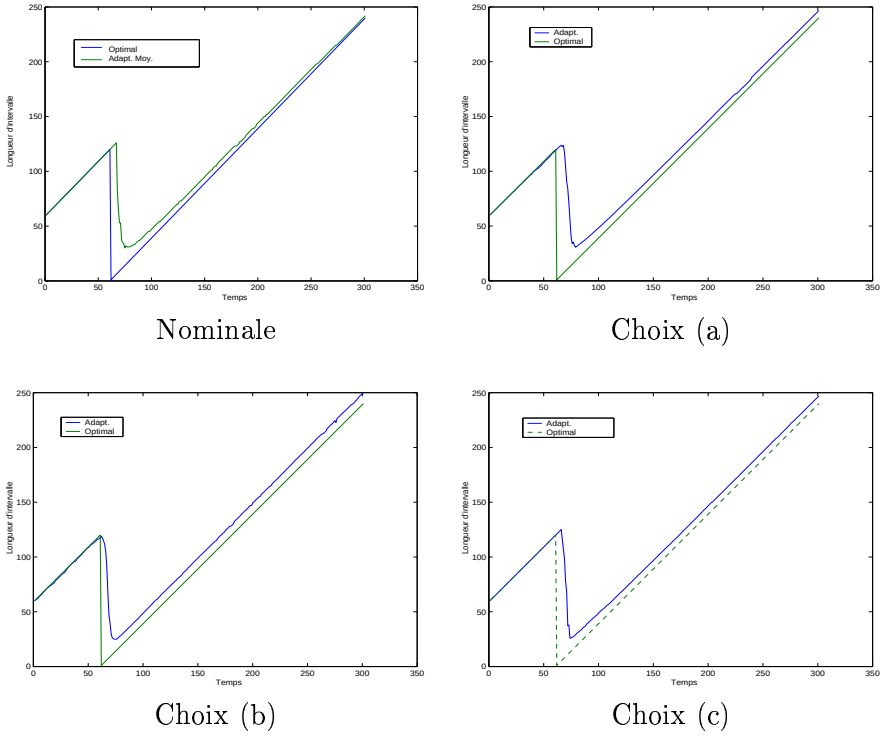


FIG. 2.7 – De gauche à droite, erreur d'estimation de la moyenne, erreur d'estimation de la matrice de covariance et comparaison entre l'intervalle optimal et l'intervalle adaptatif pour un processus Gaussien ayant une rupture dans la moyenne.

L'erreur dans l'estimation de la moyenne et de la matrice de covariance décroît rapidement après la rupture, pour retrouver le niveau d'erreur avant la rupture. Le changement dans la moyenne est rapidement détecté. En effet, en moyenne, le premier instant t pour lequel l'intervalle d'homogénéité n'est pas $[1, t]$ est $t = 122$, c.à.d deux pas de temps après la rupture. En moyenne, 15 pas de temps sont nécessaires pour obtenir un intervalle adaptatif dont l'extrémité gauche est $t = 121$. Par définition de $\mathcal{I}$, la longueur de l'intervalle adaptatif est au moins m_0. Cela signifie que $t = 135$ était le meilleur instant de détection de la rupture que notre algorithme pouvait fournir.

Nous prenons maintenant, pour l'ensemble $\mathcal{I}$, des intervalles de la forme $[\tau - m_0 - k, \tau[$, $k \in \mathbb{N}^*$ et on illustre le fait que les quatre différents choix des sous-intervalles de tests détaillés en section 2.3.2 donnent des résultats similaires. Le choix de cet ensemble $\mathcal{I}$ fournit des temps de détection beaucoup plus courts et il sera donc dorénavant adopté dans toutes les simulations de ce chapitre. On considère un changement dans la moyenne (en utilisant les mêmes valeurs des paramètres que ci-dessus, sauf pour β_1' qui vaut maintenant 1.3) et on représente les estimations de la longueur moyenne de l'intervalle adaptatif ainsi que la longueur de l'intervalle optimal, en utilisant les différentes méthodes adaptatives. Les paramètres sont choisis de façon à minimiser la MAE. Les résultats sont représentés sur la

figure ci-dessous.



Nominale | Choix (a)

Choix (b) | Choix (c)

La méthode nominale consiste à choisir comme sous-intervalles, tous les intervalles de longueur proportionnelle à m_0 et ayant la même extrémité gauche ou la même extrémité droite que l'intervalle I. Les autres choix de sous-intervalles de tests (choix (a), (b) et (c)) sont détaillés dans le théorème 2.3.4. Les différents choix d'implémentation donnent en effet des estimations proches de l'intervalle adaptatif d'homogénéité. La variance de l'estimateur est très petite dans tous les cas (l'amplitude de la rupture étant suffisamment importante pour permettre une bonne détection de l'instant de la rupture).

Les simulations restantes de ce paragraphe seront effectuées en utilisant le choix nominal pour les sous-intervalles de tests. Nous étudions alors 3 modèles à sauts. Pour le premier de ces modèles, on choisit $T_1 = 120$, $T_2 = 240$, $t_0 = 60$, $\beta_1' = 1.3$ et $\beta_1 = \beta_2 = \beta_2' = 1$ (changement dans la moyenne). Le deuxième modèle se focalise sur un changement dans la matrice de covariance seulement et est décrit par les valeurs suivantes des paramètres : $T_1 = 300$, $T_2 = 250$, $t_0 = 201$, $\beta_1 = \beta_1' = 1$ et $\beta_2 = 0.5$, $\beta_2' = 2$. On considère enfin une rupture dans la moyenne et la matrice de covariance en prenant $T_1 = 120$, $T_2 = 240$, $t_0 = 60$, $\beta_1 = 1$, $\beta_1' = 1.3$, $\beta_2 = 0.5$ et $\beta_2 = 2$. Pour ces modèles, l'erreur d'estimation de la moyenne (moyenne et quartiles), l'erreur d'estimation de la matrice de covariance (moyenne et quartiles) ainsi que la longueur de l'intervalle adaptatif (moyenne et quartiles) sont représentés sur la figure 2.8 ci-dessous. Les paramètres choisis pour implémenter l'algorithme adaptatif et qui minimisent la MAE, sont donnés dans le tableau 2.4 ci-dessous.

Nature de la rupture	λ	μ	k_1	k_2	m_0
Rupture moyenne	0.3	0.5	0.09	0.5	70
Rupture matrice cov.	1	0.25	0.5	0.025	70
Rupture mixte	0.09	0.08	0.08	0.014	70

TAB. 2.4 – Valeurs des paramètres minimisant la MAE.

Signalons que les valeurs des paramètres minimisant les différents critères d'erreur sont proches les unes des autres.

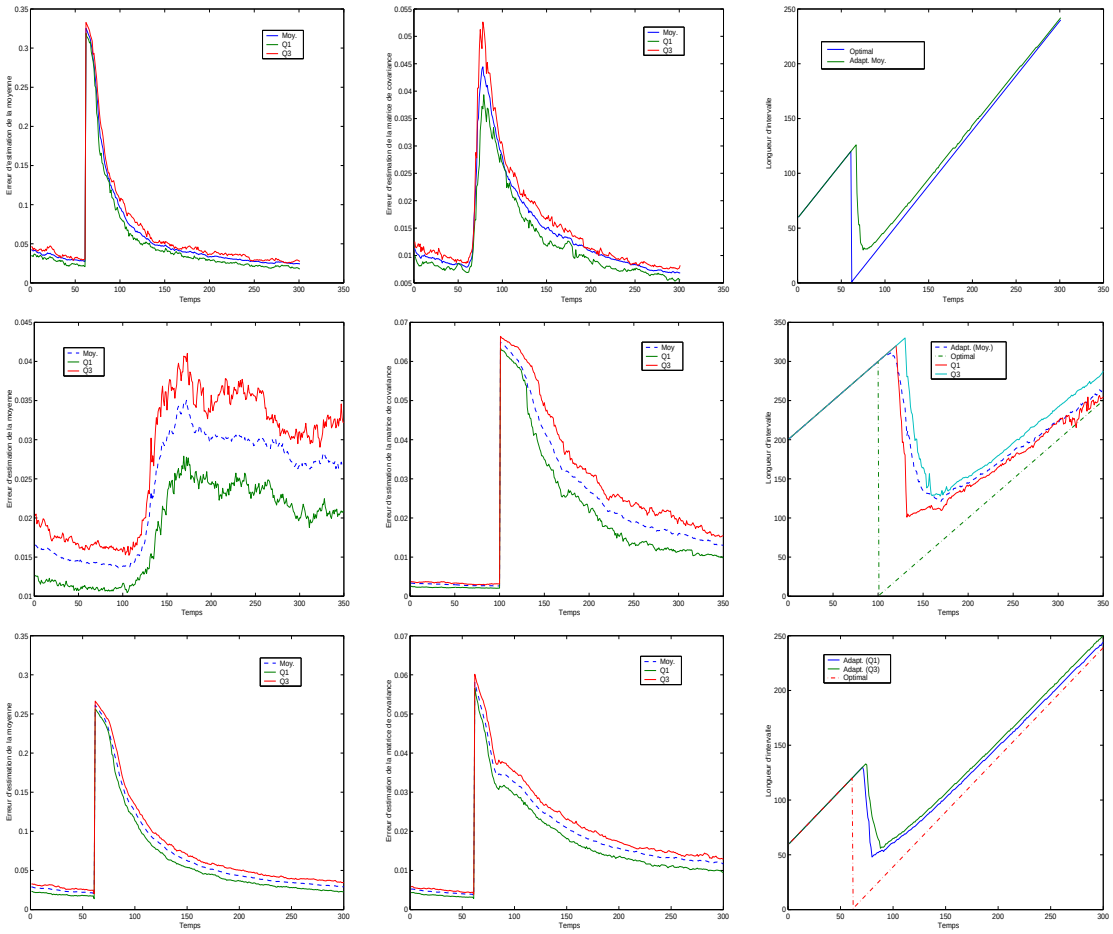


FIG. 2.8 – De haut en bas rupture dans la moyenne, dans la matrice de covariance et dans les deux. De gauche à droite, erreur d'estimation de la moyenne, erreur d'estimation de la matrice de covariance et comparaison entre la longueur de l'intervalle adaptatif et la longueur de l'intervalle optimal.

Notre algorithme peut sélectionner différents intervalles d'homogénéité pour la moyenne et les variances/covariances si la dernière rupture importante dans la moyenne s'est produite avant la dernière rupture importante dans la matrice de covariance. Nous illustrons cela en prenant un modèle Gaussien à sauts. Les $T_1 = 140$ premières observations sont

issues d'une loi normale $\mathcal{N}(\rho, 0.2Q)$, les $T_2 = 200$ observations suivantes d'une loi normale $\mathcal{N}(1.3\rho, 0.2Q)$, et les $T_3 = 300$ dernières observations d'une loi normale $\mathcal{N}(1.3\rho, 2Q)$. En choisissant la même grille de valeurs pour les paramètres de l'algorithme adaptatif, on simule pour chaque combinaison des valeurs des paramètres 100 réalisations de ce processus. Les intervalles adaptatifs sont estimés pour chaque pas de temps $t \geq 80$. Les valeurs des paramètres fournissant la meilleure MAE sont $(\lambda, \mu, k_1, k_2, m_0) = (0.09, 0.08, 0.08, 0.014, 70)$ et donnent les longueurs moyennes d'intervalles représentées sur la figure 2.9 ci-dessous.

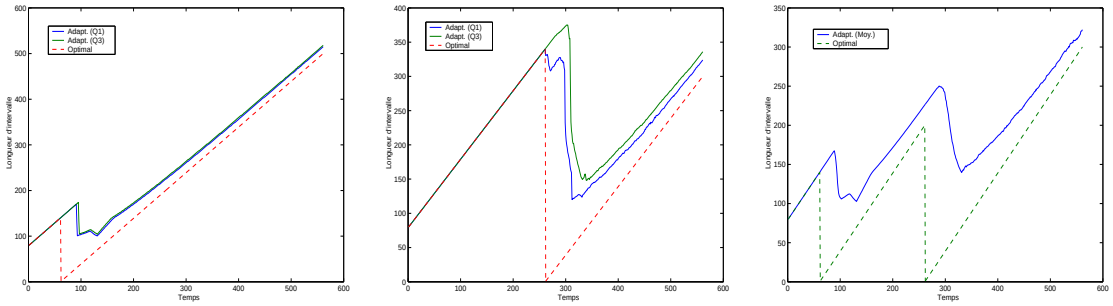


FIG. 2.9 – De gauche à droite, longueur moyenne de l'intervalle adaptatif pour respectivement la moyenne, la matrice de covariance et leur intersection comparée à la longueur de l'intervalle optimal.

Nous voyons qu' ici aussi notre algorithme détecte plus facilement des ruptures dans la moyenne que dans la matrice de covariance, le temps de détection étant raisonnable et décroissant avec l'amplitude de la rupture.

2.5.2 Simulations sur des données réelles

Nous voulons comparer l'évolution des portefeuilles obtenus en utilisant les modèles VaR et de Markowitz en rebalançant les portefeuilles à différentes dates et en utilisant les deux méthodes (adaptatives/non adaptatives) de calibration des paramètres des modèles. On s'intéressera en particulier au rendement du portefeuille et à sa volatilité. Nous utiliserons deux stratégies de calibration non adaptatives des paramètres. La première consiste à déterminer les estimations empiriques des paramètres obtenues en utilisant toutes les données disponibles le jour de l'investissement. La seconde calcule les estimations empiriques en utilisant un horizon d'estimation fixe de 100 jours. La méthode adaptative, on l'a vu, détermine les estimations empiriques obtenues en utilisant les données de l'intervalle adaptatif. Les différentes techniques de calibration diffèrent donc quant au choix de l'horizon d'estimation.
Les actifs des portefeuilles sont les 30 actifs du DOW Jones. On connaît les cours de ces actifs du 2 janvier 1992 au 30 juin 2004. Précisons qu'il s'agit de cours corrigés qui prennent en compte les fusions et augmentations de capital. On investit 1000$ (ce argent est initialement détenu dans l'actif sans risque) le 2 janvier 1995. On choisit un horizon d'investissement M. Différentes valeurs de M seront testées : $M = 90$ jours ouvrables (approximativement 4 mois et demi calendaires), $M = 60$, 30 et $M = 15$. Le portefeuille est ensuite régulièrement rebalancé tous les M jours ouvrables, en utilisant soit le modèle Value-at-Risk (2.5) soit le modèle de Markowitz (2.6). Les frais de transactions s'élèvent

à 0.5%. Le rendement cible ℓ dans le modèle de Markowitz et le rendement de l'actif sans risque sur la période d'investissement dépendent de la valeur M et de l'estimation des rendements moyens. Ils sont choisis de manière à ce que le problème soit faisable. On impose de plus que le rendement cible soit supérieur au rendement de l'actif sans risque (afin de ne pas tout investir dans cet actif sans risque). Par ailleurs, les rendements moyens de certains actifs doivent être supérieurs au rendement cible. On choisit $\kappa(\varepsilon) = 0.5$ ou $\kappa(\varepsilon) = 0.3$. Ceci conduit à des portefeuilles assez risqués, mais les simulations ont montré, qu'en utilisant nos données, des valeurs de $\kappa(\varepsilon)$ supérieures à 0.6 donnent généralement des portefeuilles où tout le capital (ou la plupart du capital) est investi dans l'actif sans risque.
On s'intéressera aussi à la moyenne et à l'écart type de la longueur des estimations d'horizon d'investissement obtenues pour les différentes dates de rebalancement en utilisant l'algorithme adaptatif. Les ruptures dans la moyenne pourraient, dans une première approximation, être visuellement déterminées en observant l'évolution des rendements des actifs. Cependant, la détermination des ruptures dans les variances/covariances doit passer par une étape calculatoire. D'autre part, afin de réduire le temps de calcul, nous utiliserons le choix des sous-intervalles de tests noté choix (c) dans le théorème 2.3.4. Les intervalles de l'ensemble $\mathcal{I}$ sont de la forme $I_k = [\tau - m_0 - k, \tau[$ où $k \in \mathbb{N}$. Nous expliquons maintenant comment les paramètres de l'algorithme adaptatif sont choisis.

Choix à postériori des paramètres.

Les paramètres de l'algorithme adaptatif sont tout d'abord déterminés a posteriori afin de montrer l'influence de l'horizon d'estimation sur la performance des portefeuilles. Nous déterminerons un intervalle d'homogénéité $\hat{I}$, l'intersection de $\hat{I}_\rho$ et $\hat{I}_Q$. Une grille de valeurs est choisie pour les paramètres k_1 et k_2. Les valeurs λ et μ sont fixées (afin de réduire le temps de calcul) à 0.5 et on fixe $m_0(\rho) = m_0(Q) = 15$. On envisage ensuite toutes les combinaisons possibles des valeurs des paramètres k_1 et k_2 et on choisit la combinaison donnant le rendement maximum sur la période d'investissement (on suppose que tous les cours sont connus sur l'ensemble de la période d'investissement). Remarquons que pour une expérience donnée, les mêmes valeurs des paramètres sont utilisées à chaque rebalancement. Un choix dynamique des paramètres pourrait encore améliorer les résultats. On donne les valeurs suivantes pour les paramètres fixés :

Horizon	ℓ	$r(n+1)$	$\kappa(\varepsilon)$
$M = 90$	1.1	1.02	0.5
$M = 60$	1.05	1.015	0.5
$M = 30$	1.015	1.007	0.5
$M = 15$	1.007	1.003	0.3

TAB. 2.5 – Valeurs fixées pour les paramètres des modèles VaR et de Markowitz.

On notera $\hat{E}(\hat{I})$ et $\hat{\sigma}(\hat{I})$ la moyenne et l'écart type de la longueur de l'intervalle d'homogénéité temporelle locale. Les résultats, pour le modèle de Markowitz, sont résumés dans le tableau 2.6 ci-après :

Horizon	k_1	k_2	R_{Ad}	R'_{Ad}	σ_{Ad}	R_{Emp}	R'_{Emp}	σ_{Emp}	$\hat{E}(\hat{I})$	$\hat{\sigma}(\hat{I})$
$M = 90$	1	1	4.625	1.068	0.142	4.625	1.068	0.142	1863	714
$M = 60$	0.4	0.1	3.867	1.039	0.091	3.071	1.031	0.084	1074	545
$M = 30$	0.31	0.1	2.444	1.011	0.027	2.345	1.011	0.026	997	560
$M = 15$	0.24	0.1	2.467	1.0056	0.0182	2.197	1.005	0.017	1058	630

TAB. 2.6 – Comparaison des méthodes empiriques et adaptatives sur des données réelles. Modèle de Markowitz. Choix statique, à posteriori des paramètres.

Dans le tableau ci-dessus, (et dans ce qui suit) R_{Ad} est le rendement du portefeuille obtenu avec la méthode adaptative sur l'ensemble de la période d'investissement (de 1995 à 2004). R_{Emp} est le rendement du portefeuille obtenu en utilisant la méthode empirique sur la même période. Connaissant l'évolution de la valeur du portefeuille sur la période d'investissement, on peut calculer l'échantillon des rendements à M jours du portefeuille. La moyenne empirique de cet échantillon (le rendement moyen à M jours) est R'_{Ad} (si la méthode adaptative est utilisée) ou R'_{Emp} (si la méthode empirique est utilisée) et son écart type empirique (qui mesure la volatilité du rendement du portefeuille) est σ_{Ad} pour la méthode adaptative et σ_{Emp} pour la méthode empirique. Les résultats, pour le modèle VaR, sont donnés dans le tableau 2.7 ci-dessous.

Horizon	k_1	k_2	R_{Ad}	R'_{Ad}	σ_{Ad}	R_{Emp}	R'_{Emp}	σ_{Emp}	$\hat{E}(\hat{I})$	$\hat{\sigma}(\hat{I})$
$M = 90$	0.15	0.1	4.430	1.075	0.188	3.937	1.064	0.159	60	30
$M = 60$	0.28	0.1	3.934	1.042	0.115	3.202	1.035	0.114	440	330
$M = 30$	0.35	0.1	3.739	1.018	0.056	3.678	1.018	0.059	1251	510
$M = 15$	0.24	0.1	5.289	1.0115	0.051	2.554	1.007	0.056	1058	630

TAB. 2.7 – Comparaison des méthodes adaptatives et empiriques sur des données réelles, modèle VaR. Choix statique à posteriori des paramètres.

À part pour le modèle de Markowitz et l'horizon $M = 90$ jours, on peut toujours trouver des valeurs des paramètres fournissant un portefeuille ayant un rendement supérieur à celui du portefeuille empirique sur la période d'investissement. Il semble que plus les durées de détention sont courtes et plus l'algorithme adaptatif est intéressant. Ceci signifierait que plus M est faible, et plus les moyennes et les variances des rendements varient. En particulier, pour $M = 15$ jours et le modèle VaR, on obtient une amélioration notoire en utilisant la méthode adaptative, le rendement du portefeuille passant de 2.554 à 5.289.

Choix à priori des paramètres.

Cependant, d'un point de vue pratique, on doit déterminer une politique de choix des paramètres en utilisant seulement l'information disponible le jour de l'investissement. On utilisera la même grille de valeurs admissibles pour les paramètres de l'algorithme adaptatif que celle utilisée précédemment pour la détermination à posteriori des paramètres. Les paramètres seront choisis de manière à optimiser un certain critère d'erreur. Nous proposons quatre différentes techniques dynamiques (les valeurs des paramètres peuvent changer d'un rebalancement à l'autre) pour résoudre ce problème. Tout d'abord, les rendements moyens peuvent être déterminés dans le passé (le nombre de données disponibles

variant d'un rebalancement à l'autre, rendant ainsi la méthode dynamique) à différentes dates en utilisant l'algorithme adaptatif. On peut ensuite déterminer les valeurs des paramètres donnant la plus petite erreur absolue moyenne de prédiction (Mean Absolute Forecast Error ou MAFE en anglais) ou la plus petite erreur quadratique moyenne de prédiction (Mean Squared Forecast Error ou MSFE en anglais). Les dates choisies sont les 15 dernières dates pour lesquelles les rendements des actifs sont connus. Si l'on cherche à déterminer deux différents intervalles $\hat{I}_\rho$ et $\hat{I}_Q$, cette méthode permet une calibration de k_1 et λ. Afin d'éviter toute ambiguïté, cette méthode sera utilisée si on détermine un seul intervalle d'homogénéité. Dans ce cas, les valeurs de k_2 et μ influencent aussi l'estimation de l'intervalle d'homogénéité et par conséquent l'estimation de la moyenne. Une deuxième manière de calibrer les paramètres de l'algorithme adaptatif, consiste à simuler différentes dates d'investissement dans la passé. On choisit 5 (arbitrairement) différentes dates d'investissement précédant le jour du rebalancement (pour lequel on veut calibrer les paramètres) et telles que les cours des actifs soient connus pour les M jours suivants ces différentes dates d'investissement. Ceci permet de calculer les portefeuilles optimaux pour chaque valeur des paramètres et d'observer l'évolution réelle des différents portefeuilles. On choisit ensuite les valeurs des paramètres donnant le rendement moyen maximal sur l'ensemble des dates testées. On obtient alors deux ensembles de paramètres étant donné que l'on teste à la fois la méthode donnant un intervalle d'homogénéité (méthode notée Rdt, $I = I_\rho \cap I_Q$ dans les tableaux ci-dessous) et la méthode fournissant deux intervalles d'homogénéité (méthode notée Rdt, I_ρ, I_Q dans les tableaux ci-dessous). Les résultats pour l'horizon $M = 60$ jours sont résumés dans les tableaux 2.8 et 2.9 ci-après.

Méthode	k_1	k_2	R_{Ad}	R'_{Ad}	σ_{Ad}	$\hat{E}(\hat{I}_\rho)$	$\hat{\sigma}(\hat{I}_\rho)$	$\hat{E}(\hat{I}_Q)$	$\hat{\sigma}(\hat{I}_Q)$
MSFE	0.358	0.1	3.30	1.033	0.071	1001	866	1001	866
MAFE	0.358	0.1	3.25	1.032	0.072	1003	864	1003	864
Rdt, $I = I_\rho \cap I_Q$	0.362	0.1	2.21	1.022	0.074	991	810	991	810
Rdt, I_ρ, I_Q	0.362	0.1	1.88	1.018	0.076	922	768	922	768

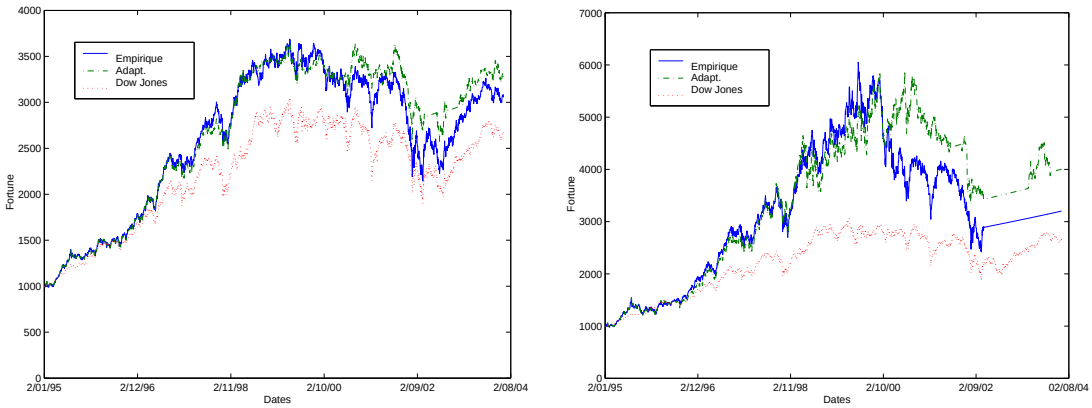
TAB. 2.8 – Méthode adaptative sur des données réelles, modèle de Markowitz. Choix dynamique, à priori, des paramètres. Horizon $M = 60$ jours.

Méthode	k_1	k_2	R_{Ad}	R'_{Ad}	σ_{Ad}	$\hat{E}(\hat{I}_\rho)$	$\hat{\sigma}(\hat{I}_\rho)$	$\hat{E}(\hat{I}_Q)$	$\hat{\sigma}(\hat{I}_Q)$
MSFE	0.358	0.1	3.39	1.037	0.118	1001	866	1001	866
MAFE	0.358	0.1	3.10	1.036	0.123	1003	864	1003	864
Rdt, $I = I_\rho \cap I_Q$	0.344	0.1	3.98	1.041	0.112	890	710	890	710
Rdt, I_ρ, I_Q	0.371	0.1	4.01	1.041	0.111	893	705	893	705

TAB. 2.9 – Méthode adaptative sur des données réelles, modèle VaR. Choix dynamique, à priori, des paramètres. Horizon $M = 60$ jours.

Le rendement de la méthode utilisant un horizon d'estimation fixe de 100 jours est 1.88. Le même test a été effectué avec les autres horizons d'investissement $M = 15$ et $M = 90$. La méthode adaptative permet d'améliorer le rendement du portefeuille ou le ratio de Sharpe (le quotient du rendement du portefeuille R_{Ad} par son risque σ_{Ad}) sur la

période d'investissement. Avec le modèle VaR, pour $M = 15$, trois méthodes de calibration des paramètres sur quatre fournissent des portefeuilles ayant un rendement supérieur et pour $M = 90$, deux méthodes sur quatre. Pour $M = 90$ et $M = 15$ et en utilisant le modèle de Markowitz, le rendement du portefeuille adaptatif est inférieur au rendement du portefeuille empirique mais on peut réduire de façon conséquente le ratio de Sharpe. En effet, pour $M = 90$ on obtient $\frac{R_{Emp}}{\sigma_{Emp}} = 7.54$ alors que $\frac{R'_{Ad}}{\sigma_{Ad}} = 18.32$ avec les méthodes MAFE et MSFE. La méthode utilisant un horizon d'estimation de 100 jours est battue (en terme de rendements moyens) par toutes les autres méthodes. Dans le cas où $M = 60$ jours, l'évolution de la valeur des portefeuilles empirique et adaptatif est représentée sur la figure 2.10 ci-dessous. On représente aussi l'évolution d'un portefeuille ayant même valeur initiale et dont les rendements suivraient ceux de l'indice Dow Jones. Pour la méthode adaptative, on ne retient que le portefeuille ayant le meilleur rendement sur la période d'investissement.



Modèle de Markowitz Modèle VaR

FIG. 2.10 – Comparaison entre le portefeuille adaptatif, le portefeuille empirique et un portefeuille indicé sur l'indice Dow Jones.

Le rendement du portefeuille adaptatif est supérieur au rendement du portefeuille empirique à la fois pour les modèles Value-at-Risk et de Markowitz, la volatilité du portefeuille adaptatif étant légèrement inférieure (voir les tableaux ci-dessus). Les portefeuilles adaptatifs et empiriques sont meilleurs, en terme de performance, que le portefeuille indicé sur le Dow Jones. Le comportement linéaire de la valeur du portefeuille à la fin de la période d'investissement pour le modèle VaR provient du fait tout est investi dans l'actif sans risque pendant cette période. D'autres choix à priori des paramètres pourraient être envisagés, en tenant compte notamment de la variabilité du rendement du portefeuille. Au lieu de ne s'intéresser qu'au rendement du portefeuille et rechercher les paramètres donnant le meilleur rendement en utilisant des données passées, on pourrait rechercher les paramètres donnant la plus petite variabilité du portefeuille (mesurée par σ_{Ad}) ou la plus grande valeur du quotient du rendement du portefeuille par la variabilité du rendement du portefeuille.

2.6 Conclusion sur l'inférence statistique pour la gestion de portefeuilles

La moyenne et la matrice de covariance entre les rendements d'actifs financiers (intervenant en particulier dans les modèles VaR et de Markowitz) sont en général calculées sous l'hypothèse de stationnarité du processus des rendements, mais cette hypothèse n'est pas vérifiée en pratique, tout particulièrement à cause des variances des rendements. Cependant, le processus peut être considéré localement stationnaire. Cela signifie qu'il existe un intervalle passé, appelé intervalle d'homogénéité temporelle locale tel que la moyenne et la matrice de covariance entre les rendements sont presque constants sur cet intervalle. Ce chapitre a montré, à la fois d'un point de vue théorique et pratique, que des gains substantiels pouvaient être obtenus si on essaie de déterminer cet intervalle d'homogénéité temporelle locale pour trouver des estimations adaptatives de la moyenne et de la matrice de covariance. Cette étude a été conduite dans un cadre général, sans hypothèses paramétriques sur le processus des rendements.

Nous avons ensuite montré comment utiliser ces estimations pour définir des problèmes VaR et de Markowitz approchés dont la précision a été étudiée. Dans le cas fortement convexe, la qualité de l'approximation est contrôlée par l'inverse de la constante de convexité forte $\beta(Q_\tau)$ par rapport à la norme L_1.

Nous avons aussi appliqué l'algorithme adaptatif à des processus à sauts Gaussiens. Nous avons montré l'efficacité de cette méthode dans ce contexte et en quoi des simulations de Monte Carlo ou d'autres techniques de calibration des paramètres pourraient nous aider à détecter les ruptures importantes pour un processus Gaussien multidimensionnel. Cette méthode de détection de rupture pourrait en fait être appliquée dans d'autres contextes paramétriques multidimensionnels. Cependant, dans ce cas, le contexte paramétrique devrait permettre, moyennant une étude théorique spécifique et inspirée de celle effectuée dans ce chapitre dans un contexte non paramétrique, d'obtenir de meilleures valeurs pour les paramètres k_1 et k_2. Ces valeurs pourraient même être directement utilisées d'un point de vue pratique.

On a supposé, dans le cadre de l'étude théorique, que le paramètre σ était connu. D'un point de vue pratique nous devons l'estimer. Nous avons relativement confiance en cette estimation dans notre contexte financier étant donné que l'on a un grand nombre de données disponibles. L'obtention de résultats théoriques sur la précision de l'algorithme adaptatif lorsque σ est estimé constitue une perspective de poursuite des travaux effectués dans ce chapitre.

Par ailleurs, remarquons que l'algorithme adaptatif pourrait être utilisé de manière récursive de façon à déterminer la suite des dernières ruptures importantes dans un processus (ruptures dans la moyenne ou les variances).

Enfin, l'estimation empirique adaptative de la matrice de covariance n'est pas nécessairement inversible. L'objet du chapitre suivant est de proposer comment améliorer les estimations adaptatives des paramètres des modèles VaR et de Markowitz fournies dans ce chapitre. En particulier, nous chercherons à déterminer des calibrations stables et robustes (ces concepts sont précisés dans le chapitre suivant) de la matrice de covariance. D'un point de vue mathématique, cette étude passera par une analyse de sensibilité.

2.7 Preuves

Nous commençons par montrer un certain nombre de lemmes qui seront utilisés pour prouver les théorèmes de ce chapitre. Ces lemmes donnent, dans le cas stationnaire, des bornes non asymptotiques sur la précision des estimateurs empiriques de la moyenne et de la matrice de covariance ainsi que sur la précision d'autres estimateurs de la moyenne et de la matrice de covariance.
On travaille sur un processus r_i satisfaisant l'hypothèse 1 ou l'hypothèse 2 (introduites dans ce chapitre). Nous introduisons l'hypothèse supplémentaire suivante.

Hypothèse 3. Les ζ_i sont bornés : $\|\zeta_i\|_\infty \leq M$ et M est choisi suffisamment grand ($M \geq \sigma$).

Lemme 2.7.1 *Supposons que pour tout t, $\rho_t = \rho$, $Q_t = Q$, que l'hypothèse 2 soit vérifiée pour $\tau - 1 = N$, p=4 et fixons $\lambda > 0$. Soient $\delta_1 = \ln 2n + \lambda \ln N$ et $\delta_2 = \ln n(n+1) + \lambda \ln N$. Si de plus l'hypothèse 3 est vérifiée, les estimateurs $\hat{\rho}_N$, $\hat{Q}_N^b$ et $\hat{Q}_N$ définis par :*

$$\hat{\rho}_N = \frac{1}{N}\sum_{i=1}^{N} r_i, \quad \hat{Q}_N^b = \frac{1}{N}\sum_{i=1}^{N}(r_i - \rho)(r_i - \rho)^T, \quad \hat{Q}_N = \frac{1}{N}\sum_{i=1}^{N}(r_i - \hat{\rho}_N)(r_i - \hat{\rho}_N)^T. \quad (2.57)$$

satisfont :

$$E\|\widehat{\rho}_N - \rho\|_\infty \leq \sqrt{2\,e}\sigma\sqrt{\frac{\ln n}{N}}, \quad (2.58)$$

$$P\left(\|\hat{\rho}_N - \rho\|_\infty \geq \frac{M\delta_1}{3N} + \sqrt{\left(\frac{M\delta_1}{3N}\right)^2 + 2\sigma^2\frac{\delta_1}{N}}\right) \leq \frac{1}{N^\lambda}, \quad (2.59)$$

$$E\|\widehat{Q}_N^b - Q\|_\infty \leq 4\sqrt{e}\sigma^2\sqrt{\frac{\ln n}{N}}, \quad E\|\widehat{Q}_N - Q\|_\infty \leq 2\sqrt{e}\sigma^2\sqrt{\frac{\ln n}{N}}(2 + \sqrt{e}\sqrt{\frac{\ln n}{N}}), \quad (2.60)$$

$$P\left(\|\hat{Q}_N^b - Q\|_\infty \geq \frac{(M^2+\sigma^2)\delta_2}{3N} + \sqrt{\left(\frac{(M^2+\sigma^2)\delta_2}{3N}\right)^2 + 2\sigma^4\frac{\delta_2}{N}}\right) \leq \frac{1}{N^\lambda}, \quad (2.61)$$

$$P\left(\|\hat{Q}_N - Q\|_\infty \geq 2\left(\frac{(M^2+\sigma^2)\delta_2}{3N} + \sqrt{\left(\frac{(M^2+\sigma^2)\delta_2}{3N}\right)^2 + 2M^2\sigma^2\frac{\delta_2}{N}}\right)\right) \leq \frac{2}{N^\lambda}. \quad (2.62)$$

Remarquons que comme $n = N^\gamma$, si N est suffisamment grand, $\frac{\delta_1}{N} \leq 1$, $\frac{\delta_2}{N} \leq 1$ et les résultats ci-dessus se simplifient :

$$P\left(\|\hat{\rho}_N - \rho\|_\infty \geq M_1\sqrt{\frac{\ln 2n + \lambda \ln N}{N}}\right) \leq \frac{1}{N^\lambda}, \text{ où } M_1 = \frac{M}{3} + \sqrt{2\sigma^2 + \frac{M^2}{9}},$$

et

$$P\left(\|\hat{Q}_N - Q\|_\infty \geq M_2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \leq \frac{2}{N^\lambda} \text{ où } M_2 = 2\left(\frac{M^2+\sigma^2}{3} + \sqrt{\left(\frac{M^2+\sigma^2}{3}\right)^2 + 2M^2\sigma^2}\right).$$

Si par contre les ζ_i (et les r_i) ne sont pas bornés et sous l'hypothèse 2, on obtient les propriétés suivantes des estimateurs empiriques $\hat{\rho}_N$ et $\hat{Q}_N$ de ρ et Q.

Lemme 2.7.2 *Supposons que pour tout t, $\rho_t = \rho$, $Q_t = Q$. Fixons $\lambda > 0$ et supposons l'hypothèse 2 vérifiée avec $\tau - 1 = N$, et $E[\frac{p}{2}] > 2 + 2\lambda$. On définit*

$$K_1(\lambda) = 2 + \left(\frac{10\,p(\gamma+\lambda)}{3e(p-2-2\lambda)}\right)^{\frac{p}{2}}, \quad K_2'(\lambda) = 1 + 2^{E[\frac{p}{2}]}\left(\frac{28\,E[\frac{p}{2}](2\gamma+\lambda)}{3\,e(E[\frac{p}{2}]-2-2\lambda)}\right)^{\frac{E[\frac{p}{2}]}{2}}.$$

Si N est assez grand i.e $3N \geq 28(\ln n(n+1) + \lambda \ln N)$, alors les estimateurs $\hat{\rho}_N$, $\hat{Q}_N^b$ et $\hat{Q}_N$ définis par (2.57) satisfont :

$$P\left(\|\hat{\rho}_N - \rho\|_\infty \geq 2\sqrt{\frac{10}{3}}\sigma\sqrt{\frac{\ln n + \lambda \ln N}{N}}\right) \leq \frac{K_1(\lambda)}{N^\lambda}, \tag{2.63}$$

$$P\left(\|\hat{Q}_N^b - Q\|_\infty \geq (2^{E[\frac{p}{2}]}+1)\sqrt{\frac{28}{3}}\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \leq \frac{K_2'(\lambda)}{N^\lambda}, \tag{2.64}$$

$$P\left(\|\hat{Q}_N - Q\|_\infty \geq (\frac{40}{3} + (2^{E[\frac{p}{2}]}+1)\sqrt{\frac{28}{3}})\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \leq \frac{K_1(\lambda) + K_2'(\lambda)}{N^\lambda}. \tag{2.65}$$

Les inégalités (2.58) et (2.60) restent valides.

La condition $3N \geq 28(\ln n(n+1) + \lambda \ln N)$ dans le lemme ci-dessus peut être supprimée mais cela conduit à des seconds membres plus complexes. Dans ce qui suit, afin d'alléger les notations, les premiers membres des inégalités du type (2.59), (2.61) et (2.62) seront écrits sous la forme $P(\|\hat{\rho}_\tau - \rho\|_\infty \geq \eta_1)$, $P(\|\hat{Q}_\tau^b - Q\|_\infty \geq \eta_2')$ et $P(\|\hat{Q}_\tau - Q\|_\infty \geq \eta_2)$.

Preuve du lemme 2.7.1. Soit *svec* l'application de $\mathcal{S}_n$ dans $\mathbb{R}^{\frac{n(n+1)}{2}}$ qui associe à une matrice symétrique A sa représentation vectorielle $(A_{11}, \ldots, A_{1n}, A_{22}, \ldots, A_{2n}, \ldots, A_{nn})$. Nous commençons par prouver (2.58). Pour $n > 2$, soit $q = 2\ln n$ et soit $W(x) = \|x\|_q^2/2$. Alors pour $x \in \mathbb{R}^n$,

$$\|x\|_\infty \leq \|x\|_q \leq n^{1/q}\|x\|_\infty \leq \sqrt{\mathrm{e}}\|x\|_\infty, \tag{2.66}$$

et $\|x\|_\infty^2 \leq 2W(x)$. La preuve est basée sur le lemme suivant.

Lemme 2.7.3 *Soit $n > 2$, et soit $q = 2\ln n$. Alors la fonction*

$$W(z) = \frac{1}{2}\|z\|_q^2 : \mathbb{R}^n \to \mathbb{R}$$

satisfait pour tout $z, d \in \mathbb{R}^n$, la relation

$$W(z+d) \leq W(z) + d^T\nabla W(z) + c^*(n)\|d\|_\infty^2, \quad c^*(n) = e\ln n.$$

Preuve. Remarquons que pour $x \neq 0$, $\nabla_i W(x) = \mathrm{sign}(x_i)|x_i|^{q-1}\|x\|_q^{2-q}$ et

$$\nabla^2 W(x) = -(q-2)XX^T + (q-1)\frac{diag(|x_i|^{q-2})}{\|x\|_q^{q-2}},$$

avec

$$X = (X_1, ..., X_n)^T, \;\; X_i = \frac{|x_i|^{q-1}\mathrm{sign}(x_i)}{\|x\|_q^{q-1}}.$$

Alors pour tout $h \in \mathbb{R}^n$ tel que $\|h\|_q < \infty$, pour un certain $0 \leq \alpha \leq 1$ et $z = x + \alpha h$,

$$\begin{aligned} W(x+h) &= W(x) + h^T\nabla W(x) + \frac{1}{2}h^T\nabla^2 W(z)h \\ &\leq W(x) + h^T\nabla W(x) + \frac{q}{2}\|h\|_q^2 \leq W(x) + h^T\nabla W(x) + \frac{q\mathrm{e}}{2}\|h\|_\infty^2. \end{aligned}$$

□

Montrons maintenant (2.58). Pour $\zeta_t = r_t - \rho$, nous avons, en utilisant le lemme 2.7.3, pour $1 \leq k \leq N-1$:

$$W(\sum_{t=1}^{k+1} \zeta_t) \leq W(\sum_{t=1}^{k} \zeta_t) + (\zeta_{k+1})^T \nabla W(\sum_{t=1}^{k} \zeta_t) + c^*(n) \|\zeta_{k+1}\|_\infty^2,$$

ainsi, en prenant l'espérance, et comme les ζ_t sont indépendants et de moyennes nulles,

$$E\,[W(\sum_{t=1}^{k+1} \zeta_t)\,] \leq E\,[W(\sum_{t=1}^{k} \zeta_t)\,] + c^*(n)\,\sigma^2.$$

Il s'ensuit que

$$E\,[W(\sum_{t=1}^{N} \zeta_t)\,] \leq N\,c^*(n)\,\sigma^2,$$

et comme $W(z) \geq \frac{1}{2}\|z\|_\infty^2$,

$$E\,[\|\sum_{t=1}^{N} \zeta_t\|_\infty^2\,] \leq 2\,\mathrm{e}\,\sigma^2\,N\,\ln n, \tag{2.67}$$

ce qui achève la preuve de (2.58). Nous prouvons maintenant (2.59).

$$\begin{aligned} P(\|\hat{\rho}_N - \rho\|_\infty \geq \eta_1) &= P(\|\sum_{i=1}^{N} \zeta_i\|_\infty \geq N\eta_1), \\ &\leq n \max_{1\leq j\leq n} P(|\sum_{i=1}^{N} \zeta_i(j)| \geq N\eta_1). \end{aligned}$$

Pour tout $1 \leq j \leq n$ les variables $(\zeta_i(j))_{1\leq i\leq N}$ sont indépendantes, de moyennes nulles et vérifient

$$|\zeta_i(j)| \leq M \text{ et } E\zeta_i^2(j) \leq E\|\zeta_i\|_\infty^2 \leq \sigma^2.$$

Ainsi, en utilisant l'inégalité de Bernstein, on obtient :

$$P(\|\hat{\rho}_N - \rho\|_\infty \geq \eta_1) \leq 2\,n\,\exp\left(-\frac{N\eta_1^2}{2\sigma^2 + \frac{2}{3}M\eta_1}\right) = \frac{1}{N^\lambda},$$

ce qui est (2.59). Soit *svec* l'application de $\mathcal{S}_n$ dans $\mathbb{R}^{\frac{n(n+1)}{2}}$ qui associe à une matrice symétrique A sa représentation vectorielle $(A_{11}, \ldots, A_{1n}, A_{22}, \ldots, A_{2n}, \ldots, A_{nn})$. Nous avons :

$$\|\hat{Q}_N^b - Q\|_\infty = \|svec(\hat{Q}_N^b) - svec(Q)\|_\infty = \|\frac{1}{N}\sum_{t=1}^{N} \xi_t\|_\infty$$

avec

$$\xi_t(j + \frac{k(2n-k+1)}{2}) = \zeta_t(k+1)\,\zeta_t(j+k) - Q(k+1, j+k), \;\; 0 \leq k \leq n-1, \;\; 1 \leq j \leq n-k. \tag{2.68}$$

Puisque les $(\xi_t)_t$ sont indépendants de moyenne nulle et que $E\|\xi_t\|_\infty^2 \leq E(\|\zeta_t\|_\infty^2 + \|Q\|_\infty)^2 \leq E(\|\zeta_t\|_\infty^2 + \sigma^2)^2 \leq 4\sigma^4$, pour prouver la première inégalité dans (2.60), il suffit ensuite de suivre le schéma de la preuve de (2.58) en changeant n en $\frac{n(n+1)}{2}$ et σ^2 en $4\sigma^4$. Nous avons ensuite :

$$E\|\hat{Q}_N - Q\|_\infty \leq E\|\hat{Q}_N - \hat{Q}_N^b\|_\infty + E\|\hat{Q}_N^b - Q\|_\infty$$

avec

$$\|\hat{Q}_N^b - \hat{Q}_N\|_\infty = \|(\rho - \hat{\rho}_N)(\rho - \hat{\rho}_N)^T\|_\infty = \|\rho - \hat{\rho}_N\|_\infty^2, \tag{2.69}$$

ce qui donne la seconde inégalité dans (2.60) en utilisant (2.67) et le majorant de $E\|\hat{Q}_N^b - Q\|_\infty$ fourni précédemment. Nous prouvons maintenant (2.61). Pour tout $i \in [1, \frac{n(n+1)}{2}]$, les variables $(\xi_t(i))_{1\leq t\leq N}$ définies par (2.68) sont indépendantes de moyenne nulle et telles que

$$\|\xi_t\|_\infty \leq \|\zeta_t\|_\infty^2 + \|Q\|_\infty \leq M^2 + \sigma^2, \quad E\xi_t^2(i) \leq E\|\zeta_t\|_\infty^4 \leq \sigma^4.$$

(2.61) peut ainsi être prouvé comme (2.59) en changeant σ^2 en σ^4, M en $M^2 + \sigma^2$ et n en $\frac{n(n+1)}{2}$. La dernière inégalité (2.62) ne peut pas se montrer de la même façon en utilisant l'inégalité de Bernstein, puisque les statistiques associées aux estimations des covariances ne s'expriment pas directement sous la forme d'une somme de variables indépendantes. Cependant,

$$\begin{aligned} P(\|\hat{Q}_N - Q\|_\infty \geq \eta_2) &\leq P(\|\hat{Q}_N^b - \hat{Q}_N\|_\infty + \|\hat{Q}_N^b - Q\|_\infty \geq \eta_2) \\ &\leq P(\|\hat{Q}_N^b - \hat{Q}_N\|_\infty \geq \frac{\eta_2}{2}) + P(\|\hat{Q}_N^b - Q\|_\infty \geq \frac{\eta_2}{2}). \end{aligned} \tag{2.70}$$

D'après (2.69), $\|\hat{Q}_N^b - \hat{Q}_N\|_\infty \leq M\|\hat{\rho}_N - \rho\|_\infty$, et en utilisant (2.70), on obtient :

$$P(\|\hat{Q}_N - Q\|_\infty \geq \eta_2) \leq P(\|\hat{\rho}_N - \rho\|_\infty \geq \frac{\eta_2}{2M}) + P(\|\hat{Q}_N^b - Q\|_\infty \geq \frac{\eta_2}{2}).$$

Si $M \geq \sigma$, le choix de η_2 donne $\frac{\eta_2}{2M} \geq \eta_1$, $\frac{\eta_2}{2} \geq \eta_2'$ et

$$P(\|\hat{Q}_N - Q\|_\infty \geq \eta_2) \leq P(\|\hat{\rho}_N - \rho\|_\infty \geq \eta_1) + P(\|\hat{Q}_N^b - Q\|_\infty \geq \eta_2') \leq \frac{2}{N^\lambda}.$$

□

Preuve du lemme 2.7.2. Puisque $3N \geq 10(\ln n + \lambda \ln N)$ et $p \geq 2$,

$$\eta_1 = 2\sqrt{\frac{10}{3}}\sigma\sqrt{\frac{\ln n + \lambda \ln N}{N}} \geq \tilde{\eta}_1 + \frac{\sigma^p}{A^{p-1}},$$

avec

$$\tilde{\eta}_1 = \sigma\sqrt{\frac{10}{3}}\sqrt{\frac{\ln n + \lambda \ln N}{N}}, \quad \text{et} \quad A = \sigma\sqrt{\frac{3}{10}}\sqrt{\frac{N}{\ln n + \lambda \ln N}}.$$

On définit ensuite $\bar{\zeta}_t = \zeta_t 1(\|\zeta_t\|_\infty \leq A)$. Nous avons :

$$\begin{aligned} P\left(\|\frac{1}{N}\sum_{t=1}^N \zeta_t\|_\infty \geq \eta_1\right) &\leq P\left(\|\frac{1}{N}\sum_{t=1}^N \bar{\zeta}_t\|_\infty \geq \eta_1\right) + P\left(\max_{1\leq t\leq N}\|\zeta_t\|_\infty > A\right) \\ &= p_1 + p_2 \end{aligned} \tag{2.71}$$

Nous avons pour p_2 :

$$p_2 \leq \sum_{t=1}^{N} P(\|\zeta_t\|_\infty > A) \leq \sum_{t=1}^{N} \frac{E\|\zeta_t\|_\infty^p}{A^p} \leq \frac{N\sigma^p}{A^p} \leq \left(\frac{10(\gamma+\lambda)}{3}\right)^{\frac{p}{2}} \frac{(\ln N)^{\frac{p}{2}}}{N^{\frac{p}{2}-1}}.$$

Puisque $N \geq 2$, $\lambda > 0$, et $p > 2(1+\lambda)$:

$$p_2 \leq \left(\frac{10\,p(\gamma+\lambda)}{3e(p-2-2\lambda)}\right)^{\frac{p}{2}} \frac{1}{N^\lambda}. \tag{2.72}$$

Estimons maintenant p_1. Remarquons tout d'abord que

$$\|E\bar{\zeta}_t\|_\infty = \|E\zeta_t 1(\|\zeta_t\|_\infty > A)\|_\infty \leq \frac{\sigma^p}{A^{p-1}},$$

et

$$\begin{aligned} p_1 &\leq P\left(\|\frac{1}{N}\sum_{t=1}^{N}\bar{\zeta}_t - E\bar{\zeta}_t\|_\infty \geq \tilde{\eta}_1 + \frac{\sigma^p}{A^{p-1}} - \|E\bar{\zeta}_t\|_\infty\right) \\ &\leq P\left(\|\sum_{t=1}^{N}\bar{\zeta}_t - E\bar{\zeta}_t\|_\infty \geq N\tilde{\eta}_1\right). \end{aligned} \tag{2.73}$$

Remarquons ensuite que les variables aléatoires $\tilde{\xi}_t(i) = \bar{\zeta}_t(i) - E\bar{\zeta}_t(i)$ satisfont

$$E\tilde{\xi}_t(i) = 0, \;\; E\tilde{\xi}_t(i)^2 \leq \sigma^2 \;\text{ et }\; |\tilde{\xi}_t(i)| \leq 2\,A.$$

Puisque pour tout $i \in [1,n]$ les variables $(\tilde{\xi}_t(i))_{1\leq t\leq N}$ sont indépendantes, en utilisant l'inégalité de Bernstein, nous avons

$$P\left(\|\sum_{t=1}^{N}\tilde{\xi}_t\|_\infty \geq N\tilde{\eta}_1\right) \leq 2\,n\,\exp\left(-\frac{1}{2}\,\frac{N\tilde{\eta}_1^2}{\sigma^2 + \frac{2}{3}A\tilde{\eta}_1}\right) = \frac{2}{N^\lambda}. \tag{2.74}$$

Il suffit alors d'injecter (2.74) dans (2.73) et (2.73) et (2.72) dans (2.71) pour achever la preuve de (2.63). Nous prouvons maintenant (2.64). Rappelons que

$$\|\hat{Q}_N^b - Q\|_\infty = \|svec(\hat{Q}_N^b) - svec(Q)\|_\infty = \|\frac{1}{N}\sum_{t=1}^{N}\xi_t\|_\infty,$$

où ξ_t est défini dans (2.68). Pour tout $i \in [1, \frac{n(n+1)}{2}]$, les variables $(\xi_t(i))_{1\leq t\leq N}$ sont indépendantes et pour $1 \leq k \leq E[\frac{p}{2}]$,

$$E\|\xi_t\|_\infty^k \leq E\big(\|\zeta_t\|_\infty^2 + \sigma^2\big)^k \leq \sum_{i=0}^{k} C_k^i\,\sigma^{2i}\sigma^{2(k-i)} = (2\sigma^2)^k.$$

Il suffit ensuite de suivre le schéma de la preuve (2.63) pour prouver (2.64). Enfin, pour montrer (2.65), on remarque que :

$$P(\|\hat{Q}_N - Q\|_\infty \geq \eta_2) \leq P(\|\hat{\rho}_N - \rho\|_\infty \geq \eta_1) + P(\|\hat{Q}_N^b - Q\|_\infty \geq \eta_2'). \tag{2.75}$$

□

Lemme 2.7.4 *Soient r_t, $t = 1, \ldots, \tau - 1 = N$ les réalisations d'un vecteur aléatoire de dimension n vérifiant l'hypothèse 1 où $\rho_t = \rho$ et $Q_t = Q$ sont constants. Fixons $\lambda > 0$ et $K_0 = \sigma\Big(\frac{N}{\ln n(n+1)+\lambda \ln N}\Big)^{\frac{1}{4}}$. On suppose que N est suffisamment grand pour avoir $\ln(n(n+1)) + \lambda \ln N \leq N$. On définit les statistiques suivantes :*

$$\hat{\rho}_\tau = \frac{1}{N}\sum_{t=1}^{N} \alpha_t, \quad \hat{Q}_\tau^b = \frac{1}{N}\sum_{t=1}^{N} (\alpha_t - \rho)(\alpha_t - \rho)^T, \quad \hat{Q}_\tau = \frac{1}{N}\sum_{t=1}^{N} (\alpha_t - \hat{\rho}_\tau)(\alpha_t - \hat{\rho}_\tau)^T,$$

avec $\alpha_t = r_t\, 1(\|r_t\|_\infty \leq K_0)$. Alors il existe des constantes $m_1 = \frac{7}{3} + \sqrt{2}$, $m_2 = \frac{25}{3} + 4\sqrt{2}$, $m_3 = m_1^2 + m_2$ telles que :

$$\|E\widehat{\rho}_\tau - \rho\|_\infty \leq \sigma(\frac{\ln n(n+1) + \lambda \ln N}{N})^{\frac{3}{4}}, \tag{2.76}$$

$$P\left(\|\hat{\rho}_\tau - \rho\|_\infty \geq m_1\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \leq \frac{1}{N^\lambda}, \tag{2.77}$$

$$\|E\widehat{Q}_\tau^b - Q\|_\infty \leq 3\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}, \tag{2.78}$$

$$\|E\widehat{Q}_\tau - Q\|_\infty \leq 4\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}} + \frac{\sigma^2}{N}, \tag{2.79}$$

$$P\left(\|\hat{Q}_\tau^b - Q\|_\infty \geq m_2\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \leq \frac{1}{N^\lambda}, \tag{2.80}$$

$$P\left(\|\hat{Q}_\tau - Q\|_\infty \geq m_3\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \leq \frac{2}{N^\lambda}. \tag{2.81}$$

Lemme 2.7.5 *Sous les hypothèses et notations du lemme 2.7.4, on définit les statistiques suivantes :*

$$\hat{\rho}_\tau = \frac{1}{N}\sum_{t=1}^{N} \alpha_t, \quad \hat{Q}_\tau^b = \frac{1}{N}\sum_{t=1}^{N} (\alpha_t - \rho)(\alpha_t - \rho)^T, \quad \hat{Q}_\tau = \frac{1}{N}\sum_{t=1}^{N} (\alpha_t - \hat{\rho}_\tau)(\alpha_t - \hat{\rho}_\tau)^T$$

où pour $i = 1, \ldots, n$:

$$\alpha_k(i) = r_k(i)\, 1(|r_k(i)| \leq K_0) + K_0\frac{r_k(i)}{|r_k(i)|}\, 1(|r_k(i)| > K_0). \tag{2.82}$$

Les estimateurs $\hat{\rho}_\tau$, $\hat{Q}_\tau^b$ et $\hat{Q}_\tau$ satisfont (2.76),(2.77),(2.80),(2.81) *(avec $m_1 = \frac{7}{3} + \sqrt{2}$, $m_2 = \frac{31}{3} + 4\sqrt{2}$, $m_3 = m_1^2 + m_2$) et :*

$$\|E\hat{Q}_\tau^b - Q\|_\infty \leq 5\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}, \quad \|E\hat{Q}_\tau - Q\|_\infty \leq 7\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}} + \frac{\sigma^2}{N}. \tag{2.83}$$

Preuve du lemme 2.7.4. Remarquons tout d'abord que pour tout $k \in [1, n]$:

$$|E\hat{\rho}_\tau(k) - \rho(k)| = |E\alpha_1(k) - Er_1(k)| = |Er_1(k)1(\|r_1\|_\infty > K_0)| \leq E\|r_1\|_\infty 1(\|r_1\|_\infty > K_0) \leq \frac{\sigma^4}{K_0^3}.$$

Ceci prouve (2.76). Si $\delta_1 = \ln 2n + \lambda \ln N$ et $\delta_2 = \ln n(n+1) + \lambda \ln N$, nous allons montrer que :

$$P\left(\|\hat{\rho}_\tau - \rho\|_\infty \geq \eta_1 = \frac{\sigma^4}{K_0^3} + \eta_1'\right) \leq \frac{1}{N^\lambda}, \tag{2.84}$$

avec

$$\eta_1' = \sqrt{2}\sigma\sqrt{\frac{\delta_1}{N}} + \frac{4}{3}\sigma\frac{\delta_1}{\delta_2^{\frac{1}{4}}N^{\frac{3}{4}}} = \sqrt{2}\sigma\sqrt{\frac{\delta_1}{N}} + \frac{4}{3}\frac{K_0\delta_1}{N},$$

ce qui prouvera (2.77) avec $m_1 = \frac{7}{3} + \sqrt{2}$. Nous avons :

$$\begin{aligned} P(\|\hat{\rho}_\tau - \rho\|_\infty \geq \eta_1) &\leq n \max_{1\leq k\leq n} P(|\hat{\rho}_\tau(k) - E\hat{\rho}_\tau(k)| \geq \eta_1 - |E\hat{\rho}_\tau(k) - \rho(k)|) \\ &\leq n \max_{1\leq k\leq n} P\left(|\sum_{i=1}^{N} \alpha_i(k) - E\alpha_i(k)| \geq N\eta_1'\right). \end{aligned}$$

Remarquons que

$$\eta_1' \geq \eta_1'' = \frac{2}{3}\frac{K_0\delta_1}{N} + \sqrt{(\frac{2}{3}\frac{K_0\delta_1}{N})^2 + 2\sigma^2\frac{\delta_1}{N}}.$$

Si $X_k^i = \alpha_i(k) - E\alpha_i(k)$, pour tout k fixé, les variables $(X_k^i)_{1\leq i\leq N}$ sont indépendantes, centrées et telles que

$$E(X_k^i)^2 \leq E\alpha_i(k)^2 \leq \sigma^2 \text{ et } |X_k^i| \leq 2\,K_0.$$

Ainsi, en utilisant l'inégalité de Bernstein, on obtient (2.77) :

$$P(\|\hat{\rho}_\tau - \rho\|_\infty \geq \eta_1) \leq 2n\exp\left(-\frac{1}{2}\frac{N\eta_1''^2}{\sigma^2 + \frac{2}{3}K_0\eta_1''}\right) = \frac{1}{N^\lambda}.$$

Ensuite, pour tout $1 \leq j,k \leq n$,

$$\begin{aligned} |E\widehat{Q}_\tau^b(j,k) - Q(j,k)| &\leq E|r_1(j)r_1(k) - \alpha_1(j)\alpha_1(k)| + |\rho(k)||E\alpha_1(j) - \rho(j)| + |\rho(j)||E\alpha_1(k) - \rho(k)| \quad (2.85) \\ &\leq E|r_1(j)r_1(k)1(\|r_1\|_\infty > K_0)| + |\rho(k)||Er_1(j)1(\|r_1\|_\infty > K_0)| + |\rho(j)|\,|Er_1(k)1(\|r_1\|_\infty > K_0)| \\ &\leq \frac{\sigma^4}{K_0^2} + 2\frac{\sigma^5}{K_0^3}; \quad (2.86) \end{aligned}$$

ce qui montre (2.78). Nous montrons maintenant que :

$$P(\|\hat{Q}_\tau^b - Q\|_\infty \geq \eta_2' = \frac{\sigma^4}{K_0^2} + 2\frac{\sigma^5}{K_0^3} + \eta_2'') \leq \frac{1}{N^\lambda}, \quad (2.87)$$

où

$$\eta_2''(N,\lambda) = \frac{4}{3}\frac{(K_0+\sigma)^2\delta_2}{N} + 4\sqrt{2}\sigma^2\sqrt{\frac{\delta_2}{N}}; \quad (2.88)$$

ce qui prouvera (2.80). En raisonnant comme ci-dessus,

$$\begin{aligned} P(\|\hat{Q}_\tau^b - Q\|_\infty \geq \eta_2') &\leq \frac{n(n+1)}{2} \max_{1\leq j\leq k\leq n} P(|\hat{Q}_\tau^b(j,k) - E\hat{Q}_\tau^b(j,k)| \geq \eta_2''), \\ &\leq \frac{n(n+1)}{2} \max_{1\leq j\leq k\leq n} P(|\sum_{i=1}^{N} A_i^{j,k} - EA_i^{j,k}| \geq N\tilde{\eta}) \end{aligned}$$

avec

$$\tilde{\eta} = \frac{2}{3}\frac{(K_0+\sigma)^2\delta_2}{N} + \sqrt{(\frac{2}{3}\frac{(K_0+\sigma)^2\delta_2}{N})^2 + 32\sigma^4\frac{\delta_2}{N}} \leq \eta_2'',$$

et $A_i^{j,k} = (\alpha_i(j)-\rho(j))(\alpha_i(k)-\rho(k))$. Si $X_i^{j,k} = A_i^{j,k} - EA_i^{j,k}$, les $(X_i^{j,k})_i$ sont indépendantes, centrées et vérifient $|X_i^{j,k}| \leq 2(K_0+\sigma)^2$, $E(X_i^{j,k})^2 \leq E(\|\alpha_i\|_\infty + \sigma)^4 \leq 16\sigma^4$. On applique alors l'inégalité de Bernstein pour terminer la preuve de (2.80). Si ρ n'est pas connu, $E\hat{Q}_\tau(j,k) = \frac{N-1}{N}(E\alpha_1(k)\alpha_1(j) - E\alpha_1(k)E\alpha_1(j))$ et donc :

$$\begin{aligned} |E\hat{Q}_\tau(j,k) - Q(j,k)| &\leq \frac{N-1}{N}|E\alpha_1(k)\alpha_1(j) - E\alpha_1(k)E\alpha_1(j) - Q(j,k)| + \frac{|Q(j,k)|}{N} \\ &\leq |E\alpha_1(j)\alpha_1(k) - r_1(j)r_1(k)| + |E\alpha_1(j)E\alpha_1(k) - Er_1(j)Er_1(k)| + \frac{\sigma^2}{N}. \end{aligned} \tag{2.89}$$

Remarquons que $|E\left(\alpha_1(j)\alpha_1(k) - r_1(j)r_1(k)\right)| \leq \frac{\sigma^4}{K_0^2}$ et si $\delta^- = 1(\|r_1\|_\infty \leq K_0)$ et $\delta^+ = 1(\|r_1\|_\infty > K_0)$,

$$|E\alpha_1(j)E\alpha_1(k) - Er_1(j)Er_1(k)| \leq E|r_1(j)|\delta^- E|r_1(k)|\delta^+ + E|r_1(j)|\delta^+ E|r_1(k)|\delta^- + E|r_1(j)|\delta^+ E|r_1(k)|\delta^+ \leq 3\frac{\sigma^4}{K_0^2},$$

ce qui montre (2.79). Ensuite, comme

$$\|\hat{Q}_\tau^b - \hat{Q}_\tau\|_\infty = \|(\rho - \hat{\rho}_\tau)(\rho - \hat{\rho}_\tau)^T\|_\infty = \|\rho - \hat{\rho}_\tau\|_\infty^2, \tag{2.90}$$

nous avons :

$$P(\|\hat{Q}_\tau - Q\|_\infty \geq \eta_2) \leq P(\|\hat{\rho}_\tau - \rho\|_\infty^2 \geq m_1^2\sigma^2\sqrt{\frac{\delta_2}{N}}) + P(\|\hat{Q}_\tau^b - Q\|_\infty \geq m_2\sigma^2\sqrt{\frac{\delta_2}{N}}) \tag{2.91}$$

Comme $\sqrt{\frac{\delta_2}{N}} \geq \frac{\delta_2}{N}$, (2.77) permet de majorer le premier terme du membre droit par $\frac{1}{N^\lambda}$ et le deuxième terme par $\frac{1}{N^\lambda}$ également (en utilisant (2.80)).

□

Preuve du lemme 2.7.5. Si l'on remarque que $|\alpha_i(k)| \leq |r_i(k)|$ on voit (en suivant la preuve du lemme 2.7.4) que (2.76),(2.77),(2.80) et (2.81) restent valables pour le lemme 2.7.5 à condition que les constantes m_1, m_2 et m_3 soient mises à jour. Par exemple, pour $1 \leq k \leq n$:

$$\begin{aligned} |E\hat{\rho}_\tau(k) - \rho(k)| &= |E[K_0\frac{r_1(k)}{|r_1(k)|} - r_1(k)]1(|r_1(k)| > K_0)| \\ &\leq E|r_1(k)|1(|r_1(k)| > K_0) \leq \frac{\sigma^4}{K_0^3}. \end{aligned}$$

Majorons $\|E\hat{Q}_\tau^b - Q\|_\infty$ et $\|E\hat{Q}_\tau - Q\|_\infty$. Notons k^+ (k^-) les quantités $1(|r_1(k)| > K_0)$ $(1(|r_1(k)| \leq K_0))$ et j^+ (j^-), $1(|r_1(j)| > K_0)$ $(1(|r_1(j)| \leq K_0))$. Tout d'abord, remarquons que (2.85) est vérifiée. Ensuite, pour tout $1 \leq j,k \leq n$:

$$|\rho(k)||Er_1(j) - \alpha_1(j)| \leq \sigma E|r_1(j)|j^+ \leq \frac{\sigma^4}{K_0^2};$$

et

$$\begin{aligned} E|\alpha_1(j)\alpha_1(k) - r_1(j)r_1(k)| &\leq E|r_1(k)||r_1(j)||\frac{K_0}{|r_1(j)|} - 1|k^-j^+ + E|r_1(k)||r_1(j)||\frac{K_0}{|r_1(k)|} - 1|k^+j^- \\ &\quad + E|r_1(j)r_1(k)||\frac{K_0^2}{|r_1(j)r_1(k)|} - 1|k^+j^+ \leq 3\frac{\sigma^4}{K_0^2}. \end{aligned}$$

La première inégalité de (2.83) s'ensuit. Finalement, (2.89) est vérifiée et

$$\begin{aligned}|E\alpha_1(j)E\alpha_1(k) - Er_1(j)Er_1(k)| &\leq E|r_1(j)|j^-E|\frac{K_0}{|r_1(k)|}-1||r_1(k)|k^+ + E|r_1(k)|k^-E|\frac{K_0}{|r_1(j)|}-1||r_1(j)|j^+ \\ &+K_0^2|E\frac{r_1(j)}{|r_1(j)|}j^+E\frac{r_1(k)}{|r_1(k)|}k^+| + |Er_1(j)j^+Er_1(k)k^+| \leq 4\frac{\sigma^4}{K_0^2}.\end{aligned}$$

□

Le lemme suivant sera aussi utile par la suite. Même si $V_I(\rho)$ et $V_I(Q)$ sont inconnus, ce lemme permet d'avoir un majorant (non asymptotique) de ces quantités : $V_I(\rho) \leq 2\sqrt{2e}\sigma\sqrt{\frac{\ln n}{|I|}}$ et $V_I(Q) \leq k_Q\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$, pour une certaine constante k_Q. Si I est un intervalle d'homogénéité temporelle locale, ce lemme fournira aussi un majorant pour $\Delta_I(\rho)$ et $\Delta_I(Q)$. Dans ce qui suit, $\bar{\rho}_I$ et $\bar{Q}_I$, représenteront respectivement les quantités $\frac{1}{|I|}\sum_{t\in I}\rho_t$ et $\frac{1}{|I|}\sum_{t\in I}Q_t$.

Lemme 2.7.6 *Soit I un intervalle non vide. Soit $\lambda > 0$ tel que $\ln n(n+1) + \lambda\ln|I| \leq |I|$ et soient $\hat{\rho}_I$ et $\hat{Q}_I$ les estimateurs définis par* (2.9). *Il existe alors une constante $k_Q = 2\left(1+4\sqrt{e}(\sqrt{2}+2\sqrt{e}+2)\right)$ telle que*

$$E\|\hat{\rho}_I - E\hat{\rho}_I\|_\infty^2 \leq 8e\sigma^2\left(\frac{\ln n}{|I|}\right), \tag{2.92}$$

$$E\|\hat{Q}_I - E\hat{Q}_I\|_\infty \leq k_Q\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}. \tag{2.93}$$

Preuve du lemme 2.7.6. Remarquons que les variables $(X_k) = (\alpha_k - E\alpha_k)$, pour $k \in I$, sont des variables aléatoires indépendantes centrées et vérifiant :

$$E\|X_k\|_\infty^2 \leq E(\|\alpha_k\|_\infty + E\|\alpha_k\|_\infty)^2 \leq E(\|r_k\|_\infty + \sigma)^2 \leq 4\sigma^2.$$

Il suffit ensuite de suivre la preuve de (2.58), lemme 2.7.1 pour prouver (2.92). Maintenant, soit $\tilde{Q}_I^b = \frac{1}{|I|}\sum_{t\in I}(\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^T$. Nous montrons tout d'abord que :

$$E\|\tilde{Q}_I^b - E\tilde{Q}_I^b\|_\infty \leq 16\sqrt{e}\sigma^2\sqrt{\frac{\ln n}{|I|}}. \tag{2.94}$$

En effet, $E\|\tilde{Q}_I^b - E\tilde{Q}_I^b\|_\infty = \frac{1}{|I|}\ E\|\sum_{t\in I}\tilde{\zeta}_t\|_\infty$, avec $\tilde{\zeta}_t = (\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^T - E(\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^T$. Les variables $(\tilde{\zeta}_t)_{t\in I}$ sont indépendantes centrées et $E\|\tilde{\zeta}_t\|_\infty^2 \leq E((\|r_t\|_\infty + \sigma)^2 + 4\sigma^2)^2 \leq 64\sigma^4$. En suivant à nouveau la preuve de (2.58), lemme 2.7.1, on obtient (2.94). Maintenant,

$$\begin{aligned}E\|\hat{Q}_I - E\hat{Q}_I\|_\infty &\leq E\|\hat{Q}_I - \tilde{Q}_I^b\|_\infty + E\|\tilde{Q}_I^b - E\tilde{Q}_I^b\|_\infty + \|E\tilde{Q}_I^b - E\hat{Q}_I\|_\infty \\ &\leq 2E\|\hat{Q}_I - \tilde{Q}_I^b\|_\infty + 16\sqrt{e}\sigma^2\sqrt{\frac{\ln n}{|I|}}.\end{aligned} \tag{2.95}$$

De plus, $\|\hat{Q}_I - \tilde{Q}_I^b\|_\infty = \|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2$ et

$$E\|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2 \leq E(\|\hat{\rho}_I - E\hat{\rho}_I\|_\infty + \|E\hat{\rho}_I - \bar{\rho}_I\|_\infty)^2 \leq E(\|\hat{\rho}_I - E\hat{\rho}_I\|_\infty + \sigma\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}})^2. \tag{2.96}$$

On utilise ensuite (2.92), (2.95) et (2.96) pour prouver (2.93).

□

Preuve du théorème 2.3.1

Supposons que r_t satisfasse l'hypothèse 1. Nous avons :

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq \|\hat{\rho}_I - \bar{\rho}_I\|_\infty + \|\bar{\rho}_I - \bar{\rho}_J\|_\infty + \|\bar{\rho}_J - \hat{\rho}_J\|_\infty. \tag{2.97}$$

Ensuite $\|\bar{\rho}_I - \bar{\rho}_J\|_\infty \leq \|\bar{\rho}_I - \rho_\tau\|_\infty + \|\rho_\tau - \bar{\rho}_J\|_\infty$ et d'après l'inégalité de Cauchy-Schwartz $\|\bar{\rho}_I - \rho_\tau\|_\infty \leq \Delta_I(\rho)$ et $\|\rho_\tau - \bar{\rho}_J\|_\infty \leq \Delta_J(\rho)$. Si I est un intervalle d'homogénéité temporelle locale, alors $\Delta_I(\rho) \leq DV_I(\rho)$ et $\Delta_J(\rho) \leq DV_J(\rho)$. Grâce au lemme 2.7.6, on obtient alors :

$$\|\bar{\rho}_I - \bar{\rho}_J\|_\infty \leq 2\sqrt{2e}D\sigma\left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}}\right). \tag{2.98}$$

On peut ensuite facilement adapter la preuve de (2.77) pour montrer que pour tout intervalle I non vide et pour tout $\lambda > 0$:

$$P(\|\hat{\rho}_I - \bar{\rho}_I\|_\infty \geq \eta_1(|I|, \lambda)) \leq \frac{1}{|I|^\lambda}, \tag{2.99}$$

avec

$$\eta_1(|I|, \lambda) = \frac{\sigma^4}{K_0^3(I)} + \sqrt{2}\sigma\sqrt{\frac{\ln 2n + \lambda \ln |I|}{|I|}} + \frac{4}{3}\frac{K_0(I)(\ln 2n + \lambda \ln |I|)}{|I|}, \tag{2.100}$$

où $K_0(I) = \sigma\left(\frac{|I|}{\ln n(n+1) + \lambda \ln |I|}\right)^{\frac{1}{4}}$. Remarquons que si $\ln n(n+1) + \lambda \ln |I| \leq |I|$, alors (2.99) est vérifié avec $\eta_1(|I|, \lambda) = m_1\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}}$. Si $\ln n(n+1) + \lambda \ln |I| \leq |I|$ et $\ln n(n+1) + \lambda \ln |J| \leq |J|$ alors on montre (2.24) en utilisant (2.97), (2.98) et (2.99). Nous avons maintenant besoin du lemme suivant pour prouver (2.25).

Lemme 2.7.7 *Soit I un intervalle d'homogénéité temporelle locale non vide, alors pour tout $\lambda > 0$,*

$$P\left(\|\hat{Q}_I - \bar{Q}_I\|_\infty \geq \eta(|I|, \lambda) = \eta_1^2(|I|, \lambda) + \eta_2''(|I|, \lambda) + (5 + 4\sqrt{2e}D)\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}}\right) \leq \frac{2}{|I|^\lambda}, \tag{2.101}$$

où les fonctions η_1 et η_2'' sont définies dans (2.100) *et* (2.88).

Preuve du lemme 2.7.7. Remarquons tout d'abord que :

$$\|\hat{Q}_I - \bar{Q}_I\|_\infty \leq \|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2 + \frac{1}{|I|}\|\sum_{t \in I}(\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^T - Q_t\|_\infty.$$

Nous devons alors majorer : $\|\Delta\|_\infty = \frac{1}{|I|}\|\sum_{t\in I} E(\alpha_t - \bar\rho_I)(\alpha_t - \bar\rho_I)^T - Q_t\|_\infty$. Nous avons, pour tout $1 \le j,k \le n$:

$$\begin{aligned}
|\Delta(j,k)| &\le \frac{1}{|I|}\sum_{t\in I}|E\alpha_t(j)\alpha_t(k) - r_t(j)r_t(k)| + \frac{|\bar\rho_I(k)|}{|I|}\sum_{t\in I}|\rho_t(j) - E\alpha_t(j)| \\
&\quad + \frac{1}{|I|}\sum_{t\in I}|\rho_t(j)||\rho_t(k) - E\alpha_t(k)| + \frac{1}{|I|}\sum_{t\in I}|E\alpha_t(k)||\rho_t(j) - \bar\rho_I(j)| \\
&\le \frac{5\sigma^4}{K_0^2(I)} + \frac{\sigma}{|I|}\left(\sum_{t\in I}|(\rho_t - \rho_\tau)(j)| + |(\rho_\tau - \bar\rho_I)(j)|\right) \le \frac{5\sigma^4}{K_0^2(I)} + 2\sigma\Delta_I(\rho) \\
&\le 5\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda\ln|I|}{|I|}} + 4\sqrt{2e}D\sigma^2\sqrt{\frac{\ln n}{|I|}} \\
&\le (5 + 4\sqrt{2e}D)\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda\ln|I|}{|I|}}.
\end{aligned}$$

On suit ensuite la preuve de (2.80) pour conclure.

□

Remarque 2.7.1 *Lorsque* $\ln n(n+1) + \lambda\ln|I| \le |I|$ *et si* $k'_Q = \frac{160}{9} + \frac{26}{3}\sqrt{2} + 4\sqrt{2e}D$, *alors le lemme précédent est vérifié avec*

$$\eta(|I|,\lambda) = k'_Q\,\sigma^2\,\sqrt{\frac{\ln n(n+1) + \lambda\ln|I|}{|I|}}. \tag{2.102}$$

Comme précédemment nous avons :

$$\|\hat Q_I - \hat Q_J\|_\infty \le \|\hat Q_I - \bar Q_I\|_\infty + \|\bar Q_I - \bar Q_J\|_\infty + \|\bar Q_J - \hat Q_J\|_\infty. \tag{2.103}$$

En utilisant l'inégalité de Cauchy-Schwartz, le lemme 2.7.6 et puisque I est un intervalle d'homogénéité temporelle locale, nous obtenons :

$$\begin{aligned}
\|\bar Q_I - \bar Q_J\|_\infty &\le D(V_I(Q) + V_J(Q)) \\
&\le k_Q D\sigma^2\left(\sqrt{\frac{\ln n(n+1) + \lambda\ln|J|}{|J|}} + \sqrt{\frac{\ln n(n+1) + \lambda\ln|I|}{|I|}}\right)
\end{aligned} \tag{2.104}$$

On conclut en utilisant (2.103), (2.104) et la remarque 2.7.1.

□

On peut remarquer que si au lieu d'utiliser les estimateurs définis par (2.9), on utilise les estimteurs empiriques (revenant à remplacer α_t par r_t) alors on peut obtenir des expressions plus fines de $\gamma_\rho(|I|,|J|,\lambda)$ et $\gamma_Q(|I|,|J|,\lambda)$. Pour cela, il suffit d'écrire pour la moyenne :

$$\|\hat\rho_I - \hat\rho_J\|_\infty \le (\frac{1}{|J|} - \frac{1}{|I|})\|\sum_{t\in J} r_t - \rho_t\|_\infty + \|\bar\rho_I - \bar\rho_J\|_\infty + \frac{1}{|I|}\|\sum_{t\in I\setminus J} r_t - \rho_t\|_\infty;$$

et pour la matrice de covariance :

$$\|\hat Q_I - \hat Q_J\|_\infty \le \|\bar Q_I - \bar Q_J\|_\infty + A(I,J) + B(I,J) + C(I,J) + D(I,J);$$

où

$$A(I,J) = (\frac{1}{|J|} - \frac{1}{|I|})\|\sum_{t\in J}(r_t - \hat{\rho}_J)(r_t - \hat{\rho}_J)^T - Q_t\|_\infty, \quad B(I,J) = \frac{|I|-|J|}{|I|}\|\hat{Q}_{I\setminus J} - \bar{Q}_{I\setminus J}\|_\infty,$$

$$C(I,J) = \frac{|J|}{|I|}\|\hat{\rho}_I - \hat{\rho}_J\|_\infty^2 \text{ et } D(I,J) = \frac{|I|-|J|}{|I|}\|\hat{\rho}_I - \hat{\rho}_{I\setminus J}\|_\infty^2.$$

On conclut ensuite en suivant la preuve du théorème 2.3.1.

Preuve du théorème 2.3.2

Tout d'abord,

$$\|\hat{\rho}_{\mathbb{I}} - \rho_\tau\|_\infty \leq \frac{1}{|\mathbb{I}|}\sum_{t\in\mathbb{I}}\|\alpha_t - \rho_t\|_\infty + \frac{1}{|\mathbb{I}|}\sum_{t\in\mathbb{I}}\|\rho_t - \rho_\tau\|_\infty$$

Ensuite, en utilisant l'inégalité de Cauchy Schwartz, la définition de $\mathbb{I}$ et le lemme 2.7.6 :

$$\frac{1}{|\mathbb{I}|}\sum_{t\in\mathbb{I}}\|\rho_t - \rho_\tau\|_\infty \leq \Delta_{\mathbb{I}}(\rho) \leq DV_{\mathbb{I}}(\rho) \leq 2\sqrt{2e}D\sigma\sqrt{\frac{\ln n}{|\mathbb{I}|}},$$

et (2.28) en résulte. De même, pour la matrice de covariance :

$$\|\hat{Q}_{\mathbb{I}} - Q_\tau\|_\infty \leq \|\hat{Q}_{\mathbb{I}} - \bar{Q}_{\mathbb{I}}\|_\infty + \Delta_{\mathbb{I}}(Q),$$

avec $\Delta_{\mathbb{I}}(Q) \leq DV_{\mathbb{I}}(Q)$. On conclut en utilisant les lemmes 2.7.6 et 2.7.7.

□

Preuve du théorème 2.3.3

On montre que l'union de événement :

$$\left\{\|\hat{\rho}_{\hat{I}} - \rho_\tau\|_\infty > 3\eta_1(|\mathbb{I}|,\lambda) + 4\sqrt{2e}D\sigma\sqrt{\frac{\ln n}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}(\rho)\right\}$$

et de l'événement :

$$\left\{\|\hat{Q}_{\hat{I}} - Q_\tau\|_\infty > 3\eta(|\mathbb{I}|,\lambda) + 2k_Q D\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda\ln|\mathbb{I}|}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}(Q)\right\}$$

implique l'événement :

$$\bigcup_{J\in\mathcal{I}_+(\mathbb{I})}\left\{\|\hat{\rho}_J - \bar{\rho}_J\|_\infty > \eta_1(|J|,\lambda) \cup \|\hat{Q}_J - \bar{Q}_J\|_\infty > \eta(|J|,\lambda)\right\},$$

où $\eta_1(|I|,\lambda) = (\frac{7}{3} + \sqrt{2})\sigma\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$ et $\eta(|I|,\lambda) = k'_Q\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$, ce qui prouvera le théorème. Supposons que pour tout J dans $\mathcal{I}_+(\mathbb{I})$, $\|\hat{\rho}_J - \bar{\rho}_J\|_\infty \leq \eta_1(|J|,\lambda)$ et $\|\hat{Q}_J - \bar{Q}_J\|_\infty \leq \eta(|J|,\lambda)$. Nous cherchons à montrer que $\|\hat{\rho}_{\hat{I}} - \rho_\tau\|_\infty \leq 3\eta_1(|\mathbb{I}|,\lambda) + 4\sqrt{2e}D\sigma\sqrt{\frac{\ln n}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}(\rho)$ et $\|\hat{Q}_{\hat{I}} - Q_\tau\|_\infty \leq 3\eta(|\mathbb{I}|,\lambda) + 2k_Q D\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|\mathbb{I}|}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}(Q)$.

Tout d'abord, remarquons que $\mathbb{I}$ n'est pas rejeté. En effet, pour tout $I \in \mathcal{I}$ tel que $I \subseteq \mathbb{I}$ et pour tout $J \in \mathcal{I}(I)$:

$$\begin{aligned}\|\hat{\rho}_I - \hat{\rho}_J\|_\infty &\leq \|\hat{\rho}_I - \bar{\rho}_I\|_\infty + \|\bar{\rho}_I - \rho_\tau\|_\infty + \|\rho_\tau - \bar{\rho}_J\|_\infty + \|\bar{\rho}_J - \hat{\rho}_J\|_\infty \\ &\leq \eta_1(|I|,\lambda) + \eta_1(|J|,\lambda) + \Delta_I(\rho) + \Delta_J(\rho).\end{aligned}$$

Par ailleurs, par définition de $\mathbb{I}$, $\Delta_I(\rho) \leq DV_I(\rho)$, $\Delta_J(\rho) \leq DV_J(\rho)$ et en utilisant le lemme 2.7.6 on obtient :

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq \eta_1(|I|,\lambda) + \eta_1(|J|,\lambda) + 2\sqrt{2e}D\sigma\left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}}\right).$$

De même on montre que :

$$\|\hat{Q}_I - \hat{Q}_J\|_\infty \leq \eta(|I|,\lambda) + \eta(|J|,\lambda) + k_Q D\sigma^2\left(\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}} + \sqrt{\frac{\ln n(n+1) + \lambda \ln |J|}{|J|}}\right).$$

Donc pour tout $I \in \mathcal{I}$ tel que $I \subseteq \mathbb{I}$, nous avons, pour tout $J \in \mathcal{I}(I)$, $\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq \gamma_\rho(|I|,|J|,\lambda)$ et $\|\hat{Q}_I - \hat{Q}_J\|_\infty \leq \gamma_Q(|I|,|J|,\lambda)$. Donc I est accepté et $\mathbb{I}$ est accepté et donc $\mathbb{I} \subseteq \hat{I}$. Ceci implique

$$\begin{aligned}\|\hat{\rho}_{\mathbb{I}} - \hat{\rho}_{\hat{I}}\|_\infty &\leq \eta_1(|\mathbb{I}|,\lambda) + \eta_1(|\hat{I}|,\lambda) + 2\sqrt{2e}D\sigma\left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|\hat{I}|}}\right), \\ &\leq 2\eta_1(|\mathbb{I}|,\lambda) + 4\sqrt{2e}D\sigma\sqrt{\frac{\ln n}{|\mathbb{I}|}},\end{aligned}$$

puisque $\eta_1(|I|,\lambda)$ est une fonction décroissante de $|I|$. De plus,

$$\begin{aligned}\|\hat{\rho}_{\hat{I}} - \rho_\tau\|_\infty &\leq \|\hat{\rho}_{\hat{I}} - \hat{\rho}_{\mathbb{I}}\|_\infty + \|\hat{\rho}_{\mathbb{I}} - \bar{\rho}_{\mathbb{I}}\|_\infty + \|\bar{\rho}_{\mathbb{I}} - \rho_\tau\|_\infty, \\ &\leq 3\eta_1(|\mathbb{I}|,\lambda) + 4\sqrt{2e}D\sigma\sqrt{\frac{\ln n}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}(\rho).\end{aligned}$$

Comme $\eta(|I|,\lambda)$ est aussi une fonction décroissante de $|I|$, on peut montrer de la même manière que

$$\|\hat{Q}_{\hat{I}} - Q_\tau\|_\infty \leq 3\eta(|\mathbb{I}|,\lambda) + 2k_Q D\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}(Q),$$

ce qui achève la preuve.

□

Preuve des théorèmes 2.2.1 et 2.2.3

Dans le cas stationnaire, soit $\lambda > 0$ tel que $\ln n(n+1) + \lambda \ln N \leq N$ et dans le cas où les paramètres varient lentement soit λ le paramètre de lissage de l'algorithme adaptatif tel que $\ln n(n+1) + \lambda \ln m_0 \leq m_0$. Dans le cas stationnaire (resp. dans le cas où les paramètres varient lentement), d'après le lemme 2.7.5 (resp. le théorème 2.3.3) on peut trouver un

ensemble aléatoire S de probabilité au moins $1-\frac{3}{N^\lambda}$ (resp. au moins $1-\sum_{J\in\mathcal{I}_+(\mathbb{I})}\frac{3}{|J|^\lambda}$) et des fonctions η_ρ et η_Q dépendant de λ,σ,n et N (resp. λ,σ,n et $|\mathbb{I}|$) telles que si $\omega\in S$:

$$\|\hat{\rho}_\tau-\rho_\tau\|_\infty\leq\eta_\rho,\ \text{et}\ \|\hat{Q}_\tau-Q_\tau\|_\infty\leq\eta_Q. \tag{2.105}$$

Soit $\hat{f}$ la fonction objectif du problème $(\widehat{\mathcal{V}})$ et f la fonction objectif du problème $(\mathcal{V})$. Remarquons que

$$\begin{aligned}
f(x^*)-f(\widehat{x}) &= f(x^*)-\widehat{f}(x^*)\\
&\quad+\widehat{f}(x^*)-\widehat{f}(\widehat{x})\\
&\quad+\widehat{f}(\widehat{x})-f(\widehat{x})\\
&\leq 2\sup_{X(\mu,\nu,x^-)}|\widehat{f}(x)-f(x)|\\
&\leq 2\sup_{X(\mu,\nu,x^-)}|(\widehat{\rho}_\tau-\rho_\tau)^Tx|+2\kappa(\epsilon)\sup_{X(\mu,\nu,x^-)}|\sqrt{x^T\widehat{Q}_\tau x}-\sqrt{x^TQ_\tau x}|\\
&= \delta_1+\delta_2.
\end{aligned}\tag{2.106}$$

D'autre part, puisque sur l'ensemble $S\subseteq\Omega$, nous avons $\|\widehat{\rho}_\tau-\rho_\tau\|_\infty\leq\eta_\rho$ et $\|\hat{Q}_\tau-Q_\tau\|_\infty\leq\eta_Q$, et comme $\|x\|_1\leq\|x^-\|_1$ pour tout $x\in X(\mu,\nu,x^-)$, nous avons sur S la majoration :

$$\delta_1\leq 2\|\widehat{\rho}_\tau-\rho_\tau\|_\infty\|x^-\|_1\leq 2\,\eta_\rho\,\|x^-\|_1.$$

Si $\hat{Q}_\tau\neq Q_\tau$ alors pour tout $\theta>0$ nous avons par ailleurs :

$$\begin{aligned}
|\sqrt{x^TQ_\tau x}-\sqrt{x^T\widehat{Q}_\tau x}| &\leq \left|\frac{x^T\widehat{Q}_\tau x-x^TQ_\tau x}{\theta}\right|1\left(\max(\sqrt{x^T\widehat{Q}_\tau x},\,\sqrt{x^TQ_\tau x})\geq\theta\right)\\
&\quad+\theta\,1\left(\max(\sqrt{x^T\widehat{Q}_\tau x},\,\sqrt{x^TQ_\tau x})<\theta\right).
\end{aligned}$$

En choisissant $\theta=\sqrt{\|\widehat{Q}_\tau-Q_\tau\|_\infty}\|x^-\|_1$, on obtient :

$$\begin{aligned}
\delta_2 &\leq 2\,\kappa(\varepsilon)\sqrt{\|\widehat{Q}_\tau-Q_\tau\|_\infty}\|x^-\|_1\\
&\leq 2\,\kappa(\varepsilon)\sqrt{\eta_Q}\|x^-\|_1,
\end{aligned}$$

sur S. En substituant les majorants ci-dessus dans (2.106) on obtient les majorations (2.15) et (2.20). De plus, si Q_τ est définie positive, on peut améliorer la majoration de δ_2. En effet, dans ce cas, $\sqrt{x^TQ_\tau x}\geq\sqrt{\beta(Q_\tau)}\,\|x\|_1$ et

$$\left|\frac{x^TQ_\tau x-x^T\widehat{Q}_\tau x}{\sqrt{x^TQ_\tau x}+\sqrt{x^T\widehat{Q}_\tau x}}\right| \leq \frac{1}{\sqrt{\beta(Q_\tau)}}\|\widehat{Q}_\tau-Q_\tau\|_\infty\|x\|_1\leq\frac{1}{\sqrt{\beta(Q_\tau)}}\|\widehat{Q}_\tau-Q_\tau\|_\infty\|x^-\|_1.$$

On obtient alors sur S la majoration suivante pour δ_2 :

$$\delta_2\leq\frac{2\kappa(\epsilon)}{\sqrt{\beta(Q_\tau)}}\|\widehat{Q}_\tau-Q_\tau\|_\infty\|x^-\|_1\leq\frac{2\kappa(\epsilon)}{\sqrt{\beta(Q_\tau)}}\,\eta_Q\|x^-\|_1,$$

ce qui prouve les majorations (2.16) et (2.21).
Remarquons que pour le théorème 2.2.3, les constantes k_1' et k_2' sont données par $k_1'=6(\frac{7}{3}+\sqrt{2}+2\sqrt{2eD})$ et $k_2'=3(k_QD+k_Q')$.

□

Preuve des théorèmes 2.2.2 et 2.2.4

Soit S l'ensemble aléatoire défini dans la preuve du théorème 2.2.1 tel que les inégalités (2.105) soient vérifiées sur S. Sur S si le problème $(\mathcal{M})$ est faisable, alors $(\widehat{\mathcal{M}})$ le sera aussi. En effet, pour $\omega \in S$, si x est faisable pour $(\mathcal{M})$ alors $x \geq 0$ et $\ell \mathbf{e}^T x^- \leq \rho_\tau^T x + r(n+1)x_{n+1} \leq (\hat{\rho}_\tau + \eta_\rho \mathbf{e})^T x + r(n+1)x_{n+1}$. De plus, la solution $(\widehat{x}, \widehat{x}_{n+1})$ de $(\widehat{\mathcal{M}})$ satisfait la contrainte

$$(\widehat{\rho}_\tau + \eta_\rho \mathbf{e})^T \widehat{x} + r(n+1)\widehat{x}_{n+1} \geq \ell\, \mathbf{e}^T x^-,$$

et ainsi sur S,

$$\begin{aligned} \rho_\tau^T \widehat{x} + r(n+1)\widehat{x}_{n+1} &\geq \ell\, \mathbf{e}^T x^- + (\rho_\tau - (\widehat{\rho}_\tau + \eta_\rho \mathbf{e}))^T \widehat{x} \\ &\geq \ell\, \mathbf{e}^T x^- - (\|\rho_\tau - \widehat{\rho}_\tau\|_\infty + \eta_\rho)\|\widehat{x}\|_1 \qquad (2.107) \\ &\geq \ell\, \mathbf{e}^T x^- - 2\eta_\rho \|x^-\|_1. \qquad (2.108) \end{aligned}$$

D'autre part, si $\hat{f}$ est la fonction objectif du problème $(\widehat{\mathcal{M}})$, alors

$$\begin{aligned} f(\widehat{x}) - f(x^*) &\leq f(\widehat{x}) - \hat{f}(\widehat{x}) \\ &\quad + \hat{f}(\widehat{x}) - \hat{f}(x^*) \\ &\quad + \hat{f}(x^*) - f(x^*) \\ &\leq 2 \sup_{X(\mu,\nu,x^-)} |f(x) - \hat{f}(x)| \leq \|\widehat{Q}_\tau - Q_\tau\|_\infty \|x^-\|_1^2. \qquad (2.109) \end{aligned}$$

Cette dernière quantité est majorée sur S par $\eta_Q \|x^-\|_1^2$, ce qui prouve (2.17) et (2.22). Pour prouver (2.18) et (2.23), nous allons montrer, sur l'ensemble S, comment raffiner la majoration $f(\widehat{x}) - f(x^*)$. Notons $\bar{x}$ la solution optimale du problème

$$(\mathcal{M}') \begin{cases} \min \frac{1}{2} x^T Q_\tau x \\ (\widehat{\rho}_\tau + \eta_\rho \mathbf{e})^T x + r(n+1)x_{n+1} \geq \ell\, \mathbf{e}^T x^-, \\ (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-). \end{cases} \qquad (2.110)$$

Remarquons que sur S, l'ensemble de faisabilité de $(\mathcal{M})$ est contenu dans celui de $(\mathcal{M}')$, et donc $f(\bar{x}) \leq f(x^*)$ sur S. Considérons la fonction :

$$g(x) = \frac{x^T(\widehat{Q}_\tau - Q_\tau)x}{2}.$$

Évidemment, g est Lipschitzienne sur la boule $B_1 = \{x \in \mathbb{R}^n : \ \|x\|_1 \leq \|x^-\|_1\}$ avec la constante de Lipschitz

$$L_g \leq \sup_{\|x\|_1 \leq \|x^-\|_1} \|(\widehat{Q}_\tau - Q_\tau)x\|_\infty \leq \|\widehat{Q}_\tau - Q_\tau\|_\infty \|x^-\|_1.$$

On peut majorer L_g sur S comme suit : $L_g \leq \eta_Q \|x^-\|_1$. Comme $\widehat{x}$ et $\bar{x}$ appartiennent à la boule B_1,

$$f(\widehat{x}) - f(\bar{x}) \leq |g(\widehat{x}) - g(\bar{x})| \leq L_g \|\widehat{x} - \bar{x}\|_1. \qquad (2.111)$$

D'autre part, on peut écrire $f(\hat{x})$ sous la forme $f(\bar{x}) + (\hat{x} - \bar{x})^T \nabla f(\bar{x}) + \frac{1}{2}(\hat{x} - \bar{x})^T f''(\bar{x})(\hat{x} - \bar{x})$. Puisque $\hat{x}$ est dans l'ensemble de faisabilité de $(\mathcal{M}')$, $(\hat{x} - \bar{x})^T \nabla f(\bar{x}) \geq 0$. De plus, par définition de $\beta(Q_\tau)$,

$$\frac{1}{2}(\hat{x} - \bar{x})^T f''(\bar{x})(\hat{x} - \bar{x}) = \frac{1}{2}(\hat{x} - \bar{x})^T Q_\tau (\hat{x} - \bar{x}) \geq \frac{\beta(Q_\tau)}{2} \|\hat{x} - \bar{x}\|_1^2,$$

ce qui donne

$$\frac{\beta(Q_\tau)}{2}\|\widehat{x}-\bar{x}\|_1^2 \leq f(\widehat{x})-f(\bar{x}) \leq L_g\|\widehat{x}-\bar{x}\|_1,$$

et

$$\|\widehat{x}-\bar{x}\|_1 \leq \frac{2L_g}{\beta(Q_\tau)} \leq \frac{2\,\eta_Q\,\|x^-\|_1}{\beta(Q_\tau)}.$$

En utilisant à nouveau la majoration (2.111) on obtient :

$$\begin{aligned} f(\widehat{x})-f(x^*) &\leq f(\widehat{x})-f(\bar{x}) \\ &\leq \frac{2\,\eta_Q^2\,\|x^-\|_1^2}{\beta(Q_\tau)}. \end{aligned}$$

Remarquons que pour le théorème 2.2.4, les constantes k_1' et k_2' sont données par $k_1' = 6(\frac{7}{3}+\sqrt{2}+2\sqrt{2e}D)$ et $k_2' = 3(k_Q D + k_Q')$.

□

Preuve du théorème 2.3.5

En utilisant les notations introduites dans la section 2.3.1, on cherche à majorer :

$$p = P\left(\exists\, J \in \mathcal{I}(I) \mid \|\hat{\rho}_J-\hat{\rho}_I\|_\infty > \gamma_\rho(|I|,|J|,\lambda) \bigcup \|\hat{Q}_J-\hat{Q}_I\|_\infty > \gamma_Q(|I|,|J|,\lambda)\right).$$

Nous avons :

$$\begin{aligned} p &\leq \sum_{J\in\mathcal{I}(I)} P(\|\hat{\rho}_J-\hat{\rho}_I\|_\infty > \gamma_\rho(|I|,|J|,\lambda)) + P(\|\hat{Q}_J-\hat{Q}_I\|_\infty > \gamma_Q(|I|,|J|,\lambda)) \\ &\leq \sum_{J\in\mathcal{I}(I)} (\frac{3}{|I|^\lambda}+\frac{3}{|J|^\lambda}) \leq \frac{6\;Card(\mathcal{I}(I))}{m_0^\lambda}. \end{aligned}$$

□

Preuve du théorème 2.3.6

Si l'intervalle I est accepté comme intervalle d'homogénéité temporelle, alors on peut considérer que les moyennes et variances/covariances sont proches sur J et I. Ainsi, la probabilité p que I soit accepté est majorée par :

$$p = P\left(\|\hat{\rho}_J-\hat{\rho}_I\|_\infty \leq \gamma_\rho(|I|,|J|,\lambda) \bigcap \|\hat{Q}_J-\hat{Q}_I\|_\infty \leq \gamma_Q(|I|,|J|,\lambda)\right)$$

qui est majorée par la somme des probabilités. Par ailleurs,

$$\|\hat{\rho}_J-\hat{\rho}_I\|_\infty \geq \|\bar{\rho}_I-\bar{\rho}_J\|_\infty - \|\hat{\rho}_J-\bar{\rho}_J+\bar{\rho}_I-\hat{\rho}_I\|_\infty, \quad \|\hat{Q}_J-\hat{Q}_I\|_\infty \geq \|\bar{Q}_I-\bar{Q}_J\|_\infty - \|\hat{Q}_J-\bar{Q}_J+\bar{Q}_I-\hat{Q}_I\|_\infty. \tag{2.112}$$

Remarquons que (2.112), (2.33) et (2.34) donnent

$$\begin{aligned} p &\leq P(\|\hat{\rho}_J-\bar{\rho}_J+\bar{\rho}_I-\hat{\rho}_I\|_\infty \geq \eta_1(|J|,\lambda)+\eta_1(|I|,\lambda)) + P(\|\hat{Q}_J-\bar{Q}_J+\bar{Q}_I-\hat{Q}_I\|_\infty \geq \eta(|J|,\lambda)+\eta(|I|,\lambda)), \\ &\leq P(\|\hat{\rho}_J-\bar{\rho}_J\|_\infty \geq \eta_1(|J|,\lambda)) + P(\|\hat{\rho}_I-\bar{\rho}_I\|_\infty \geq \eta_1(|I|,\lambda)) \\ &\quad + P(\|\hat{Q}_J-\bar{Q}_J\|_\infty \geq \eta(|J|,\lambda)) + P(\|\hat{Q}_I-\bar{Q}_I\|_\infty \geq \eta(|I|,\lambda)), \\ &\leq 3\left(\frac{1}{m^\lambda}+\frac{1}{(m+m')^\lambda}\right). \end{aligned}$$

□

Preuve du théorème 2.3.7

Si $I_1 = [T_{cp} - m + m_0, \tau - 1]$ et $I_2 = [T_{cp} - m, T_{cp} - m + m_0 - 1]$, la probabilité que I soit accepté est la probabilité que $\|\hat{\rho}_{I_1} - \hat{\rho}_{I_2}\|_\infty \leq \gamma_\rho(|I_1|, |I_2|, \lambda)$. Nous avons $\bar{\rho}_{I_2} = \rho_2$ et $\bar{\rho}_{I_1} = \frac{m'}{m+m'-m_0}(\rho_1 - \rho_2)$. Il suffit ensuite de suivre la preuve du théorème 2.3.6 (en remplaçant I et J par I_1 et I_2) pour conclure.

□

Preuve du théorème 2.3.8

D'après les lemmes 2.7.1 et 2.7.2, nous avons, pour tout intervalle I non vide, et pour tout $\lambda > 0$ tel que $10(\ln n + \lambda \ln |I|) \leq 3|I|$:

$$P\left(\|\hat{\rho}_I - \bar{\rho}_I\|_\infty \geq f_1(|I|, \lambda) = 2\sqrt{\frac{10}{3}}\sigma\sqrt{\frac{\ln n + \lambda \ln |I|}{|I|}}\right) \leq \frac{K_1(\lambda)}{|I|^\lambda}, \tag{2.113}$$

$$E\|\hat{\rho}_I - E\hat{\rho}_I\|_\infty \leq \sqrt{2e}\sigma\sqrt{\frac{\ln n}{|I|}}. \tag{2.114}$$

D'autre part, nous avons

$$\|\hat{Q}_I - \bar{Q}_I\|_\infty \leq \|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2 + \|\tilde{Q}_I^b - \bar{Q}_I\|_\infty, \tag{2.115}$$

avec maintenant $\tilde{Q}_I^b = \frac{1}{|I|}\sum_{t\in I}(r_t - \bar{\rho}_I)(r_t - \bar{\rho}_I)^T$. On peut écrire $\|\tilde{Q}_I^b - \bar{Q}_I\|_\infty$ sous la forme $\|\frac{1}{|I|}\sum_{t\in I}\xi_t\|_\infty$ avec $\xi_t = (r_t - \bar{\rho}_I)(r_t - \bar{\rho}_I)^T - Q_t$. En remarquant que $\forall 1 \leq k \leq E[\frac{p}{2}]$, $E\|\xi_t\|_\infty^k \leq \sigma''^k$ avec $\sigma'' = \sigma^2 + (\sigma + 2\sigma')^2$, on peut, en posant $A = \sigma''\sqrt{\frac{3|I|}{10(\ln n(n+1) + \lambda \ln |I|)}}$, suivre la preuve de (2.63) pour montrer que si $10(\ln n(n+1) + \lambda \ln |I|) \leq 3|I|$:

$$P\left(\|\tilde{Q}_I^b - \bar{Q}_I\|_\infty \geq 2\sqrt{\frac{10}{3}}\sigma''\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}}\right) \leq \frac{K_2(\lambda)}{|I|^\lambda} \tag{2.116}$$

avec $K_2(\lambda) = 1 + \left(\frac{10\,E[\frac{p}{2}](2\gamma+\lambda)}{3\,e(E[\frac{p}{2}]-2-2\lambda)}\right)^{\frac{E[\frac{p}{2}]}{2}}$. En utilisant (2.113), (2.115) et (2.116) on obtient :

$$P\left(\|\hat{Q}_I - \bar{Q}_I\|_\infty \geq f_2(|I|, \lambda) = 2\sqrt{\frac{10}{3}}(2\sqrt{\frac{10}{3}}\sigma^2 + \sigma'')\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}}\right) \leq \frac{K_1(\lambda) + K_2(\lambda)}{|I|^\lambda}. \tag{2.117}$$

Enfin

$$\begin{aligned} E\|\hat{Q}_I - E\hat{Q}_I\|_\infty &\leq 2E\|\hat{Q}_I - \tilde{Q}_I^b\|_\infty + E\|\tilde{Q}_I^b - E\tilde{Q}_I^b\|_\infty, \\ &\leq 2E\|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2 + \|\frac{1}{|I|}\sum_{t\in I}\tilde{\xi}_t\|_\infty, \end{aligned} \tag{2.118}$$

en posant $\tilde{\xi}_t = (r_t - \bar{\rho}_I)(r_t - \bar{\rho}_I)^T - E(r_t - \bar{\rho}_I)(r_t - \bar{\rho}_I)^T$. Puisque $E\|\tilde{\xi}_t\|_\infty^2 \leq \sigma'''^2$ avec $\sigma''' = 2(\sigma + 2\sigma')^2$, en suivant la preuve de (2.58) on montre que

$$E\|\tilde{Q}_I^b - E\tilde{Q}_I^b\|_\infty \leq 2\sqrt{e}\sigma'''\sqrt{\frac{\ln n}{|I|}}. \tag{2.119}$$

En utilisant (2.114), (2.118) et (2.119) on obtient :

$$E\|\hat{Q}_I - E\hat{Q}_I\|_\infty \leq (4e\sigma^2 + 2\sqrt{e}\sigma''')\sqrt{\frac{\ln n}{|I|}}.$$

Ceci justifie les règles d'acceptation données en (2.36) et (2.38). Notons que (2.36) est vérifiée avec une probabilité supérieure à $1 - \frac{K_1(\lambda)}{|I|^\lambda} - \frac{K_1(\lambda)}{|J|^\lambda}$ et (2.38) avec une probabilité supérieure à $1 - (K_1(\mu) + K_2(\mu))(\frac{1}{|I|^\mu} + \frac{1}{|J|^\mu})$. Si l'algorithme adaptatif est implémenté en utilisant les estimations empiriques de la moyenne et de la matrice de covariance, on peut donc utiliser les règles d'acceptation (2.36) et (2.38). Le théorème 2.3.8 se prouve ensuite comme le théorème 2.3.3 et en utilisant (2.113) et (2.117).

□

Preuve du théorème 2.4.1

Nous montrons le résultat lorsque l'hypothèse 1 est vérifiée (ce qui justifiera la remarque 2.4.1 faisant suite au théorème 2.4.1). Si l'hypothèse 2 est vérifiée, on montre le théorème 2.4.1 de façon analogue.
D'après (2.100), pour tout $\lambda > 0$ et pour tout intervalle non vide I et de longueur h :

$$P(\|\hat{\rho}_I - \bar{\rho}_I\|_\infty \geq \eta_1(h,\lambda)) \leq \frac{1}{h^\lambda}. \tag{2.120}$$

On considère quatre cas.

- Si $h \leq t \leq T_{cp} - h + 1$, $\bar{\rho}_g^t(h) = \bar{\rho}_d^t(h) = \rho_1$ et

$$\begin{aligned} P(\|\rho_g^t(h) - \rho_d^t(h)\|_\infty > 2\eta_1(h,\lambda)) &\leq P(\|\rho_g^t(h) - \bar{\rho}_g^t(h)\|_\infty + \|\rho_d^t(h) - \bar{\rho}_d^t(h)\|_\infty > 2\eta_1(h,\lambda)) \\ &\leq P(\|\rho_g^t(h) - \bar{\rho}_g^t(h)\|_\infty > \eta_1(h,\lambda)) + P(\|\rho_d^t(h) - \bar{\rho}_d^t(h)\|_\infty > \eta_1(h,\lambda)) \\ &\leq \frac{2}{h^\lambda}. \end{aligned}$$

  Ainsi, il y a un ensemble $S_t \subset \Omega$ de probabilité au moins $1 - \frac{2}{h^\lambda}$ tel que pour $\omega \in S_t$, $\|\rho_g^t(h) - \rho_d^t(h)\|_\infty \leq 2\eta_1(h,\lambda)$.
- De même, pour $T_{cp} + h \leq t \leq N - h + 1$ il y a un ensemble $S_t \subset \Omega$ de probabilité au moins $1 - \frac{2}{h^\lambda}$ tel que pour $\omega \in S_t$, $\|\rho_g^t(h) - \rho_d^t(h)\|_\infty \leq 2\eta_1(h,\lambda)$.
- Pour $T_{cp} + 1 \leq t \leq T_{cp} + h - 1$, nous avons

$$\bar{\rho}_d^t(h) = \rho_2, \text{ et } \bar{\rho}_g^t(h) = \frac{(t - T_{cp})\rho_2 + (h - t + T_{cp})\rho_1}{h}.$$

  Remarquons que

$$\|\rho_g^t(h) - \rho_d^t(h)\|_\infty \geq -\|\rho_g^t(h) - \bar{\rho}_g^t(h) - \rho_d^t(h) + \bar{\rho}_d^t(h)\|_\infty + \left(\frac{h - t + T_{cp}}{h}\right)\|\rho_2 - \rho_1\|_\infty,$$

  et que $P(\|\rho_g^t(h) - \bar{\rho}_g^t(h) - \rho_d^t(h) + \bar{\rho}_d^t(h)\|_\infty > 2\eta_1(h,\lambda)) \leq \frac{2}{h^\lambda}$. Ainsi, il y a un ensemble $S_t \subset \Omega$ de probabilité au moins $1 - \frac{2}{h^\lambda}$ tel que pour $\omega \in S_t$,

$$\|\rho_g^t(h) - \rho_d^t(h)\|_\infty \geq -2\eta_1(h,\lambda) + \left(\frac{h - t + T_{cp}}{h}\right)\|\rho_2 - \rho_1\|_\infty.$$

– De même, pour $T_{cp} - h + 2 \leq t \leq T_{cp}$, on peut trouver un ensemble $S_t \subset \Omega$ de probabilité au moins $1 - \frac{2}{h^\lambda}$ tel que pour $\omega \in S_t$,

$$\|\rho_g^t(h) - \rho_d^t(h)\|_\infty \geq -2\eta_1(h,\lambda) + \left(\frac{h + t - T_{cp} - 1}{h}\right) \|\rho_2 - \rho_1\|_\infty.$$

Ainsi, si $w \in S_{T_{cp}}$ et $\|\rho_2 - \rho_1\|_\infty > \frac{4h}{h-1}\eta_1(h,\lambda)$ alors $\|\rho_g^{T_{cp}}(h) - \rho_d^{T_{cp}}(h)\|_\infty > 2\eta_1(h,\lambda)$. Finalement, si $w \in \left(\bigcap_{t=h}^{T_{cp}-h+1} S_t\right) \bigcap \left(\bigcap_{t=T_{cp}+h}^{N-h+1} S_t\right) \bigcap S_{T_{cp}}$ alors $\hat{T}_{cp}(h) \in [T_{cp} - h + 1, T_{cp} + h - 1]$. On obtient donc :

$$P(\hat{T}_{cp}(h) \in [T_{cp}-h+1, T_{cp}+h-1]) \geq P\left(w \in \left(\bigcap_{t=h}^{T_{cp}-h+1} S_t\right) \bigcap \left(\bigcap_{t=T_{cp}+h}^{N-h+1} S_t\right) \bigcap S_{T_{cp}}\right) \geq 1 - \frac{2(N-4h+5)}{h^\lambda}.$$

□

Preuve du théorème 2.4.2

Nous montrons le résultat lorsque l'hypothèse 1 est vérifiée (ce qui justifiera la remarque 2.4.2 faisant suite au théorème 2.4.2). Si l'hypothèse 2 est vérifiée, on peut montrer le théorème 2.4.2 de façon analogue.
En suivant la preuve du théorème 2.4.1, on peut trouver des ensembles S_t tels que :

– Si $h \leq t \leq T_{cp} - h + 1$ ou $T_{cp} + h \leq t \leq N - h + 1$ et $\omega \in S_t$, alors $\|\rho_g^t(h) - \rho_d^t(h)\|_\infty \leq 2\eta_1(h,\lambda)$.
– Si $T_{cp} + 2 \leq t \leq T_{cp} + h - 1$ ou $T_{cp} - h + 2 \leq t \leq T_{cp} - 1$ et $\omega \in S_t$, alors $\|\rho_g^t(h) - \rho_d^t(h)\|_\infty \leq 2\eta_1(h,\lambda) + \frac{h-2}{h}\|\rho_2 - \rho_1\|_\infty$.
– Si $t = T_{cp}$ ou $t = T_{cp} + 1$ et $\omega \in S_t$,

$$\|\rho_g^t(h) - \rho_d^t(h)\|_\infty \geq -2\eta_1(h,\lambda) + \frac{h-1}{h}\|\rho_2 - \rho_1\|_\infty > 2\eta_1(h,\lambda) + \frac{h-2}{h}\|\rho_2 - \rho_1\|_\infty.$$

Ainsi, si $\omega \in \bigcap_{t=h}^{N-h+1} S_t$ et $\|\rho_2 - \rho_1\|_\infty > 4h\eta_1(h,\lambda)$ alors $\hat{T}_{cp}(h) \in \{T_{cp}\} \cup \{T_{cp} + 1\}$. □

Preuve des théorèmes 2.4.3 et 2.4.4

Les théorèmes 2.4.3 et 2.4.4 se montrent respectivement en suivant les preuves des théorèmes 2.4.1 et 2.4.2.

□

Chapitre 3

Analyse de sensibilité et calibration stable et robuste de matrices de covariances

Nota Bene : Les résultats de ce chapitre seront consignés dans un rapport de recherche en cours ([54]).

Dans ce chapitre, nous présentons une mise en oeuvre du problème de Markowitz conduisant à des portefeuilles stables par rapport aux perturbations des paramètres, ainsi que des contreparties robustes de ce problème. Pour obtenir la stabilité des portefeuilles, une nouvelle calibration de la matrice de covariance entre les rendements est proposée. Cette calibration est la solution d'un problème d'optimisation semidéfini positif et nous la comparons avec d'autres méthodes de calibration de la matrice de covariance. Pour justifier notre calibration, on détermine deux propriétés que la matrice de covariance devrait vérifier. Tout d'abord, une analyse de sensibilité du modèle de Markowitz montre le rôle joué par la plus petite valeur propre de cette matrice. Cette étude passe par une analyse de sensibilité des solutions d'un problème quadratique. Par ailleurs, en utilisant les résultats du chapitre précédent, on peut localiser la matrice de covariance et la moyenne dans des ensembles bornés. Pour ce qui est de la robustification du modèle, on utilise aussi les résultats du chapitre précédent qui permettent de mettre en oeuvre facilement différentes contreparties robustes du modèle de Markowitz. Les contreparties robustes sont des problèmes résolus aussi efficacement que le problème nominal.

3.1 Introduction

Dans ce chapitre, nous nous penchons donc sur la stabilité des portefeuilles solutions du problème de Markowitz ([69]). Le modèle de Markowitz (que nous présenterons dans ce chapitre sous deux formes légèrement différentes de celle présentée au chapitre précédent) reste aujourd'hui le modèle de gestion de portefeuilles à la fois le plus simple et le plus général. Cependant, ce modèle est difficile à utiliser en pratique car, comme nous l'avons souligné au chapitre précédent, il intervient après une analyse statistique délicate : la prédiction des paramètres du modèle (le vecteur des rendements moyens ρ_τ et la matrice de

covariance Q_τ entre ces rendements sur la période d'investissement). De plus, comme le soulignent [15] et [14], si on utilise les estimations empiriques des paramètres, la composition des portefeuilles est typiquement très sensible aux variations des rendements. Le contrôle de la stabilité de la composition des portefeuilles nécessite donc une estimation adéquate des paramètres du modèle. Dans ce chapitre, nous montrons en fait qu'il existe une mise en oeuvre simple et automatique du modèle de Markowitz, qui permet d'obtenir les propriétés de stabilité recherchées. Pour cela, nous devons tenir compte du risque numérique lié à la première étape de prédiction des grandeurs statistiques. Notre approche introduit une étape intermédiaire entre la première étape d'estimation et la deuxième étape de sélection, étape intermédiaire pouvant être vue comme un filtre ou une régularisation numérique des prédictions statistiques. La régularisation correspond à une nouvelle calibration de la matrice de covariance.
Plus précisément, le plan du chapitre est le suivant. Nous commençons par introduire dans la deuxième section les deux variantes du modèle de Markowitz que nous étudions dans ce chapitre. Nous en profitons pour donner quelques propriétés du modèle de Markowitz qui seront utilisées par la suite. Afin de contrôler la stabilité des portefeuilles, étant donnés deux portefeuilles x_1 et x_2 obtenus pour les valeurs (ρ_1, Q_1) et (ρ_2, Q_2) des paramètres du modèle, nous cherchons à majorer $\|x_2 - x_1\|_1$ en fonction de $\|Q_2 - Q_1\|_\infty$ et $\|\rho_2 - \rho_1\|_\infty$. Ainsi, la troisième section est consacrée à une analyse de sensibilité des solutions optimales. Deux différentes versions du modèle de Markowitz sont étudiées. Puisqu'elles sont toutes des problèmes d'optimisation quadratique satisfaisant l'hypothèse de Slater, on sait ([32]) que les solutions sont radialement lipschitziennes. Ceci nous garantit l'existence d'une constante de lipschitz mais pour (ρ_1, Q_1) fixés, ne dit rien sur sa dépendance en fonction de (ρ_1, Q_1). Ainsi, notre analyse de sensibilité vise à déterminer des bornes précises et globales. Pour la version du problème où la contrainte de rendement est agrégée dans la fonction objectif, on montre que les solutions sont lipschitz radiales par rapport aux paramètres. La constante de lipschitz obtenue est plus simple que celle de [14] et résulte de l'application de la proposition 3.8.1 donnée en fin de ce chapitre [[18]]. On étudie ensuite une autre version du problème intégrant une contrainte de rendement. Notre analyse de sensibilité tend à montrer que les portefeuilles générés en utilisant le modèle de Markowitz seront stables si la plus petite valeur propre de la matrice de covariance estimée est suffisamment grande. On donne enfin une analyse de sensibilité des solutions du modèle de Markowitz général (2.6) introduit dans le chapitre précédent et tenant compte des frais de transactions.
Dans la quatrième section, on utilise les résultats du chapitre précédent et on propose des contreparties robustes du problème de Markowitz. Le modèle statistique utilisé pour les rendements est plus général et conduit à des contreparties robustes plus simples que [46]. Dans cette même section, nous proposons deux calibrations stables de la matrice de covariance. D'autres calibrations existantes de la matrice de covariance préoccupées par la stabilité des portefeuilles sont ensuite présentées. La dernière section est consacrée à des simulations numériques.

3.2 Modèle de Markowitz et sources d'instabilité

3.2.1 Modèle moyenne variance de Markowitz

Nous présentons deux déclinaisons du modèle de Markowitz introduit dans le chapitre précédent. Nous ne tenons pas compte, dans un premier temps et afin de simplifier la pré-

sentation des résultats de ce chapitre, des frais de transaction. Le cas plus général présenté dans le chapitre précédent sera ensuite brièvement traité. Nous désignerons toujours par n le nombre d'actifs risqués, τ la date de l'investissement, M l'horizon d'investissement, ρ_τ le vecteur des rendements moyens des actifs risqués sur la période d'investissement et Q_τ la matrice de covariance entre les rendements de ces actifs sur la période d'investissement. Le rendement de l'actif sans risque sera noté $r(n+1)$. Dans ce chapitre, un portefeuille x sera un vecteur $x = (x_1, \ldots, x_n)$ donnant les poids des actifs risqués. Le poids de l'actif sans risque est alors $x_0 = 1 - x^T\mathbf{e}$. Une première approche consiste à rechercher le portefeuille solution du problème $P(k, \rho_\tau, Q_\tau)$ suivant paramétré par k, ρ_τ et Q_τ :

$$P(k, \rho_\tau, Q_\tau) \begin{cases} \min & \frac{1}{2}\, x^T Q_\tau x - k\, x^T(\rho_\tau - r(n+1)\mathbf{e}) \\ x \in \Delta_n; \end{cases} \tag{3.1}$$

où $k \geq 0$ dépend de l'aversion au risque de l'investisseur. On essaie alors simultanément de minimiser la variance du rendement du portefeuille et de maximiser le rendement moyen du portefeuille sur la période d'investissement. L'investisseur peut aussi se fixer un rendement cible. Dans ce cas on cherche on portefeuille dont le rendement moyen ${\rho_\tau}^T x + r(n+1)x_0$ est supérieur à une valeur cible donnée ℓ et dont le risque $x^T Q_\tau x$ est minimum parmi tous les portefeuilles admissibles. Si l'on n'autorise pas la vente à découvert, on cherchera alors à résoudre le problème $P'(\ell, \rho_\tau, Q_\tau)$:

$$P'(\ell, \rho_\tau, Q_\tau) \begin{cases} \min & \frac{1}{2}\, x^T Q_\tau x \\ x^T(\rho_\tau - r(n+1)\mathbf{e}) \geq \ell - r(n+1), & x \in \Delta_n. \end{cases} \tag{3.2}$$

Hypothèses de travail sur les problèmes $P(k, \rho_\tau, Q_\tau)$ et $P'(\ell, \rho_\tau, Q_\tau)$.

Nous ferons, dans tout ce chapitre, les hypothèses suivantes :

- **H1.** En tant que matrice de covariance, Q_τ est semidéfinie positive. Cependant, nous supposerons qu'elle est définie positive. Ceci est cohérent avec l'hypothèse de non arbitrage des marchés. En effet, si Q_τ avait une valeur propre nulle, nous pourrions construire un portefeuille sans risque $x = \dfrac{v}{v^T\mathbf{e}}$ (rappelons qu'en toute généralité un vecteur x est un portefeuille s'il satisfait la contrainte $x^T\mathbf{e} = 1$) à partir du vecteur propre v associé à cette valeur propre. Nous aurions ainsi l'illusion de pouvoir investir sans risque dans des actifs risqués.
- **H2.** Le rendement cible pour le problème P', noté ℓ, est supérieur à $r(n+1)$. En effet, le cas $r(n+1) \geq \ell$ est résolu immédiatement (avec la solution correspondante $x_i^* = 0$ pour $i \geq 1$: on investit tout dans l'actif sans risque).
- **H3.** Pour que le problème P' soit faisable, il existe un actif i dont le rendement moyen est supérieur ou égal à ℓ. On suppose qu'il existe un actif i dont le rendement moyen $\rho_\tau(i)$ vérifie $\rho_\tau(i) > \ell$. Enfin, le vecteur des rendements moyens ρ_τ et $\mathbf{e}$ ne sont pas colinéaires.

Par la suite, nous dirons que le problème $P(k, \rho_1, Q_1)$ (ou $P'(\ell, \rho_1, Q_1)$) satisfait les hypothèses **H1**, **H2** et **H3** si les hypothèses de travail ci-dessus sont satisfaites en remplaçant ρ_τ par ρ_1 et Q_τ par Q_1.

3.2.2 Quelques propriétés du modèle de Markowitz

Nous allons donner quelques propriétés du problème de Markowitz que nous utiliserons par la suite. Tout d'abord, remarquons que la fonction objectif du problème $P'(\ell, \rho_\tau, Q_\tau)$

étant minorée sur l'ensemble de faisabilité de P' qui est défini par des contraintes linéaires, le problème primal P' et son dual sont équivalents. Nous pourrons ainsi faire un voyage dans l'espace dual pour étudier le problème P'. On trouvera en annexe du chapitre 3 la preuve du lemme suivant.

Lemme 3.2.1 *Si une contrainte d'un problème strictement convexe n'est pas active, on peut l'enlever sans changer la solution optimale.*

Proposition 3.2.1 *La contrainte* $(\rho_\tau - r(n+1)\,\mathbf{e})^T x \geq \ell - r(n+1)$ *est active pour* P'.

Preuve. En utilisant le lemme 3.2.1, si la contrainte de rendement $(\rho_\tau - r(n+1)\,\mathbf{e})^T x \geq \ell - r(n+1)$ n'était pas active, on pourrait l'enlever et la solution du problème serait $x = 0$ qui ne satisfait pas la contrainte de rendement. Ainsi, la contrainte de rendement est bien active.

□

Remarquons que si la solution optimale x^* de $P'(\ell, \rho, Q)$ satisfait $x_i^* > 0$ pour $1 \leq i \leq n$ alors il suffit d'appliquer le théorème de KKT (p.305-306 de [64]) pour obtenir une expression explicite de la solution x^*.

Lemme 3.2.2 *Sous les hypothèses* ***H1***, ***H2***, ***H3***, *l'hypothèse de Slater est vérifiée pour* $P'(\ell, \rho_\tau, Q_\tau)$.

Preuve. Puisqu'il y a un actif i de rendement moyen $\rho_\tau(i) > \ell$, alors il existe $\varepsilon > 0$ et $1 \leq i \leq n$, tels que $\rho_\tau(i) \geq \ell + 2\varepsilon$. Soit x défini par $x_i = \dfrac{\ell + \varepsilon}{\rho_\tau(i)}$ et $x_k = \dfrac{\varepsilon}{4(n-1)\max(\|\rho_\tau\|_\infty, 1)}$ pour $k \neq i$. Nous avons $x > 0$. De plus, en choisissant ε inférieur ou égal à $1 - \frac{\ell}{2}$, nous avons $x_i \leq 1 - \frac{\varepsilon}{2}$. Donc $x^T\,\mathbf{e} \leq 1 - \frac{\varepsilon}{4} < 1$. Enfin, $x^T(\rho_\tau - r(n+1)\,\mathbf{e}) \geq \ell + \frac{3\varepsilon}{4} - r(n+1) > \ell - r(n+1)$ donc aucune contrainte n'est active en x.

□

On conclut cette section en soulignant les sources d'instabilité de la composition, du rendement et de la variance des portefeuilles.

3.2.3 Sources d'instabilité

Les sources d'instabilité sont les paramètres du modèle : le vecteur des rendements moyens ρ_τ et la matrice de covariance Q_τ.
A partir de l'observation de rendements $(r_i)_{i=1,\ldots,N}$, calculables à partir des cours des actifs disponibles, on peut calculer les estimations empiriques et empiriques adaptatives du vecteur des rendements moyens comme nous l'avons expliqué au chapitre précédent. Cependant, la prédiction des rendements moyens, devrait, si possible, non seulement tenir compte des historiques des cours des actifs, mais aussi de données exogènes. Des informations telles que l'attente d'une reprise ou d'un coup dur pour une certaine compagnie, doivent être prises en compte. Ainsi, l'estimation des rendements moyens ne devrait pas être uniquement traitée par une procédure mathématique et automatique. Au contraire, les gestionnaires ont une contribution à apporter dans cette étape. Nous verrons dans la section consacrée à la robustification du modèle de Markowitz, une manière d'allier ces deux outils. Par contre, les variances et covariances peuvent uniquement être calculées par des outils statistiques. C'est pourquoi, dans ce chapitre, nous nous pencherons plus particulièrement sur la calibration de la matrice de covariance et en particulier sur des méthodes

de correction de la matrice de covariance empirique adaptative. Nous utiliserons comme estimation des rendements moyens, l'estimation empirique adaptative.
L'utilisation de la matrice de covariance empirique ou même de la matrice de covariance empirique adaptative ne semble pas propice à générer des portefeuilles stables et, d'autre part, peut poser des problèmes numériques puisque ces matrices peuvent être dégénérées. Nous développerons ces questions en section 3.4 et nous expliquerons comment on pourrait corriger les estimations empiriques et empiriques adaptatives de la matrice de covariance. L'objet de la section suivante est d'effectuer une analyse de sensibilité du problème de Markowitz.

3.3 Analyse de sensibilité du modèle de Markowitz

On fixe des valeurs nominales ℓ (ou k) et (ρ_1, Q_1) pour les paramètres du problème de Markowitz. Nous appellerons alors problème d'optimisation non perturbé, le problème de Markowitz associé à ces valeurs des paramètres. Soit x_1 la solution de ce problème. On considère ensuite une perturbation (ρ_2, Q_2) du couple de paramètres (ρ_1, Q_1) ; le paramètre ℓ (ou k) étant fixé. On appellera problème perturbé le problème de Markowitz associé aux valeurs perturbées des paramètres. Soit x_2 la solution du problème perturbé. Les fonctions objectifs des problèmes non perturbés et perturbés sont notées f et g (dont les expressions peuvent différer, en fonction du problème de Markowitz étudié). On pose enfin $\bar{\rho}_1 = \rho_1 - r(n+1)\,\mathbf{e}$, $\bar{\rho}_2 = \rho_2 - r(n+1)\,\mathbf{e}$, et $\bar{\ell} = \ell - r(n+1)$.

3.3.1 Quelques résultats connus

3.3.1.1 Sur la sensibilité des changements de ρ pour P

Dans [15], Best et Grauer étudient une formulation du problème de Markowitz du type $P(k, \rho, Q)$. Plus précisément, ils travaillent sur le problème quadratique paramétré (Parametric Quadratic Programming ou PQP en anglais) suivant :

$$\max \{\, T(\mu + tq)^T x - \frac{1}{2} x^T Q x \mid Ax \leq b \,\},$$

T étant un paramètre d'aversion au risque, μ et q sont des vecteurs de taille n fixée et la paramètre t varie entre des bornes inférieures et supérieures données. Le principal résultat qu'ils présentent et commentent dans [15] est qu'il existe un ensemble d'intervalles $[t_0, t_1], \ldots, [t_{m-1}, t_m]$ et des vecteurs h_{0i}, h_{1i}, g_{0i} et g_{1i} tels que les portefeuilles optimaux $x_i(t)$ et les multiplicateurs de Lagrange $\lambda_i(t)$ vérifient

$$x_i(t) = h_{0i} + t\, h_{1i} \;\text{ et }\; \lambda_i(t) = g_{0i} + t\, g_{1i},$$

pour tout $t_i \leq t \leq t_{i+1}$, $\quad 0 \leq i \leq m-1$. Chaque intervalle correspond à un ensemble de contraintes actives différent. Le problème est que les quantités t_i, h_{ji}, g_{ji} ne sont pas spécifiées. Cependant, dans [14] un majorant sur h_{ji} est donné. Dans la section suivante, nous étudions l'impact de perturbations à la fois dans le vecteur ρ et dans la matrice Q sur la composition des portefeuilles optimaux pour le problème P. L'approche diffère de celle utilisée dans [14]. Néanmoins, comme dans [14], on montre le rôle joué par la plus petite valeur propre de la matrice Q_1 lorsqu'on cherche à majorer $\|x_2 - x_1\|_2$.

3.3.1.2 Sur la sensibilité liée à des perturbations générales

[32] montre que les solutions d'un problème d'optimisation quadratique sont localement lipschitz radiales sous l'hypothèse de Slater. Dans la suite de cette section, on cherche à proposer une analyse de sensibilité plus précise et globale pour un type de problèmes quadratiques particuliers et pour les deux versions du problème de Markowitz présentées dans l'introduction. Les bornes sont globales et explicites en fonction des paramètres du problème.

3.3.2 Sensibilité de la valeur optimale du problème P

Proposition 3.3.1 *Soit x_i la solution de $P(k, \rho_i, Q_i)$ alors*

$$|g(x_2) - f(x_1)| \leq \frac{1}{2} \|Q_2 - Q_1\|_\infty + k\|\rho_2 - \rho_1\|_\infty$$

Preuve. On suppose que $g(x_2) \geq f(x_1)$ (l'autre cas est symétrique). Dans ce cas, $|g(x_2) - f(x_1)| = g(x_2) - f(x_1) = g(x_2) - g(x_1) + g(x_1) - f(x_1)$. Mais comme $x_1 \in \Delta_n$, par définition de x_2, $g(x_2) - g(x_1) \leq 0$. Ainsi,

$$|g(x_2) - f(x_1)| \leq \frac{1}{2} x_1^T (Q_2 - Q_1) x_1 - k(\rho_2 - \rho_1)^T x_1 \leq \frac{1}{2} \|x_1\|_1^2 \|Q_2 - Q_1\|_\infty + k\|\rho_2 - \rho_1\|_\infty \|x_1\|_1.$$

avec $\|x_1\|_1 \leq 1$.

□

D'un point de vue pratique, la valeur optimale du problème P n'est pas très sensible aux variations des paramètres de ce problème.

3.3.3 Analyse de sensibilité des solutions de $P(k, \rho, Q)$

L'ensemble de faisabilité du problème P est fixé quand Q et ρ varient. Comme f satisfait une condition de croissance du second ordre sur Δ_n, il suffit d'appliquer la proposition 3.8.1 rappelée en fin de ce chapitre, pour obtenir une analyse de sensibilité des solutions du problème P. Rappelons que pour une matrice symétrique Q, $\beta(Q)$ a été introduit dans la définition 2.2.1.

Théorème 3.3.1 *Soit x_i la solution de $P(k, \rho_i, Q_i)$. Alors si $P(k, \rho_1, Q_1)$ et $P(k, \rho_2, Q_2)$ satisfont les hypothèses **H1**, **H2** et **H3** :*

$$\|x_2 - x_1\|_1 \leq \frac{2}{\beta(Q_1)} (\|Q_2 - Q_1\|_\infty + k\|\rho_2 - \rho_1\|_\infty). \tag{3.3}$$

Preuve. Tout d'abord, remarquons que la fonction objectif f du problème de Markowitz $P(k, \rho_1, Q_1)$ satisfait une condition de croissance du second ordre sur Δ_n :

$$\exists\ c > 0,\ \ \forall x \in \Delta_n, \quad f(x) \geq f(x_1) + c\|x - x_1\|_1^2, \ \text{ avec } \ c = \frac{\beta(Q_1)}{2}. \tag{3.4}$$

En effet, pour tout $x \in \Delta_n$, un développement de Taylor au second ordre de f en x_1 donne :

$$f(x) = f(x_1) + (x - x_1)^T \nabla f(x_1) + \frac{1}{2}(x - x_1)^T \nabla^2 f(x_1)(x - x_1), \tag{3.5}$$

où $\nabla f(x_1) = Q_1 x_1 - k\bar{\rho}_1$ et $\nabla^2 f(x_1) = Q_1$. Les conditions d'optimalité du premier ordre donnent $(x - x_1)^T \nabla f(x_1) \geq 0$ pour tout $x \in \Delta_n$. D'autre part, pour tout $x \in \Delta_n$:

$$(x - x_1)^T \nabla^2 f(x_1)(x - x_1) \geq \beta(Q_1)\|x - x_1\|_1^2. \tag{3.6}$$

Ainsi, (3.4) est vérifiée et $c > 0$ puisque $Q_1 \succ 0$. Il reste à montrer que la fonction $h(.) = g(.) - f(.)$ est lipschitzienne sur Δ_n de module β ce qui est immédiat. En effet, comme h est continue et différentiable, on peut utiliser le théorème des accroissements finis pour obtenir :

$$\forall\ (x, y) \in \Delta_n,\ |h(x) - h(y)| \leq \sup_{x \in \Delta_n} (\|\nabla h(x)\|_\infty)\|x - y\|_1.$$

Et pour tout $x \in \Delta_n$:

$$\|\nabla h(x)\|_\infty = \|(Q_2 - Q_1)x - k(\rho_2 - \rho_1)\|_\infty \leq \|Q_2 - Q_1\|_\infty + k\|\rho_2 - \rho_1\|_\infty = \beta. \tag{3.7}$$

Il suffit ensuite d'appliquer la proposition 3.8.1 rappelée en fin de ce chapitre.

□

Dans le cadre de l'analyse de sensibilité des solutions des problèmes P et P', la norme $\|.\|_1$ a plus de sens que les autres puisqu'elle représente ici la variation de la composition du portefeuille et les frais de transactions dépendent de cette variation. Cependant, d'un point de vue de l'analyse de sensibilité, d'autres normes pourraient être utilisées. Si on utilise la norme $\|.\|_2$, on voit que $\|x_2 - x_1\|_2$ peut être majoré par une quantité dépendant de $\lambda_{\min}(Q_1)$. Certaines calibrations de la matrice de covariance en finance ([76],[63]) essaient précisément de corriger le spectre de la matrice de covariance empirique qui est mal conditionnée ([63]) afin de stabiliser le comportement des portefeuilles dans le temps. Typiquement, ces calibrations essaient d'élever les plus petites valeurs propres de la matrice de covariance empirique et de réduire les plus grandes. Les majorations que l'on peut donner pour $\|x_2 - x_1\|_2$ font intervenir des termes du type $\dfrac{1}{\lambda_{\min}(Q_1)}$. Ces majorations pourraient ainsi apporter une justification à de telles calibrations. C'est pourquoi nous les donnerons, tout en donnant les majorations pour $\|x_2 - x_1\|_1$. Enfin, l'utilisation de la norme $\|.\|_1$ donne typiquement des majorations faisant intervenir $\beta(Q_i)$ au lieu de $\lambda_{\min}(Q_i)$ qui est calculé facilement. Par exemple, pour le problème P :

$$\|x_2 - x_1\|_2 \leq \frac{2}{\lambda_{\min}(Q_1)}(\max_i\ \|C_i(Q_2 - Q_1)\|_2 + k\|\rho_2 - \rho_1\|_2), \tag{3.8}$$

où $C_i(Q)$ est la i-ème colonne de la matrice Q. Pour vérifier ce résultat, il suffit de suivre la preuve du théorème 3.3.1 et d'appliquer le lemme suivant.

Lemme 3.3.1 *Soit $Q \in \mathcal{S}_n$, alors* $\sup_{x \in \Delta_n}\ \|Qx\|_2 = \max_i\ \|C_i(Q)\|_2$.

Preuve. Soient $\tilde{q}_{ij}$ les éléments de la matrice $Q^T Q$. Nous avons $\tilde{q}_{ii} = \sum_j\ {q_{ji}}^2 = \|C_i(Q)\|_2^2$. Ainsi :

$$\sup_{x \in \Delta_n}\ \|Qx\|_2 = \sup_{x \in \Delta_n}\ (x^T Q^T Q x)^{\frac{1}{2}} = \max_i\,(e_i^T Q^T Q e_i)^{\frac{1}{2}} = \max_i\,(\tilde{q}_{ii})^{\frac{1}{2}} = \max_i \|C_i(Q)\|_2.$$

La deuxième égalité vient de la convexité du problème : le maximum est atteint en un point extrémal du domaine.

□

3.3.4 Analyse de sensibilité d'un problème quadratique et de P'

On trouvera la preuve des théorèmes de cette section à la fin de ce chapitre. Considérons la classe suivante de problèmes d'optimisation :

$$\begin{cases} \min \quad \frac{1}{2}\, x^T Q x + c^T\, x \\ x^T f_j = b_j, \;\; 1 \leq j \leq m_1, \\ x \in X, \end{cases} \tag{3.9}$$

où X est un ensemble non vide, fermé, convexe et les éléments $(f_j)_{1\leq j\leq m_1}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^{m_1}$ et $Q \succ 0$ des paramètres des problèmes de cette classe. On suppose que l'ensemble X peut être décrit par un ensemble d'inégalités du type $h_j(x) \leq 0$, $1 \leq j \leq m_2$ où les fonctions h_j sont des fonctions différentiables données . On suppose aussi qu'il existe $M > 0$ tel que pour tout $x \in X, \|\nabla h_j(x)\|_\infty \leq M$. L'ensemble X n'est décrit par aucune contrainte d'égalité. Le théorème suivant donne une analyse de sensibilité des solutions d'un problème de la classe ci-dessus.

Théorème 3.3.2 *Soient P_1 et P_2 deux problèmes appartenant à la classe décrite ci-dessus et correspondant aux paramètres $(Q_1, c_1, (f_j^1)_{1\leq j\leq m_1}, b^1)$ et $(Q_2, c_2, (f_j^2)_{1\leq j\leq m_1}, b^2)$. Supposons que l'hypothèse de Slater soit vérifiée pour P_1 et P_2 ce qui signifie que pour $i = (1,2)$ les $(f_j^i)_{1\leq j\leq m_1}$ sont indépendants et qu'il existe $x_0^i \in int(X)$ tels que $f_j^T x_0^i = b_j^i$ pour tout $1 \leq j \leq m_1$. Pour le problème P_i, la somme des multiplicateurs de Lagrange μ_j^i associés aux contraintes $h_j(x) \leq 0$ est alors majorée par :*

$$M_i = \frac{\frac{1}{2}\|Q_i\|_\infty D_1^2(X) + \|c_i\|_\infty D_1(X) - \bar{f}^i}{\min(-h_j(x_0^i), j = 1, \ldots, m_2)},$$

où $\bar{f}_i$ est une borne inférieure sur la valeur optimale du problème P_i et $D_1(X) = \max_{x\in X} \|x\|_1$. Pour $i = 1, 2$, soit U_i la matrice définie par $U_i(j,k) = {f_j^i}^T f_k^i$, $1 \leq j, k, \leq m_1$. Alors si x_i est la solution du problème P_i :

$$\|x_2 - x_1\|_1 \leq \frac{\beta}{2\,\beta(Q_1)} + \frac{\sqrt{\beta^2 + 4\,\beta(Q_1) D_1(X)(A^1 + A^2) \sum_{j=1}^{m_1} \|f_j^2 - f_j^1\|_\infty}}{2\,\beta(Q_1)} \tag{3.10}$$

où

$$\beta = \|c_2 - c_1\|_\infty + \|Q_2 - Q_1\|_\infty D_1(X) \quad et \quad A^i = \frac{1}{\|f_1^i\|_\infty} \left(\|c_i\|_\infty + D_1(X)\|Q_i\|_\infty + M M_i\right),$$

s'il y a une contrainte d'égalité et

$$A^i = \|U_i^{-1}\|_\infty \left(\|c_i\|_\infty + D_1(X)\|Q_i\|_\infty + M M_i\right) \sum_k \|f_k^i\|_1 \quad sinon.$$

L'idée de la preuve est de dualiser certaines contraintes de sorte que le sous problème de minimisation interne résolu pour calculer la fonction duale ait un ensemble de faisabilité fixe. On écrit les conditions d'optimalité du premier ordre pour ce problème et on majore les multiplicateurs de Lagrange. Remarquons que l'hypothèse de Slater pour le problème P_i (voir théorème 2.3.2 p.312 de [64]) est une condition nécessaire et suffisante pour que

l'ensemble des multiplicateurs de Lagrange soit borné. Le théorème ci-dessus peut ensuite être appliqué au problème de Markowitz P' puisque la contrainte de rendement est active pour P', mais cela fournit des bornes compliquées et x_0^i n'est pas nécessairement explicite. De plus, l'hypothèse de Slater n'a pas à priori d'interprétation financière. Par contre, l'hypothèse **H3**, spécifiant qu'il existe un actif i dont le rendement est strictement supérieur à ℓ, permet de donner une majoration plus simple pour $\|x_2 - x_1\|_1$ que celle fournie par le théorème 3.3.2. De plus, elle assure aussi que l'hypothèse de Slater sera satisfaite pour le problème P' (pour lequel la contrainte de rendement est écrite comme une contrainte d'inégalité). Remarquons que si tous les actifs ont des rendements moyens strictement supérieurs à ℓ, alors on investit forcément dans l'actif sans risque.
Définissons à présent $B = 2(B^1 + B^2)$ avec

$$B^i = \min_{\rho_i(j)>\ell} \frac{Q_i(j,j)}{\rho_i(j) - \ell} - \frac{\bar{\ell}^2}{\bar{\rho}_i^T {Q_i}^{-1} \bar{\rho}_i},\ i = 1, 2, \tag{3.11}$$

un majorant du multiplicateur de Lagrange associé à la contrainte de rendement pour $P'(\ell, \rho_i, Q_i)$ (ce résultat est prouvé dans la dernière section de ce chapitre).

Théorème 3.3.3 *Si $P'(\ell, \rho_1, Q_1)$ et $P'(\ell, \rho_2, Q_2)$ satisfont les hypothèses **H1**, **H2** et **H3** et si x_i est l'unique solution de $P'(\ell, \rho_i, Q_i)$, la sensibilité des solutions P' est donnée par :*

$$\|x_2 - x_1\|_1 \quad \leq \quad \frac{\|Q_2 - Q_1\|_\infty}{2\,\beta(Q_1)} + \frac{\sqrt{\|Q_2 - Q_1\|_\infty^2 + B\,\beta(Q_1)\|\rho_2 - \rho_1\|_\infty}}{2\,\beta(Q_1)}. \tag{3.12}$$

Une majoration pour $\|x_2 - x_1\|_2$ peut s'obtenir en utilisant la majoration précédente de $\|x_2 - x_1\|_1$ et en remplaçant $\beta(Q_1)$ par $\lambda_{\min}(Q_1)$ et $\|Q_2 - Q_1\|_\infty$ par $\|Q_2 - Q_1\|_2$. De plus, pour fixer les idées, si ε est tel que $\min(\|\rho_1\|_\infty, \|\rho_2\|_\infty) - \ell > \varepsilon$, on peut remplacer B par $\frac{2(\|Q_1\|_\infty + \|Q_2\|_\infty)}{\varepsilon}$. La proposition 4.37, p.291 de [18] donne une analyse de sensibilité locale pour un problème d'optimisation générique où à la fois la fonction objectif et l'ensemble de faisabilité varient. Si $C(\ell, \rho)$ est l'ensemble de faisabilité de $P'(\ell, \rho, Q)$, la majoration fournie pour $\|x_2 - x_1\|$ par cette proposition dépend de la distance de Hausdorff $Haus(C(\ell, \rho_1), C(\ell, \rho_1) \cap C(\ell, \rho_2))$. En utilisant la borne de Hoffman ([59]), on obtient alors une majoration du type $\tau(\rho_1, \rho_2)\|\rho_2 - \rho_1\|$ pour la distance de Hausdorff mais comme $\tau(\rho_1, \rho_2)$ est inconnu, la majoration est locale et non explicite. Pour le problème P', l'hypothèse de Slater implique la condition de qualification de Robinson. La proposition 4.41 de [18] peut ainsi être appliquée pour obtenir :

$$\exists K > 0, \ \mid \ Haus(C(\ell, \rho_1), C(\ell, \rho_1) \cap C(\ell, \rho_2)) \leq K\|\rho_2 - \rho_1\|,$$

mais K n'est toujours pas explicite et l'analyse locale.
Nous donnons maintenant une analyse de sensibilité des solutions du problème de Markowitz (2.6) (introduit au chapitre précédent) tenant compte des frais de transactions. Nous considérons alors un problème $(\mathcal{M}_1)$ (resp. $(\mathcal{M}_2)$) du type (2.6) associé aux valeurs ρ_1 (resp. ρ_2) de la moyenne et Q_1 (resp. Q_2) de la matrice de covariance. Pour chacun de ces problèmes $(\mathcal{M}_i, i = 1, 2)$, nous faisons les hypothèses suivantes :

- **H1'.** Puisque le cas $r(n+1) \geq \tilde{\ell} = \frac{\ell\, \mathbf{e}^T x^-}{\mathbf{e}^T x^- - \sum_{i=1}^n \mu_i x_i^-}$ est résolu immédiatement (la solution est alors $y_i = x_i^-$ et $z_i = 0$) on supposera que $r(n+1) < \tilde{\ell}$.

- **H2'.** Pour que le problème de Markowitz soit faisable, un actif doit avoir un rendement moyen supérieur ou égal à ℓ. Nous supposerons en fait qu'il y a au moins un actif j dont le rendement moyen est strictement supérieur à $\frac{(1+\nu_j)\,\ell\,\mathbf{e}^T\,x^-}{\nu_j\,x_j^- + \mathbf{e}^T\,x^- - \sum_{k\neq j}\,\mu_k\,x_k^-}$. Si les frais de transaction sont nuls, il suffit qu'il existe un actif dont le rendement moyen soit strictement supérieur à ℓ.
- **H3'.** Finalement, nous supposerons que $Q_i \succ 0$.

Pour $1 \leq i \leq n$, posons

$$f(i) = x_i^- + \frac{x_{n+1}^- + \sum_{k\neq i}\,x_k^- - \sum_{k\neq i}\,\mu_k\,x_k^-}{1+\nu_i},$$

$J(i) = \{k \in \{1,\ldots,n\} \mid f(k)\,\rho_i(k) > \ell\,\mathbf{e}^T\,x^-\}$ ($\neq \emptyset$ d'après **H2'**), et $A = 2(A_1 + A_2)$ avec $A_i = \min\limits_{j\,\in\,J(i)} \dfrac{f(j)^2\,Q_i(j,j)}{f(j)\,\rho_i(j) - \ell\,\mathbf{e}^T\,x^-}$.

Théorème 3.3.4 *Si $(\mathcal{M}_1)$ et $(\mathcal{M}_2)$ satisfont les hypothèses **H1'**, **H2'**, et **H3'** et si x_i est l'unique solution de $(\mathcal{M}_i)$ alors :*

$$\|x_2 - x_1\|_1 \;\leq\; \frac{\|Q_2 - Q_1\|_\infty\,\|x^-\|_1}{2\,\beta(Q_1)} + \frac{\sqrt{\|Q_2 - Q_1\|_\infty\,\|x^-\|_1^2 + A\,\beta(Q_1)\|\rho_2 - \rho_1\|_\infty\,\|x^-\|_1}}{2\,\beta(Q_1)}.$$

3.4 Contreparties robustes et stables

3.4.1 Pourquoi corriger les estimations adaptatives

Dans le chapitre précédent, nous avons montré, qu'on peut déterminer des estimations $\hat{\rho}_\tau$ et $\hat{Q}_\tau$ de ρ_τ et Q_τ et des fonctions $\eta_\rho(\lambda,n,N)$ et $\eta_Q(\lambda,n,N)$ (N est le nombre données utilisées pour déterminer les estimations) telles que :

$$\|\hat{\rho}_\tau - \rho_\tau\|_\infty \leq \eta_\rho(\lambda,n,N), \quad \|\hat{Q}_\tau - Q_\tau\|_\infty \leq \eta_Q(\lambda,n,N), \tag{3.13}$$

avec des probabilités fonction de λ. En particulier, on sait choisir λ de telle sorte que les inégalités (3.13) soient vérifiées avec une probabilité arbitrairement grande. De plus, on a aussi montré dans le chapitre précédent que la solution du problème $P'(\ell, \hat{\rho}_\tau + \eta_\rho(\lambda,n,N)\mathbf{e}, \widehat{Q}_\tau)$ fournit une bonne approximation de la solution du problème $P'(\ell,\rho_\tau,Q_\tau)$. Néanmoins, on peut émettre quelques griefs contre l'utilisation des estimateurs empiriques (dans le cas stationnaire) ou des estimateurs empiriques adaptatifs (dans le cas général). En effet, le rang de $\hat{Q}_\tau$ est inférieur ou égal à N et donc si $n \geq N+1$, $\hat{Q}_\tau$ est dégénérée. En pratique, on réalise que même si le nombre de données utilisées est bien supérieur au nombre d'actifs, l'estimation empirique $\hat{Q}_\tau$ utilisant ces données est mal conditionnée. En effet, nous avons conduit des expériences sur 28 actifs du Dow Jones (en utilisant des cours d'actifs de janvier 1999 à janvier 2002) et dans la plupart des cas, la moitié des valeurs propres de la matrice de covariance empirique (calculée en utilisant différents jeux de données) était presque nulle et le conditionnement autour de 10^7. Cependant, si les estimations empiriques ou empiriques adaptatives ont des défauts, elles contiennent de l'information et permettent, non seulement d'obtenir des majorations sur les erreurs faites en les utilisant, mais aussi d'obtenir une estimation raisonnable de la solution. L'objet de cette section est donc de proposer des méthodes de correction des estimations empiriques ou empiriques

adaptatives de la moyenne et de la matrice de covariance. Tout d'abord, les estimations empiriques étant sujettes à des erreurs, des contreparties robustes sont envisagées. Ensuite, nous présenterons une manière de déterminer une estimation de la matrice de covariance conduisant à des portefeuilles plus stables.

3.4.2 Contreparties robustes du problème de Markowitz

3.4.2.1 Formulation générale du problème

On cherche à déterminer des contreparties robustes du problème de Markowitz P'. Une étude similaire pourrait être faite pour P. Remarquons que, pour des raisons techniques, dans la section précédente, la fonction objectif du problème de Markowitz P' était $\frac{1}{2}x^T Q_\tau x$. Dans cette section seulement, pour alléger les notations, nous utiliserons $x^T Q_\tau x$ comme fonction objectif du problème P', ce qui ne change pas le problème. Les estimations statistiques que l'on peut proposer pour les paramètres du modèle de Markowitz (comme les estimations empiriques adaptatives) sont sujettes à des erreurs. Cependant, des procédures statistiques, comme celle présentée dans le chapitre précédent, permettent d'obtenir des régions de confiance sur la moyenne et la matrice de covariance. Les contreparties robustes visent à utiliser cette information. Elles permettent au gestionnaire de voir l'influence de changements continus dans les paramètres. Ainsi, on n'utilise pas une estimation ponctuelle (ρ_τ, Q_τ) des paramètres du modèle mais on suppose qu'ils appartiennent à un ensemble convexe donné $\mathcal{U}$ de $\mathbb{R}^n \times \mathcal{S}_n$. En supposant que $\mathcal{U}$ est le produit direct de deux ensembles : $\mathcal{U} = \mathcal{U}_\rho \times \mathcal{U}_Q$ où $\mathcal{U}_\rho \subseteq \mathbb{R}^n$ et $\mathcal{U}_Q \subseteq \mathcal{S}_n$, la contrepartie robuste du problème de Markowitz est :

$$\begin{cases} \min f_{wc}(x) \\ (\rho_\tau - r(n+1)\mathbf{e})^T x \geq \ell - r(n+1), \quad \forall\, \rho_\tau \in \mathcal{U}_\rho \\ x \in \Delta_n, \end{cases} \tag{3.14}$$

où

$$f_{wc}(x) = \max\ x^T Q_\tau x \quad \text{sous la contrainte} \quad Q_\tau \in \mathcal{U}_Q, \tag{3.15}$$

est la fonction objectif pire cas de la contrepartie robuste. Nous allons maintenant nous concentrer sur différents ensembles d'incertitude $\mathcal{U}$.

3.4.2.2 Contraintes de boîte sur la moyenne et la matrice de covariance

Pour un choix approprié de $\lambda > 0$, et sous l'hypothèse 1 sur les rendements faite dans le chapitre précédent, nous avons une grande probabilité que :

$$\|\hat{\rho}_\tau - \rho_\tau\|_\infty \leq \eta_\rho(\lambda, n, N) \ \text{ et } \ \|\hat{Q}_\tau - Q_\tau\|_\infty \leq \eta_Q(\lambda, n, N).$$

En utilisant cette information, on peut calculer le portefeuille de Markowitz pire cas minimisant $f^1_{wc}(x)$ sous les contraintes :

$$x \in \Delta_n \ \text{ et } \min_{\|\rho_\tau - \hat{\rho}_\tau\|_\infty \leq\, \eta_\rho} (\rho_\tau - r(n+1)\mathbf{e})^T x \geq \ell - r(n+1),$$

avec $f^1_{wc}(x) = \max\ \ x^T Q_\tau x$ sous contrainte $\|Q_\tau - \hat{Q}_\tau\|_\infty \leq \eta_Q(\lambda, n, N)$. Si E est une matrice de 1, alors $f^1_{wc}(x) = x^T \tilde{Q}_\tau x$ avec $\tilde{Q}_\tau = \hat{Q}_\tau + \eta_Q(\lambda, n, N)E$. On peut ainsi voir $\tilde{Q}_\tau$ comme une matrice de covariance "robuste" calculée facilement à partir de $\hat{Q}_\tau$ (la matrice Q_τ est en fait une variable dans (3.15)).

La contrainte de rendement, initialement donnée par $x^T(\hat{\rho}_\tau - r(n+1)\mathbf{e}) \geq \ell - r(n+1)$ (si on utilise les estimations empiriques adaptatives de la moyenne) devient maintenant :

$$x^T(\hat{\rho}_\tau - (r(n+1) + \eta_\rho(\lambda, n, N))\mathbf{e}) \geq \ell - r(n+1).$$

Ainsi, cela revient à réduire l'estimation du rendement moyen de chaque actif risqué d'une quantité $\eta_\rho(\lambda, n, N)$.

3.4.2.3 Incertitude polytopique

On se focalise à présent sur le cas où l'on sait simplement que (ρ_τ, Q_τ) appartient à un polytope $\mathcal{U}$ défini par ses sommets. Plus précisément, (ρ_τ, Q_τ) appartient à l'enveloppe convexe des sommets (ρ_i, Q_i) $1 \leq i \leq m$ (sans perte de généralité, nous avons supposé, par souci de simplicité, que le nombre de sommets de chaque polytope est le même) :

$$\mathcal{U} = \mathbf{Co}\{(\rho_1, Q_1), \ldots, (\rho_m, Q_m)\}.$$

Les sommets (ρ_i, Q_i) pourraient être différentes estimations des rendements moyens et de la matrice de covariance. Ainsi, $Q_i \succeq 0$. Par exemple, parmi les matrices Q_i nous pourrions choisir la matrice de covariance empirique adaptative, une estimation multi-facteurs de la matrice de covariance ou une matrice de covariance expliquée par un seul index en utilisant un modèle CAPM (Roll,1977). De plus, $(\rho_\tau, Q_\tau) \in \mathcal{U}$ si et seulement si il existe α tel que :

$$(\rho_\tau, Q_\tau) = \sum_{i=1}^{m} \alpha_i(\rho_i, Q_i), \quad \alpha^T\mathbf{e} = 1, \quad \alpha \geq 0.$$

Pour entrer dans le cadre de travail spécifié en (3.14), on considère la cas d'incertitude polytopique indépendante. Les polytopes $\mathcal{U}_\rho$ et $\mathcal{U}_Q$ sont ainsi définis par

$$\mathcal{U}_\rho = \mathbf{Co}\{\rho_1, \ldots, \rho_m\}, \quad \mathcal{U}_Q = \mathbf{Co}\{Q_1, \ldots, Q_m\}.$$

Nous pourrions faire un parallèle entre cette approche robuste et la méthode développée dans [16] et [66] pour décrire le modèle de gestion de portefeuilles de Black Littermann. L'estimation finale des rendements moyens ρ_τ se fait à partir d'un ensemble d'estimations ρ_i de ces rendements moyens. Un certain nombre de ces 'vues' sont données par des experts, chacun d'entre eux ayant une perception particulière de l'évolution des rendements sur la période d'investissement. L'idée est ensuite d'assigner à la i-ème 'vue' ρ_i un niveau de confiance $\alpha_i \geq 0$. L'estimation des rendements moyens est alors $\sum_{i=1}^m \alpha_i\rho_i$ qui appartient à $\mathcal{U}_\rho$.
Comme $\min_{\rho\in\mathcal{U}_\rho} (\rho - r(n+1)\mathbf{e})^T x = \min_{1\leq i\leq m} (\rho_i - r(n+1)\mathbf{e})^T x$ et $\max_{Q\in\mathcal{U}_Q} x^TQx = \max_{1\leq i\leq m} \|Q_i{}^{\frac{1}{2}}x\|_2^2$, le modèle de Markowitz pire cas avec incertitude polytopique revient à résoudre le problème conique du second ordre suivant :

$$\min t \quad : \quad t \geq \|Q_i{}^{\frac{1}{2}}x\|_2, \quad x \in \Delta_n, \quad (\rho_i - r(n+1)\mathbf{e})^T x \geq \ell - r(n+1), \quad 1 \leq i \leq m.$$

Ce résultat est utilisable si le nombre de scénarios n'est pas trop élevé.

3.4.3 Calibration stable de la matrice de covariance

Les idées de cette section font suite à l'étude faite dans la section 3.3 et aux remarques du paragraphe 3.4.1. Nous commençons par expliquer ce que l'on entend par calibration 'stable'.

3.4.3.1 Motivations

Nous pouvons voir l'étape du choix du portefeuille comme une boîte noire dont les entrées sont la prédiction des rendements moyens sur la période d'investissement et la prédiction de la matrice de covariance, et fournissant en sortie un portefeuille. La composition du portefeuille sera stable par rapport aux entrées, si de faibles perturbations des rendements et donc des entrées produisent de faibles changements dans la composition du portefeuille. Un tel comportement est souhaitable pour essentiellement trois raisons :

- Tout d'abord, c'est intéressant en soi, puisque les gestionnaires de portefeuilles préfèrent les portefeuilles stables.
- D'autre part, lorsque les portefeuilles sont rebalancés, plus ils sont stables, moins il y a de frais de transactions. Bien sûr, un compromis doit être trouvé entre efficacité et stabilité. Lorsque les portefeuilles sont rebalancés régulièrement, si x_i est le portefeuille déterminé pour la i-ème période, une idée simple serait d'ajouter la contrainte $\|x_{i+1} - x_i\| \leq \varepsilon$ pour déterminer x_{i+1}. Cependant cette approche n'est pas automatique et pose un autre problème : celui du choix de ε. De plus, elle rend le modèle trop rigide.
- Enfin, si les entrées que l'on utilise sont proches des vraies entrées inconnues et si l'étape de sélection est stable, la composition du portefeuille produit devrait être proche du portefeuille optimal correspondant aux vraies entrées.

On cherche donc à obtenir une telle stabilité tout en déterminant une matrice de covariance 'raisonnable'. D'autre part, il serait souhaitable que le problème de Markowitz approché soit alors faisable.

3.4.3.2 Maximisation de la plus petite valeur propre

D'après (3.8), la plus petite valeur propre de la matrice de covariance utilisée pour résoudre P devrait avoir une valeur significative afin de générer des portefeuilles dont la composition soit stable. On peut aussi adapter les preuves des théorèmes 3.3.3 et 3.3.4 pour majorer $\|x_2 - x_1\|_2$ par des quantités du type $\frac{f(\rho_1,\rho_2,Q_1,Q_2)}{\lambda_{\min}(Q_1)}$ où $f(\rho_1, \rho_2, Q_1, Q_2)$ est borné. En particulier, la stabilité n'est pas garantie si l'estimation de la matrice de covariance est dégénérée. Dans le but de corriger la dégénérescence de la matrice de covariance empirique, [63] propose d'imposer un seuil $\alpha > 0$ devenant la valeur de la plus petite valeur propre de la matrice de covariance empirique corrigée. Cependant, la calibration de cet α est difficile. Notre idée, est de maximiser la plus petite valeur propre de la matrice de covariance, tout en imposant des contraintes de boîte sur les composantes de la matrice recherchée. Nous choisirons les contraintes de boîte obtenues dans le chapitre 2, centrées sur la matrice de covariance empirique adaptative, et rappelées dans les inégalités (3.13). Comme Q est une matrice de covariance, on impose aussi $Q \succeq 0$. Ainsi, la matrice de covariance Q que l'on recherche est solution du problème suivant :

$$\begin{cases} \max \lambda_{\min}(Q) \\ \|Q - \hat{Q}_\tau\|_\infty \leq \eta_Q(\lambda, n, N), \quad Q \succeq 0. \end{cases} \tag{3.16}$$

C'est un problème d'optimisation convexe non différentiable. Ce problème est équivalent au problème SDP (3.17) suivant qui peut être efficacement résolu par des méthodes de

points intérieurs :

$$\begin{cases} \min -u \\ v_{ij} + u\delta_{ij} + Y_{ij} = \eta_Q + \hat{Q}_\tau(i,j), \\ w_{ij} - u\delta_{ij} - Y_{ij} = \eta_Q - \hat{Q}_\tau(i,j),\ v_{ij} \geq 0,\ \ w_{ij} \geq 0,\ \ Y \succeq 0. \end{cases} \tag{3.17}$$

Q est ensuite donnée par $Y^* + u^* I$ avec Y^* et u^* les valeurs optimales de Y et u dans (3.17).

3.4.3.3 Optimisation du conditionnement

Si la vente à découvert est autorisée, on a une expression explicite des solutions optimales et on peut alors majorer $\|x_2 - x_1\|_2$ par une quantité du type $f(\rho_1, \rho_2, Q_1, Q_2)\frac{\lambda_{\max}(Q_1)}{\lambda_{\min}(Q_1)}$ où $f(\rho_1, \rho_2, Q_1, Q_2)$ est borné. De plus, on sait que la matrice de covariance empirique des rendements d'actifs financiers a tendance à surestimer les plus grandes valeurs propres et sous-estimer les plus petites ([63]). On peut alors essayer de trouver le meilleur conditionnement possible pour la matrice de covariance, tout en imposant les mêmes contraintes de boîte que précédemment sur les composantes de cette matrice. La matrice de covariance est alors solution du problème :

$$\begin{cases} \min \dfrac{\lambda_{\max}(Q)}{\lambda_{\min}(Q)} \\ \|Q - \hat{Q}_\tau\|_\infty \leq \eta_Q(\lambda, n, N), \quad Q \succeq 0. \end{cases} \tag{3.18}$$

C'est un problème quasi convexe. Il est équivalent de résoudre :

$$\begin{cases} \min t \\ s \leq \lambda_{\min}(Q) \\ v \geq \lambda_{\max}(Q) \\ v \leq ts \\ \|Q - \hat{Q}_\tau\|_\infty \leq \eta_Q(\lambda, n, N), \quad Q \succeq 0. \end{cases} \tag{3.19}$$

On peut alors trouver une solution de ce problème par dichotomie.

3.5 Autres calibrations de la matrice de covariance

Nous donnons ici d'autres méthodes de calibration de la matrice de covariance basées sur des techniques de moindres carrés semidéfinis positifs et préoccupées par la stabilité des portefeuilles. L'idée commune à ces approches consiste à rechercher la matrice la plus proche de la matrice de covariance empirique adaptative satisfaisant un certain nombre de contraintes, assurant en particulier, que la matrice résultante soit définie positive. Nous considérons les contraintes suivantes. Tout d'abord, nous introduisons une contrainte de conservation de la trace de la matrice de covariance empirique adaptative $\hat{Q}_\tau$: $< X, I >=< \hat{Q}_\tau, I >$. Cette contrainte est introduite et commentée dans [63]. Ensuite, on utilise m portefeuilles $(q_i)_{i=1,\ldots,m}$. On peut estimer la variance $\hat{\sigma}_i^2$ du rendement du portefeuille q_i et imposer que l'estimation $q_i^T X q_i$ de la variance du rendement du portefeuille q_i, obtenue en utilisant la matrice de covariance X, soit égale à $\hat{\sigma}_i^2$. Si on suppose le processus des rendements stationnaire, toutes les données seront utilisées pour calculer $\hat{\sigma}_i^2$.

Sous l'hypothèse d'homogénéité temporelle locale, seules les données de l'intervalle d'homogénéité temporelle locale sont utilisées pour calculer $\hat{\sigma}_i^2$. La variance $\hat{\sigma}_i^2$ du portefeuille q_i devrait alors être calculée en utilisant les rendements des actifs sur l'intervalle d'homogénéité. Finalement, on impose une dernière contrainte précisant que la matrice solution X doit vérifier $X \succeq \alpha I$. Cela conduit au problème suivant :

$$\begin{cases} \min \quad \|X - \hat{Q}_\tau\| & \\ < X, I >=< \hat{Q}_\tau, I >, & (a) \\ < X, q_i q_i^T >= \hat{\sigma}_i^2, \quad 1 \leq i \leq m, & (b) \\ X \succeq \alpha I. & (c) \end{cases} \tag{3.20}$$

Dans ce qui suit, cette méthode de correction de la matrice $\hat{Q}_\tau$ sera désignée par C_1. On peut aussi s'intéresser à des cas particuliers de cette méthode. Si les contraintes (a) et (b) sont ôtées (calibration C_2) et si la décomposition spectrale de $\hat{Q}_\tau$ s'écrit $\hat{Q}_\tau = \sum_{i=1}^{n} \lambda_i(\hat{Q}_\tau) v_i v_i^T$, où v_i est le i-ème vecteur propre de la matrice $\hat{Q}_\tau$ associé à la valeur propre $\lambda_i(\hat{Q}_\tau)$, alors la solution du problème (3.20) s'écrit $X = \sum_{i=1}^{n} \max(\lambda_i(\hat{Q}_\tau), \alpha) v_i v_i^T$. Un autre cas particulier pour lequel nous avons une solution explicite est le cas où (a) est ôtée, $\alpha = 0$ et les portefeuilles choisis pour les contraintes (c) constituent une base orthonormale de vecteurs propres de la matrice $\hat{Q}_\tau$ (calibration C_3).

Proposition 3.5.1 *Considérons le problème d'optimisation* (3.20) *où (a) est ôtée,* $m = n$ *est la dimension de la matrice* $\hat{Q}_\tau$, $\alpha = 0$ *et les vecteurs* q_i *constituent une base orthonormale de vecteurs propres de la matrice* $\hat{Q}_\tau$. *Alors la solution de* (3.20) *est donnée par :* $X^* = \sum_{i=1}^{n} \hat{\sigma}_i^2 \, q_i q_i^T$.

Preuve. L'hypothèse de Slater étant satisfaite, $(X^*, Z^*, (\mu_i^*)_{1 \leq i \leq n})$ constituent une solution primale-duale du problème (3.20) si et seulement si :

$$\begin{cases} X^* \succeq 0, \;\; Z^* \succeq 0, \;\; < X^*, Z^* >= 0, & (a') \\ < X^*, q_i q_i^T >= \hat{\sigma}_i^2, & (b') \\ X^* = \hat{Q}_\tau + Z^* - \sum_{i=1}^{n} \mu_i^* q_i q_i^T. & (c') \end{cases} \tag{3.21}$$

Les conditions (a') donnent $X^* Z^* = 0$ et comme $X^* \succ 0$, nous avons $Z^* = 0$. La condition (c') est ainsi satisfaite avec $\mu_i^* = \lambda_i(\hat{Q}_\tau) - \hat{\sigma}_i^2$ où $\lambda_i(\hat{Q}_\tau)$ est la valeur propre de la matrice $\hat{Q}_\tau$ associée au vecteur propre q_i. Finalement, (b') est vérifiée :

$$< X^*, q_i q_i^T >= \sum_{j=1}^{n} \hat{\sigma}_j^2 Tr(q_j q_j^T q_i q_i^T) = \sum_{j=1}^{n} \hat{\sigma}_j^2 \delta_{ij} Tr(q_j q_i^T) = \hat{\sigma}_i^2 Tr(q_i q_i^T) = \hat{\sigma}_i^2 \|q_i\|_2^2 = \hat{\sigma}_i^2.$$

□.

Nous pourrions aussi enlever les contraintes (b) de (3.20) (calibration C_4).
Nous présentons une dernière calibration utilisant la théorie des groupes de Kato. Les valeurs propres $(\lambda_i(\hat{Q}_\tau))_{1 \leq i \leq n}$ (triées par ordre croissant) de la matrice de covariance empirique adaptative $\hat{Q}_\tau$ peuvent se concevoir comme des niveaux de risque. Si ε est la distance

minimale entre deux niveaux de risque, on peut partitionner les valeurs propres en différents groupes appelés groupes de Kato. Ces groupes de Kato G_j sont définis de la façon suivante :

Initialisation. $i = j = 1$, $G_j = \{\lambda_i(\hat{Q}_\tau)\}$.

Tant que (i<n)
si ($\lambda_i(\hat{Q}_\tau)$-$\lambda_{i+1}(\hat{Q}_\tau) > \varepsilon$) alors j=j+1, **Fin si,**
i=i+1, $\lambda_i(\hat{Q}_\tau) \in G_j$,
Fin tant que

On associe ensuite au groupe de Kato G_j le risque :

$$\hat{\lambda}_j = \frac{1}{n_j} \sum_{i \in G_j} \lambda_i(\hat{Q}_\tau),$$

où n_j est le nombre d'éléments dans le groupe G_j. Soit q_j le vecteur propre normé de la matrice $\hat{Q}_\tau$ associé à la valeur propre $\lambda_j(\hat{Q}_\tau)$ et soit $P_j(\hat{Q}_\tau)$ le projecteur sur l'espace engendré par les vecteurs propres du groupe de Kato G_j :

$$P_j(\hat{Q}\tau) = \sum_{i \in G_j} q_i q_i^T.$$

L'opérateur P_j est lipschitzien. En utilisant les notations ci-dessus, une correction possible de la matrice de covariance consiste à déterminer la matrice semidéfinie positive X la plus proche de la matrice $\hat{Q}_\tau$ et telle que la somme des risques (estimés avec la matrice de covariance X) des portefeuilles du groupe G_j soit la somme des risques de ces portefeuilles estimés avec la matrice de covariance $\hat{Q}_\tau$:

$$\begin{cases} \min \quad \|X - \hat{Q}_\tau\| \\ < P_j(\hat{Q}\tau), X >= n_j \hat{\lambda}_j, \quad j = 1, \ldots, K, \\ X \succeq 0, \end{cases} \tag{3.22}$$

s'il y a K groupes de Kato. Cette calibration sera notée C_5. Dans le même esprit, une façon simple de procéder (calibration C_6), revient à remplaçer $\hat{Q}_\tau = \sum_{i=1}^{n} \lambda_i(\hat{Q}_\tau) q_i q_i^T$ par $\hat{Q}_\tau = \sum_{i=1}^{n} \tilde{\lambda}_i(\hat{Q}_\tau) q_i q_i^T$, où $\tilde{\lambda}_i = \hat{\lambda}_j$ si $i \in G_j$. Nous noterons enfin C_7 la première calibration de la matrice de covariance présentée en section 3.4.3.

3.6 Résultats numériques

Les simulations numériques seront effectuées en utilisant des données réelles (les cours des actifs du Dow Jones utilisés dans le chapitre 2) et des données simulées.

3.6.1 Tests de stabilité

Le but de cette section est d'illustrer, via des simulations sur les données réelles, l'influence de l'augmentation de la plus petite des valeurs propres de la matrice de covariance

empirique $\hat{Q}_N$ sur la sensibilité de la composition du portefeuille optimal aux perturbations des données. Nous confronterons les comportements des portefeuilles optimaux obtenus en utilisant la matrice de covariance empirique $\hat{Q}_N$ et ses corrections C_2, C_6 et C_7.

3.6.1.1 Réduction du conditionnement

Nous commençons par illustrer dans quelle mesure le conditionnement de $\hat{Q}_\tau$ est réduit en utilisant les calibrations de Q proposées en section 3.4.3. Nous prenons une matrice de covariance empirique dont le conditionnement est 1.1078×10^6. Nous calculerons ensuite le conditionnement de différentes matrices Q obtenues à partir de cette matrice de covariance empirique en résolvant (3.17) et (3.19) (notée 'Min Cond' par la suite) pour les valeurs suivantes de η_Q : $\eta_Q^1 = 0.01\lambda_{\max}(\hat{Q}_\tau)$, $\eta_Q^2 = 0.05\lambda_{\max}(\hat{Q}_\tau)$, $\eta_Q^3 = 0.1\lambda_{\max}(\hat{Q}_\tau)$. Les résultats sont donnés dans le tableau 3.1 ci-après.

Méthode	η_Q^1	η_Q^2	η_Q^3
C_7	80.24	16.25	6.91
Min Cond	70.85	9.06	2.29

TAB. 3.1 – Influence des corrections (3.17) et (3.19) sur le conditionnement.

Le conditionnement diminue donc fortement même si on n'autorise que de petites variations des entrées de $\hat{Q}_\tau$.

3.6.1.2 Évolution de la composition du portefeuille dans le temps

Afin d'observer l'influence de l'augmentation de $\lambda_{\min}(\hat{Q}_N)$ sur la composition des portefeuilles, nous faisons l'expérience suivante. Un premier investissement est fait en date du 2 janvier 1999 (nous notons cette date t_0) ; l'horizon d'investissement est de 60 jours (ce qui correspond approximativement à 3 mois calendaires), le taux sans risque annuel est 5% et le rendement cible pour ces 60 jours est 2.5%. On analyse l'influence du paramètre α de la méthode C_2 sur la stabilité de la composition des portefeuilles. On donne à α les valeurs $\alpha(i) = 10^i\,\lambda_{\max}(\hat{Q}_N)$ pour $-6 \leq i \leq -1$ et pour chaque valeur $\alpha(i)$ de α on détermine les portefeuilles optimaux $x^j_{\alpha(i)}$ aux dates $t_0 + 60j$, $0 \leq j \leq 11$, et on calcule $p(i) = \frac{1}{11}\sum_{j=0}^{10} \|x^{j+1}_{\alpha(i)} - x^j_{\alpha(i)}\|_1$. L'évolution de $p(i)$ en fonction de i est représentée sur la figure 3.1 ci-après. On constate ainsi que l'augmentation de $\lambda_{\min}(Q)$ tend à stabiliser la composition des portefeuilles. Cela a été observé en utilisant différentes dates initiales t_0, différents rendements cibles et différents taux sans risque. Cependant, nous avons remarqué que si on utilise des valeurs trop grandes de α (disons plus grandes que $\lambda_{\max}(\hat{Q}_N)$) il est possible, pour certaines valeurs des paramètres, que la variation moyenne de la composition des portefeuilles (en tant que fonction de α) augmente légèrement avant de se stabiliser.

Nous comparons maintenant les méthodes 'empiriques', C_2, C_6, et C_7. On choisit $\alpha = 0.01\lambda_{\max}(\hat{Q}_N)$. En outre, on choisit $\varepsilon = \alpha$ pour le paramètre ε de la méthode C_6. On choisit aussi $\eta_Q = \alpha$. La date du premier investissement est le 2 janvier 1999 (date que nous notons t_0), l'horizon d'investissement est toujours 60 jours, le rendement cible 4% et le taux

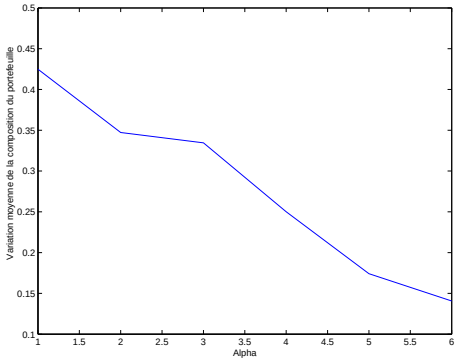


FIG. 3.1 – Sensibilité de la variation moyenne de la composition du portefeuille (ordonnée) en fonction de α (abscisse).

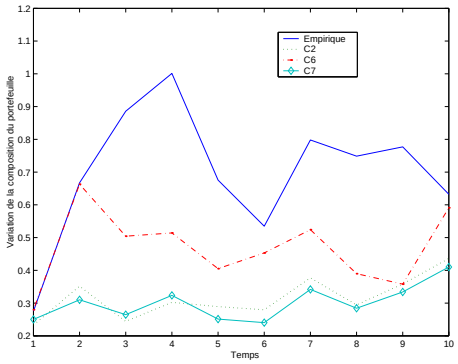


FIG. 3.2 – Variation de la composition du portefeuille dans le temps pour 4 différentes méthodes.

sans risque annuel 5%. Les portefeuilles sont régulièrement rebalancés tous les 60 jours à partir de la date t_0. Pour le i-ème rebalancement, on détermine un portefeuille x_M^i pour chaque méthode M. La figure 3.2 ci-dessus représente l'évolution de $(\|x_M^i - x_M^{i-1}\|_1)_{i\geq 2}$ en fonction de i et pour chaque méthode. Ceci tend aussi à montrer que l'augmentation de $\lambda_{\min}(\hat{Q}_N)$ permet de stabiliser la composition des portefeuilles. Les méthodes C_2 et C_7 semblent particulièrement stables dans cet exemple. Pour ces méthodes, la modification de la composition du portefeuille est toujours moins importante que pour la méthode 'Empirique'. La méthode C_6 est aussi plus stable que la méthode 'Empirique'. La même expérience a été conduite en utilisant différentes valeurs pour les paramètres du modèle de Markowitz. On a en particulier utilisé différentes dates initiales t_0 (du 2 janvier 1999 au 2 janvier 2000 par pas de 10 jours), différents horizons d'investissement (40 et 60 jours) et différents rendements cibles (2, 3 et 4%). Dans toutes les simulations, les méthodes C_2 et C_7 étaient les plus stables, conduisant toujours à des modifications de la composition des portefeuilles moins importantes que la méthode 'Empirique'. En terme de stabilité, la méthode C_6 se place troisième en moyenne.

3.6.1.3 Influence de la perturbation des moyennes sur la composition du portefeuille optimal

On se fixe une date t_0 (le 2 janvier 1999) et pour chaque méthode M, on estime le couple (ρ, Q) par $(\hat{\rho}_N, \hat{Q}_M)$ [$\hat{\rho}_N$ est la moyenne empirique des rendements et $\hat{Q}_M$ la correction de la matrice de covariance empirique par la méthode M]. À partir de ces estimations, on peut calculer le portefeuille optimal x_M associé à la méthode M. On fait ensuite n (i.e 30) itérations. À l'itération i, on envisage 4 perturbations qui consistent à remplacer $\hat{\rho}_N(i)$ par $\hat{\rho}_N(i) \pm 0.05|\hat{\rho}_N(i)|$, $\hat{\rho}_N(i) \pm 0.1|\hat{\rho}_N(i)|$. Chaque perturbation j produit un portefeuille x_M^{ij} pour la méthode M. Une comparaison de $\frac{1}{30*4}\sum_{i,j} \|x_M - x_M^{ij}\|_1$ peut ensuite être effectuée pour toutes les méthodes M. Cette expérience a été répétée 400 fois et a donné les résultats moyens suivants :

Empirique	C_2	C_7	C_6
0.0119	0.0060	0.0058	0.0085

Nous avons observé que dans ces expériences, la perturbation de ρ n'avait pas d'impact significatif sur la composition du portefeuille. Une fois de plus, la méthode C_7 semble la plus stable par rapport aux perturbations du vecteur des rendements moyens.

3.6.2 Comparaison des précisions

On s'intéresse maintenant à la qualité des problèmes Value-At-Risk et de Markowitz approchés (définis par (2.13) et (2.14) dans la chapitre précédent) en utilisant des données simulées. On observera également l'apport des corrections de la matrice de covariance adaptative présentées dans les sections 3.4 et 3.5. On choisit la même grille de valeurs que dans le chapitre précédent pour les paramètres de l'algorithme adaptatif. On suppose que les rendements sont Gaussiens et on simule, pour chaque combinaison des valeurs des paramètres de l'algorithme adaptatif, $N = 400$ réalisations d'un processus Gaussien ayant une rupture dans la moyenne et/ou la matrice de covariance (chaque réalisation correspond à un ensemble de valeurs possibles pour les rendements passés jusqu'à la date de l'investissement). Les rendements ne sont pas Gaussiens par morceaux en pratique (même si certains modèles font cette hypothèse) mais cela nous permettra de contrôler à nouveau le comportement de l'algorithme adaptatif, de mesurer la précision des problèmes approchés et d'observer la qualité des calibrations de la matrice de covariance présentées dans les sections 3.4 et 3.5. Les $T_1 = 360$ premières observations sont issues d'une loi normale $\mathcal{N}(\rho_1, Q_1)$ et les T_1 dernières observations d'une loi normale $\mathcal{N}(\rho_2, Q_2)$. La moyenne et la matrice de covariance sur la période d'investissement sont supposées prendre les valeurs ρ_2 et Q_2. On connaît ainsi les portefeuilles théoriques x^* obtenus en résolvant les problèmes Value-At-Risk et de Markowitz avec les valeurs (ρ_2, Q_2) des paramètres. Pour chaque réalisation i d'un tel processus, on peut calculer, pour chaque méthode M de calibration des paramètres, le portefeuille optimal x_i^M. On peut ensuite mesurer la précision des différentes méthodes (la précision est définie dans le chapitre précédent en section 2.2). On donne le quantile d'ordre 0.9 de la précision de chaque méthode (noté $A_{0.9}$). Notre définition de la précision ne tient pas compte de la proximité entre le portefeuille estimé et le portefeuille optimal x^*. C'est pourquoi, nous considérons aussi le critère d'erreur :

$$\Delta'_M = \frac{1}{N}\sum_{i=1}^{N} \|x_i^M - x^*\|_1.$$

	m_{theo}		σ_{theo}		$\hat{m}$	$\hat{\sigma}$		Δ'_M		Δ		$A_{0.9}$	
Method	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$
Emp	1.033	2.32	1.62	88	1.94	8.5	395	34	1010	6.62	295	1.68	40.1
Adap	1.025	2.30	0.67	59	2.3	0.7	57	3	320	1.23	120	0.35	2.5
Emp C_1	1.034	2.283	1.56	73	1.90	1.9	89	31	1	0.07	6	1.42	42
Adap C_1	1.025	2.30	1.07	82	2.31	0.3	26	22	1100	0.06	101	0.85	45.2
Emp C_2	1.034	2.296	1.53	74	1.91	8.6	393	31	1100	0.94	44	1.5	33.7
Adap C_2	1.025	2.299	0.70	59	2.30	0.8	61	9	450	0.85	77	0.35	4.14
Emp C_3	1.034	2.27	1.57	70	1.89	1.6	70	31	1380	0.08	3	1.42	46.4
Adap C_3	1.025	2.30	1.18	93	2.30	0.3	26	24	1370	0.54	63	1.02	59.7
Emp C_4	1.034	2.283	1.56	73	1.90	1.9	89	31	1	0.07	6	1.42	42
Adap C_4	1.025	2.297	0.74	60	2.30	0.7	55	10	540	0.95	82	0.38	7.41
Emp C_5	1.033	2.323	1.62	89	1.94	8.5	395	33	1010	6.62	295	1.7	40.1
Adap C_5	1.025	2.299	0.67	59	2.30	0.7	57	3	320	1.23	120	0.35	2.47
Emp C_6	1.033	2.325	1.78	102	1.94	8.4	396	32	1150	6.64	312	2.12	76
Adap C_6	1.025	2.294	0.78	60	2.30	0.7	53	11	620	1.19	96	0.45	12.9
Emp C_7	1.034	2.275	1.50	69	1.90	3.1	140	30	1340	0.38	19	1.37	43.3
Adap C_7	1.025	2.292	0.78	59	2.29	0.7	52	13	730	0.55	50	0.42	10.5

TAB. 3.2 – Comparaison des calibrations sur des données simulées. Rupture dans la moyenne.

Comme on s'y attendait, nous verrons que la méthode adaptative fournit aussi des portefeuilles plus proches du portefeuille optimal (i.e donnant une valeur de Δ'_M plus petite) que les portefeuilles obtenus en utilisant la méthode empirique. On calcule ensuite une estimation de la moyenne théorique m_{theo} (l'espérance empirique $\hat{E}\rho_2^T\hat{x}$ de $\rho_2^T\hat{x}$) du portefeuille, de l'écart type théorique σ_{theo} du portefeuille (l'estimation empirique $\hat{E}\sqrt{\hat{x}^T Q_2 \hat{x}}$) ainsi qu'une estimation $\hat{m}$ de la moyenne (l'espérance empirique $\hat{E}\hat{\rho}^T\hat{x}$) et $\hat{\sigma}$ de l'écart type (l'espérance empirique $\hat{E}\sqrt{\hat{x}^T\hat{Q}\hat{x}}$) du portefeuille. Finalement, nous mesurerons la variation de la composition des portefeuilles par :

$$\Delta^2 = \frac{1}{N}\, Tr(X - \bar{X})(X - \bar{X})^T, \tag{3.23}$$

où $X = [x_1, \ldots, x_N]$ est la matrice stockant les portefeuilles obtenus pour une méthode donnée, $\bar{X} = \bar{x}\,\mathbf{e}$, où $\mathbf{e}$ est un vecteur de 1 de taille N et $\bar{x}$ est la barycentre des portefeuilles. Δ^2 est ainsi la somme des variances des composantes. Les estimations des paramètres utilisées seront les estimations empiriques (utilisant les $2T_1$ données disponibles), les estimations empiriques adaptatives et leurs corrections définies en sections 3.4.3 et 3.5. On choisit 3 portefeuilles aléatoirement pour les contraintes (b) dans (3.20) et $\alpha = \varepsilon$ (pour définir les groupes de Kato) $= \eta_Q = 0.05\ \lambda_{\max}(\hat{Q}_\tau)$. On considère tout d'abord une rupture dans la moyenne : $Q_1 = Q_2 = Q$ et $\rho_1 = 1.5\,\rho$, $\rho_2 = 2.25\,\rho$ (ρ et Q sont respectivement la moyenne et la matrice de covariance empirique des rendements en date du 2 janvier 1999). Le rendement cible pour le modèle de Markowitz est $\ell = 1.025$, le taux sans risque $r(n+1) = 1.005$, les frais de transactions $\mu = \nu = 0.005$ et $\kappa(\varepsilon) = 1.64$. L'argent est initialement détenu dans l'actif sans risque.

Method	m_{theo}		σ_{theo}		$\hat{m}$	$\hat{\sigma}$		Δ'_M		Δ		$A_{0.9}$	
Method	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$
Emp	1.025	1.47	11.8	277	1.48	11.3	267	20.3	477	3.98	151	6.2	63
Adap	1.024	1.47	10.4	259	1.48	9.8	243	12.5	246	3.8	90	3.1	37
Emp C_1	1.025	1.49	13.5	334	1.49	3.5	86	30.0	700	1.1	34.5	9.0	142
Adap C_1	1.024	1.49	12.9	338	1.50	4.1	108	31.4	709	1.8	55.9	8.5	154
Emp C_2	1.025	1.47	11.8	277	1.48	11.3	267	20.3	477	3.98	151	6.2	63
Adap C_2	1.024	1.47	10.4	260	1.48	9.8	245	12.5	244	3.7	87.3	3.1	37
Emp C_3	1.025	1.49	13.6	329	1.49	8.2	197	29.0	652	5.7	133	10.9	189
Adap C_3	1.024	1.48	12.0	303	1.48	7.7	194	23.6	513	6.0	147	8.1	140
Emp C_4	1.025	1.47	11.8	278	1.48	11.3	266	20.3	477	2.0	150	6.2	63
Adap C_4	1.024	1.47	10.4	260	1.48	9.8	245	12.5	244	3.7	87.3	3.1	37
Emp C_5	1.025	1.47	11.8	278	1.48	11.3	266	20.2	477	2.0	150	6.2	63
Adap C_5	1.024	1.47	10.4	260	1.48	9.8	243	12.5	246	3.8	90.2	3.1	37
Emp C_6	1.025	1.47	11.9	280	1.48	11.3	267	20.9	492	2.1	152	6.3	65
Adap C_6	1.024	1.47	10.5	265	1.48	9.9	247	14.7	297	4.5	106	3.5	53
Emp C_7	1.025	1.46	12.0	276	1.47	11.6	267	21.9	545	1.7	206	6.5	65
Adap C_7	1.024	1.46	10.6	257	1.47	10.7	258	16.0	385	2.9	214	3.9	51

TAB. 3.3 – Comparaison des calibrations sur des données simulées. Rupture dans la matrice de covariance.

Dans le tableau 3.2 ci-avant où sont consignés les résultats de cette expérience, Emp désigne la méthode empirique et Adap la méthode adaptative. La valeur $\hat{m}$ de la moyenne estimée en utilisant le modèle de Markowitz n'est pas fournie. En effet, comme $r(n+1) \leq \ell$, la contrainte de rendement est active et la moyenne estimée est $\ell = 1.025$ pour toutes les méthodes. La moyenne théorique obtenue en utilisant la méthode empirique est légèrement supérieure à la moyenne obtenue en utilisant la méthode adaptative. Cependant, le vrai risque σ_{theo} du portefeuille est bien supérieur à la fois pour le modèle VaR et le modèle de Markowitz. La précision de la méthode adaptative est, comme prévu, bien meilleure que la précision de la méthode empirique. La méthode adaptative fournit une meilleure estimation des paramètres et des portefeuilles plus proches du portefeuille optimal x^*. Les rendements des portefeuilles obtenus en utilisant les différentes corrections de la matrice de covariance sont proches du rendement du portefeuille obtenu avec la méthode nominale. D'autre part, l'écart type du rendement de ce portefeuille est en général plus petit si on corrige l'estimation empirique et plus grand si on corrige l'estimation adaptative. Finalement, on observe une variation de la composition des portefeuilles moins importante (Δ est en général plus petit) si on utilise les différentes corrections de la matrice de covariance empirique (ou adaptative).

On considère ensuite une rupture dans la matrice de covariance : $Q_1 = I_{15}$, $Q_2 = \text{diag}([0.2 : 0.2 : 3])$ et $\rho_1 = \rho_2 = 1.5\ \rho(1:15)$. Les autres paramètres ont les mêmes valeurs que dans l'expérience conduite précédemment où l'on a envisagé une rupture dans la moyenne. Les résultats sont représentés dans le tableau 3.3 ci-dessus. Enfin, on considère une rupture à la fois dans la moyenne et la matrice de covariance : $\rho_1 = 1.5\ \rho$, $\rho_2 = 2.25\ \rho$, $Q_1 = Q$ et la

Method	m_{theo}		σ_{theo}		$\hat{m}$	$\hat{\sigma}$		Δ'_M		Δ		$A_{0.9}$	
Method	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$	$(\mathcal{M})$	$(\mathcal{V})$
Emp	1.034	2.3	2.6	128	1.9	8.6	393	26.4	849	3.4	123	2.4	45
Adap	1.025	2.28	1.38	95	2.28	1.26	88	4.9	356	1.1	85	0.42	11
Emp C_1	1.034	2.28	2.2	103	1.9	2.0	89	19	594	0.1	24	1.8	22
Adap C_1	1.025	2.28	1.5	111	2.29	0.43	33	9.7	667	0.08	76	0.76	41
Emp C_2	1.040	2.29	2.3	111	1.9	8.7	393	19.6	552	1.4	54	1.9	24
Adap C_2	1.025	2.28	1.38	95	2.28	1.3	88	4.9	356	1.1	85	0.4	11
Emp C_3	1.034	2.27	2.2	100	1.9	1.6	71	19.3	606	0.15	6	1.8	22
Adap C_3	1.025	2.29	1.8	130	2.29	0.8	52.3	12.8	857	1.4	92	1.4	78
Emp C_4	1.034	2.29	2.23	109	1.91	3.0	137	18.8	591	0.13	17	1.8	24
Adap C_4	1.025	2.28	1.4	95	2.28	1.3	88	4.9	356	1.1	85	0.4	11
Emp C_5	1.034	2.3	2.6	128	1.9	8.6	393	26	849	3.4	123	2.4	45
Adap C_5	1.025	2.28	1.4	95	2.3	1.3	88	4.9	356	1.1	85	0.4	11
Emp C_6	1.034	2.30	2.57	132	1.9	8.6	394	24	781	2.9	113	2.4	58
Adap C_6	1.025	2.28	1.4	100	2.28	1.3	91	7.3	498	0.9	72	0.6	17
Emp C_7	1.034	2.27	1.99	89	1.89	4.3	194	14.9	356	0.7	28	1.5	9
Adap C_7	1.032	2.28	1.36	94	2.28	1.3	87	4.1	301	0.74	63	0.39	8

TAB. 3.4 – Comparaison des calibrations sur des données simulées. Rupture dans la moyenne et dans la matrice de covariance.

matrice Q_2 a les mêmes covariances que Q_1 et des variances 9 fois supérieures. Les autres paramètres ont la même valeur que pour l'expérience conduite précédemment où l'on a envisagé une rupture dans la moyenne. Les résultats sont représentés dans le tableau 3.4 ci-dessus. Les remarques faites sur la première simulation restent valables pour ces deux derniers exemples.

3.6.3 Comparaison de la diversification des portefeuilles

Nous avons remarqué que l'utilisation des matrices de covariance corrigées tend à diversifier notablement plus les portefeuilles que si les matrices de covariance empiriques ou empiriques adaptatives étaient utilisées. Ce phénomène est le plus spectaculaire pour les corrections C_1 et C_3. Pour illustrer cela, on choisit quatre dates d'investissement. Pour chaque date, le portefeuille optimal est calculé en utilisant le modèle de Markowitz (2.6) avec $\ell = 1.025$, le modèle Value-At-Risk (2.3) avec $\kappa(\varepsilon) = 0.5$ et toutes les calibrations de la matrice de covariance présentées dans les sections 3.4.3 et 3.5. L'horizon d'investissement est $M = 60$ jours, le taux sans risque 1.01 et les frais de transaction 0.5%. On représente ci-après, sur les figures 3.3 et 3.4 ci-après la composition des quatre différents portefeuilles en utilisant l'estimation empirique de la matrice de covariance et les corrections C_1, C_3 et C_7. Nous voyons tout spécialement l'effet de la diversification en utilisant les corrections C_1 et C_3. De nombreuses simulations du même type ont été effectuées, conduisant aux mêmes conclusions. Pour obtenir des portefeuilles diversifiés, les gestionnaires de portefeuilles introduisent traditionnellement des contraintes de boîte sur certaines composantes du portefeuille. Il est intéressant de remarquer que les corrections C_1 et C_3 tendent à donner naturellement des portefeuilles diversifiés sans changer les contraintes du problème.

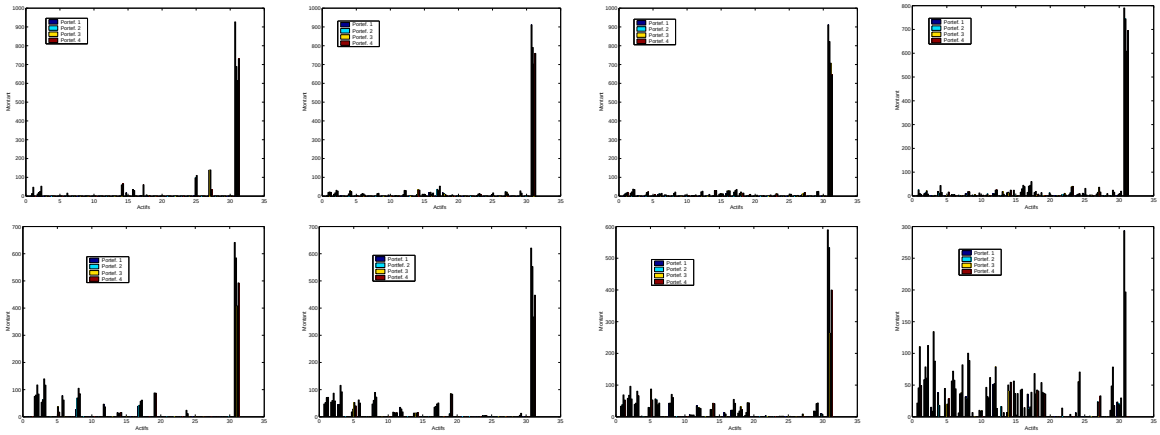


FIG. 3.3 – De gauche à droite, méthode empirique et calibrations C_7, C_1 et C_3. Les portefeuilles ont été obtenus en utilisant le modèle de Markowitz, quatre dates d'investissement (différentes pour les courbes du haut et du bas), $\ell = 1.03$ (quatre courbes du haut) et $\ell = 1.02$ (quatre courbes du bas).

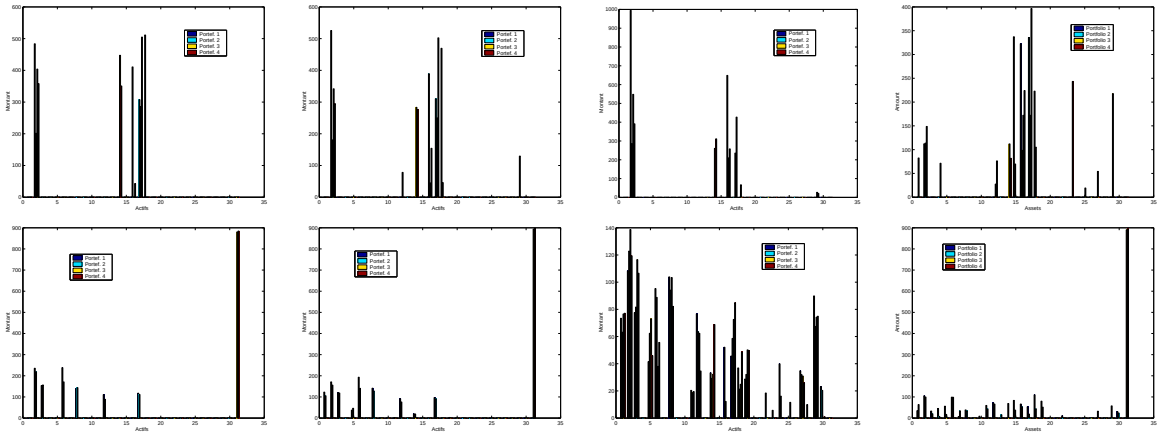


FIG. 3.4 – De gauche à droite, méthode empirique et calibrations C_7, C_1 et C_3. De haut en bas, portefeuilles obtenus en utilisant le modèle VaR, et deux différents ensembles de quatre dates d'investissement.

3.6.4 Comparaison des calibrations de la matrice de covariance sur des données réelles

On utilise les valeurs des paramètres utilisés dans la section 3.6.3 ci-dessus pour calculer les portefeuilles optimaux obtenus en investissant dans les actifs du Dow Jones du 2 janvier 1995 au 30 juin 2004. On rebalance le portefeuille tous les M jours. On mesure l'influence des corrections de la matrice de covariance adaptative introduites dans les sections 3.4.3 et 3.5. Les paramètres de la méthode adaptative sont choisis à posteriori. Dans le tableaux 3.5 à 3.8 on appellera *Rdt* le rendement de chaque méthode sur la période d'investissement. Enfin, $\bar{R}$ et σ désignent la moyenne et l'écart type empiriques de l'échantillon des rendements à M jours du portefeuille sur la période d'investissement. Finalement, on s'intéresse à la variabilité Δ de la composition des portefeuilles (voir (3.23)). Les résultats de ces expériences, conduites en utilisant les modèles VaR et de Markowitz et différentes valeurs de M, sont donnés dans les tableaux 3.5 à 3.8 ci-après.

	$M = 15$ *jours*				$M = 30$ *jours*			
Méthode	Rdt	$\bar{R}$	σ	Δ	Rdt	$\bar{R}$	σ	Δ
Adaptative	2.47	1.0057	0.0184	0.4876	2.4444	1.0113	0.0253	0.5629
C_1	2.63	1.0061	0.0210	0.6128	2.8044	1.0131	0.0304	0.5821
C_2	2.50	1.0057	0.0184	0.4888	2.5363	1.0117	0.0257	0.5149
C_3	2.64	1.0062	0.0257	0.7105	2.7134	1.0130	0.0387	0.7237
C_4	2.52	1.0058	0.0183	0.4891	2.5591	1.0118	0.0257	0.5173
C_5	2.47	1.0057	0.0184	0.4876	2.4444	1.0113	0.0253	0.5629
C_6	2.52	1.0058	0.0185	0.4927	2.5855	1.0120	0.0256	0.5434
C_7	2.58	1.0059	0.0185	0.4975	2.6058	1.0121	0.0262	0.5258

TAB. 3.5 – Comparaison des différentes calibrations de la matrice de covariance en utilisant les actifs du Dow Jones (de janvier 1995 à juin 2004) et un actif sans risque en utilisant le modèle de Markowitz.

	$M = 60$ *jours*			
Méthode	Rdt	$\bar{R}$	σ	Δ
Adaptative	3.8672	1.0386	0.1082	1.1957
C_1	4.1250	1.0409	0.1138	1.1219
C_2	4.0898	1.0398	0.1045	1.2525
C_3	4.1549	1.0414	0.1152	1.2634
C_4	4.1487	1.0401	0.1044	1.2726
C_5	3.8672	1.0386	0.1082	1.1957
C_6	4.1805	1.0404	0.1054	1.3096
C_7	4.4440	1.0421	0.1075	1.2899

TAB. 3.6 – Comparaison des différentes calibrations de la matrice de covariance en utilisant les actifs du Dow Jones (de janvier 1995 à juin 2004) et un actif sans risque en utilisant le modèle de Markowitz.

	$M = 15\ jours$				$M = 30\ jours$			
Méthode	Rdt	$\bar{R}$	σ	Δ	Rdt	$\bar{R}$	σ	Δ
Adaptative	5.288	1.0115	0.0526	2.70	3.7388	1.0177	0.0544	1.7128
C_1	5.714	1.0119	0.0502	2.68	3.910	1.0185	0.0581	1.7468
C_2	5.429	1.0116	0.0511	2.56	3.8310	1.0180	0.0555	1.7417
C_3	3.749	1.0087	0.0369	1.70	3.601	1.0168	0.0463	1.5439
C_4	5.071	1.0113	0.0534	2.53	3.823	1.0180	0.0561	1.7494
C_5	5.289	1.0115	0.0526	2.69	3.739	1.0177	0.0544	1.7128
C_6	5.208	1.0114	0.0531	2.5822	4.0425	1.0189	0.0591	1.8710
C_7	4.736	1.0109	0.0534	2.28	4.2956	1.0195	0.0573	1.9400

TAB. 3.7 – Comparaison des différentes calibrations de la matrice de covariance en utilisant les actifs du Dow Jones (de janvier 1995 à juin 2004) et un actif sans risque en utilisant le modèle Value-At-Risk.

	$M = 60\ jours$				$M = 90\ jours$			
Méthode	Rdt	$\bar{R}$	σ	Δ	Rdt	$\bar{R}$	σ	Δ
Adaptive	3.9341	1.0404	0.1169	2.0874	4.4297	1.0722	0.1887	3.56
C_1	5.3755	1.0471	0.1067	2.4992	3.6837	1.0652	0.1902	3.02
C_2	4.2586	1.0421	0.1138	2.0720	3.9766	1.0682	0.1902	3.18
C_3	3.4632	1.0348	0.0909	1.5025	4.9669	1.0746	0.1756	3.38
C_4	3.5040	1.0384	0.1248	2.0743	3.5209	1.0642	0.1937	3.03
C_5	3.9351	1.0404	0.1169	2.0879	4.4297	1.0722	0.1887	3.56
C_6	4.3873	1.0433	0.1180	2.2423	4.1713	1.0697	0.1881	3.36
C_7	3.7474	1.0387	0.1120	1.7759	3.8967	1.0676	0.1909	3.10

TAB. 3.8 – Comparaison des différentes calibrations de la matrice de covariance en utilisant les actifs du Dow Jones (de janvier 1995 à juin 2004) et un actif sans risque en utilisant le modèle Value-At-Risk.

On remarque que les corrections de la méthode adaptative tendent à fournir des portefeuilles dont les rendements sont plus élevés et un écart type voisin. Cependant, la variabilité de la composition du portefeuille n'est pas nécessairement plus petite si on utilise les différentes corrections de la matrice de covariance comme c'était le cas pour les expériences conduites sur les données simulées de la section 3.6.2.

3.7 Conclusion sur la calibration stable et robuste de la matrice de covariance

Pour conclure, ce chapitre visait, dans la veine du modèle de Black et Litterman, à donner un second souffle au modèle de Markowitz. Nous avons tout d'abord présenté une analyse de sensibilité pour différentes versions du modèle de Markowitz. En utilisant un modèle assez général pour les rendements, nous avons proposé des stratégies pour calculer des portefeuilles stables ou robustes en utilisant le modèle de Markowitz.

La calibration de la matrice de covariance proposée en section 3.4.3 a montré son efficacité d'un point de vue numérique, en battant toutes les autres méthodes dans la plupart des tests de stabilité effectués. Nous pensons que cela montre l'attention qui doit être portée au conditionnement de la matrice de covariance estimée. En effet, une plus petite valeur propre de la matrice de covariance très proche de zéro (comme c'est le cas pour la matrice de covariance empirique) est d'un point de vue financier absurde et pose des problèmes numériques pour résoudre le problème de Markowitz. Finalement, notre matrice de covariance est bien une matrice semidéfinie (et même en fait une matrice définie positive) puisque les contraintes l'imposent naturellement. Les méthodes robustes n'ont pas été intégrées aux tests de stabilité. En fait, les méthodes robustes ont aussi tendance à produire des portefeuilles stables mais c'est parce qu'elles investissent une bonne partie de l'argent dans l'actif sans risque.

3.8 Preuves

Preuve du lemme 3.2.1

Écrivons le problème strictement convexe sous la forme :

$$(P_1)\begin{cases} \min\ h(x) \\ g_i(x) \leq 0, \quad i \in J. \end{cases}$$

Notons X_1 l'ensemble de faisabilité de P_1, x_1 le minimum de h sur X_1 et h_1 la valeur optimale de P_1. Soit $i_0 \in J$ tel que $g_{i_0}(x) \leq 0$, soit une contrainte non active. On a donc $g_{i_0}(x_1) < 0$. On montre alors que le problème P_1 est équivalent au problème consistant à minimiser h sur l'ensemble $X_2 = \{x \in \mathbb{R}^n \ \ g_i(x) \leq 0 \ \ i \in J \backslash i_0\}$.
Comme $X_1 \subseteq X_2$, le minimum h_2 de h sur X_2 est clairement inférieur ou égal à h_1. On montre qu'en fait, pour tout $x \in X_2, h(x) \geq h_1$ (donc $h_2 \geq h_1$, et les deux problèmes ont les mêmes valeurs optimales atteintes au même point). Soit $x \in X_2$. Si $g_{i_0}(x) \leq 0$, alors $x \in X_1$ et $h(x) \geq h_1$ par définition de x_1. Par contre, si $g_{i_0}(x) > 0$, comme $g_{i_0}(x_1) < 0$ et comme g_{i_0} est continue, le théorème des valeurs intermédiaires donne l'existence de $t^* \in]0,1[$ tel que $g_{i_0}(t^* x_1 + (1-t^*)x) = 0$. En outre, par convexité de l'ensemble X_2, il s'ensuit que $x_0 = t^* x_1 + (1-t^*)x \in X_2$ (comme x_1 et x appartiennent à X_2). Donc $x_0 \in X_1$ et $h(x_0) \geq h_1$. On conclut par convexité de h, ce qui donne $h_1 \leq h(x_0) \leq t^* h_1 + (1-t^*)h(x)$.

□

Preuve des théorèmes 3.3.2 et 3.3.3

Nous commençons par prouver le théorème 3.3.2. Soit $\lambda \in \mathbb{R}^{m_1}$, soit la fonction duale

$$\theta_i(\lambda) = \begin{cases} \inf\ \frac{1}{2}\, x^T Q_i x + x^T c_i + \sum_{j=1}^{m_1} \lambda_j (x^T f_j^i - b_j^i) \\ x \in X, \end{cases} \tag{3.24}$$

du problème P_i et soit $\lambda^i = \text{argmax}_{\lambda \in \mathbb{R}^{m_1}}\ \theta_i(\lambda)$ le problème dual. Comme l'hypothèse de Slater est vérifiée pour le problème convexe P_i, le problème P_i et son problème dual sont équivalents et ont la même valeur optimale. Puisque Q_i est définie positive, si λ^i est

une solution optimale du problème dual et si $x(\lambda)$ est la solution optimale de (3.35), alors $x_i = x(\lambda^i)$). En écrivant les conditions d'optimalité du premier ordre pour x_i, nous avons :

$$\forall x \in X, \ \ (x - x_i)^T (Q_i x_i + c_i + \sum_{j=1}^{m_1} \lambda_j^i f_j^i) \geq 0.$$

Comme x_1 et x_2 sont dans X, on peut utiliser l'inégalité précédente avec $x = x_2, i = 1$ et $x = x_1, i = 2$, ce qui donne :

$$\begin{cases} (x_2 - x_1)^T (Q_1 x_1 + c_1 + \sum_{j=1}^{m_1} \lambda_j^1 f_j^1) \geq 0 \\ (x_1 - x_2)^T (Q_2 x_2 + c_2 + \sum_{j=1}^{m_1} \lambda_j^2 f_j^2) \geq 0. \end{cases} \tag{3.25}$$

En ajoutant les inégalités (3.25) et en réarrangeant les termes, on obtient :

$$(x_2 - x_1)^T Q_1 (x_2 - x_1) \ \ \leq \ \ (x_2 - x_1)^T (Q_1 x_2 - Q_2 x_2 + c_1 - c_2) + R \tag{3.26}$$

avec $R = (x_2 - x_1)^T \sum_{j=1}^{m_1} (\lambda_j^1 f_j^1 - \lambda_j^2 f_j^2)$. Comme pour $1 \leq j \leq m_1$ et $i = 1, 2$, $x_i^T f_j^i = b_j^i$, nous avons $(x_2 - x_1)^T (\lambda_j^1 f_j^1 - \lambda_j^2 f_j^2) = (f_j^2 - f_j^1)^T (\lambda_j^2 x_1 - \lambda_j^1 x_2)$. En injectant ce résultat dans (3.36) on a :

$$\beta(Q_1) \|x_2 - x_1\|_1{}^2 \leq \beta \|x_2 - x_1\|_1 + \sum_{j=1}^{m_1} \|f_j^2 - f_j^1\|_\infty (|\lambda_j^2| + |\lambda_j^1|) D_1(X), \tag{3.27}$$

avec $\beta = \|Q_2 - Q_1\|_\infty D_1(X) + \|c_1 - c_2\|_\infty$. Il reste à majorer les multiplicateurs de Lagrange λ_j^i. Comme l'hypothèse de Slater est vérifiée pour P_i, si μ_j^i est le multiplicateur de Lagrange associé à la contrainte $h_j(x) \leq 0$ pour le problème P_i alors (voir [64]) $\sum_{j=1}^{m_2} \mu_j^i \leq M_i$ où M_i est défini dans le théorème 3.3.2. Puisque

$$\sum_{j=1}^{m_1} \lambda_j^i f_j^i = -Q_i x_i - c_i - \sum_{j=1}^{m_2} \mu_j^i \nabla_x h_j(x_i) = w^i, \tag{3.28}$$

s'il y a une contrainte d'égalité ($m_1 = 1$), on obtient immédiatement $|\lambda_1^i| \leq A^i = \frac{1}{\|f_1^i\|_\infty} (\|c^i\|_\infty + \|Q_i\|_\infty D_1(X) + M M_i)$. Si $m_1 > 1$, alors on multiplie (3.28) par ${f_k^i}^T$ pour $1 \leq k \leq m_1$, ce qui donne $U^i \lambda^i = \tilde{w}_i$ où $\tilde{w}_i(k) = {w^i}^T f_k^i$. Comme les $(f_k^i)_{1 \leq k \leq m_1}$ sont indépendants, U^i est inversible et $\|\lambda^i\|_\infty \leq \|U_i^{-1}\|_\infty \|w^i\|_\infty \sum_{j=1}^{m_1} \|f_j^i\|_1 \leq A^i$. En injectant ces majorations de λ_j^i dans (3.37) on a $P(\|x_2 - x_1\|_1) \leq 0$, P étant le polynôme du second degré $P(x) = \beta(Q_1) x^2 - \beta x - (A^1 + A^2) D_1(X) \sum_{j=1}^{m_2} \|f_j^2 - f_j^1\|_\infty$. Ainsi, $\|x_2 - x_1\|_1$ est majorée par la plus grande des racines de P, ce qui achève la preuve du théorème 3.3.2. Pour montrer

le théorème 3.3.3, il suffit de suivre la preuve du théorème 3.3.2 et de montrer que pour le problème $P'(\ell, \rho_i, Q_i)$, $\frac{B^i}{2}$ est en effet une majoration du multiplicateur de Lagrange associé à la contrainte de rendement (où B^i est défini dans (3.11)). On peut minorer la valeur optimale de $P'(\ell, \rho_i, Q_i)$ par $\frac{1}{2} {y_i}^T Q_i y_i$ où y_i est la solution du problème relâché suivant :

$$\begin{cases} \min \ \frac{1}{2} y^T Q_i y \\ \bar{\rho}_i^T y = \bar{\ell}. \end{cases} \tag{3.29}$$

Ainsi, si $\theta_i(\lambda)$ est la fonction duale du problème $P'(\ell, \rho_i, Q_i)$ où seulement la contrainte de rendement a été dualisée, nous avons :

$$\theta_i(\lambda_i) \geq \frac{1}{2} y_i^T Q_i y_i. \tag{3.30}$$

Par définition de la fonction duale, nous avons pour tout j tel que $\bar{\rho}_i(j) > \bar{\ell}$:

$$\forall \lambda \quad \theta_i(\lambda) \leq \frac{1}{2} {e_j}^T Q_i e_j + \lambda(\bar{\ell} - \bar{\rho}_i^T e_j) \tag{3.31}$$

En utilisant (3.30) et (3.38) avec $\lambda = \lambda^i$ on a :

$$\lambda^i \leq \frac{1}{2} \min_{\rho_i(j) > \ell} \frac{1}{\rho_i(j) - \ell} (Q_i(j,j) - {y_i}^T Q_i y_i). \tag{3.32}$$

La solution de (3.29) est donnée par $y_i = \dfrac{\bar{\ell}}{\bar{\rho}_i^T {Q_i}^{-1} \bar{\rho}_i} {Q_i}^{-1} \bar{\rho}_i$. En injectant cette expression de y_i dans (3.39), on obtient la majoration suivante pour λ^i :

$$\lambda^i \leq \frac{B^i}{2} = \frac{1}{2} \min_{\rho_i(j) > \ell} \frac{Q_i(j,j)}{\rho_i(j) - \ell} - \frac{\bar{\ell}^2}{\bar{\rho}_i^T {Q_i}^{-1} \bar{\rho}_i}. \tag{3.33}$$

□

Preuve du théorème 3.3.4

On prouve le résultat en utilisant la formulation $(\mathcal{M}')$ suivante du problème $(\mathcal{M})$:

$$(\mathcal{M}') \begin{cases} \min \ \frac{1}{2} x^T Q_\tau x \\ \rho_\tau^T x + r(n+1) \left(x_{n+1}^- + (\mathbf{e} - \mu)^T y - (\mathbf{e} + \nu)^T z \right) \geq \ell \left(\mathbf{e}^T x^- + x_{n+1}^- \right) \\ x + y - z = x^-, \\ (\mathbf{e} + \nu)^T z - (\mathbf{e} - \mu)^T y \leq x_{n+1}^-, \\ x \geq 0, \ y \geq 0, \ z \geq 0; \end{cases} \tag{3.34}$$

donnant le montant des actifs risqués dans le portefeuille. Le montant détenu dans l'actif sans risque est alors :

$$x_{n+1} = x_{n+1}^- + (\mathbf{e} - \mu)^T y^* - (\mathbf{e} + \nu)^T z^*,$$

si y^* et z^* sont les valeurs optimales de y et z pour le problème (3.34). L'ensemble de faisabilité de $(\mathcal{M}')$ est l'intersection de l'hyperplan défini par la contrainte de rendement

et d'un ensemble défini par les contraintes restantes et que nous noterons $Y(\mu, \nu, x^-)$. Soit $\bar{\ell} = \ell\,(\mathbf{e}^T x^- + x^-_{n+1}) - r_{n+1}\, x^-_{n+1}$, $W = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, le vecteur de variables de décision, W_i la solution de $(\mathcal{M}')$ pour les valeurs (ρ_i, Q_i) de ses paramètres, $\lambda \in \mathbb{R}$ et soit

$$\theta_i(\lambda) = \begin{cases} \inf\ \frac{1}{2}\, x^T Q_i x + \lambda \left(\bar{\ell} - x^T\, \rho_i - r_{n+1}\, (\mathbf{e} - \mu)^T\, y + r_{n+1}\, (\mathbf{e} + \nu)^T\, z\right) \\ W = (x, y, z)^T \in Y(\mu, \nu, x^-), \end{cases} \tag{3.35}$$

la fonction duale du problème $(\mathcal{M}')$(pour les valeurs (ρ_i, Q_i) de ses paramètres) où seule la contrainte de rendement a été dualisée. On définit aussi λ_i une solution du problème dual $\lambda_i = \text{argmax}_{\lambda \geq 0}\, \theta_i(\lambda)$. Le problème primal $(\mathcal{M}')$ et son dual sont équivalents et ont la même valeur optimale atteinte au même point (si $W(\lambda)$ est la solution optimale de (3.35) alors $W_i = W(\lambda_i)$). Par optimalité de W_i nous avons :

$$\forall\, W = (x, y, z)^T \in Y(\mu, \nu, x^-),\ \ (W - W_i)^T \begin{pmatrix} Q_i\, x - \lambda_i \rho_i \\ \lambda_i r_{n+1}(\mu - \mathbf{e}) \\ \lambda_i r_{n+1}(\nu + \mathbf{e}) \end{pmatrix} \geq 0.$$

En utilisant l'inégalité précédente pour $W = W_2, i = 1$ et $W = W_1, i = 2$, on obtient :

$$\begin{cases} (x_2 - x_1)^T (Q_1\, x_1 - \lambda_1\, \rho_1) + \lambda_1\, r_{n+1}\, \left((y_2 - y_1)^T\, (\mu - \mathbf{e}) + (\nu + \mathbf{e})^T\, (z_2 - z_1)\right) \geq 0 \\ (x_1 - x_2)^T (Q_2\, x_2 - \lambda_2\, \rho_2) + \lambda_2\, r_{n+1}\, \left((y_1 - y_2)^T\, (\mu - \mathbf{e}) + (\nu + \mathbf{e})^T\, (z_1 - z_2)\right) \geq 0. \end{cases}$$

En additionnant les deux inégalités précédentes et en réarrangeant les termes on a :

$$(x_1 - x_2)^T\, Q_1\, (x_1 - x_2) \leq (x_2 - x_1)^T\, (Q_1 - Q_2)\, x_2 + (x_2 - x_1)^T (\lambda_2 \rho_2 - \lambda_1 \rho_1) + M \tag{3.36}$$

avec

$$M = r_{n+1}(\lambda_1 - \lambda_2)\, \left((y_2 - y_1)^T\, (\mu - \mathbf{e}) + (z_2 - z_1)^T (\nu + \mathbf{e})\right).$$

Comme la contrainte de rendement est active, nous avons, pour $i = \{1, 2\}$,

$$x_i^T\, \rho_i + r_{n+1}\, \left(x^-_{n+1} + (\mathbf{e} - \mu)^T\, y_i - (\nu + \mathbf{e})^T\, z_i\right) = \ell\, \mathbf{e}^T x^-.$$

Ainsi, $M = (\lambda_1 - \lambda_2)\, (x_2^T\, \rho_2 - x_1^T\, \rho_1)$. En injectant ce résultat dans (3.36), nous avons :

$$\beta(Q_1) \|x_2 - x_1\|_1^{\,2} \leq \|x_2 - x_1\|_1\, \|Q_2 - Q_1\|_\infty\, \|x^-\|_1 + (\lambda_1 + \lambda_2)\, \|\rho_2 - \rho_1\|_\infty\, \|x^-\|_1. \tag{3.37}$$

Il reste à majorer les multiplicateurs λ_i. On peut minorer la valeur optimale du problème de Markowitz par 0. Ainsi, nous avons $\theta_i(\lambda_i) \geq 0$. Par définition de la fonction duale, pour tout $i = \{1, 2\}$ et pour chaque $j \in J(i)$, $\theta_i(\lambda)$ est inférieur à la valeur de la fonction objectif (3.35) en $W = (x, y, z)^T$ tel que $x_i = 0, y_i = x_i^-, z_i = 0$ pour $i \neq j$ et

$$x_j = f(j),\ \ y_j = 0,\ \ z_j = f(j) - x_j^-.$$

Ainsi, pour chaque $j \in J(i)$:

$$\forall\, \lambda, \quad \theta_i(\lambda) \leq \frac{1}{2}\, Q_i(j, j)\, f(j)^2 + \lambda(\ell\ \mathbf{e}^T\ x^- - f(j)\ \rho_i(j)). \tag{3.38}$$

En utilisant (3.38) avec $\lambda = \lambda_i$ et comme $\theta_i(\lambda_i) \geq 0$ on a :

$$\lambda_i \leq \frac{1}{2} \min_{j\, \in\, J(i)}\ \frac{f(j)^2\, Q_i(j, j)}{f(j)\ \rho_i(j) - \ell\ \mathbf{e}^T\ x^-} = \frac{A_i}{2}. \tag{3.39}$$

Si $A = 2(A_1 + A_2)$, d'après (3.37), nous avons $P(\|x_2 - x_1\|_1) \leq 0$, P étant le polynôme du second degré $P(x) = \beta(Q_1)x^2 - \|Q_2 - Q_1\|_\infty \|x^-\|_1 x - \frac{1}{4} A \|\rho_2 - \rho_1\|_\infty \|x^-\|_1$. Ainsi, $\|x_2 - x_1\|_1$ est majoré par la plus grande racine de P, ce qui achève la preuve de (3.3.3). □

Proposition 3.8.1 *(proposition 4.32, p.287 de [18]) Considérons deux problèmes d'optimisation*

$$\begin{array}{l} \min\ f(x) \\ x \in X \end{array}$$

et

$$\begin{array}{l} \min\ g(x) \\ x \in X \end{array}$$

où $f, g : X \to \mathbb{R}$. Soit S_0 l'ensemble des solutions du premier problème associé à la fonction objectif f et x_2 une solution du deuxième problème.
Supposons que (i) la fonction f satisfasse une condition de croissance du second ordre sur X ($\exists\, c > 0$ tel que pour tout $x \in X$ et $x_1 \in S_0$, $f(x) \geq f(x_1) + c\|x - x_1\|^2$) et que (ii) la fonction différence $g(.) - f(.)$ est Lipschitz continue, Lipschitzienne, de module β sur X, alors

$$dist(x_2, S_0) \leq \frac{\beta}{c}. \tag{3.40}$$

Chapitre 4

Analyse de problèmes de gestion de production

Nota Bene : Les résultats présentés dans ce chapitre pourront se retrouver dans l'article [55] soumis à IEEE Transaction on Automatic and Control et dans deux articles en cours ([53] et [52].)

L'objet de ce chapitre est de proposer de nouvelles modélisations pour deux problèmes de gestion de deux grands systèmes industriels : d'une part la gestion du parc de production d'électricité en France et d'autre part la gestion d'un réseau de transport et de distribution du gaz. On étudie également les aspects liés à la résolution de ces problèmes.

Nous nous intéressons dans une première partie à des problèmes d'optimisation permettant d'assister la gestion du parc de production d'électricité en France. Etant donné un certain nombre de centrales électriques (centrales nucléaires, thermiques et hydroélectriques ainsi que des contrats de gestion des pics de demande modélisés par une unité de production virtuelle baptisée EJP), il s'agit de minimiser les coûts de production sur la période de gestion, tout en satisfaisant des contraintes opérationnelles liées aux unités de production et en assurant l'équilibre entre la production et la demande à chaque pas de temps. En pratique, la modélisation dépend grandement de l'horizon de gestion du problème d'optimisation : pour de courtes périodes de gestion, typiquement jounalières voire hebdomadaires, on considère en général que le problème est déterministe (cf. [2],[31]), alors que pour des périodes de gestion plus longues, une attention toute particulière est portée sur la nature stochastique des paramètres du problème et des décisions à prendre. En particulier, pour un horizon de gestion annuel (qui est celui que l'on considère dans ce chapitre), on ne peut pas considérer que les apports hydrauliques des réservoirs des centrales hydrauliques, la consommation électrique, la disponibilité des usines ainsi que les prix de l'électricité sont connus à l'avance. En France, par exemple, à cause du chauffage électrique, une baisse d'un degré Celsius de la température moyenne conduit à une croissance de consommation de plus de 1GW, alors que les pics de consommation sont autour de 80GW. Les problèmes de gestion de production (et tout particulièrement les problèmes de gestion de production électrique) ont fait l'objet de nombreuses études s'intéressant à la fois aux aspects de modélisation et de résolution de ces problèmes ([22],[35],[25],[79],[48] [1],[49],[89],[77],[21]).

En général, l'évolution des paramètres incertains sur la période de gestion est modélisée par un arbre de scénarios et le but est de minimiser le coût moyen de production sur cet ensemble de scénarios. Les méthodes de résolution proposées récemment (dans les papiers [1],[49],[89] par exemple), envisagent une résolution du problème par relaxation lagrangienne. Les méthodes de résolution présentées dans ces articles diffèrent par les méthodes d'optimisation non différentiables utilisées et les outils de résolution des sous-problèmes.

D'un point de vue de la modélisation, nous présenterons différentes alternatives à la modélisation du problème sous la forme d'un problème d'optimisation sur un arbre de scénarios. Notre objectif sera de mettre en oeuvre des méthodes de gestion stables ou robustes. Nous dirons qu'une méthode de gestion est stable si elle fournit des solutions relativement stables en terme de coût (face aux variations de la consommation mais aussi des apports hydrauliques et des taux de disponibilité) sans que ces coûts ne soient prohibitifs. Les modélisations retenues jusqu'à présent ne se sont pas intéressées à l'obtention de méthodes stables. D'autre part, à l'exception de [22], elles ne proposent pas (à notre connaissance) de méthodes de gestion robustes et pourraient conduire à des pertes importantes si des scénarios difficiles se produisent. Dans [22] on propose une approche consistant à classer les scénarios de l'arbre des scénarios afin de trouver ceux qui sont les plus défavorables. Le problème d'optimisation déterministe correspondant au pire scénario $\mathcal{S}_{det}$ est alors résolu et on note $\mathcal{C}_{det}$ le coût minimal sur ce scénario. Une optimisation stochastique sur l'arbre de scénario est ensuite effectuée en ajoutant une contrainte spécifiant que le coût sur le scénario $\mathcal{S}_{det}$ doit être proche de $\mathcal{C}_{det}$, ce qui s'écrit :

$$\mathcal{C}_{sto} \leq \mathcal{C}_{det} + \varepsilon,$$

où $\mathcal{C}_{sto}$ est le coût sur le scénario $\mathcal{S}_{det}$ résultant de l'optimisation globale sur l'arbre de scénarios et ε est un niveau de tolérance.
Ce chapitre est précisément dédié à la présentation de méthodes de résolution robustes ou stables du problème de gestion de production d'électricité à l'horizon annuel. Notre contribution est de montrer comment des régularisations de la fonction duale, permettent de mettre en place des méthodes stables et d'apporter les avantages suivants : (i) diminution significative de la variance des coûts simulés - variance divisée par 4 pour la plus performante des méthodes pour un coût moyen comparable sur les scénarios difficiles - ; (ii) utilisation parcimonieuse de la plus grosse des réserves hydrauliques ; (iii) diminution du nombre de scénarios de coût très élevé. De plus, nous montrerons que ces modifications du problème dual préservent le schéma de décomposition par les prix. On construit ensuite des contreparties robustes ajustées du problème de gestion de production électrique fournissant des solutions robustes et adaptatives.

Du point de vue de la résolution des problèmes induits par ces modélisations, nous utilisons soit des méthodes de résolution directe soit la méthode de décomposition par les prix revenant à dualiser les contraintes couplantes du problème. On propose aussi une méthode originale de résolution du problème de gestion de production électrique par programmation dynamique dans le cas où l'on dispose d'une seule centrale hydraulique et d'un nombre arbitraire de centrales thermiques. On montre ensuite comment utiliser cette méthode pour calculer efficacement la valeur optimale de la fonction duale dans la méthode de décomposition par les prix.

Afin de tester l'application des méthodes de gestion de production électrique présentées dans ce chapitre sur des données de consommation électrique réelles, une analyse statistique des courbes de consommations d'électricité en France sur la période 1996-2003 est mise en oeuvre dans la section 4.6.

La section 4.7.1 met ensuite les méthodes de gestion proposées dans ce chapitre à l'épreuve sur des données réelles et simulées.

La deuxième partie, beaucoup plus succinte, et regroupée dans la section 4.9, propose des outils de modélisation robustes pour le problème de transport et de distribution du gaz.

Nous commençons par présenter le problème de gestion de production électrique.

4.1 Modélisation du problème de gestion de production électrique

L'objectif est de décider des instants de mise en route/arrêt des usines et de leurs niveaux de production à chaque pas de temps de façon à satisfaire la consommation électrique à chaque pas de temps tout en minimisant les coûts de production. Le modèle physique considéré est un système dynamique stochastique pour lequel les paramètres incertains sont la consommation électrique, la disponibilité des centrales thermiques et la quantité de précipitation reçue par les différents réservoirs des centrales hydroélectriques. La modélisation retenue reprend certains aspects de [35] et [48] et introduit des taux de disponibilité pour les centrales thermiques, des fonctions de valorisation des réserves hydrauliques et des contrats EJP. Soit T le nombre de sous-intervalles obtenus par discrétisation de l'horizon de gestion. Cette discrétisation peut être choisie uniforme (journalière, hebdomadaire, mensuelle) ou non. Les variables de décision repéreront soit les états x soit les commandes u du système. Soit $\mathcal{L} = \mathcal{L}_T \cup \mathcal{L}_H \cup \mathcal{L}_J$, une partition de l'ensemble des unités de production dans laquelle $\mathcal{L}_T$ regroupe l'ensemble des centrales thermiques, $\mathcal{L}_H$ l'ensemble des centrales hydrauliques et $\mathcal{L}_J$ l'ensemble des contrats EJP. Par ailleurs on notera $Duree(t)$, la durée du pas de temps t.

4.1.1 Modélisation du fonctionnement des centrales thermiques

Pour toute centrale thermique $\ell \in \mathcal{L}_T$, nous désignons par $(u_t^\ell) = (y_t^\ell, z_t^\ell)$ la commande de la centrale thermique ℓ au pas de temps t, où $y_t^\ell \in \{0,1\}$ et z_t^ℓ correspondent respectivement à la décision de mise en route/arrêt et au niveau de production (en MWh) de l'usine ℓ au pas de temps t. Ainsi, $y_t^\ell = 0$ (resp. $y_t^\ell = 1$) signifie que la centrale thermique ℓ est arrêtée (resp. fonctionne) au pas de temps t. Les puissances minimales et maximales (en MW) de production de la centrale thermique ℓ sont respectivement $P_{\min}^\ell$ et $P_{\max}^\ell$. Chaque centrale thermique est constituée d'un certain nombre de groupes thermiques dont certains peuvent ne pas fonctionner au pas de temps t. Ainsi, pour chaque pas de temps t et chaque centrale thermique ℓ, les puissances théoriques minimales $P_{\min}^\ell$ et maximales $P_{\max}^\ell$ doivent être corrigées en tenant compte du taux de disponibilité τ_t^ℓ de la centrale thermique ℓ au pas de temps t. Les puissances réelles minimales et maximales au pas de temps t pour la

centrale thermique ℓ sont donc respectivement $\tau_t^\ell P_{\min}^\ell$ et $\tau_t^\ell P_{\max}^\ell$. Les contraintes fixées sur les niveaux de production des centrales thermiques s'expriment donc par :

$$\tau_t^\ell \, P_{\min}^\ell \, Duree(t) \, y_t^\ell \leq z_t^\ell \leq \tau_t^\ell \, P_{\max}^\ell \, Duree(t) \, y_t^\ell, \;\; y_t^\ell \in \{0,1\}, \; \ell \in \mathcal{L}_T, \; t = 1, \ldots, T. \quad (4.1)$$

On prendra par la suite $P_{\min}^\ell = 0$. D'autre part, en général, lorsqu'une centrale thermique est mise en route, elle doit fonctionner pendant au moins $\underline{t}_i$ pas de temps. De même, lorsqu'une centrale est arrêtée, on ne peut pas la faire redémarrer avant $\bar{t}_i$ pas de temps. Ces contraintes s'expriment sous la forme :

$$\begin{array}{ll} y_{t-1}^\ell - y_t^\ell \leq 1 - y_k^\ell, & k = t+1, \ldots, \min\{t + \bar{t}_i - 1, T\}, \; t = 1, \ldots, T, \\ y_t^\ell - y_{t-1}^\ell \leq y_k^\ell, & k = t+1, \ldots, \min\{t + \underline{t}_i - 1, T\}, \; t = 1, \ldots, T. \end{array} \quad (4.2)$$

Les coûts de fonctionnement des centrales thermiques sont de deux types : les coûts de mise en route de ces centrales et les coûts de production. Une description simple des coûts de démarrage ([35], [79], [25]) consiste à supposer que le coût de mise en route de chaque centrale thermique ℓ est fixe et vaut C_ℓ^f. Si $y_\ell(t) = (y_1^\ell, \ldots, y_t^\ell)$, le coût de mise en route de la centrale thermique ℓ au pas de temps t s'exprime alors par :

$$CM_{\ell\, t}(y_\ell(t)) = C_\ell^f \max(y_t^\ell - y_{t-1}^\ell, 0), \; t = 2, \ldots, T.$$

Cependant, cette modélisation ne reflète pas le fait que le coût de démarrage dépend du temps pendant lequel la centrale a été arrêtée. Afin de prendre en compte le fait que le coût de démarrage est d'autant plus élevé que la centrale a été arrêtée longtemps d'autres modélisations ont été proposées (voir [35]).
Quant aux coûts de production, ce sont des fonctions linéaires des niveaux de production : le coût de production de la centrale thermique ℓ au pas de temps t est donc :

$$CP_{\ell\, t}(y_t^\ell, z_t^\ell) = c_\ell \; z_t^\ell,$$

où c_ℓ est un coût fixe (en Euros/MWh).

4.1.2 Modélisation du fonctionnement des centrales hydrauliques

Le complexe hydroélectrique est constitué d'un ensemble de centrales hydrauliques et de réservoirs interconnectés comme le montre la figure 4.1. Chaque usine peut comporter une ou plusieurs turbines, chaque turbine pouvant recevoir de l'eau provenant de différents réservoirs. Nous ferons cependant l'hypothèse que chaque réservoir alimente une seule centrale hydraulique (le cas général peut aussi être facilement modélisé). Pour toute centrale hydraulique $\ell \in \mathcal{L}_H$, soit $(u_t^\ell) = (v_t^\ell, w_t^\ell, dev_t^\ell)$, la commande de la centrale hydraulique ℓ au pas de temps t, où v_t^ℓ désigne la production (en MWh) de la centrale hydraulique ℓ au pas de temps t, w_t^ℓ le pompage (en MWh) effectué par cette centrale au pas de temps t et dev_t^ℓ l'eau éventuellement déversée du réservoir alimentant la centrale ℓ (s'il y a un débordement). Les commandes de pompage permettent, notamment dans les périodes de creux de consommation comme la nuit, de faire remonter de l'eau pour alimenter un réservoir situé au dessus de la centrale hydraulique effectuant le pompage. Cette eau pourra ensuite être utilisée lorsque la consommation sera plus importante. L'état x_t^ℓ de la centrale hydraulique ℓ au pas de temps t correspond au volume d'eau (exprimé en MWh) de la réserve alimentant cette centrale hydraulique au début de ce pas de temps. En dehors du

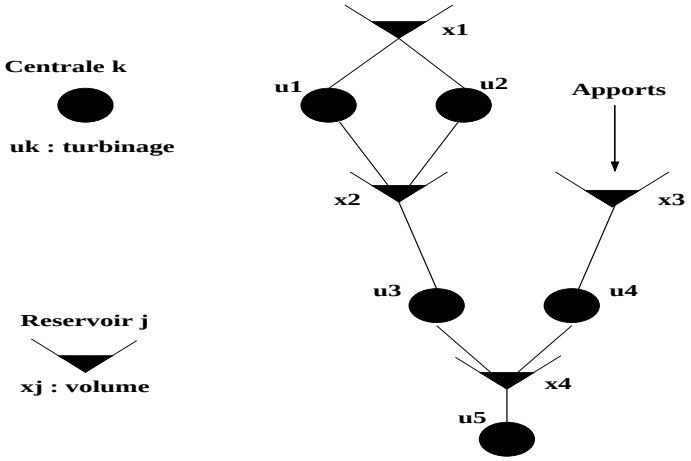


FIG. 4.1 – Représentation d'un complexe hydroélectrique.

pompage, deux types d'apports hydrauliques contribuent à accroître le volume des réservoirs. D'une part les apports hydrauliques naturels dûs aux précipitations : on notera a_t^ℓ l'apport naturel (en MWh) du réservoir ℓ au pas de temps t. D'autre part, pour chaque réservoir ℓ, de l'eau peut provenir des centrales hydrauliques situées en amont. Pour toute centrale hydraulique $\ell \in \mathcal{L}_H$, on désignera par $f(\ell)$ l'indice du réservoir situé au dessous de la centrale hydraulique ℓ (si une telle centrale existe) ; l'eau mettant un temps $d(\ell, f(\ell))$ pour parcourir le trajet séparant la centrale hydraulique ℓ du réservoir $f(\ell)$. On suppose par ailleurs que les centrales hydrauliques sont toujours disponibles et que le coût de mise en fonctionnement d'une centrale hydraulique est négligeable. Les contraintes liées au fonctionnement des centrales hydrauliques sont de deux types : (i) des contraintes de boîte sur les variables de commandes et le volume des réservoirs ; (ii) des équations de conservation de l'énergie pour chaque réservoir. Ces contraintes s'expriment sous la forme :

$$\begin{array}{ll}
x_{t+1}^\ell = x_t^\ell + a_t^\ell + \eta_\ell\, w_t^\ell - v_t^\ell + \displaystyle\sum_{m \mid f(m)=\ell} v_{t-d(m,f(m))}^m - dev_t^\ell, & \forall\, \ell \in \mathcal{L}_H,\ \ \forall\, t \in 1,\ldots,T, \\
0 \le v_t^\ell \le Duree(t)\, P_{\max}^\ell,\ \ 0 \le w_t^\ell \le w_{t\ell}^{\max}, & \forall\, \ell \in \mathcal{L}_H,\ \ \forall\, t \in 1,\ldots,T, \\
x_{\min}^\ell \le x_t^\ell \le x_{\max}^\ell, & \forall\, \ell \in \mathcal{L}_H,\ \ \forall\, t \in 1,\ldots,T,
\end{array} \tag{4.3}$$

où $x_{\min}^\ell$ et $x_{\max}^\ell$ sont les niveaux minimum et maximum (en MWh) du volume du réservoir ℓ ; $P_{\max}^\ell$ est la puissance maximale de la centrale hydraulique ℓ, η_ℓ est l'efficacité du pompage et $w_{t\ell}^{\max}$ est l'énergie de pompage maximale possible pour la centrale ℓ au pas de temps t. Le stock initial de chacune des réserves hydrauliques est connu. Afin d'utiliser l'eau de façon parsimonieuse (plus on aura d'eau en fin de période de gestion et plus on pourra s'en servir sur la période à venir), deux stratégies sont couramment employées. La première ([35], [48], [25]) consiste à imposer que le niveau des réservoirs à la fin de la période de gestion soit supérieur ou égal à ce qu'il était au début de la période de gestion. Une alternative consiste à associer à chaque réservoir ℓ, une fonction $V_H^\ell(.)$ de valorisation de son stock d'eau au dernier pas de temps. Cette fonction associe à chaque valeur admissible du stock d'eau une valeur en Euros. C'est une fonction croissante, qui est supposée quadratique dans [22] et linéaire dans [79]. Nous la considérerons concave, linéaire par morceaux. On représente ci-après, sur la figure 4.2, un exemple d'une telle fonction. On s'intéressera alors pour chaque

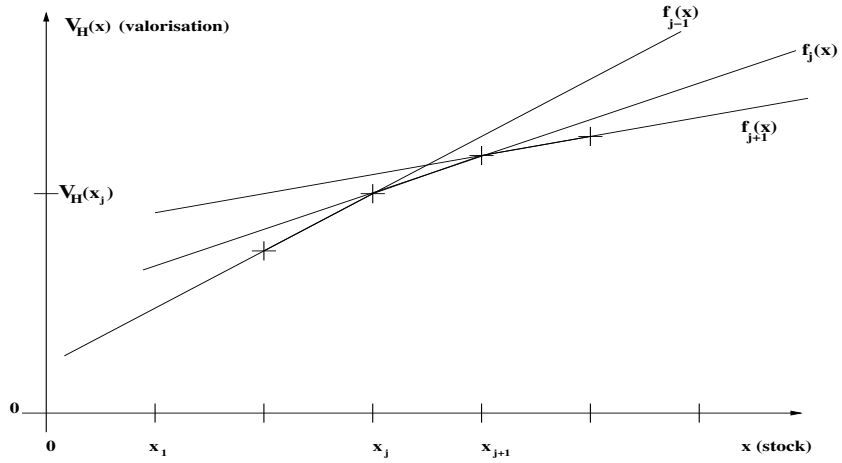


FIG. 4.2 – Fonction de valorisation d'un stock d'eau.

réservoir ℓ, à la valeur $V_H(x^\ell_{T+1})$ de son stock d'énergie en fin de jeu (induisant un coût $-V^\ell_H(x^\ell_{T+1})$).

4.1.3 Modélisation du fonctionnement des contrats EJP

Les contrats EJP peuvent être vus comme des réserves indépendantes, disposant d'une capacité de production limitée, ne pouvant pas être utilisées plus d'un certain nombre de pas de temps fixés par le contrat et telles que pour chaque pas de temps, soit la capacité totale de production est utilisée, soit cette capacité n'est pas utilisée du tout. La modélisation de ces contrats EJP est alors la suivante. Pour tout contrat EJP $\ell \in \mathcal{L}_J$, soit u^ℓ_t la commande appliquée à ce contrat ℓ au pas de temps t. Si la puissance attachée au contrat ℓ est $P^\ell_{\max}$, alors si le contrat EJP est utilisé au pas de temps t, $u^\ell_t = Duree(t)P^\ell_{\max}$ et sinon $u^\ell_t = 0$. L'état du contrat EJP ℓ au pas de temps t est repéré par la variable x^ℓ_t fournissant le stock d'énergie encore disponible sur ce contrat au début du pas de temps t. Le stock d'énergie x^ℓ_1 disponible sur chaque contrat ℓ pour la période d'étude est connue. Les équations régissant l'évolution des commandes et des états sont donc :

$$(u^\ell_t - Duree(t)P^\ell_{\max})\, u^\ell_t = 0, \quad x^\ell_{t+1} = x^\ell_t - u^\ell_t, \quad x^\ell_t \geq 0, \ \ \ell \in \mathcal{L}_J, \ \ 1 \leq t \leq T. \qquad (4.4)$$

D'autre part, comme pour les réserves hydrauliques, on associe à chaque contrat ℓ une fonction $V^\ell_J(.)$ de valorisation de son stock d'énergie en fin de jeu. Cette fonction est concave linéaire par morceaux. On s'intéressera alors pour chaque contrat ℓ à la valeur $V^\ell_J(x^\ell_{T+1})$ de son stock d'énergie en fin de jeu (induisant un coût $-V^\ell_J(x^\ell_{T+1})$).

Pour être complet, il reste à tenir compte des centrales nucléaires ; dont la modélisation des coûts de gestion peut être calquée sur celle des coûts de gestion des centrales thermiques [voir [21]].

4.2 Formulation déterministe du problème

Il s'agit ensuite d'utiliser ces unités de production de façon à satisfaire la demande. Si d_t est la consommation électrique au pas de temps t, les contraintes de satisfaction de la demande s'écrivent :

$$\sum_{\ell \in \mathcal{L}_T} z_t^\ell + \sum_{\ell \in \mathcal{L}_J} u_t^\ell + \sum_{\ell \in \mathcal{L}_H} (v_t^\ell - w_t^\ell) \geq d_t, \quad t = 1, \ldots, T. \tag{4.5}$$

Dans le but de pouvoir pallier des pics de consommation, on peut aussi imposer ([35]) que la capacité totale du système thermique à chaque pas de temps soit suffisamment plus grande que la somme des productions thermiques :

$$\sum_{\ell \in \mathcal{L}_T} (\tau_t^\ell \, P_{\max}^\ell \, Duree(t) \, y_t^\ell - z_t^\ell) \geq r_t > 0, \quad t = 1, \ldots, T, \tag{4.6}$$

les paramètres r_t étant fixés arbitrairement. Le problème de gestion de production électrique consiste alors à minimiser :

$$\sum_{\ell \in \mathcal{L}_T} \sum_{t=1}^{T} (c_\ell \, u_t^\ell + CM_{\ell t}(y_\ell(t))) - \sum_{\ell \in \mathcal{L}_H} V_H^\ell(x_{T+1}^\ell) - \sum_{\ell \in \mathcal{L}_J} V_J^\ell(x_{T+1}^\ell), \tag{4.7}$$

sous les contraintes (4.1), (4.2), (4.3), (4.4) (4.5), (4.6). La fonction $V_H^\ell(.)$ peut s'exprimer sous la forme : $V_H^\ell(x) = \min\limits_{0 \leq k \leq m_H^\ell - 1} f_{H,k}^\ell(x)$, où les fonctions $f_{H,k}^\ell(x)$ sont des fonctions affines. Plus précisément, entre des valeurs que nous noterons $g_{H,k}^\ell$ et $g_{H,k+1}^\ell$, $(0 \leq k \leq m_H^\ell - 1)$, la fonction $V_H^\ell(x)$ coïncide avec la fonction $f_{H,k}^\ell(x)$. Ainsi, on peut remplacer la contribution $-\sum\limits_{\ell \in \mathcal{L}_H} V_H^\ell(x_{T+1}^\ell)$ des centrales hydrauliques à la fonction objectif par $-\sum\limits_{\ell \in \mathcal{L}_H} a_\ell$ en ajoutant les contraintes :

$$a_\ell \leq f_{H,k}^\ell(x_{T+1}^\ell), \quad \ell \in \mathcal{L}_H, \quad 0 \leq k \leq m_H^\ell - 1. \tag{4.8}$$

Si on définit

$$c_{H,k}^\ell = \frac{V_H^\ell(g_{H,k+1}^\ell) - V_H^\ell(g_{H,k}^\ell)}{g_{H,k+1}^\ell - g_{H,k}^\ell}, \quad d_{H,k}^\ell = V_H^\ell(g_{H,k}^\ell) - c_{H,k}^\ell \, g_{H,k}^\ell, \tag{4.9}$$

(4.8) se réécrit :

$$a_\ell \leq c_{H,k}^\ell \, x_{T+1}^\ell + d_{H,k}^\ell, \quad \ell \in \mathcal{L}_H, \quad 0 \leq k \leq m_H^\ell - 1. \tag{4.10}$$

Pour chaque contrat EJP $\ell \in \mathcal{L}_J$, la fonction $V_J^\ell(.)$ est également concave et linéaire par morceaux et peut s'écrire sous la forme, $V_J^\ell(x) = \min\limits_{0 \leq k \leq m_J^\ell - 1} f_{J,k}^\ell(x)$, où les fonctions $f_{J,k}^\ell(x)$ sont des fonctions affines. On peut alors introduire les quantités $c_{J,k}^\ell$ et $d_{J,k}^\ell$ (définies en remplaçant H par J dans (4.9)). Le problème de gestion de production électrique revient alors à minimiser :

$$\sum_{\ell \in \mathcal{L}_T} \sum_{t=1}^{T} (c_\ell \, u_t^\ell + CM_{\ell t}(y_\ell(t))) - \sum_{\ell \in \mathcal{L}_H} a_\ell - \sum_{\ell \in \mathcal{L}_J} b_\ell,$$

sous les contraintes (4.1), (4.2), (4.3), (4.4) (4.5), (4.6), (4.10) et :

$$b_\ell \leq c^\ell_{J,k}\, x^\ell_{T+1} + d^\ell_{J,k}, \;\; \ell \in \mathcal{L}_J, \;\; 0 \leq k \leq m^\ell_J - 1.$$

Par la suite nous ferons les hypothèses suivantes :

- Le complexe hydroélectrique est constitué d'un ensemble de couples centrales-réservoirs non reliés entre eux. On suppose également qu'il n'y a pas de pompage ; les équations de conservation de l'énergie pour les réservoirs s'écrivent donc :

$$x^\ell_{t+1} = x^\ell_t + a^\ell_t - dev^\ell_t - v^\ell_t, \quad \forall\, \ell \in \mathcal{L}_H, \;\; \forall\, t \in 1, \ldots, T. \tag{4.11}$$

- Les unités thermiques fonctionnent toutes au début de la période de gestion et restent en fonctionnement pendant toute la période de gestion ($y^\ell_t = 1, \;\; \forall t = 1, \ldots, T, \;\; \forall\, \ell \in \mathcal{L}_T$). Par contre, certains groupes thermiques peuvent tomber en panne et être réparés ensuite, ce qui a motivé l'introduction de taux de disponibilité. La modélisation du comportement de ces taux de disponibilité est expliquée plus loin. Enfin, on ne tiendra pas compte des contraintes (4.6).

Les équations (4.4) et (4.11) de conservation de l'énergie des réserves hydrauliques et des contrats EJP peuvent alors se réécrire :

$$\forall t = 1, \ldots, T, \;\; x^\ell_{t+1} = x^\ell_1 + \sum_{k=1}^{t} (a^\ell_k - v^\ell_k - dev^\ell_k), \;\; \ell \in \mathcal{L}_H, \;\; x^\ell_{t+1} = x^\ell_1 - \sum_{k=1}^{t} u^\ell_k, \;\; \ell \in \mathcal{L}_J.$$

Sous les hypothèses ci-dessus, si tous les paramètres du problème étaient connus, en injectant ces contraintes d'égalité dans les contraintes $x^\ell_{\min} \leq x^\ell_t \leq x^\ell_{\max}, \;\; \ell \in \mathcal{L}_H$, et $x^\ell_t \geq 0, \;\; \ell \in \mathcal{L}_J$, sur les états des réserves, on pourrait alors exprimer le problème de gestion de production électrique sous la forme :

$$\left\{ \begin{array}{l} \min \;\; \displaystyle\sum_{t=1}^{T} \sum_{\ell \in \mathcal{L}_T} c_\ell\, u^\ell_t - \sum_{\ell \in \mathcal{L}_H} a_\ell - \sum_{\ell \in \mathcal{L}_J} b_\ell \\ a_\ell + c^\ell_{H,k} \displaystyle\sum_{t=1}^{T} (v^\ell_t + dev^\ell_t) \leq c^\ell_{H,k}\, (x^\ell_1 + \sum_{t=1}^{T} a^\ell_t) + d^\ell_{H,k}, \;\; \ell \in \mathcal{L}_H, 0 \leq k \leq m^\ell_H - 1, \\ b_\ell + c^\ell_{J,k} \displaystyle\sum_{t=1}^{T} u^\ell_t \leq c^\ell_{J,k} u^\ell_1 + d^\ell_{J,k}, \;\; \ell \in \mathcal{L}_J, 0 \leq k \leq m^\ell_J - 1, \\ x^\ell_1 + \displaystyle\sum_{k=1}^{t} a^\ell_k - x^\ell_{\max} \leq \sum_{k=1}^{t} v^\ell_k + dev^\ell_k \leq x^\ell_1 + \sum_{k=1}^{t} a^\ell_k - x^\ell_{\min}, \;\; \ell \in \mathcal{L}_H, 2 \leq t \leq T, \\ 0 \leq v^\ell_t \leq Duree(t)\, P^\ell_{\max}, \;\; \ell \in \mathcal{L}_H, 1 \leq t \leq T, \\ 0 \leq dev^\ell_t, \;\; \ell \in \mathcal{L}_H, 1 \leq t \leq T, \\ u^\ell_t \in \{0, Duree(t) P^\ell_{\max}\}, \;\; \ell \in \mathcal{L}_J, 1 \leq t \leq T, \\ 0 \leq u^\ell_t \leq Duree(t)\, \tau^\ell_t\, P^\ell_{\max}, \;\; \ell \in \mathcal{L}_T, 1 \leq t \leq T, \\ \displaystyle\sum_{k=1}^{T} u^\ell_k \leq x^\ell_1, \;\; \ell \in \mathcal{L}_J, \\ d_t \leq \displaystyle\sum_{\ell \in \mathcal{L}_T} u^\ell_t + \sum_{\ell \in \mathcal{L}_H} v^\ell_t + \sum_{\ell \in \mathcal{L}_J} u^\ell_t, \;\; 1 \leq t \leq T. \end{array} \right. \tag{4.12}$$

Dans ce problème, notons que la contrainte de satisfaction de la demande est active. Néanmoins, afin de mettre en place des CR, CRA et CRAA de ce problème, nous garderons les contraintes de satisfaction de la demande écrites sous la forme de contraintes d'inégalités. Les variables de déversement *dev* ont été introduites pour assurer que le problème soit faisable (sous réserve que le parc de production soit de capacité suffisante) quelles que soient les valeurs des apports hydrauliques.

Proposition 4.2.1 *Si pour tout réservoir hydraulique* $\ell \in \mathcal{L}_H$, $x_1^\ell \in [x_{\min}^\ell, x_{\max}^\ell]$ *et si*

$$\sum_{\ell \in \mathcal{L}_T} \tau_t^\ell \, Duree(t) \, P_{\max}^\ell \geq d_t, \quad t = 1, \ldots, T, \tag{4.13}$$

alors le problème de gestion de production qui vient d'être présenté est faisable.

Preuve. On vérifie immédiatement que le plan de production consistant à utiliser les moyens de production thermique par ordre croissant de leur coût de fonctionnement jusqu'à satisfaction de la demande est faisable. Les contraintes sur les variables de commande d'un tel plan de production sont satisfaites. D'autre part, on vérifie par récurrence, que pour tout réservoir hydraulique ℓ et pour tout pas de temps t, $x_t^\ell \in [x_{\min}^\ell, x_{\max}^\ell]$. Soit ℓ un réservoir hydraulique. Par hypothèse, la relation est bien vérifiée pour ce réservoir ℓ et pour $t = 1$. Supposons maintenant que pour $1 \leq t \leq T-1$, on ait $x_t^\ell \in [x_{\min}^\ell, x_{\max}^\ell]$. Si $x_t^\ell + a_t^\ell > x_{\max}^\ell$ alors on définit $dev_t^\ell = x_t^\ell + a_t^\ell - x_{\max}^\ell$. On a bien $dev_t^\ell \geq 0$, et puisque $v_t^\ell = 0$, $x_{t+1}^\ell = x_t^\ell + a_t^\ell - dev_t^\ell = x_{\max}^\ell \in [x_{\min}^\ell, x_{\max}^\ell]$. Sinon si $x_t^\ell + a_t^\ell \leq x_{\max}^\ell$, on définit $dev_t^\ell = 0$, et donc $x_{t+1}^\ell = x_t^\ell + a_t^\ell \geq x_t^\ell$, et $x_{t+1}^\ell \in [x_{\min}^\ell, x_{\max}^\ell]$.

□

On pourrait donner d'autres conditions moins restrictives mais plus compliquées à énoncer sur la faisabilité du problème. Notons cependant que la condition (4.13) revient à dire que l'on peut satisfaire la demande en utilisant les unités de production thermiques et nucléaires (rappelons que les centrales nucléaires sont en général modélisées, du point de vue des contraintes de fonctionnement, comme des centrales thermiques), ce qui est en général le cas.
La modélisation qui vient d'être présentée suppose par ailleurs que tous les paramètres incertains du système (les taux de disponibilité, les apports hydrauliques des réservoirs et les consommations électriques) sont connus. Auquel cas, le problème d'optimisation déterministe (4.12) que l'on vient de présenter fournit effectivement un planning de production optimal. Cependant, en pratique, ces paramètres sont incertains. L'objectif est alors de déterminer des stratégies de gestion c'est à dire des plans de productions adaptatifs (qui s'adapteront aux réalisations des paramètres) et permettant, en suivant cette stratégie, de satisfaire la consommation quelle que soit cette consommation. Ces stratégies de gestion dépendront des hypothèses faites sur les paramètres incertains du modèle. Nous présenterons 3 approches différentes pour tenir compte de l'incertitude des paramètres du modèle. Nous montrerons ensuite, comment, dans chaque cas, mettre en place des stratégies de gestion stables ou robustes.

4.3 Modélisation des paramètres incertains

Dans cette section, et dans la suite de ce chapitre, on notera les variables aléatoires avec les mêmes lettres que leur réalisation, mais surmontée d'une tilde.

4.3.1 Modélisation des taux de disponibilité

Soit ℓ, une centrale thermique comportant n_ℓ groupes thermiques. Soit $\alpha_{j,\ell}(t)$ la probabilité que le groupe j de la centrale thermique ℓ fonctionne au pas de temps t et $U_{j,\ell}^t$ la variable aléatoire telle que $U_{j,\ell}^t = 1$ si le groupe j fonctionne au pas de temps t et $U_{j,\ell}^t = 0$ sinon. On suppose que les groupes thermiques sont régulièrement contrôlés et, le cas échéant

réparés tous les m_0 pas de temps. Entre deux contrôles consécutifs, on suppose que les taux de disponibilité des centrales ne changent pas. Cela signifie qu'entre deux contrôles consécutifs, un groupe donné va soit fonctionner soit être hors d'état de fonctionnement durant toute la période. Si $t_0 = 1$ et $t_k = m_0 k$ pour $k \in \mathbb{N}^*$, alors les probabilités $\alpha_{j,\ell}(t)$, pour $t = t_k, \ldots, t_{k+1} - 1$ sont les mêmes et on a seulement besoin d'évaluer $\alpha_{j,\ell}(t_k), k \geq 0$. Cette probabilité $\alpha_{j,\ell}(t_k)$ qu'un groupe j de la centrale ℓ fonctionne au pas de temps $m_0 k$ dépendra de l'évolution passée de la disponibilité de ce groupe. Si au pas de temps t_{k-1}, le groupe ne fonctionnait pas, il y a une grande probabilité (disons $1 - \beta_1^\ell$ avec β_1^ℓ petit) qu'il fonctionne au pas de temps t_k, et une petite probabilité β_1^ℓ qu'il soit encore hors d'état de fonctionnement au pas de temps t_k. Par contre, si le groupe fonctionnait pour les m dernières périodes délimitées par les $m+1$ dernières dates de contrôle, on peut supposer que plus il a fonctionné sans avoir de panne (plus m est grand) et plus il a de chances de tomber en panne au pas de temps t_k. Ainsi, il existe une fonction décroissante de m, $\beta_2^\ell(m)$ telle que pour tout groupe j de l'unité ℓ :

$$P(U_{j,\ell}^{t_k} = 1 | \text{le groupe j fonctionnait de } t_{k-m} \text{ à } t_{k-1}) = \beta_2^\ell(m).$$

Un cas particulier est le cas où le processus des états d'un groupe donné est une chaîne de Markov homogène où l'espace d'état est {P,F} où P représente l'état de panne et F l'état de fonctionnement. Dans ce cas, $\beta_2^\ell(m) = \beta_2^\ell$ est fixé et correspond à la probabilité qu'un groupe de la centrale ℓ fonctionne sur une période donnée sachant qu'il fonctionnait la période précédente. La matrice de transition pour les groupes de la centrale ℓ est donnée par :

$$P_\ell = \begin{pmatrix} \beta_1^\ell & 1 - \beta_1^\ell \\ 1 - \beta_2^\ell & \beta_2^\ell \end{pmatrix}.$$

La probabilité $\alpha_{j,\ell}(t_k)$ est alors donnée pour $k \geq 1$ par :

$$\alpha_{j,\ell}(t_k) = p_F^\ell(j) P_\ell^k(1,2) + p_W^\ell(j) P_\ell^k(2,2);$$

où $p_F^\ell(j) = 1 - p_P^\ell(j)$ et $p_F^\ell(j)$ est la probabilité que le groupe j de la centrale ℓ fonctionne au premier pas de temps. Afin de simplifier la présentation du modèle, nous supposons que pour une centrale ℓ donnée, au premier pas de temps soit tous les groupes thermiques fonctionnent soit tous les groupes sont en panne. Ainsi, $\alpha_{j,\ell}(t_k)$ est indépendant de j et $\alpha_\ell(t_k)$ représentera la probabilité qu'un groupe de la centrale ℓ fonctionne au pas de temps t_k. Soit $\bar{P}_{\max}^\ell$ la puissance maximale d'un groupe de la centrale ℓ. La puissance théorique maximale disponible de la centrale thermique ℓ est alors donnée par $P_{\max}^\ell = n_\ell \bar{P}_{\max}^\ell$. La puissance maximale disponible pour l'unité ℓ au pas de temps t_k est alors :

$$\tilde{P}_{\max}^{\ell,k} = \sum_{j=1}^{n_\ell} U_{j,\ell}^{t_k} \bar{P}_{\max}^\ell = n_\ell \bar{P}_{\max}^\ell \frac{\sum_{j=1}^{n_\ell} U_{j,\ell}^{t_k}}{n_\ell} = P_{\max}^\ell \tilde{\tau}_\ell(k).$$

Sous les hypothèses ci-dessus, la variable aléatoire $n_\ell \tilde{\tau}_\ell(k)$ suit une loi binomiale $\mathcal{B}(n_\ell, \alpha_\ell(t_k))$. De plus, $E[\tilde{\tau}_\ell(k)] = \alpha_\ell(t_k)$ et la variance de $\tilde{\tau}_\ell(k)$, $\mathrm{Var}(\tilde{\tau}_\ell(k)) = \frac{\alpha_\ell(t_k)(1-\alpha_\ell(t_k))}{n_\ell}$.

Nous ne faisons pas d'hypothèses paramétriques sur les lois des autres paramètres incertains du problème. Afin tenir compte de l'incertitude des paramètres sur la période de gestion, 3 voies différentes, que nous détaillons maintenant, peuvent être empruntées. Chacune d'entre elles se prête à une résolution particulière du problème de gestion de production électrique à l'horizon annuel.

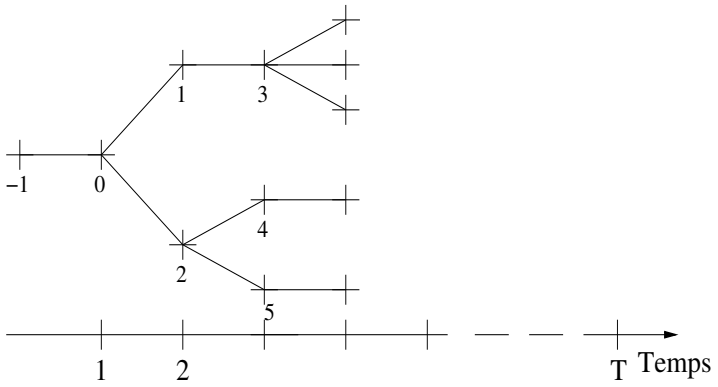


FIG. 4.3 – Arbre de scénarios.

4.3.2 Modélisation par arbre de scénarios

L'approche couramment utilisée pour résoudre le problème de gestion de production électrique consiste à modéliser l'évolution des paramètres incertains sur la période de gestion par un arbre de scénarios (figure 4.3). À chaque noeud n de l'arbre de scénarios sont attachées les réalisations des paramètres suivants :

- la consommation électrique $d(n)$ au noeud n;
- pour tout $\ell \in \mathcal{L}_H$, l'apport a_n^ℓ du réservoir ℓ;
- pour toute centrale thermique $\ell \in \mathcal{L}_T$, le taux de disponibilité τ_n^ℓ.

Le noeud racine correspond au premier pas de temps et les noeuds feuilles (les noeuds n'ayant pas de successeurs) aux derniers pas de temps. Plus généralement, chaque noeud n est associé à un pas de temps $\tau(n)$. On désignera par $\mathcal{O}$, l'ensemble des noeuds de l'arbre et par $\mathcal{O}_T$ l'ensemble des feuilles de l'arbre. Le nombre de scénarios représenté par cet arbre de scénarios est donc le nombre de feuilles $|\mathcal{O}_T|$, chaque scénario correspondant à un chemin dans l'arbre allant du noeud racine à une feuille. On utilisera également les notations suivantes :

- $F(n)$ est l'ensemble des noeuds fils du noeud n, et $P(n)$ le noeud père du noeud n.
- π_n est la probabilité de se trouver en un noeud n de l'arbre ($\sum_{\tau(n)=t} \pi_n = 1, \quad \forall\, t = 1, \ldots, T,$) et $\pi_T(n)$ est la probabilité de transition entre les noeuds $P(n)$ et n.

Adoptant une telle représentation des paramètres incertains sous la forme d'un arbre de scénarios, nous pouvons naturellement formuler le problème d'optimisation comme un problème de contrôle stochastique à temps discret et horizon fini sur cet arbre de scénarios représentant le comportement des états et entrées aléatoires. On chercher alors à minimiser le coût de production moyen sur l'arbre de scénarios. Les états et commandes qui étaient définis pour chaque pas de temps dans la version déterministe du problème de gestion de production électrique sont maintenant indexés par les noeuds n de l'arbre. En particulier, les commandes de la centrale hydraulique ℓ au noeud n sont $u_n^\ell = (v_n^\ell, dev_n^\ell)$, où v_n^ℓ et dev_n^ℓ sont respectivement les productions électriques de l'usine ℓ et la quantité d'eau déversée du réservoir ℓ pour ce noeud. Pour toute centrale hydraulique ℓ, soit x_n^ℓ le volume du réservoir ℓ au noeud n (au début du pas de temps $\tau(n)$). Les contraintes liées au fonctionnement

des centrales hydrauliques s'expriment alors sous la forme :

$$\begin{cases} x_n^\ell = x_{P(n)}^\ell + a_{P(n)}^\ell - v_{P(n)}^\ell - dev_{P(n)}^\ell & \forall\ n\ \in \mathcal{O},\ \ \forall\, \ell\ \ \in \mathcal{L}_H, \\ x_{min}^\ell \leq x_n^\ell + a_n^\ell - v_n^\ell - dev_n^\ell \leq x_{\max}^\ell & \forall\ n \in \mathcal{O}_\mathcal{T}, \\ 0 \leq v_n^\ell \leq P_{max}^\ell\, Duree(\tau(n)) & \forall\ (n,\ell), \\ 0 \leq dev_n^\ell & \forall\ (n,\ell), \\ x_{min}^\ell \leq x_n^\ell \leq x_{max}^\ell & \forall\ (n,\ell). \end{cases} \tag{4.14}$$

Pour chaque contrat EJP ℓ, x_n^ℓ est le stock d'énergie encore disponible au noeud n (au début du pas de temps $\tau(n)$). On connaît la valeur x_1^ℓ, du stock d'énergie disponible sur le contrat ℓ au début de la période de gestion. Les contraintes portant sur les contrats EJP sont alors :

$$\begin{cases} u_n^\ell(u_n^\ell - P_{\max}^\ell\, Duree(\tau(n))) = 0 & \forall\, n \in \mathcal{O}, \quad \forall \ell \in \mathcal{L}_J, \\ x_{P(n)}^\ell - u_{P(n)}^\ell = x_n^\ell & \forall\, n \in \mathcal{O}, \quad \forall \ell \in \mathcal{L}_J, \\ x_n^\ell \geq 0 & \forall\, n \in \mathcal{O}, \quad \forall \ell \in \mathcal{L}_J, \\ x_n^\ell - u_n^\ell \geq 0 & \forall\, n \in \mathcal{O}_T, \quad \forall \ell \in \mathcal{L}_J. \end{cases} \tag{4.15}$$

Le coût de production de la centrale thermique ℓ au noeud n est donné par $c_\ell\, u_n^\ell$, où u_n^ℓ est la production thermique de la centrale thermique ℓ au noeud n. Les contraintes sur les niveaux de production des centrales thermiques sont données par :

$$0 \leq u_n^\ell \leq \tau_n^\ell\, P_{\max}^\ell\, Duree(\tau(n)),\ \ \forall\, n \in \mathcal{O},\ \ \forall \ell \in \mathcal{L}_T.$$

Les contraintes de satisfaction de la demande s'écrivent alors :

$$d(n) \leq \sum_{\ell \in \mathcal{L}_T} u_n^\ell + \sum_{\ell \in \mathcal{L}_H} v_n^\ell + \sum_{\ell \in \mathcal{L}_J} u_n^\ell,\ \ \forall n \in \mathcal{O};$$

et la fonction objectif à minimiser est :

$$\sum_{\ell \in \mathcal{L}_T} \sum_{n \in \mathcal{O}_T} \pi_n\, c_\ell\, u_n^\ell - \sum_{\ell \in \mathcal{L}_J} \sum_{n \in \mathcal{O}_T} \pi_n V_J^\ell(x_n^\ell - u_n^\ell) - \sum_{\ell \in \mathcal{L}_H} \sum_{n \in O_T} \pi_n\, V_H^\ell(x_n^\ell + a_n^\ell - v_n^\ell + dev_n^\ell).$$

Pour un scénario répertorié dans l'arbre de scénarios, on dispose donc d'un planning de production. Pour un scénario non représenté dans l'arbre, nous expliquerons plus loin comment exploiter une solution du problème dual obtenu en dualisant les contraintes couplantes du problème.
Des variantes pourraient être introduites dans cette modélisation du problème sous la forme d'un problème d'optimisation sur un arbre de scénarios. Dans [35], par exemple, un dédoublement de toutes les variables de commandes est introduit de façon à présenter le problème sous la forme d'un problème d'optimisation stochastique bi-étapes avec recours ; les variables de dédoublement introduites étant des variables de recours permettant d'ajuster les commandes réelles finales aux réalisations des paramètres. Un problème d'optimisation stochastique bi-étapes est également formulé dans [25]. Dans ce problème, les variables de décision de la première étape correspondent aux décisions de mise en route/arrêt de certaines centrales et les variables de décision de la deuxième étape aux décisions de mise en route/arrêt des centrales restantes et aux niveaux de production des usines.

4.3.3 Modèle statistique pour la demande et les apports

Une première façon de s'affranchir de l'utilisation d'arbre de scénarios est de supposer que la demande et les apports pour chaque réservoir sur la période de gestion sont des vecteurs aléatoires pour lesquels on supposera que la moyenne et la matrice de covariance sont connues. Il s'agit ici d'un modèle minimaliste, se proposant d'utiliser seulement ces deux caractéristiques alors que l'utilisation des arbres de scénarios nécessite d'estimer des distributions de probabilité. Ainsi, nous écrirons l'apport $\tilde{a}_t^\ell$ de la réserve $\ell \in \mathcal{L}_H$:

$$\tilde{a}_t^\ell = \bar{a}_t^\ell + \varepsilon_t^\ell, \;\; t = 1, \ldots, T,$$

où $\bar{a}_t^\ell$ est l'apport moyen pour le réservoir ℓ au pas de temps t et $\varepsilon^\ell = (\varepsilon_t^\ell)_{t=1}^T$ est un bruit centré de matrice de covariance Γ_ℓ.
De même, nous écrirons la demande $\tilde{d}_t$ au pas de temps t :

$$\tilde{d}_t = \bar{d}_t + \varepsilon'_t, \;\; t = 1, \ldots, T,$$

où $\bar{d}_t$ est la demande moyenne au pas de temps t et $\varepsilon' = (\varepsilon'_t)_{t=1}^T$ est un bruit centré de matrice de covariance Γ.

4.3.4 Ensembles d'incertitude

Une autre manière de tenir compte de l'incertitude sur les paramètres du problème est de considérer que bien qu'inconnus, ces paramètres appartiennent à un ensemble d'incertitude $\mathcal{Z}$ convexe compact et non vide qui lui est connu. Aucune hypothèse statistique n'est faite sur ces paramètres. On retrouve ici le contexte décrit dans la section 1.5 grâce auquel nous pourrons proposer des CR, des CRA et des CRAA du problème de gestion de production électrique. Remarquons que d'un point de vue pratique, l'ensemble $\mathcal{Z}$ sera estimé et l'hypothèse faite dans cette section ne sera pas forcément vérifiée ; à savoir que les valeurs des paramètres sur la période de gestion n'appartiendront pas nécessairement à l'ensemble $\mathcal{Z}$.

4.4 Contreparties robustes et stables

Nous avons proposé, dans l'article [55] deux méthodes stables de régularisation de la fonction duale que nous comparerons, dans la section 4.7, avec les méthodes de gestion développées dans cette section. La première approche proposée dans cet article (que nous noterons VaR_{CA}) revient à envisager une relaxation primale de la demande : comment contrôler le chiffre d'affaire des ventes avec une demande incertaine ? La seconde approche présentée dans cet article (que nous noterons VaR_{benef}), provient d'une relaxation duale : comment contrôler les coûts de production alors que la disponibilité des centrales thermiques est aléatoire ? Les deux approches proposées dans l'article [55] supposent que l'évolution des paramètres incertains est modélisée par un arbre de scénarios.
Dans cette section, nous travaillons sur une relaxation convexe du problème de gestion de production électrique. Les contraintes $u_t^\ell \in \{0, Duree(t)\ P_{\max}^\ell\}$ sur les commandes appliquées au pas de temps t au contrat EJP $\ell \in \mathcal{L}_J$, seront remplacées par

$$0 \leq u_t^\ell \leq Duree(t)\ P_{\max}^\ell, \quad \ell \in \mathcal{L}_J. \tag{4.16}$$

Cela signifie que pour chaque contrat EJP, on dispose d'une certaine réserve d'énergie, qui, chaque jour, et tant que cette réserve n'est pas épuisée, peut être utilisée tout ou partie de la journée. Cette modélisation revient à considérer qu'un contrat EJP est une réserve hydraulique particulière sans apports. Cette relaxation permettra notamment de se placer dans le cadre de travail de la section 1.6 pour pouvoir mettre en place des CRA et des CRAA. D'autre part, le problème dual du problème (4.12) (utilisé dans la méthode de régularisation de la fonction duale présentée en section 4.4.1) est alors équivalent au problème primal.

4.4.1 Régularisation de la fonction duale

Dans cette section, les hypothèses faites sur les paramètres incertains sont celles de la section 4.3.3. Nous présentons une dernière méthode de régularisation de la fonction duale permettant de stabiliser les solutions duales et de réduire l'écart type des coûts de gestion simulés (par rapport aux méthodes VaR_{CA} et VaR_{benef}). Nous ne tenons pas compte dans ce nouveau modèle de l'incertitude sur les apports hydrauliques. Soit $\theta(\lambda, \tau, d)$ (on insiste dans cette écriture sur la dépendance de cette fonction par rapport aux taux de disponibilité τ et à la demande d), la fonction duale obtenue en ayant dualisé les contraintes couplantes de satisfaction de la demande (λ est donc le vecteur des multiplicateurs de Lagrange associés à cette dualisation) du problème de gestion de production électrique. Si la demande et les taux de disponibilité étaient connus, on aurait à résoudre :

$$\max_{\lambda \in \mathbb{R}^T} \theta(\lambda, \tau, d). \tag{4.17}$$

Une manière simple de tenir compte de l'incertitude serait de remplacer chaque paramètre incertain par une estimation de sa moyenne sur la période de gestion. Cette approche revient à résoudre le problème dual $\max_{\lambda \in \mathbb{R}^T} E[\theta(\lambda, \tilde{\tau}, \tilde{d})]$. Afin de tenir compte de la variance de $\tilde{d}$ et $\tilde{\tau}$, on propose de remplacer ce problème par le problème :

$$\max_{\lambda \in \mathbb{R}^T} E[\theta(\lambda, \tilde{\tau}, \tilde{d})] - \kappa(\varepsilon) \sqrt{Var(\theta(\lambda, \tilde{\tau}, \tilde{d}))}, \tag{4.18}$$

où $\kappa(\varepsilon) > 0$ est un facteur de risque fixé. Il s'agit de pénaliser la fonction objectif par l'écart type de $\theta(\lambda, \tilde{\tau}, \tilde{d})$. Une autre manière de limiter cet écart type est d'ajouter une contrainte du type $\sqrt{Var(\theta(\lambda, \tilde{\tau}, \tilde{d}))} \leq \alpha$. Le multiplicateur de Lagrange optimal associé à la dualisation de cette contrainte peut fournir une estimation de $\kappa(\varepsilon)$. On peut remarquer aussi que l'approche que nous proposons ici revient à mettre en place une approche Value-at-Risk sur le problème dual (pour lequel $\tilde{d}$ et $\tilde{\tau}$ sont aléatoires). Cette remarque permet de choisir $\kappa(\varepsilon) = \sqrt{\frac{1-\varepsilon}{\varepsilon}}$ (voir section 2.1.2.1). La fonction duale $\theta(\lambda, \tilde{\tau}, \tilde{d})$ peut se décomposer ainsi :

$$\theta(\lambda, \tilde{\tau}, \tilde{d}) = \lambda^T \tilde{d} + \theta_T(\lambda, \tilde{\tau}) + \theta_H(\lambda) + \theta_J(\lambda),$$

où $\theta_T(\lambda, \tilde{\tau})$ est la contribution des centrales thermiques au problème dual, $\theta_H(\lambda)$ la contribution des centrales hydrauliques au problème dual et $\theta_J(\lambda)$ la contribution des contrats EJP (les centrales nucléaires sont modélisées comme des centrales thermiques). La fonction

$\theta_T(\lambda, \tilde{\tau})$ est explicite et est la seule à dépendre des taux de disponibilité :

$$\theta_T(\lambda, \tilde{\tau}) = \sum_{k=1}^{T/m_0} \sum_{t=(k-1)m_0+1}^{km_0} Duree(t) \sum_{c_\ell < \lambda_t} \tilde{\tau}_\ell(k) P^\ell_{\max}(c_\ell - \lambda_t).$$

La fonction $\theta_H(\lambda)$ ne dépend ni de la demande ni des taux de disponibilité et se calcule facilement en résolvant $\mathcal{L}_H$ problèmes linéaires de taille $2T+1$ (le i-ème problème linéaire correspondant à la i-ème centrale hydraulique). De façon similaire, la fonction $\theta_J(\lambda)$ associée aux sous-problèmes EJP ne dépend ni de la demande ni des taux de disponibilité et se calcule en résolvant $\mathcal{L}_J$ problèmes linéaires de taille $T+1$ (le i-ème problème linéaire correspondant au i-ème contrat EJP). Rappelons que l'on considère dans cette section que $\tilde{d}$ est un vecteur aléatoire de moyenne $\bar{d}$ et de matrice de covariance $Cov(d) = \Gamma$. Nous avons alors :

$$E[\theta(\lambda, \tilde{\tau}, \tilde{d})] = \theta_H(\lambda) + \theta_J(\lambda) + \lambda^T \bar{d} + \sum_{t=1}^{T} \sum_{c_\ell < \lambda_t} Duree(t) \alpha_\ell (c_\ell - \lambda_t) P^\ell_{\max},$$

et

$$Var(\theta(\lambda, \tilde{\tau}, \tilde{d})) = \lambda^T \Gamma \lambda + \sum_{k=1}^{T/m_0} \sum_{\ell \in \mathcal{L}_H} (f_{k,\ell} P^\ell_{\max})^2 \frac{\alpha_\ell (1-\alpha_\ell)}{n_\ell},$$

où

$$f_{k,\ell} = \sum_{\substack{(k-1)m_0 + 1 \leq t \leq km_0 \\ c_\ell \leq \lambda_t}} Duree(t)(c_\ell - \lambda_t).$$

Les solutions d'un problème linéaire sont instables. L'intérêt de la modélisation présentée dans cette section est d'introduire un terme permettant de régulariser le problème et de stabiliser les solutions duales. Lorsque seule la demande est incertaine, le nouveau problème dual correspond à une robustification du problème dual (4.17) en choisissant comme ensemble d'incertitude pour les consommations l'ellipsoïde :

$$\mathcal{E} = \{d \in \mathbb{R}^n, \ \ (d - \bar{d})^T \Gamma^{-1} (d - \bar{d}) \leq \kappa(\varepsilon)^2\}.$$

Ce nouveau problème dual correspond au problème primal initial auquel on a rajouté la contrainte $d \in \mathcal{E}$.
Dans les deux sections suivantes, on suppose que les paramètres incertains suivent les hypothèses de la section 4.3.4.

4.4.2 Contreparties robustes ajustées

L'objectif de ce cette section est d'expliquer comment mettre en place des contreparties robustes ajustées du problème de gestion de production électrique (4.12) pour lequel les contraintes sur les commandes EJP ont été relâchées et remplacées par (4.16).
Il s'agit ensuite de déterminer dans ce problème, les variables qui sont ajustables et celles qui ne le sont pas. Cette décision est un choix de modélisation et différentes réponses peuvent être apportées à ce problème. On peut voir les variables de commandes comme des variables non ajustables et les variables restantes (en particulier les variables d'état)

comme des variables ajustables. Néanmoins, au début de l'horizon de gestion, aucun des paramètres incertains n'a été observé et donc seules les commandes à prendre le premier jour (ou au premier pas de temps) ne sont réellement pas ajustables. Nous allons alors adopter le choix de modélisation suivant : nous considérons que les variables de commande du premier jour de la période de gestion sont non ajustables et que toutes les autres variables sont ajustables. Si l'ensemble d'incertitude des paramètres incertains est l'enveloppe convexe de S scénarios $(d_t(s), a_t^\ell(s), \tau_t^\ell(s))_{s=1}^S$, alors si $S' = \{1, 2, \ldots, S\}$, la contrepartie robuste ajustée du problème de gestion de production électrique étudié est alors le problème (4.53) donné à la fin de ce chapitre.

On peut se demander si la contrepartie robuste ajustée est équivalente à la contrepartie robuste du problème (4.12) obtenue en choisissant le même ensemble d'incertitude polytopique ou si la contrepartie robuste ajustée permet de fournir de meilleurs résultats. Remarquons tout d'abord, que du fait des apports hydrauliques, l'hypothèse B de la section 1.5 n'est pas satisfaite. Par conséquent, les hypothèses du théorème [11] fournissant des conditions sous lesquelles la CRA et la CR sont équivalentes, ne sont pas satisfaites. En fait, si l'ensemble d'incertitude est polytopique, on peut donner des exemples de problèmes de gestion de production électrique pour lesquels la CR et la CRA sont équivalentes (même lorsque ce n'est pas le même scénario qui pour chaque pas de temps a reçu le moins de précipitations depuis le début de la période de gestion). Par exemple, considérons une centrale thermique, une centrale hydraulique et 3 pas de temps. La demande vaut $d = (1, 3, 2)^T$ et les apports sont dans l'enveloppe convexe des 2 scénarios d'apports $(1, 3, 2)$ et $(1, 2, 1)^T$. Le coût unitaire de production thermique est de 1 et la fonction de valorisation du stock d'eau l'identité. Dans ce cas, la CR et la CRA sont équivalentes. Nous montrerons aussi, dans la section 4.7, à travers divers exemples, que la CRA peut fournir de meilleurs résultats. Cependant, même lorsque les valeurs optimales de la CR et de la CRA sont les mêmes, la solution de la CRA fournit une solution, qui, en moyenne, sera très certainement meilleure que la solution robuste qui n'est pas adaptative. Si l'ensemble d'incertitude est une enveloppe convexe de scénarios $(\xi_1, \ldots, \xi_S)$, et si l'on note $(u, v_1, \ldots, v_S)$ la solution de la CRA, alors lorsque se produit le scénario $\sum_{i=1}^{S} \alpha_i \, \xi_i$, la solution de la CRA nous fournit un point $(u, \sum_{i=1}^{S} \alpha_i \, v_i)$ faisable et dont le coût associé est inférieur ou égal à la valeur optimale de la CRA. La valeur optimale de la CRA permet de connaître le pire coût que l'on pourrait obtenir en utilisant la solution de la CRA (et ainsi avoir une idée des budgets maximums qu'il faudra débloquer sur la période de gestion).

Par contre, l'utilisation de cette méthode d'un point de vue pratique pose un problème. En effet, pour un pas de temps donné, la dépendance des commandes en fonction des réalisations passées des paramètres n'est pas connue. Afin de connaître les commandes préconisées par la CRA, il faudrait connaître les valeurs $(\alpha_1, \ldots, \alpha_S)$, telles que le vecteur des paramètres sur la période de gestion est $\sum_{i=1}^{S} \alpha_i \, \xi_i$. Auquel cas on peut utiliser les commandes $(u, \sum_{i=1}^{S} \alpha_i \, v_i)$. Cependant, tant que tous les pas de temps ne sont pas écoulés,

on ne peut pas déterminer ces coefficients α_i. Un moyen d'utiliser, d'un point de vue pratique, la méthodologie de la CRA, pour robustifier le problème (4.12), consiste à mettre en place cette CRA à chaque pas de temps avec un horizon de gestion de plus en plus court. Nous appelerons cette méthode la CRA glissante. Plus précisément, au début de la période de gestion, afin de connaître les commandes à appliquer pour le premier pas de temps, on met en place une CRA avec ensemble d'incertitude polytopique en prenant un horizon de gestion de T pas de temps. Les variables non ajustables (et donc connues au début de la période de gestion) de cette CRA correspondent aux commandes à appliquer au premier pas de temps. Plus généralement, pour déterminer les commandes au pas de temps $t \leq T$, on résoud la CRA avec ensemble d'incertitude polytopique pour le problème (4.12) correspondant aux $T-t+1$ derniers pas de temps. Les variables non ajustables de ce problème correspondent aux commandes à appliquer au pas de temps t. Cette méthode sera testée sur un problème réel de grande taille en section 4.7.1.

4.4.3 Contreparties robustes ajustées affinement

Afin de connaître la dépendance des variables ajustables en fonction des réalisations passées des paramètres un moyen simple consiste à choisir une forme paramétrique particulière pour cette dépendance. Lorsque les variables ajustables sont des fonctions linéaires des paramètres incertains on aboutit à la notion de CRAA présentée dans la section 1.6. Potentiellement, tous les paramètres incertains sont susceptibles d'influencer les variables ajustables. Néanmoins, afin d'éviter une explosion de la taille de la CRAA, on cherche, pour une variable ajustable donnée, les paramètres incertains les plus susceptibles de l'influencer. Cette étape fait partie intégrante de la modélisation et des tests seront nécessaires pour savoir si un choix donné conduit à une amélioration significative par rapport à l'utilisation d'une CR classique. Les paramètres influençant le plus les commandes sont les consommations électriques d_t aux différents pas de temps t. Au pas de temps t, on peut donc raisonnablement supposer que les commandes thermiques $u_t^\ell, \ell \in \mathcal{L}_T$, hydrauliques $v_t^\ell, \ell \in \mathcal{L}_H$, et les commandes des contrats EJP, $u_t^\ell, \ell \in \mathcal{L}_J$, dépendent linéairement de consommations électriques d_j observées à des pas de temps $j \in I_t$, où I_t est un sous-ensemble donné de $\{1, \ldots, t\}$:

$$\begin{aligned} u_t^\ell &= p_t^{\ell,0} + \sum_{j \in I_t} p_t^{\ell,j}\, d_j, \quad \ell \in \mathcal{L}_J, \quad u_t^\ell = q_t^{\ell,0} + \sum_{j \in I_t} q_t^{\ell,j}\, d_j, \quad \ell \in \mathcal{L}_T, \\ v_t^\ell &= r_t^{\ell,0} + \sum_{j \in I_t} r_t^{\ell,j}\, d_j, \quad \ell \in \mathcal{L}_H. \end{aligned} \tag{4.19}$$

Les nouvelles variables non ajustables sont maintenant $p_t^{\ell,j}$, $q_t^{\ell,j}$, et $r_t^{\ell,j}$. On va donc chercher les meilleures combinaisons linéaires définies par (4.19) en choisissant différents ensembles d'incertitude pour la demande. Il semble aussi raisonnable de penser que les commandes appliquées à une centrale hydraulique ℓ dépendent des apports hydrauliques du réservoir alimentant cette centrale. On pourra donc aussi, si cela conduit à une CRAA soluble en temps raisonnable, enrichir l'information à partir de laquelle les commandes hydrauliques seront calculées sous la forme :

$$v_t^\ell = r_t^{\ell,0} + \sum_{j \in I_t} r_{t,1}^{\ell,j}\, d_j + \sum_{j \in I_t} r_{t,2}^{\ell,j}\, a_j^\ell, \quad \ell \in \mathcal{L}_H. \tag{4.20}$$

On peut bien évidemment combiner ces schémas de dépendance. On peut décider que certaines variables de commande (par exemple les variables des problèmes thermiques et

hydrauliques) sont ajustables et gouvernées par les équations (4.19) ou (4.20) et que les autres variables sont non ajustables.
L'ensemble I_t dépend de l'information disponible à la date t. En pratique, on connaît, pour chaque pas de temps, toutes les réalisations passées des paramètres du problème et donc le choix le plus naturel pour I_t est $I_t = \{1, \ldots, t-1\}$. Si l'information sur les valeurs des paramètres arrive avec un décalage de m pas de temps, on a alors au mieux, $I_t = \{1, \ldots, t-m\}$. Afin de réduire la taille des CRAA et afin de privilégier les informations récentes, on peut choisir de ne s'intéresser qu'aux valeurs des paramètres pour les m derniers pas de temps en prenant $I_t = \{t-1, \ldots, t-m\}$. On peut imaginer aussi des choix dynamiques du sous-ensemble I_t dont la taille dépende du pas de temps t et des données récoltées jusqu'à la date t. Par la suite, nous fixons $1 \leq m \leq T-1$, et si $t \geq m+1$ alors on choisit $I_t = \{1, \ldots, t-m\}$ et sinon $I_t = \emptyset$.
Contrairement à la CRA présentée précédemment, la CRAA fournit une solution exploitable directement d'un point de vue pratique. En effet, au pas de temps t, les valeurs des paramètres incertains pour tous les pas de temps $j \in I_t$ sont connus. On peut donc utiliser la solution de la CRAA et les équations (4.19) et/ou (4.20) afin de déterminer les commandes $((u_t^\ell)_{\ell \in \mathcal{L}_T \cup \mathcal{L}_J}, (v_t^\ell)_{\ell \in \mathcal{L}_H})$ fournies par la CRAA. Une question qui se pose ensuite naturellement est de savoir si une CRAA donnée (obtenue avec un choix particulier pour l'ensemble d'incertitude des paramètres incertains et une politique donnée de dépendance des variables ajustables en fonction de ces paramètres incertains) et la CR utilisant le même ensemble d'incertitude pour les paramètres incertains sont équivalentes. En fait, même dans le cas où les valeurs optimales de la CRAA et de la CR sont les mêmes, l'utilisation des commandes données par (4.19) et/ou (4.20) (et donc l'utilisation de la solution de la CRAA) sera à privilégier car elle a de grandes chances de fournir une meilleure solution en moyenne (même si les pires scénarios pour la CR et la CRAA ont les mêmes coûts).

4.4.3.1 Ensemble d'incertitude polytopique

Supposons que toutes les variables de commande sont ajustables et déterminées par les équations (4.19) pour les commandes thermiques et EJP et (4.20) pour les commandes hydrauliques. L'ensemble d'incertitude le plus simple que l'on peut calibrer est un ensemble polytopique. Supposons que ce polytope soit l'enveloppe convexe des scénarios $((d_t(s), a_t^\ell(s), \tau_t^\ell(s))_{s=1}^S$, et notons $S' = \{1, \ldots, S\}$. La CRAA du problème (4.12) est alors, sous ces choix, le problème d'optimisation linéaire (4.54) donné à la fin de ce chapitre.

4.4.3.2 Contraintes de boîte sur les paramètres incertains

On traite maintenant le cas où les commandes thermiques et EJP sont déterminées par les équations (4.19) ; les commandes hydrauliques par les équations (4.20) et les ensembles d'incertitude sont définis par des contraintes de boîte. Plus précisément on suppose qu'il existe des vecteurs $d^{\min}$, $d^{\max}$, $(a^{\ell,\min})_{\ell \in \mathcal{L}_H}$, $(a^{\ell,\max})_{\ell \in \mathcal{L}_H}$, $(\tau^{\ell,\min})_{\ell \in \mathcal{L}_T}$, $(\tau^{\ell,\max})_{\ell \in \mathcal{L}_T}$, tels que :

$$\begin{array}{ll} \forall t = 1, \ldots, T, & d_t^{\min} \leq d_t \leq d_t^{\max}, \\ \forall t = 1, \ldots, T, & \tau_t^{\ell,\min} \leq \tau_t^\ell \leq \tau_t^{\ell,\max}, \ \forall \ell \in \mathcal{L}_T. \end{array} \qquad a_t^{\ell,\min} \leq a_t^\ell \leq a_t^{\ell,\max}, \ \forall \ell \in \mathcal{L}_H, \tag{4.21}$$

Ce cas présente un grand intérêt pratique puisque de "bons" ensembles d'incertitude de ce type peuvent être obtenus en utilisant des procédures statistiques (voir par exemple la section 4.6).

Remarque 4.4.1 *Soit $n \in \mathbb{N}^*$ et soient $(x, y_{\min}, y_{\max}) \in \mathbb{R}^n \times \mathbb{R}^n_+ \times \mathbb{R}^n_+$ alors :*

$$\min_{y \in [y_{\min}, y_{\max}]} x^T y = y_{\min}^T x^+ + y_{\max}^T x^-; \quad \max_{y \in [y_{\min}, y_{\max}]} x^T y = y_{\max}^T x^+ + y_{\min}^T x^-.$$

Nous supposerons que $m \neq 0$; la CRAA du problème (4.12) pouvant aussi se calculer facilement lorsque $m = 0$. En injectant les schémas de dépendance (4.19) et (4.20) dans (4.12) et en utilisant la remarque 4.4.1, on peut montrer que la CRAA du problème (4.12) en choisissant pour ensemble d'incertitude l'ensemble défini par (4.21), est le problème linéaire (4.52) fourni à la fin de ce chapitre. En pratique, cette approche conduit à des CRAA de grande taille ; le nombre de contraintes et le nombre de variables étant de l'ordre de $O(T^2\mathcal{L})$. Plus précisément le nombre de variables est :

$$1+|\mathcal{L}_{\mathcal{H}}|+|\mathcal{L}_J|+3m(1+|\mathcal{L}_J|)+T(3+m+3|\mathcal{L}_J|+2|\mathcal{L}_H|)+\frac{(T-m)(T-m+1)(3+3|\mathcal{L}_T|+12|\mathcal{L}_H|+|\mathcal{L}_J|)}{2}.$$

4.4.3.3 Ensemble d'incertitude ellipsoïdal sur les consommations

Nous mettons ici en place la CRAA du problème (4.12) lorsque les commandes thermiques sont des fonctions de la demande explicitées en (4.19), les commandes hydrauliques suivent le schéma général donné en (4.20) et enfin, les ensembles d'incertitude sont ellipsoïdaux pour la demande, et définis par les contraintes de boîte (4.21) pour les apports et les taux de disponibilité. Par contre, afin de simplifier la présentation des résultats, les commandes EJP ne sont pas considérées ajustables. Le vecteur des demandes peut donc être n'importe quel point de l'ellipsoïde suivant :

$$\mathcal{E} = \{d \in \mathbb{R}^T, \ | \ (d - \bar{d})^T \Gamma^{-1} (d - \bar{d}) \leq \kappa^2\};$$

où le vecteur $\bar{d}$ représente la demande moyenne aux différents pas de temps de la période de gestion ; Γ est une matrice définie positive et κ un niveau de confiance dont dépend la taille de l'ellipsoïde. Dans ces conditions, après quelques calculs élémentaires, on montre que la CRAA du problème (4.12) est le problème conique du second ordre (4.55) donné en fin de ce chapitre.

On pourrait aussi envisager de choisir comme ensemble d'incertitude pour les apports hydrauliques de la réserve $\ell \in \mathcal{L}_H$, un ellipsoïde $\mathcal{E}_\ell$ centré sur le vecteur $\bar{a}^\ell$ des apports moyens aux différents pas de temps et défini par :

$$\mathcal{E}_\ell = \{a^\ell \in \mathbb{R}^T, \ (a^\ell - \bar{a}^\ell)^T \Gamma_\ell^{-1} (a^\ell - \bar{a}^\ell) \leq \kappa_\ell^2\},$$

où Γ_ℓ est une matrice définie positive et $\kappa_\ell > 0$ un facteur de risque.

Finalement, les CRAA du problème (4.12) obtenues en utilisant :

1. des ensembles d'incertitude polytopiques, définis par des contraintes de boîte ou ellipsoïdaux pour les paramètres incertains ;
2. les schémas de dépendance (4.19) (ou éventuellement (4.20) pour les commandes hydrauliques) pour les variables ajustables ;

ont une formulation explicite traitable qui est un problème d'optimisation linéaire ou conique du deuxième ordre.

4.5 Méthodes de résolution

Les CRA et CRAA présentées dans les sections 4.4.2 et 4.4.3 sont des problèmes linéaires ou coniques du deuxième ordre que nous résolvons directement par des méthodes de points intérieurs.
Nous allons montrer comment le problème d'optimisation déterministe (4.12) peut se résoudre par la technique de la décomposition par les prix en utilisant la relaxation lagrangienne (en dualisant les contraintes couplantes de satisfaction de la demande). La même méthodologie pourra ensuite être appliquée afin de résoudre le problème d'optimisation stochastique multi-étapes présenté dans la section 4.3.2 lorsque l'évolution des paramètres incertains sur la période de gestion est représentée par un arbre de scénarios. Des précisions sur les spécificités de l'implémentation de la méthode dans ce contexte seront données. Ce schéma de décomposition qui pourrait être comparé à la méthode de Programmation Dynamique Stochastique Duale [77] revient à adapter [26] au problème de gestion de production électrique annuelle.

4.5.1 Méthodologie

Nous noterons (x, u) le vecteur des variables de décision du problème (4.12), x étant l'ensemble des variables d'état et u l'ensemble des variables de commandes. Soit $\lambda \in \mathbb{R}^T$, soit $d \in \mathbb{R}^T$ le vecteur des consommations électriques et soit L le Lagrangien obtenu en dualisant les contraintes couplantes de satisfaction de la demande :

$$L(x,u,\lambda) = \lambda^T d + \sum_{\ell \in \mathcal{L}_T} \sum_{t=1}^{T} (c_\ell - \lambda_t) u_t^\ell + \sum_{\ell \in \mathcal{L}_H} f_H^\ell(x,u,\lambda) + \sum_{\ell \in \mathcal{L}_J} f_H^\ell(x,u,\lambda),$$

avec

$$f_H^\ell(x,u,\lambda) = -a_\ell - \sum_{t=1}^{T} \lambda_t \, v_t^\ell, \quad \ell \in \mathcal{L}_H,$$

et

$$f_J^\ell(x,u,\lambda) = -b_\ell - \sum_{t=1}^{T} \lambda_t \, u_t^\ell, \quad \ell \in \mathcal{L}_J.$$

En notant χ l'ensemble de faisabilité du problème (4.12), afin de résoudre le problème d'optimisation primal (4.12), on doit résoudre $\min\limits_{(x,u)\in\chi} \max\limits_{\lambda} L(x,u,\lambda)$. Ce problème revient à résoudre le problème dual $\max\limits_{\lambda} \theta(\lambda)$ avec $\theta(\lambda) = \min\limits_{(x,u)\in\chi} L(x,u,\lambda)$, si seules des unités thermiques, nucléaires et hydrauliques sont prises en compte. En effet, dans ce cas, (4.12) est un problème d'optimisation linéaire sur un ensemble borné et dans ce cas le problème primal et le problème dual sont équivalents (dans le sens où ils ont même valeur optimale). Si l'on tient compte des contrats EJP, l'ensemble des contraintes n'est plus convexe et le saut de dualité est strictement positif. Néanmoins le théorème de la dualité faible nous donne :

$$\min_{(x,u)\in\chi} \max_{\lambda} L(x,u,\lambda) \geq \max_{\lambda} \theta(\lambda). \tag{4.22}$$

De plus, des simulations numériques faites sur le problème de gestion de production électrique tenant compte des contrats EJP, montrent que le saut de dualité est faible. Le problème dual nous permet alors d'approximer la valeur optimale primale.

On remarque que la fonction duale θ est non différentiable, concave et séparable par rapport aux unités de production. En effet, si χ_ℓ est l'ensemble des contraintes portant sur les commandes et les états de l'unité de production ℓ, la fonction duale s'écrit : $\theta(\lambda) = \lambda^T d + \sum_{\ell \in \mathcal{L}} \theta^\ell(\lambda)$ avec

$$\theta^\ell(\lambda) = \min_{(x_\ell, u_\ell) \in \chi_\ell} f_H^\ell(x_\ell, u_\ell, \lambda), \quad \text{si } \ell \in \mathcal{L}_H, \tag{4.23}$$

$$\theta^\ell(\lambda) = \min_{(x_\ell, u_\ell) \in \chi_\ell} f_J^\ell(x_\ell, u_\ell, \lambda), \quad \text{si } \ell \in \mathcal{L}_J, \tag{4.24}$$

et

$$\theta^\ell(\lambda) = \min_{u_\ell \in \chi_\ell} \sum_{t=1}^{T} (c_\ell - \lambda_t)\, u_t^\ell, \quad \text{si } \ell \in \mathcal{L}_T. \tag{4.25}$$

Cette méthode est particulièrement intéressante pour traiter des problèmes de grande taille tel que celui auquel nous nous intéressons. Pour maximiser θ (ou ce qui revient au même pour minimiser la fonction convexe $-\theta$) nous utilisons une méthode de faisceaux décrite dans [65]. Ceci nécessite la construction d'une boîte noire, qui, pour tout $\lambda \in \mathbb{R}^T$, est capable de calculer $-\theta(\lambda)$ et de donner un sous-gradient arbitraire $s(\lambda) \in \partial(-\theta(\lambda))$. Soit L_ℓ le Lagrangien partiel associé à la fonction duale partielle correspondant au moyen de production ℓ. Le calcul de $-\theta(\lambda)$ est effectué en résolvant les différents problèmes d'optimisation associés aux différentes unités de production. À chaque itération de la méthode des faisceaux, on doit donc résoudre $\mathcal{L}$ sous-problèmes : le nombre de variables d'un sous-problème étant de l'ordre du nombre de variables de commandes et d'états de l'unité de production associée à ce sous-problème. Concernant le calcul d'un sous-gradient, si $(x_\ell^*(\lambda), u_\ell^*(\lambda)) = \text{argmin}_{(x,u)\chi_\ell} L_\ell(x, u, \lambda)$, alors

$$s(\lambda) = -d + \sum_{\ell \in \mathcal{L}} u_\ell^*(\lambda) \in \partial(-\theta(\lambda)).$$

Plus précisément, en se référant à [19], une résolution globale à l'aide d'un schéma itératif peut se décrire en 4 étapes en partant d'un prix de référence λ utilisé pour initialiser l'algorithme avec l'index $k = 1$ et $\lambda_1 = \lambda$:

1. À l'iteration k, décomposition en sous-problèmes et calcul de la solution locale du sous-problème ℓ : $(x_\ell^*(\lambda_k), u_\ell^*(\lambda_k))$;
2. Évaluation de la fonction duale θ au point λ_k et calcul d'un sous-gradient $s(\lambda_k)$;
3. Mise à jour des multiplicateurs de Lagrange par le coordinateur en utilisant une boîte noire (i.e calcul de λ_{k+1}) ;
4. Mise à jour de l'index k : $k \leftarrow k + 1$ et retour à l'étape 1.

4.5.2 Résolution directe des différents sous-problèmes

Nous décrivons brièvement la première étape de l'algorithme de décomposition par les prix que l'on vient de présenter. Les sous-problèmes thermiques duaux ont des solutions évidentes et les sous-problèmes hydrauliques duaux sont des problèmes d'optimisation linéaires résolus en utilisant une méthode de points intérieurs. Nous verrons dans la section suivante, une méthode efficace de résolution de ces sous-problèmes par programmation dynamique. Le sous-problème EJP associé au contrat EJP ℓ consistant à calculer $\theta^\ell(\lambda)$ est un

problème d'optimisation NP complet, non convexe résolu par programmation dynamique stochastique. Soit λ la valeur courante du multiplicateur de Lagrange. On connaît, au pas de temps $T+1$, pour chaque contrat EJP ℓ, et pour chaque pas de stock x, les valeurs de Bellman $V_J^\ell(x, T+1)$ de la réserve EJP ℓ. Entre deux pas de stocks, rappelons que la fonction de Bellman est linéaire. On en déduit, en utilisant les équations de HJB, (phase arrière) les valeurs de Bellman $V^\ell(x,t)$ pour tout contrat ℓ, tout pas de temps t, et tout pas de stock x en utilisant les formules de récurrence :

$$\begin{cases} V^\ell(x,t) = \max\ V^\ell(x - u_t^\ell, t+1) + \lambda_t\ u_t^\ell \\ u_t^\ell(u_t^\ell - Duree(t)\ P_{\max}^\ell) = 0, \quad x - u_t^\ell \geq 0. \end{cases} \tag{4.26}$$

En connaissant le stock d'énergie sur chaque contrat ℓ au début de la période de gestion, la phase avant consiste ensuite à déduire les commandes EJP optimales $u_\ell^*(\lambda)$ à appliquer à chaque pas de temps en utilisant les valeurs de Bellman calculées dans la phase arrière. Lorsque la méthode de décomposition par les prix est utilisée pour résoudre le problème d'optimisation stochastique multi-étapes sur un arbre de scénarios décrit en section 4.3.2, les sous-problèmes EJP se résolvent aussi par programmation dynamique de la façon suivante. On modifie l'arbre de scénarios de façon à ajouter à chaque feuille de l'arbre un noeud fils qui est une duplication de son père. Les valeurs de Bellman sont dans un premier temps indexées par le pas de stock x et le noeud n. Pour chaque contrat EJP ℓ, on connaît les valeurs de Bellman $V^\ell(x,n)$ pour tous les pas de stock x et toutes les feuilles n du nouvel arbre construit. On en déduit, pour toute réserve EJP ℓ, la valeur de Bellman $V^\ell(x,n)$ au pas de stock x et au noeud n par :

$$V^\ell(x,n) = \begin{cases} \max\ (\sum_{m \in F(n)}\ \pi_T(m)(V^\ell(x - u_n^\ell, m) + \lambda_n\ u_n^\ell)) \\ x - u_n^\ell \geq 0, \quad u_n^\ell \in \{0, Duree(\tau(n))\ P_{\max}^\ell\}. \end{cases} \tag{4.27}$$

Une phase avant permet ensuite d'utiliser ces valeurs de Bellman pour calculer les commandes EJP en chaque noeud de l'arbre. Les deux types de méthodes utilisées (programmation dynamique et méthodes de points intérieurs) ont une complexité dépendant de la dimension de l'espace dual. Un défaut reconnu de la programmation dynamique est que sa complexité croît exponentiellement avec la dimension de l'espace d'état.

4.5.3 Résolution des sous-problèmes hydrauliques et EJP par programmation dynamique

L'objet de cette section est de présenter une méthode originale de résolution du problème de gestion de production électrique par programmation dynamique lorsqu'on ne dispose que d'une centrale hydraulique et de centrales thermiques. On applique ensuite cette méthodologie pour calculer efficacement les contributions $\theta^\ell(\lambda)$ (définies par (4.23) et (4.24)) de chaque centrale hydraulique ℓ et chaque contrat EJP ℓ à la fonction duale $\theta(\lambda)$ définie dans la section 4.5.1.

4.5.3.1 Préliminaires

Supposons dans un premier temps que nous ne disposions que de centrales thermiques. Notons alors pour tout $\ell \in \mathcal{L}_T$, $M^\ell = (\tau_1^\ell Duree(1) P_{\max}^\ell, \ldots, \tau_T^\ell Duree(T) P_{\max}^\ell)^T$, le vecteur spécifiant les capacités de production maximales de la centrale ℓ aux différents pas de temps et M_j le vecteur spécifiant les quantités maximales de production au pas de temps j pour

les différents moyens de production. On suppose aussi que la dernière centrale thermique a une capacité de production infinie et donc que $P_{\max}^{\mathcal{L}_T} = +\infty$ ou encore que $M^{\mathcal{L}_T} = +\infty \mathbf{e}$. Quitte à réordonner les moyens de production, on suppose enfin que les composantes du vecteur coût thermique $c = (c_1, \ldots, c_{\mathcal{L}_T})^T$, sont ordonnées :

$$0 \leq c_1 \leq c_2 \leq \ldots \leq c_{\mathcal{L}_T}.$$

Ainsi, au pas de temps j, si on n'utilise que des centrales thermiques, le coût de production optimal permettant de satisfaire une demande de t MWh, revient à calculer le coût optimal $\psi_{M_j}(t)$ suivant :

$$\psi_{M_j}(t) = \min_u \{c^T u \ : \ 0 \leq u \leq M_j, \ \sum_{\ell=1}^{\mathcal{L}_T} u_\ell = t \ \}.$$

Ainsi, pour tout $1 \leq j \leq T$, $\psi_{M_j}(t)$ est une fonction linéaire par morceaux, convexe et telle que $\psi_{M_j}(0) = 0$. Elle a des 'kinks' aux points :

$$t_k(M_j) = \sum_{\ell=1}^{k} M_j^\ell, \ \ k = 0, \ldots, \mathcal{L}_T - 1.$$

Pour tout $1 \leq j \leq T$, posons naturellement $t_{\mathcal{L}_T}(M_j) = +\infty$. Les fonctions $(\psi_{M_j}(.))_{j=1}^T$ sont différentiables en tout point différent des 'kinks' :

$$\begin{array}{rcll} \psi'_{M_j}(t) & = & 0, & t < 0, \\ \psi'_{M_j}(t) & = & c_{k+1}, & t_\ell(M_j) < t < t_{\ell+1}(M_j) \quad \ell = 0, \ldots, \mathcal{L}_T - 1. \end{array} \tag{4.28}$$

Dans ce qui suit, nous utiliserons les fonctions duales :

$$\psi^*_{M_j}(\tau) = \max_\alpha [\alpha\tau - \psi_{M_j}(\alpha)].$$

Remarquons que dom $\psi^*_{M_j} = [0, c_{\mathcal{L}_T}]$. Cette fonction est linéaire par morceaux. Un calcul simple fournit :

$$\psi^*_{M_j}(\tau) = \sum_{\ell=1}^{\mathcal{L}_T - 1} M_j^\ell (\tau - c_\ell)_+. \tag{4.29}$$

Nous allons montrer comment le problème de gestion de production électrique peut se résoudre par programmation dynamique lorsque l'on rajoute un moyen de production hydroélectrique. On notera $v_j^1 = v_j$ la production de la centrale hydroélectrique au pas de temps j et les contraintes sur les commandes hydrauliques s'écrivent :

$$0 \leq v_j \leq V_j, \ \ j = 1, \ldots, T; \tag{4.30}$$

si pour tout $j = 1, \ldots, T$, $V_j = Duree(j) P_{\max}$, où $P_{\max}$ est la puissance maximale de la centrale hydraulique. Nous supposerons que la consommation électrique d_j est supérieure ou égale à V_j. Le coût de production pour satisfaire la demande d_j au pas de temps j est $\psi_{M_j}(d_j - v_j)$. Afin de tenir compte de la valorisation du stock d'eau, au lieu d'introduire une fonction de valorisation de ce stock d'eau, nous imposons les contraintes suivantes sur les niveaux x_j du réservoir hydraulique au début des différents pas de temps j :

$$x_j \geq \underline{x}_j \geq 0, \quad j = 2, \ldots, T+1.$$

Nous notons x_1 le volume initial du réservoir de la centrale hydraulique et pour tout $1 \leq t \leq T$, $a_t^1 = a_t$, l'apport hydraulique de ce réservoir au pas de temps t. Si les consommations d_t et les apports a_t sont connus, nous devons résoudre le problème d'optimisation suivant :

$$\begin{array}{rcl} C^* & = & \min\limits_{x,v\in\mathbb{R}^T} \{ \quad \sum\limits_{j=1}^{T} \psi_{M_j}(d_j - v_j) \ : \quad 0 \leq v \leq V, \ \ x \geq \underline{x}, \\ & & \qquad\qquad x_{j+1} = x_j + a_j - v_j, \ \ j = 1, \ldots, T \ \}. \end{array} \tag{4.31}$$

Nous envisageons alors une formulation duale de ce problème en introduisant les multiplicateurs de Lagrange $\lambda \in \mathbb{R}_+^T$ pour les contraintes d'inégalité $x \geq \underline{x}$:

$$\begin{array}{rcl} C^* & = & \min\limits_{x,v\in\mathbb{R}^T} \max\limits_{\lambda\in\mathbb{R}_+^T} \{ \quad \lambda^T(\underline{x} - x) + \sum\limits_{j=1}^{T} \psi_{M_j}(d_j - v_j) \ : \ 0 \leq v \leq V, \\ & & \qquad\qquad x_{j+1} = x_j + a_j - v_j, \ j = 1, \ldots, T\}, \\ & = & \max\limits_{\lambda\in\mathbb{R}_+^T} \min\limits_{x,v\in\mathbb{R}^T} \{ \quad \lambda^T(\underline{x} - x) + \sum\limits_{j=1}^{T} \psi_{M_j}(d_j - v_j) \ : \ 0 \leq v \leq V, \\ & & \qquad\qquad x_{j+1} = x_j + a_j - v_j, \ j = 1, \ldots, T\}. \end{array}$$

On exprime ensuite, comme nous l'avons fait dans la section 4.4.2, les variables x en fonction des variables v : $x_{k+1} = x_1 + \sum\limits_{j=1}^{k} a_j - v_j, \ \ k = 1, \ldots, T$. Ainsi nous avons :

$$\begin{array}{rcl} \lambda^T(\underline{x} - x) & = & \sum\limits_{k=1}^{T} \lambda_k \left(\underline{x}_{k+1} - x_1 + \sum\limits_{j=1}^{k} (v_j - a_j) \right), \\ & = & \lambda^T \underline{x} - x_1 \sum\limits_{k=1}^{T} \lambda_k + \sum\limits_{k=1}^{T} \lambda_k \sum\limits_{j=1}^{k} (v_j - a_j), \\ & = & \lambda^T \underline{x} - x_1 \sum\limits_{k=1}^{T} \lambda_k + \sum\limits_{j=1}^{T} (v_j - a_j) \sum\limits_{k=j}^{T} \lambda_k. \end{array}$$

Si pour tout $j = 1, \ldots, T$, on note $s_j = \sum\limits_{k=j}^{T} \lambda_k$, alors :

$$\lambda_j = s_j - s_{j+1}, \quad j = 1, \ldots, T-1, \quad \lambda_T = s_T.$$

Ainsi,

$$\begin{array}{rcl} \lambda^T \underline{x} & = & s_T \, \underline{x}_{T+1} + \sum\limits_{j=1}^{T-1} (s_j - s_{j+1}) \bar{x}_{j+1} \quad = \quad s_T \, \underline{x}_{T+1} + \sum\limits_{j=1}^{T-1} s_j \, \underline{x}_{j+1} - \sum\limits_{j=2}^{T} s_j \, \underline{x}_j, \\ & = & s_1 \, \underline{x}_2 + \sum\limits_{j=2}^{T} s_j \, (\underline{x}_{j+1} - \underline{x}_j). \end{array}$$

Soit $\underline{d}_j = \underline{x}_{j+1} - \underline{x}_j, \ j = 1, \ldots, T$, avec $\underline{x}_1 = x_1$. Il vient :

$$\lambda^T(\underline{x} - x) = s^T \underline{d} + s^T(v - a).$$

Remarquons que $s_1 \geq s_2 \geq \ldots, \geq s_T \geq 0$. Par conséquent nous obtenons la représentation suivante :

$$\begin{aligned} C^* &= \min_{0\leq v\leq V} \max_{0\leq s_T\leq\ldots\leq s_1} \left(s^T(\underline{d}+v-a) + \sum_{j=1}^{T} \psi_{M_j}(d_j - v_j) \right) \\ &= \max_{0\leq s_T\leq\ldots\leq s_1} \left(s^T(\underline{d}-a) + \sum_{j=1}^{T} \min_{0\leq v_j\leq V_j} \left[\psi_{M_j}(d_j - v_j) + v_j s_j\right] \right). \end{aligned}$$

Soit

$$\phi_j(\tau) = \min_{\alpha} \left[\alpha\tau + \psi_{M_j}(d_j - \alpha) \ : \ 0 \leq \alpha \leq V_j \ \right], \ \ j = 1, \ldots, T.$$

Le problème à résoudre est alors :

$$C^* = \max_{s} \left(s^T(\underline{d}-a) + \sum_{j=1}^{T} \phi_j(s_j) \ : \ 0 \leq s_T \leq \ldots \leq s_1 \right). \tag{4.32}$$

La complexité du problème (4.32) dépend de la structure des fonctions concaves $\phi_j(.)$. Puisque

$$\phi_j(\tau) = \min_{\alpha} \left[\alpha\tau + \psi_{M_j}(d_j - \alpha) \ : \ 0 \leq \alpha \leq V_j\right],$$

pour $\tau \geq \psi'_{M_j}(d_j)$, (où $\psi'_{M_j}(d_j)$ représente la dérivée à gauche de ψ_{M_j} en d_j), la solution optimale $\alpha(\tau)$ du problème ci-dessus est zéro. Ainsi,

$$\phi_j(\tau) = \psi_{M_j}(d_j), \ \ \tau \geq \psi'_{M_j}(d_j).$$

De façon similaire,

$$\phi_j(\tau) = V_j\tau + \psi_{M_j}(d_j - V_j), \ \ \tau \leq \psi'_{M_j}(d_j - V_j),$$

où $\psi'_{M_j}(d_j - V_j)$ représente ici la dérivée à droite de ψ_{M_j} en $d_j - V_j$. Pour des valeurs intermédiaires de τ,

$$\psi'_{M_j}(d_j - V_j) \leq \tau \leq \psi'_{M_j}(d_j),$$

nous avons $\tau = \psi'_{M_j}(d_j - \alpha(\tau))$. Ainsi, en utilisant (4.29), nous avons,

$$\phi_j(\tau) \ = \ d_j\tau - \psi^*_{M_j}(\tau) \ = \ d_j\tau - \sum_{\ell=1}^{\mathcal{L}_T-1} M_j^{\ell}(\tau - c_\ell)_+.$$

Les dérivées directionnelles à droite et à gauche introduites ci-dessus sont données par :

$$\begin{aligned} \psi'_{M_j}(d_j - V_j) &= \max_{\ell} \{c_{\ell+1} \ : \ t_\ell(M_j) \leq d_j - V_j\}, \\ \psi'_{M_j}(d_j) &= \min_{\ell} \{c_{\ell+1} \ : \ t_{\ell+1}(M_j) \geq d_j\}. \end{aligned}$$

Ainsi, on peut en conclure que toutes les fonctions $(\phi_j(.))_{j=1}^T$, appartiennent à $\mathcal{S}(c)$, la classe des fonctions réelles concaves linéaires par morceaux, et pouvant avoir des 'kinks'

uniquement aux points c_ℓ, $\ell = 1, \ldots, \mathcal{L}_T$. L'ensemble $\mathcal{S}(c)$ est un cône convexe. Introduisons les fonctions $(f_j(.))_{j=1}^T$ définies par :

$$f_j(\tau) = (\underline{d}_j - a_j)\tau + \phi_j(\tau), \quad j = 1, \ldots, T.$$

Définissons enfin récursivement les deux suites de fonctions univariées $\Phi_j(.)$ et $\Psi_j(.)$ suivantes :

$$\begin{array}{rcl} \Phi_{T+1}(\tau) & = & 0, \\ \Psi_k(\tau) & = & \Phi_{k+1}(\tau) + f_k(\tau), \\ \Phi_k(\tau) & = & \max\limits_{0 \leq \alpha \leq \tau} \Psi_k(\alpha), \quad k = T, \ldots, 1. \end{array} \tag{4.33}$$

Théorème 4.5.1 *Pour tout k=T,...,1, nous avons*

$$\begin{array}{rcl} \Phi_k(\tau) & = & \max\limits_s \left(\sum\limits_{j=k}^T f_j(s_j) \ : \ 0 \leq s_T \leq \ldots \leq s_k \leq \tau \right), \\ \max\limits_{\tau \geq 0} \Psi_k(\tau) & = & \max\limits_s \left(\sum\limits_{j=k}^T f_j(s_j) \ : \ 0 \leq s_T \leq \ldots \leq s_k \right). \end{array} \tag{4.34}$$

Preuve. Montrons la première égalité dans (4.34) par induction. Pour $k = T$, elle est valide par définition. Supposons qu'elle soit vraie pour un certain k, $2 \leq k \leq T$. Alors :

$$\begin{array}{rcl} \Phi_{k-1}(\tau) & = & \max\limits_{0 \leq \alpha \leq \tau} [\Phi_k(\alpha) + f_{k-1}(\alpha)] \\ & = & \max\limits_{0 \leq \alpha \leq \tau} \left[f_{k-1}(\alpha) + \max\limits_s \left(\sum\limits_{j=k}^T f_j(s_j) \ : \ 0 \leq s_T \leq \ldots \leq s_k \leq \alpha \right) \right] \\ & = & \max\limits_s \left(\sum\limits_{j=k-1}^T f_j(s_j) \ : \ 0 \leq s_T \leq \ldots \leq s_k \leq s_{k-1} \leq \tau \right). \end{array}$$

Ainsi, la première égalité dans (4.34) est valable pour tout $k = 1, \ldots, T$. Par conséquent

$$\begin{array}{rcl} \max\limits_{\tau \geq 0} \Psi_k(\tau) & = & \max\limits_{\tau \geq 0} [\Phi_{k+1}(\tau) + f_k(\tau)] \\ & = & \max\limits_{\tau \geq 0} \left[f_k(\tau) + \max\limits_s \left(\sum\limits_{j=k+1}^T f_j(s_j) \ : \ 0 \leq s_T \leq \ldots \leq s_{k+1} \leq \tau \right) \right] \\ & = & \max\limits_s \left(\sum\limits_{j=k}^T f_j(s_j) \ : \ 0 \leq s_T \leq \ldots \leq s_{k+1} \leq s_k \right). \end{array}$$

□

En utilisant le théorème 4.5.1, nous avons $C^* = \max\limits_{\tau \geq 0} \Psi_1(\tau)$. Soit

$$\tau_k^* = \min \{\tau \ : \ \Psi_k(\tau) \geq \Psi_k(\alpha), \forall \alpha \in \mathbb{R}\}, \quad k = 1, \ldots, T.$$

Un algorithme de calcul de $C^* = \Psi_1(\tau_1^*)$ peut alors être déduit du lemme suivant.

Lemme 4.5.1 *Toutes les fonctions Ψ_k et Φ_k, $k = 1, \ldots, T$, appartiennent à $\mathcal{S}(c)$.*

Preuve. Remarquons que $f_j(\tau) \in \mathcal{S}(c)$, $j = 1, \ldots, T$. Nous allons prouver le lemme par induction. En effet, $\Phi_{T+1} \in \mathcal{S}(c)$. Supposons que $\Phi_{k+1} \in \mathcal{S}(c)$ pour un certain $k \leq T$. Alors, de façon immédiate, $\Psi_k \in \mathcal{S}(c)$. Ainsi, τ_k^* coïncide avec l'une des valeurs c_ℓ, $\ell = 1, \ldots, \mathcal{L}_T$. Puisque

$$\Phi_k(\tau) = \max_{0 \leq \alpha \leq \tau} \Psi_k(\alpha) = \begin{cases} \Psi_k(\tau), & \tau \leq \tau_k^*, \\ \Psi_k(\tau_k^*), & \tau > \tau_k^*, \end{cases}$$

nous concluons que $\Phi_k \in \mathcal{S}(c)$.

□

Nous pouvons ainsi mettre à jour de manière récursive et explicite les fonctions Φ_k et Ψ_k. D'autre part, chacune des opérations suivantes se fait en $O(|\mathcal{L}_T|)$ opérations :

- addition de deux fonctions de $\mathcal{S}(c)$;
- calcul de la valeur maximale d'une fonction de $\mathcal{S}(c)$;
- calcul d'une représentation explicite de la fonction Φ_k à partir d'une représentation explicite de la fonction Ψ_k.

Nous arrivons alors à la conclusion suivante.

Théorème 4.5.2 *La valeur optimale C^* du problème* (4.32) *peut se calculer par programmation dynamique en $O(|\mathcal{L}_T|\, T)$ opérations arithmétiques.*

Nous montrons maintenant comment générer des solutions explicites pour les problèmes (4.31) et (4.32).

Lemme 4.5.2 *Définissons la suite suivante :*

$$s_1^* = \tau_1^*, \quad s_k^* = \min\{s_{k-1}^*, \tau_k^*\}, \quad k = 2, \ldots, T.$$

Alors le vecteur $s^* = (s_1^*, \ldots, s_T^*)$ fournit une solution optimale pour le problème (4.32).

Preuve. Montrons par induction que

$$C^* = \Phi_{k+1}(s_k^*) + \sum_{j=1}^{k} f_j(s_j^*), \quad k = 1, \ldots, T. \tag{4.35}$$

D'après le théorème 4.5.1, nous avons

$$C^* = \max_{\tau \geq 0} \Psi_1(\tau) = f_1(s_1^*) + \Phi_2(s_1^*).$$

Ainsi, (4.35) est vérifiée pour $k = 1$. Supposons qu'elle soit vérifiée pour un certain k, $1 \leq k \leq T - 1$. Alors :

$$C^* = \sum_{j=1}^{k} f_j(s_j^*) + \max_{0 \leq s_{k+1} \leq s_k^*} \Psi_{k+1}(s_{k+1}).$$

Par conséquent,

$$\begin{aligned} \max_{0 \leq s_{k+1} \leq s_k^*} \Psi_{k+1}(s_{k+1}) &= \begin{cases} \Psi_{k+1}(\tau_{k+1}^*), & \text{si } \tau_{k+1}^* \leq s_k^*, \\ \Psi_{k+1}(s_k^*), & \text{sinon,} \end{cases} \\ &= \Psi_{k+1}(s_{k+1}^*) = \Psi_{k+2}(s_{k+1}^*) + f_{k+1}(s_{k+1}^*). \end{aligned}$$

Ainsi, nous avons prouvé (4.35) pour tout $k = 1, \ldots, T$. Remarquons que $\Phi_{T+1} = 0$. Puisque s^* est faisable pour le problème (4.32), nous concluons que c'est une solution optimale de ce problème.

□

Nous venons donc de montrer comment résoudre le problème de gestion de production électrique par programmation dynamique lorsqu'on prend en compte un nombre quelconque de centrales thermiques et une centrale hydraulique et lorsque l'on suppose que les demandes, les apports et les taux de disponibilité sont connus.

4.5.3.2 Application pour résoudre les sous-problèmes hydrauliques et EJP

La méthode que l'on vient d'exposer peut en particulier être utilisée pour résoudre les sous-problèmes hydrauliques duaux obtenus en dualisant la contrainte de satisfaction de la demande du problème de gestion de production électrique (pour lequel on remplace la fonction de valorisation du stock d'eau par une contrainte du type $x_{T+1} \geq \underline{x}_{T+1}$). Cette méthode permet de traiter des problèmes comportant un grand nombre de pas de temps, et fournit la valeur optimale plus rapidement que les méthodes (telles que les méthodes de points intérieurs) qui ont été utilisées (à notre connaissance) jusqu'à présent pour résoudre ces sous-problèmes hydrauliques. Plus précisément, le ℓ-ème sous-problème hydraulique revient, pour λ fixé, à calculer la fonction duale partielle associée au ℓ-ème sous-problème hydraulique en résolvant :

$$\left\{ \begin{array}{l} \min \ -\sum_{j=1}^{T} \lambda_j \, v_j^\ell \\ 0 \leq v_j^\ell \leq V_j^\ell, \\ x^\ell \geq \underline{x}^\ell, \ \ x_{j+1}^\ell = x_1^\ell + \sum_{k=1}^{j} a_k^\ell - v_k^\ell, \ \ j = 1, \ldots, T; \end{array} \right. \tag{4.36}$$

ou encore, en reprenant les calculs qui viennent d'être faits :

$$\max_{0 \leq s_T^\ell \leq \ldots \leq s_1^\ell} \sum_{j=1}^{T} f_j^\ell(s_j^\ell); \tag{4.37}$$

avec $f_j^\ell(\tau) = (\underline{d}_j^\ell - a_j^\ell)\tau - (\lambda_j - \tau)_+ V_j^\ell$, où $\underline{d}_j^\ell = \underline{x}_{j+1}^\ell - \underline{x}_j^\ell$, $j = 1, \ldots, T$. Les relations de récurrence (4.33) peuvent ensuite être utilisées pour résoudre le problème (4.37) par programmation dynamique. Si on envisage de mettre en place une relaxation convexe (définie en section 4.4) des contraintes EJP, le sous-problème EJP dual peut aussi se résoudre par programmation dynamique. D'autre part, puisque le sous-problème thermique dual a une solution explicite, on sait déterminer de manière très efficace la valeur de la fonction duale $\theta(\lambda)$ en tout point λ.

Remarquons que l'on obtient ainsi la valeur optimale et une solution duale du problème (4.36). Il reste ensuite à déterminer une solution primale pour pouvoir exploiter cette méthode de résolution des sous-problèmes hydrauliques duaux afin de résoudre le problème de gestion de production électrique par un schéma de décomposition par les prix. En effet, pour résoudre le problème de gestion de production électrique par une méthode de décomposition par les prix, nous avons vu que l'on doit être capable de fournir, à chaque étape k, une valeur $\theta(\lambda_k)$ de la fonction duale en λ_k et un sous-gradient $s_k \in \partial\theta(\lambda_k)$

de θ en λ_k ; λ_k étant l'estimation de la solution duale à l'étape k. La contribution du ℓ-ème sous-problème hydraulique à la valeur $\theta(\lambda_k)$ est fournie par la valeur optimale du problème (4.36) en prenant $\lambda = \lambda_k$, cette valeur pouvant donc être obtenue par le schéma de programmation dynamique qui vient d'être décrit. Par contre, la contribution de ce sous-problème à un sous-gradient $s \in \partial\theta(\lambda_k)$ est $-\sum_{j=1}^{T} \lambda_k(j)\, v^{*\ell}(j)$ où $v^{*\ell}$ est une solution optimale du problème (4.36).
En pratique, les contraintes sur les niveaux des lacs sont $x_j^\ell \geq 0,\ j = 1, \ldots, T,\ x_{T+1}^\ell \geq \underline{x}_{T+1}^\ell > 0$; et les apports a_j^ℓ pour chaque pas de temps j et chaque réservoir ℓ sont inférieurs à la capacité de production V_j^ℓ de la centrale hydraulique ℓ au pas de temps j. D'autre part, les apports sur la période de gestion sont en général inférieurs au volume initial du réservoir, à savoir que pour tout réservoir ℓ :

$$x_1^\ell > \sum_{j=1}^{T} a_j^\ell.$$

Sous ces hypothèses nous pouvons énoncer le résultat suivant.

Lemme 4.5.3 *Pour toute centrale hydraulique $\ell \in \mathcal{L}_H$ et pour tout pas de temps t tel que $2 \leq t \leq T$:*

- $\forall \tau \geq 0, \Psi_t^\ell(\tau) = \Phi_t^\ell(\tau)$; *les fonctions Ψ_t^ℓ et Φ_t^ℓ sont linéaires par morceaux, concaves et ont des 'kinks' aux points $(\lambda_t, \lambda_{t+1}, \ldots, \lambda_T)$.*
- *Si $\tau \geq \max(\lambda_t, \lambda_{t+1}, \ldots, \lambda_T)$, alors $\Psi_t^\ell(\tau) = (x_1^\ell - \sum_{j=t}^{T} a_j^\ell)\tau$.*
- *Si $\tau \leq \min(\lambda_t, \lambda_{t+1}, \ldots, \lambda_T)$, Ψ_t^ℓ est une droite de pente supérieure ou égale à $x_1^\ell \geq 0$.*

Preuve. Fixons $\ell \in \mathcal{L}_H$ et montrons ce résultat par induction. Pour $t = T$, $\Phi_t^\ell(\tau) = f_T^\ell(\tau) = (x_1^\ell - a_T^\ell)\tau - (\lambda_T - \tau)_+ V_T^\ell$. Si $\tau \leq \lambda_T$, $\Phi_t^\ell(\tau)$ est donc une droite de pente $x_1^\ell + V_T^\ell - a_T^\ell \geq x_1$; et si $\tau \geq \lambda_T$, alors $\Phi_t^\ell(\tau) = (x_1^\ell - a_T^\ell)\tau$. Donc Ψ_T^ℓ est concave, linéaire par morceaux, croissante et a un 'kink' au point λ_T. Par conséquent, $\Phi_T^\ell(\tau) = \Psi_T^\ell(\tau)$, pour tout $\tau \geq 0$.
Supposons le résultat vrai pour un certain t tel que $3 \leq t \leq T$. Nous avons :

$$\begin{aligned} \Psi_{t-1}^\ell(\tau) &= \Phi_t^\ell(\tau) + f_{t-1}^\ell(\tau) \\ &= \Phi_t^\ell(\tau) - a_{t-1}^\ell \tau - (\lambda_{t-1} - \tau)_+ V_{t-1}^\ell. \end{aligned}$$

Donc si $\tau \geq \max(\lambda_{t-1}, \lambda_t, \ldots, \lambda_T)$ alors puisque $\Phi_t^\ell(\tau) = (x_1^\ell - \sum_{j=t}^{T} a_j^\ell)\tau$, on a bien $\Psi_t^\ell(\tau) = (x_1^\ell - \sum_{j=t-1}^{T} a_j^\ell)\tau$ ($\Psi_t^\ell(.)$ est une droite de pente positive). Si $\tau \leq \min(\lambda_t, \lambda_{t+1}, \ldots, \lambda_T)$, alors Ψ_{t-1}^ℓ est une droite de pente supérieure ou égale à $x_1^\ell - a_{t-1}^\ell + V_{t-1}^\ell \geq x_1^\ell$. $\Psi_{t-1}^\ell(.)$ est donc linéaire par morceaux, croissante, concave et a des kinks aux points $(\lambda_{t-1}, \lambda_t, \ldots, \lambda_T)$. Par conséquent, par définition de Φ_{t-1}^ℓ, alors pour tout $\tau \geq 0$, $\Phi_{t-1}^\ell(\tau) = \Psi_{t-1}^\ell(\tau)$.

□

Il en découle alors le résultat suivant.

Théorème 4.5.3 *On suppose, pour le problème* (4.36)*, que pour tout* $1 \leq j \leq T$, $a_j^\ell \leq V_j^\ell$, $\underline{x}_j^\ell = 0$ *si* $1 \leq j \leq T$ *et* $\underline{x}_{T+1}^\ell > 0$. *La solution optimale du problème* (4.37) *vérifie alors :*

$$\tau_1^* = s_1^* = \ldots = s_k^*,$$

et τ_1^* *est l'un des* T *scalaires* $(\lambda_1, \ldots, \lambda_T)$.

Preuve. Pour tout $\tau \geq 0$,

$$\Psi_1(\tau) = \Phi_2(\tau) - (x_1^\ell + a_1^\ell)\tau - (\lambda_1 - \tau)_+ V_1^\ell. \tag{4.38}$$

En utilisant le lemme 4.5.3, si $0 \leq \tau \leq \min(\lambda_1, \ldots, \lambda_T)$, alors Ψ_1 est une droite de pente supérieure ou égale à $x_1^\ell - x_1^\ell - a_1^\ell + V_1^\ell \geq V_1^\ell - a_1^\ell > 0$. D'autre part, si $\tau \geq \max(\lambda_1, \ldots, \lambda_T)$, alors $\Psi_1(\tau) = -\sum_{j=1}^{T} a_j^\ell \tau$, et $\Psi_1(.)$ coïncide avec une droite de pente négative. La fonction $\Psi_1(.)$ est donc linéaire par morceaux, concave, les changements de pentes intervenant aux points $(\lambda_1, \ldots, \lambda_T)$. Par conséquent τ_1^* est l'un de ces scalaires $(\lambda_1, \ldots, \lambda_T)$. On conclut en utilisant le lemme 4.5.2.

□

Pour résoudre le problème d'optimisation (4.37) il suffit donc d'évaluer la fonction coût du problème (4.37) pour les T vecteurs candidats fournis par ce théorème.

Il semble plus délicat d'arriver à résoudre le problème de gestion de production électrique par programmation dynamique en utilisant la méthodologie développée dans la section 4.5.3.1 lorsqu'on introduit des majorants sur les niveaux des lacs aux différents pas de temps et que l'on considère un nombre quelconque de centrales hydrauliques.
Dans le cas où les consommations électriques sont aléatoires, on est amené à définir de nouvelles fonctions $\phi_j(.)$ par :

$$\phi_j(\tau) = \min_{0 \leq \alpha \leq V_j} [\alpha\tau + E\psi_{M_j}(d_j - \alpha)]. \tag{4.39}$$

Lorsque la consommation électrique au pas de temps t suit une loi exponentielle ou une loi de Laplace de paramètre λ_t, on peut obtenir des expressions explicites des fonctions $\phi_j(.), f_j(.)$ et $\Psi_j(.)$ et calculer les τ_k^* par dichotomie.

Par ailleurs, pour conclure cette section consacrée aux méthodes de résolution des problèmes d'optimisation issus de nos modélisations, rappelons que le modèle proposé dans la section 4.4.1 consiste à régulariser la fonction duale (introduite en section 4.4.1) du problème de gestion de production électrique pour lequel les paramètres incertains sont remplacés par leur moyenne. La nouvelle fonction duale reste séparable par rapport aux différents moyens de production et peut être maximisée par une méthode de faisceaux. La t-ème composante d'un sous-gradient (on ne redonne pas les contributions de l'hydraulique et de l'EJP qui ne sont pas modifiées par rapport à la méthode nominale) est donnée par :

$$\bar{d}(t) - \sum_{\ell | c_\ell < \lambda_t} Duree(t)\alpha_\ell P_{\max}^\ell - \frac{\kappa(\varepsilon)}{\sqrt{Var(\theta(\lambda, \tau, d))}}((\Gamma\lambda)(t) +$$

$$Duree(t) \sum_{\ell | c_\ell < \lambda_t} (P_{max}^\ell)^2 \frac{\alpha_\ell(1 - \alpha_\ell)}{n_\ell} \sum_{m \in I_t^\ell} Duree(m)(\lambda_m - c_\ell)),$$

où pour tout $\ell \in \mathcal{L}_T$, si $\frac{t}{m_0} \in \mathbb{N}$ alors $I_t^\ell = \{m \mid (\frac{t}{m_0} - 1)m_0 + 1 \leq m \leq t, \text{ et } \lambda_m \geq c_\ell\}$ et sinon $I_t^\ell = \{m \mid E[\frac{t}{m_0}]m_0 + 1 \leq m \leq (E[\frac{t}{m_0}] + 1)m_0, \text{ et } \lambda_m \geq c_\ell\}$.

4.5.4 Stratégies de gestion

Une fois le problème résolu, il s'agit de mettre en place une stratégie de gestion afin de satisfaire les consommations électriques quelles que soient leurs valeurs sur cette période de gestion.
Si la méthodologie de la CRA est utilisée, nous avons vu en quoi une méthode glissante permet de déterminer les commandes de façon à satisfaire les demandes pour chaque pas de temps.
Si la méthodologie de la CRAA est utilisée, et si les réalisations des consommations électriques sont bien dans les ensembles d'incertitude prévus, alors on obtient les commandes en fonction des réalisations passées des consommations.
Si on utilise une méthode de décomposition par les prix, on obtient une solution duale λ^*. Le problème de reconstruction d'une solution primale approchée à partir d'une solution du problème dual pour le problème de gestion de production électrique a été étudié notamment dans [49] et [75].
La solution duale λ^* est indexée par les pas de temps si on met en place la régularisation de la fonction duale proposée dans la section 4.4.1 et par les noeuds de l'arbre de scénarios si on résoud le problème d'optimisation sur un arbre de scénarios défini dans la section 4.3.2 ou si on résoud ses variantes VaR_{CA} et VaR_{benef} décrites dans l'article [55]. La stratégie mise en place prend alors la forme de fonctions de Bellman en chaque pas de temps de la période de gestion ; ces fonctions étant calculées en utilisant la version suivante du principe de Bellman. Deux cas se présentent.
Le premier cas est celui où la solution duale est indexée par les noeuds de l'arbre de scénarios. On modifie encore l'arbre de scénarios de façon à ajouter à chaque feuille de l'arbre un noeud fils qui est une duplication de son père. Soit alors ℓ une réserve hydraulique ou un contrat EJP. Pour l'unité de production ℓ, les fonction de Bellman $V^\ell(x,n)$ sont connues pour chaque niveau de stock x d'un maillage de l'ensemble des stocks admissibles et pour chaque feuille n de l'arbre de scénarios. Entre deux points du maillage, la fonction de Bellman est supposée linéaire. On a en fait pour toute feuille n et tout point x du maillage de la réserve ℓ : $V^\ell(x,n) = V_H^\ell(x)$, si ℓ est une réserve hydraulique et $V^\ell(x,n) = V_J^\ell(x)$, si ℓ est un contrat EJP. Les valeurs de Bellman $V^\ell(x,n)$ pour tous les noeuds n de l'arbre de scénarios et tous les pas de stock x sont ensuite calculées en utilisant les formules de récurrence suivantes :

$$\begin{cases} V^\ell(x,n) = \max(\sum\limits_{m \in F(n)} \pi_T(m)(V^\ell(x + a_n^\ell - v_n^\ell + dev_n^\ell, m) \\ \qquad\qquad +\lambda_n^*\, v_n^\ell)) \\ x + a_n^\ell - x_{\max}^\ell \leq v_n^\ell + dev_n^\ell \leq x + a_n^\ell \\ 0 \leq v_n^\ell \leq Duree(\tau(n))\ P_{\max}^\ell, \quad 0 \leq dev_n^\ell; \end{cases} \tag{4.40}$$

si ℓ est un réservoir hydraulique et

$$\begin{cases} V^\ell(x,n) = \max(\sum\limits_{m \in F(n)} \pi_T(m)(V^\ell(x - u_n^\ell, m) + \lambda_n^* u_n^\ell)) \\ u_n^\ell(u_n^\ell - Duree(\tau(n))\ P_{\max}^\ell) = 0, \quad x - u_n^\ell \geq 0, \end{cases} \tag{4.41}$$

si ℓ est un contrat EJP. La fonction de Bellman pour le pas de temps t est alors donnée par : $V^\ell(x,t) = \sum_{n\in\, t}\ \pi(n)\, V^\ell(x,n)$ pour le pas de stock x et l'unité de production ℓ. Cet algorithme est simplement une programmation dynamique stochastique (PDS) effectuée sur les valeurs marginales.
Le deuxième cas est celui où la solution duale est indexée par les pas de temps. On procède de la même façon mais en calculant cette fois ci directement les valeurs $V^\ell(x,t)$. Si ℓ est une réserve hydraulique, la valeur de la fonction de Bellman $V^\ell(x,t)$ au pas de stock x et pas de temps t est donnée par :

$$\begin{cases} V^\ell(x,t) = \max\ V^\ell(x + a_t^\ell - v_t^\ell - dev_t^\ell, t+1) + \lambda_t^*\ v_t^\ell \\ x + a_t^\ell - x_{\max}^\ell \leq v_t^\ell + dev_t^\ell \leq x + a_t^\ell \\ 0 \leq v_t^\ell \leq Duree(t)\ P_{\max}^\ell, \quad 0 \leq dev_t^\ell; \end{cases} \tag{4.42}$$

et si ℓ est une réserve EJP par (les contraintes de l'EJP ayant été relâchées pour le problème étudié en section 4.4.1) :

$$\begin{cases} V^\ell(x,t) = \max\ V^\ell(x - u_t^\ell, t+1) + \lambda_t^*\ u_t^\ell \\ 0 \leq u_t^\ell \leq Duree(t)\ P_{\max}^\ell, \quad x - u_t^\ell \geq 0. \end{cases} \tag{4.43}$$

Les fonctions de Bellman sont calculées ici exactement à partir de la solution optimale λ^* du problème dual et non approximées comme précédemment. Une fois ces fonctions de Bellman calculées, il est possible d'effectuer une simulation de Monte-Carlo des plans de production, en utilisant $\delta_x V^\ell(x,t)$ comme un coût de production de l'énergie stockée dans la réserve ℓ et en utilisant les moyens de production par ordre croissant de leur coût de production.

4.6 Calibration des ensembles d'incertitude

Les méthodes robustes développées dans la section 4.4.2 et 4.4.3 nécessitent de déterminer des ensembles d'incertitude pour :

1. les apports en eau des réservoirs des centrales hydrauliques ;
2. la consommation électrique aux différents pas de temps ;
3. les taux de disponibilité des usines.

On utilisera les lois sur les taux de disponibilité des centrales thermiques afin de déterminer des ensembles d'incertitude sur ces taux. D'autre part, les ensembles d'incertitude sur les apports seront posés à priori. Notons cependant que le problème de l'étude statistique des apports hydrauliques pour des réservoirs d'usines hydrauliques a été abondamment étudié ([90] par exemple). Les séries de consommations électriques ont quant à elles été moins largement étudiées.
Afin de tester les méthodes robustes de gestion de production électrique des sections 4.4.2 et 4.4.3 en utilisant les historiques des consommations journalières d'électricité en France de 1996 à 2003 inclus, nous allons conduire, dans cette section, une analyse statistique de la série chronologique correspondant à ces données. Ces données, disponibles sur le site internet de RTE fournissent donc 8 années d'historiques de consommations électriques. Les méthodes robustes seront mises en place pour déterminer des stratégies de gestion sur l'année $x+1$ (pour $3 \leq x \leq 7$) en utilisant les x premières années d'historiques. Notre objectif est donc de déterminer des ensembles d'incertitude pour les consommations

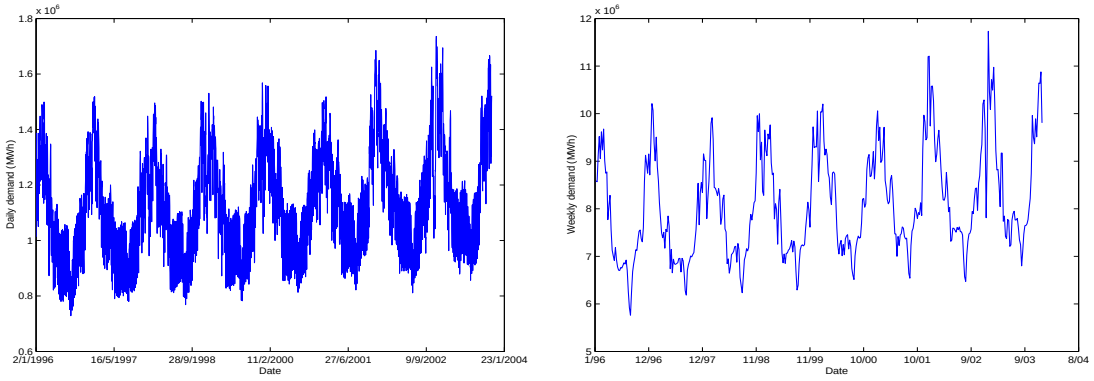


FIG. 4.4 – Courbes des consommations électriques (en MWh) journalières (à gauche) et hebdomadaires (à droite) en France de 1996 à 2003 inclus.

électriques de l'année $x+1$ (pour $3 \leq x \leq 7$) en utilisant les x premières années d'historiques de la série étudiée.

Notons qu'en optimisation robuste, lorsqu'on détermine des ensembles d'incertitude, on ne cherche pas en général des estimations du support des variables ou vecteurs aléatoires sous-jacents. On cherche plutôt à déterminer des intervalles de prédiction. D'une manière générale, si les paramètres incertains du problème étudié sont associés à un vecteur aléatoire X de taille n, alors on cherchera idéalement un ensemble E tel que $P(X \in E) \geq 1-\varepsilon$, ou, à défaut, des ensembles $E_i, 1 \leq i \leq n$, tels que $P(X_i \in E_i) \geq 1-\varepsilon$, ε étant un niveau de confiance donné. Remarquons que la méthodologie robuste a été jusqu'à présent appliquée dans trois contextes (relativement aux ensembles d'incertitude) :

- le cas où les lois des variables aléatoires X_i sous-jacentes sont connues (des ensembles d'incertitude simples tels que des ensembles d'incertitude définis par des contraintes de boîte étant alors souvent choisis) ;
- le cas où l'ensemble d'incertitude est posé à priori ;
- enfin le cas où l'ensemble d'incertitude est déterminé à partir d'historiques des paramètres incertains mais sans effectuer une véritable étude statistique (utilisation de l'enveloppe convexe des scénarios en général).

En revanche, si on souhaite déterminer les ensembles d'incertitude à partir de données réelles et si l'on choisit de ne pas prendre l'enveloppe convexe des scénarios disponibles, mais des ensembles d'incertitude fournissant des contreparties robustes moins conservatives, une étude statistique s'impose. Cet aspect a souvent été occulté dans la littérature jusqu'à présent. Pourtant, une méthode robuste sera d'autant plus intéressante que les ensembles d'incertitude sont calibrés finement. Nous disposons des données de consommation journalières ; cependant nous étudierons la série correspondant aux consommations hebdomadaires. D'une part car cette série est plus facile à étudier (on obtient de meilleures prévisions) mais aussi et surtout car la gestion de production électrique à l'horizon annuel ne nécessite pas un maillage aussi fin que le maillage journalier. Nous reviendrons sur ce problème dans la section 4.7. La figure 4.4 représente ces consommations électriques journalières et hebdomadaires. Notre objectif sera plus précisément d'utiliser les x premières années de ces historiques ($3 \leq x \leq 7$), afin de déterminer un ensemble d'incertitude défini par des contraintes de boîte pour la consommation électrique des $8-x$ dernières années.

Autrement dit, on fournira des bornes inférieures $d_t^{\min}$ et supérieures $d_t^{\max}$ pour la consommation électrique en France pour chacune des semaines des $8-x$ années à venir. Les ensembles d'incertitude seront des intervalles de prédiction au niveau 95% pour chacune des demandes d_t, i.e satisfaisant $P(\tilde{d}_t \in [d_t^{\min}, d_t^{\max}]) \geq 0.95$. Différentes méthodes statistiques seront envisagées et comparées. En particulier, on comparera, pour chaque méthode, les prédictions faites par les différents modèles avec les demandes réelles.
On observe d'une part une tendance à la croissance de la consommation au cours du temps (et ce phénomène aurait été sûrement plus marqué si une période plus longue était étudiée) et d'autre part un aspect cyclique annuel (forte consommation en hiver avec une pointe en général en janvier et faible consommation en été notamment). On constate qu'il existe aussi un cycle hebdomadaire (consommation relativement stable les 5 jours ouvrables et le week-end où la consommation diminue). Notons qu'il existe enfin aussi un cycle journalier. Les différents modèles que nous proposons pour expliquer cette série chronologique intègrent ces deux aspects de tendance à la croissance et de saisonnalité annuelle. Le premier de ces modèles étudie l'effet sur la consommation de facteurs contrôlés qualitatifs et quantitatifs. Un modèle SARIMA est ensuite testé. Enfin les résidus d'un modèle saisonnier simple avec tendance linéaire sont analysés.

4.6.1 Méthode de régression

La consommation électrique est influencée essentiellement par les facteurs suivants :

- La consommation d'électricité destinée à l'éclairage est liée à la nébulosité qui mesure le taux de couverture nuageuse. Cette grandeur se mesure en octa et varie de 0 à 8. 0 correspond à un ciel complètement dégagé et 8 à un ciel couvert.
- La consommation d'électricité destinée au chauffage en hiver et à la climatisation en été dépend de la température extérieure.

Une partie de la consommation d'électricité est aussi liée à l'activité économique, ce qui explique les baisses de consommation le week-end, les jours fériés et au moment des vacances d'été par exemple. La tendance à la croissance de la consommation électrique peut aussi en partie être expliquée par le développement de l'activité économique. Pour expliquer la demande d_t la semaine t, nous modélisons la tendance à la croissance par une droite d'équation $\alpha' t + \beta'$ et l'effet saisonnier par un processus S_t rigoureusement périodique de période $p = 52$. On note $\alpha > 0$ le gradient de température lié au chauffage et on suppose que le chauffage électrique est utilisé en dessous d'une température seuil de $T_0{}^oC$. De même, on note $\beta > 0$ le gradient de température lié à la climatisation et on suppose que la climatisation est utilisée en dessus d'une température seuil de $T_1{}^oC$. On désignera par $a_k,\ 0 \leq k \leq s$, la valeur moyenne de la consommation électrique lorsque la nébulosité prend la valeur k. Enfin, on note $\bar{N}_t$ la valeur moyenne de la nébulosité pour la période t et $\bar{T}_t$ la température moyenne sur la période t. Si ε_t capture le bruit résiduel correspondant à la partie de la consommation électrique non expliquée par les facteurs qui viennent d'être présentés, le modèle que l'on vient de construire pour la consommation électrique d_t la semaine t s'écrit sous la forme :

$$\tilde{d}_t = \sum_{k=0}^{s} a_k\, 1(\bar{N}_t = k) + \alpha\,(T_0 - \bar{T}_t) 1(\bar{T}_t \leq T_0) + \beta\,(\bar{T}_t - T_1) 1(\bar{T}_t \geq T_1) + \alpha' t + \beta' + S_t + \varepsilon_t,\ \ S_t = S_{t+p}, \quad (4.44)$$

pour tout $t = 1, \ldots, T$, T étant le nombre d'observations successives des triplets $(d_t, \bar{N}_t, \bar{T}_t)$ disponibles. Dans ce qui suit on choisira $T_0 = 15^oC$ et $T_1 = 30^oC$. Un tel modèle sera d'autant plus pertinent que les consommations électriques, les températures et les nébulosités correspondent à une zone géographique de petite taille où la température et la nébulosité influencent directement la consommation électrique locale. Les consommations électriques

que nous étudions étant relatives à la France, les températures moyennes qui devraient intervenir dans le modèle (4.44) sont les températures moyennes en France pour les différentes semaines considérées. Néanmoins, il est difficile d'obtenir une telle température moyenne, dont l'utilisation (si tant est que l'on puisse déterminer une valeur pour cette température moyenne) serait discutable. Pour la nébulosité, une valeur nationale a encore moins de sens. Cependant, pour montrer que le modèle fait sens, nous le testons de la façon suivante. Les températures et nébulosités dont nous disposons sont les températures et nébulosités journalières relevées de 1996 à 2002 inclus à la station Météo France de Montsouris à Paris. Ces températures et nébulosités permettent donc d'expliquer la consommation d'électricité de la région parisienne. Or d'une part, la région parisienne est celle qui contribue le plus à la consommation nationale, et d'autre part on peut supposer que la consommation d'électricité de la région parisienne s'obtient par homothétie à partir de la consommation nationale (pour cette consommation, il y a aussi une tendance à la croissance et la même saisonnalité [jours fériés en même temps,...]). Sous, cette hypothèse, les coefficients obtenus, en calant le modèle (4.44) en utilisant les données dont nous disposons, fournissent, à un facteur d'homothétie près, les coefficients que l'on obtiendrait si l'on cherchait à expliquer la consommation électrique de la région parisienne. À l'inverse, sous cette même hypothèse, l'utilisation du modèle (4.44) permet bien d'expliquer la consommation d'électricité en France et l'on verra qu'il permet d'obtenir des prévisions de bonnes qualités. Les estimations $((\hat{a}_k)_{0\leq k\leq 8}, \hat{\alpha}, \hat{\beta}, \hat{\alpha}', \hat{\beta}', (\hat{S}_t)_{1\leq t\leq p})$ des paramètres $((a_k)_{0\leq k\leq 8}, \alpha, \beta, \alpha', \beta', (S_t)_{1\leq t\leq p})$ du modèle (4.44) sont obtenues par une méthode de moindres carrés et en utilisant l'algorithme BFGS de quasi-newton. Un test d'adéquation du χ^2à la loi normale (obtenu en prenant 12 classes) permet d'accepter l'hypothèse de normalité des résidus. En effet, la statistique de test du χ^2 vaut 0.238 et la p-valeur du test est donc voisine de 1. L'estimation de la moyenne du bruit étant 0, la prévision $\bar{d}_t$ faite par un tel modèle pour la demande en électricité à la date t est :

$$\bar{d}_t = \sum_{k=0}^{s} \hat{a}_k \, 1(\bar{N}_t = k) + \hat{\alpha} \, (T_0 - \bar{T}_t) 1(\bar{T}_t \leq T_0) + \hat{\beta} \, (\bar{T}_t - T_1) 1(\bar{T}_t \geq T_1) + \hat{\alpha}' t + \hat{\beta}' + \hat{S}_t.$$

Connaissant les températures et nébulosités moyennes $\bar{T}_t$ et $\bar{N}_t$, un intervalle de prédiction de niveau $1-\alpha\%$ pour la demande d_t est alors estimé par :

$$\bar{d}_t \pm \sqrt{\frac{T+1}{T-1}} \, \hat{\sigma} \, t_{1-\alpha/2}^{T-1};$$

où $t_{1-\alpha/2}^{T-1}$ est le quantile d'ordre 1-$\alpha/2$ de la loi de Student à $T-1$ degrés de liberté et $\hat{\sigma}$ l'estimation empirique de l'écart type du bruit. La construction de tels intervalles nécessite donc de connaître les températures et nébulosités moyennes sur la période de prévision. Afin de montrer l'efficacité du modèle dans un tel contexte, nous avons mesuré (voir la section 4.6.4) la qualité des prédictions et des constructions d'intervalles de prédiction pour la consommation électrique en supposant connues températures et nébulosités sur la période de prévision. Cette méthode conduirait certainement à des résultats encore meilleurs si les consommations, tout comme les températures et les nébulosités, correspondaient à une zone géographique de petite taille. D'un point de vue pratique, pour appliquer cette méthode afin de prédire la consommation électrique sur un territoire de grande taille, il convient de découper ce territoire en petites zones et d'appliquer le modèle (4.44) dans chaque zone. Afin de calibrer le modèle (4.44) dans une zone donnée, des historiques de températures,

nébulosités et consommations électriques doivent être disponibles dans cette zone. D'autre part, d'un point de vue pratique, Météo France peut fournir de bonnes prévisions sur la température et la nébulosité pour un horizon de quelques jours dans une zone géographique de petite taille. Ce modèle peut donc être utilisé pour obtenir des prédictions fines de la demande sur un horizon de quelques jours. Cependant, dans le cas où des prévisions de la température et de la nébulosité ne sont pas disponibles, il convient de construire un modèle permettant d'expliquer la consommation électrique en utilisant uniquement les historiques de cette consommation. Nous étudions dans ce qui suit deux modèles allant dans ce sens.

4.6.2 Modèle SARIMA

Dans [48], un modèle SARIMA est utilisé pour expliquer une série chronologique de consommations électriques relevées toutes les $1/2$ heure sur une semaine. Nous allons aussi chercher, dans cette section, à caler un modèle SARIMA$(p,d,q) \times (P,D,Q)_S$ (voir par exemple [20] pour une présentation des modèles SARIMA) sur nos données de consommations hebdomadaires. Afin d'identifier un modèle SARIMA adéquat pour la série d_t des consommations électriques hebdomadaires étudiée, les ordres de différentiation d et D, les ordres p, P, q et Q du modèle et la durée S de la composante saisonnière doivent être identifiés. La saisonnalité S est ici connue, elle vaut $S = 52$ semaines. Soit B l'opérateur de retard défini par $B^\ell X_t = X_{t-\ell}$ pour $t, \ell \in \mathbb{Z}$. Par exemple, pour expliquer la série des consommations électriques sur les 7 premières années, en suivant la méthodologie présentée dans [20], le modèle finalement retenu, parmi l'ensemble des modèles de la classe des modèles SARIMA est un modèle s'écrivant sous la forme :

$$(I - B)\,(I - B^{52})\,d_t = \theta_q(B)\,(I - \Theta B^{52})\,a_t,$$

où $\{a_t\}_{t\in\mathbb{Z}}$ est un bruit blanc gaussien de variance σ^2, $\theta_q(z) = 1 - \sum_{i=1}^{4} \theta_i z^i$ et $\Theta \in \mathbb{R}$. Les coefficients θ_i et Θ du modèle et la variance du bruit blanc σ^2 sont estimés par maximisation de la vraissemblance conditionnelle. On peut alors ([47], p.202) construire des intervalles de prédiction au niveau $1 - \alpha = 95\%$, pour chacune des consommations de la dernière année.

4.6.3 Modèle saisonnier simple

Dans la veine de la méthode proposée en section 4.6.1, nous modélisons la tendance à la croissance par une droite et l'effet saisonnier par un processus périodique de période $p = 52$ connue. Néanmoins, nous ne supposons pas connaître les températures et les nébulosités sur la période de prédiction. La consommation électrique la semaine t s'exprime alors par :

$$\tilde{d}_t = \alpha\, t + \beta + S_t + \varepsilon_t, \quad S_t = S_{t+p}, \quad t = 1, \ldots, T, \tag{4.45}$$

pour les T observations dont on dispose. Supposons que la série est observée pendant n années de p semaines. Notons alors d_{ij} la consommation électrique l'année i la semaine j, et introduisons les quantités suivantes :

$$\bar{d}_{..} = \frac{1}{np}\sum_{i,j} d_{ij}, \quad \bar{d}_{i.} = \frac{1}{p}\sum_{j=1}^{p} d_{ij}, i = 1, \ldots, n, \quad \bar{d}_{.j} = \frac{1}{n}\sum_{i=1}^{n} d_{ij}, j = 1, \ldots, p.$$

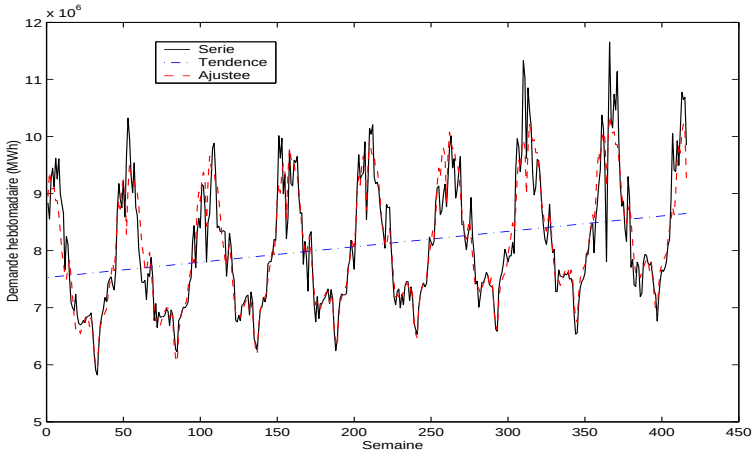


FIG. 4.5 – Série des consommations hebdomadaires, tendance et série ajustée par le modèle (4.45).

$\bar{d}_{..}$ est donc la consommation moyenne, $\bar{d}_{i.}$ la consommation moyenne l'année i et $\bar{d}_{.j}$ la consommation moyenne la semaine j. Les estimations des coefficients de ce modèle s'obtiennent par la méthode des moindres carrés. On obtient les valeurs explicites suivantes :

$$\hat{\alpha} = \frac{12}{np(n^2-1)} \left(\sum_{i=1}^{n} i\, \bar{d}_{i.} - \frac{n(n+1)}{2}\, \bar{d}_{..} \right), \quad \hat{\beta} = \bar{d}_{..} - \frac{np+1}{2}\, \hat{\alpha},$$

et

$$\hat{S}_j = \bar{d}_{.j} - \bar{d}_{..} - (j - \frac{p+1}{2})\, \hat{\alpha}, \quad j = 1, \ldots, p.$$

Nous représentons sur la figure 4.5, la série initiale des consommations hebdomadaires, la droite associée à la tendance, et la série ajustée par le modèle (4.45) (la série des $(\hat{\alpha}\, t + \hat{\beta} + \hat{S}_t)_t$). Pour illustrer la qualité de la prévision obtenue avec le modèle (4.45) (lorsque d_t est estimée par $\hat{\alpha}\, t + \hat{\beta} + \hat{S}_t$), nous comparons, sur la figure 4.6 ci-après, les prévisions des consommations hebdomadaires de la 6-ème année obtenues avec ce modèle (en utilisant donc 5 ans d'historique de données pour calibrer le modèle) et les consommations hebdomadaires réelles. Les taux d'erreur (de prédiction) relatifs en utilisant différentes longueurs d'historiques sont donnés en section 4.6.4. Afin de raffiner ce modèle, il conviendrait d'étudier les résidus. Il semblerait que le comportement du bruit présente des périodes homogènes liées aux saisons. On s'attend à ce que les fluctuations (et donc la moyenne du bruit) soient plus importantes en hiver qu'en été. On a donc cherché à modéliser le comportement des résidus par une chaîne de Markov cachée. Au vu de la distribution de probabilité empirique des résidus, une loi de mélange gaussienne semble plausible pour ces résidus. Par conséquent, nous avons choisi des lois d'émission gaussiennes. Afin, dans la mesure du possible, d'interpréter les états cachés, leur nombre K ne doit pas être choisi trop grand. Nous choisissons, parmi un nombre d'états cachés candidats $K = 3, \ldots, 7$, ceux qui maximisent le BIC. Rappelons que le BIC est défini par $BIC = \ln L - \frac{d}{2} \ln T$, où L est la vraisemblance du modèle choisi et d le nombre de paramètres indépendants du modèle. Nous prenons ici $d = K^2 + K$ et le nombre d'états cachés maximisant le BIC est 3. Les 3 états cachés se distinguent par la valeur de la moyenne et de la variance des lois conditionnelles gaussiennes.

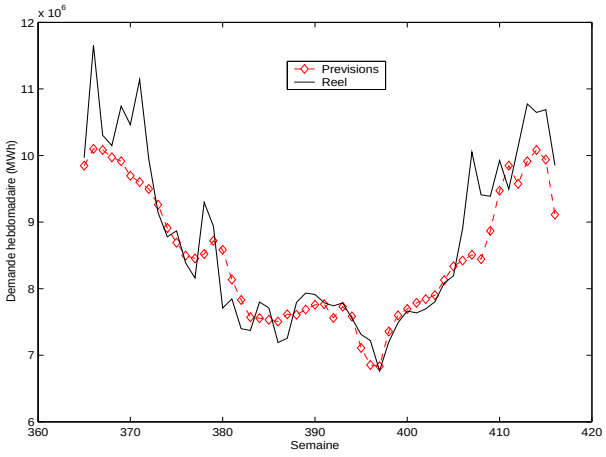


FIG. 4.6 – Consommations hebdomadaires de la 6-ème année et leurs prévisions obtenue en utilisant le modèle (4.45) en fin de 5-ème année.

Les moyennes estimées dans chacun des 3 états sont $(\mu_1, \mu_2, \mu_3) = (-2, 5.5, -0.09) \times 10^5$ et les variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (13.04, 15.5, 1.8) \times 10^{10}$. Les états cachés les plus probables sont restaurés par l'algorithme de Viterbi. Cet algorithme détermine les états cachés maximisant la probabilité conditionnelle de la chaîne cachée entière, sachant les observations. Sur la figure 4.7 ci-après, on peut observer (cadre de gauche) les résidus et les états cachés. Pour chaque observation on repère la moyenne de l'état caché, la moyenne plus l'écart type et la moyenne moins l'écart type. Le processus caché est une chaîne de Markov irréductible, stationnaire (de distribution stationnaire π) et homogène. Les variables cachées discrètes sont associées à un processus C et on a pour tout $1 \leq k \leq K$, $\pi_k = P(C_1 = k)$. Par ailleurs, le nombre d'états cachés s'interprète comme le nombre de composantes de la loi marginale de ε_t. On dispose de l'estimation $\hat{f}_{\varepsilon_t}(y) = \sum_{j=1}^{K} \hat{\pi}_j \; f_{\hat{\theta}_j}(y)$ de cette densité où $f_{\hat{\theta}_j}(y)$ est l'estimation de la loi d'émission gaussienne dans l'état j et $\hat{\pi}_j$ l'estimation de π_j. Sur la figure 4.7, cette densité est superposée à l'histogramme des résidus observés (ε_t) (cadre de droite).
On constate que la moyenne et la variance sont bien plus faibles dans l'état 3. La moyenne est plus de 20 fois inférieure (en valeur absolue) à ce qu'elle est dans les deux autres états et la variance plus de 7 fois inférieure. Cet état correspond aux périodes où la consommation est moins importante telles que l'été. En revanche dans les deux autres états, les variances sont fortes et prennent des valeurs voisines. Ces deux états se distinguent plutôt par leur moyenne. Dans l'état 1, cette moyenne est fortement négative et dans l'état 2, elle est fortement positive et plus de deux fois plus élevée en valeur absolue. On constate que ces deux états correspondent plutôt à des périodes où la consommation est assez forte et notamment l'hiver. L'état 2 est beaucoup moins fréquent ; il pourrait correspondre à des pics de consommation exceptionnels (la valeur de la moyenne de la loi d'émission dans cet état est fortement positive). L'erreur relative globale entre la série ajustée et la série initiale diminue logiquement en prenant en compte la modélisation venant d'être spécifiée pour les résidus ; cette erreur passant de 3% à 2.3%.

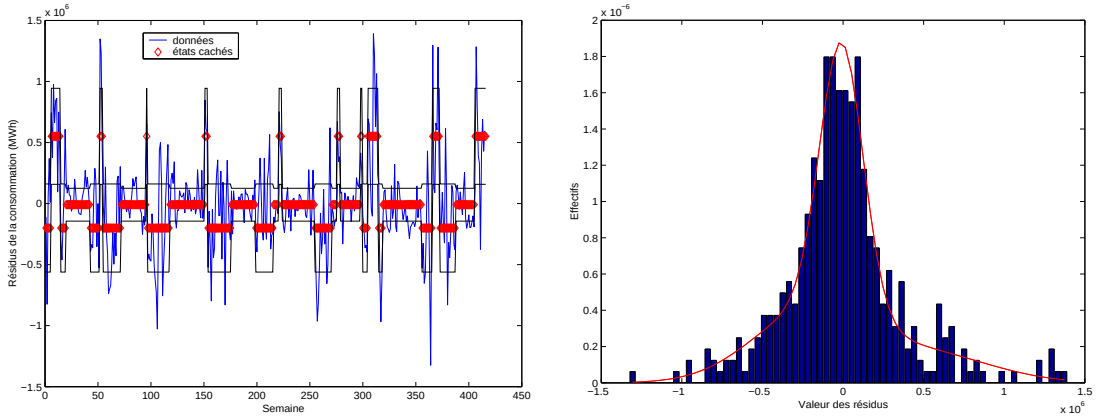


FIG. 4.7 – Etude des résidus du modèle (4.45) pour la série des consommations hebdomadaires. Cadre de gauche : restauration des états cachés. Cadre de droite : densité marginale théorique et empirique des résidus.

4.6.4 Comparaison des méthodes

Nous commençons par comparer, pour les différentes méthodes qui viennent d'être présentées, la moyenne des erreurs relatives (en %) entre les consommations prédites et les consommations réelles. Si $\hat{d}_t$ est la prédiction faite sur la demande au pas de temps t, on définit l'erreur relative de prédiction (en %) sur les semaines de l'ensemble I par :

$$\frac{100}{|I|} \sum_{t \in I} \frac{|\hat{d}_t - d_t|}{d_t}.$$

Pour diverses valeurs du nombre x d'années d'historiques de consommations électriques utilisé, les prédictions de la consommation électrique sont effectuées pour les $8 - x$ années restantes. Nous comparons aussi ces méthodes à une méthode dite 'simple' par la suite, qui consiste à estimer les consommations pour une semaine donnée par la moyenne des consommations électriques des années passées relatives à la même semaine (repérée par sa position dans l'année). Dans le tableau ci-dessous 'Régression 1' désigne le modèle présenté en section 4.6.3 et 'Régression 2' le modèle présenté en section 4.6.1 utilisant les données météorologiques.

Nombre d'années x utilisées pour calibrer le modèle	Simple	'Régression 1'	SARIMA	'Régression 2'
3	6.9	3.0	4.5	2.5
4	7.0	3.2	4.0	2.8
5	7.1	3.6	5.4	2.4
6	6.9	3.8	5.6	2.7
7	7.9	4.2	5.5	-

TAB. 4.1 – Erreurs relatives de prédictions (en %) des consommations hebdomadaires sur les $8 - x$ années à venir, en utilisant différentes valeurs du nombre x d'années utilisé pour calibrer le modèle.

Notons que pour la méthode 'Régression 2', pour certaines semaines, l'erreur de prédiction est quasi nulle (elle est descendue jusqu'à 0.02% sur l'ensemble des semaines pour lesquelles des prévisions ont été calculées). On observe en particulier les performances bien supérieures des méthodes 'Régression 1', 'SARIMA' et 'Régression 2' sur la méthode dite 'simple', ce qui justifie nos efforts de modélisation.

Nous comparons aussi la précision des intervalles de prédiction au niveau 95% fournis par les méthodes SARIMA et 'Régression 2'. Les pourcentages des consommations réelles en dehors des intervalles de prédiction (sur l'ensemble des intervalles de prédiction déterminés lorsque le nombre d'années d'historiques utilisé varie de $x = 3$ à $x = 7$) sont les suivants : 4.6% pour la méthode SARIMA et 3.5% pour la méthode 'Régression 2'.

4.6.5 D'autres méthodes de calibration

On peut aussi proposer d'utiliser des méthodes non paramétriques afin de déterminer les ensembles d'incertitude. Etant donné un échantillon de taille T, $(X_1, \ldots, X_T)$ d'un vecteur aléatoire X, on peut choisir comme ensemble d'incertitude pour le vecteur X, l'ellipsoïde de plus petit volume (ou la sphère de plus petit volume) qui contient tous les points $(X_1, \ldots, X_T)$, ce qui est un problème SDP. On peut aussi rechercher l'ellipsoïde ou la sphère la plus proche (en un sens à préciser) du nuage de points $(X_1, \ldots, X_T)$. Dans ce but, on peut mesurer l'éloignement entre l'ellipsoïde (ou la sphère) et le nuage de points par le carré de la différence entre la distance des points au centre et la distance entre le centre et la projection des points sur l'ellipsoïde (ou la sphère) [voir [24]]. De nombreuses études ont été menées sur la recherche d'ellipsoïdes approchant un nuage de points. Dans [73], par exemple, on présente une méthode efficace pour la recherche d'ellipsoïdes inscrits et circonscrits à un ensemble convexe C et 'proches' l'un de l'autre. Typiquement, dans l'optique de la calibration d'un ensemble d'incertitude pour des paramètres incertains, on prendra alors comme ensemble C l'enveloppe convexe des scénarios disponibles (voir aussi [24] pour des résultats récents à ce sujet).

4.6.6 Conclusion sur la calibration des ensembles d'incertitude

Même si des méthodes non paramétriques comme celles qui viennent d'être décrites peuvent être envisagées pour calibrer les ensembles d'incertitude (utilisation de l'enveloppe convexe de scénarios, plus petit ellipsoïde contenant les données), seule une étude fine des variables aléatoires mises en jeu permet d'obtenir des ensembles d'incertitude de bonne qualité. Les méthodes mises en jeu dans une telle étude dépendront donc des situations pratiques étudiées.
Dans le cas de la prévision des consommations électriques qui nous intéressait ici, diverses méthodes prenant en compte certaines spécificités de la série étudiée (sa tendance, sa saisonnalité, l'influence de facteurs extérieurs) ont été mises en compétition. Au final, les prévisions des consommations électriques devront être un compromis entre les prévisions données par le modèle et l'utilisation d'autres informations non prises en compte par le modèle. Ces informations peuvent être de natures diverses et seront d'autant plus fournies que les prédictions sont à faire pour un horizon de courte durée.

4.7 Comparaison des méthodes de gestion

Dans cette section, nous comparons les différentes politiques de gestion qui viennent d'être présentées dans ce chapitre. L'horizon de gestion est l'année. Deux jeux de données sont utilisés.
D'une part un ensemble de 456 scénarios simulés fournis par EDF, qui peuvent être vus comme des historiques donnant l'évolution des paramètres incertains au cours de différentes années. Ces scénarios sont utilisés pour construire trois arbres de scénarios de taille raisonnable (et correspondant à des constructions de l'arbre plus ou moins 'optimistes') utilisés pour résoudre le problème de gestion de production électrique par programmation stochastique multi-étapes. On pourra se référer à [26] pour plus de détails sur la construction de ces arbres. Nous appellerons ces arbres, arbre facile, arbre médian et arbre dur avec d'évidentes interprétations des prédictions des paramètres incertains sur ces arbres. Les arbres de scénarios sont des arbres de profondeur 364 et comportant 5227 noeuds. Chaque noeud est donc associé à une journée.
Nous comparons ensuite de façon détaillée les résultats obtenus en appliquant cette méthode et les régularisations VaR_{CA} et VaR_{benef} de cette méthode que nous avons présentées dans l'article [55]. En particulier, on s'intéressera au coût moyen et à l'écart type des coûts obtenus sur les 456 scénarios utilisés pour construire l'arbre. Plus généralement, les distributions empiriques des coûts de gestion simulés seront fournies pour chaque méthode. On comparera aussi l'évolution de la plus grosse des réserves hydrauliques en utilisant ces différentes méthodes.
On présentera ensuite la moyenne et l'écart type des coûts de ces 456 scénarios en utilisant :

- la méthode de régularisation de la fonction duale présentée en section 4.4.1 ;
- la contrepartie robuste ajustée (définie en section 4.4.2) en choisissant pour ensemble d'incertitude l'enveloppe convexe des 456 scénarios disponibles ;
- les contreparties robustes ajustées affinement (définies en section 4.4.3) en choisissant successivement comme ensemble d'incertitude l'enveloppe convexe des scénarios disponibles et un ensemble défini par des contraintes de boîte ;
- la contrepartie robuste du problème étudié en choisissant comme ensemble d'incertitude l'enveloppe convexe des scénarios disponibles.

Pour les CRAA, nous nous intéresserons aussi à l'influence du choix de l'intervalle I_t et à l'influence de la complexité des dépendances des variables ajustables en fonction des paramètres incertains sur la qualité des solutions fournies.

D'autre part, nous comparons les méthodes de gestion sur des données réelles en utilisant la série chronologique des consommations électriques en France de 1996 à 2003 inclus.

Enfin, signalons que nous utiliserons, pour chaque jeu de données, les unités de production suivantes :

- Onze centrales thermiques. Chaque unité de production thermique ℓ est décrite par son coût de production unitaire, ses puissances maximale et minimale, le nombre de groupes thermiques et la probabilité α_ℓ qu'un groupe thermique fonctionne.
- Deux usines hydroélectriques indépendantes. Chaque centrale hydroélectrique est connectée à un réservoir différent. On connaît le stock maximal (en GWh) et le stock initial de chaque réservoir et la puissance maximale (en MW) de chaque centrale hydraulique. Le stock maximal du plus grand des réservoirs est environ 30 fois su-

périeur à celui de l'autre réservoir. Ceci explique pourquoi nous serons en particulier intéressés par l'évolution du stock du plus gros réservoir sur l'année.
- Un contrat EJP de 22 jours.

4.7.1 Simulations de scénarios

4.7.1.1 Optimisation sur un arbre de scénarios

Dans cette section, nous testons 3 méthodes d'optimisation sur un arbre de scénarios : la première (que nous appellerons 'Nominale') est définie en section 4.3.2 ; les deux autres sont les méthodes VaR_{CA} et VaR_{benef}.

Caractéristiques centrales et de dispersion des coûts

Afin d'étudier des caractéristiques centrales et de dispersion de la distribution empirique des coûts des 456 scénarios, nous fournissons la moyenne et l'écart type empiriques de ces coûts simulés. Nous donnons aussi les quantiles empiriques d'ordre 0.95 (noté VaR 5%) et d'ordre 0.99 (noté VaR 1%) de la distribution de ces coûts. Nous donnons les résultats en utilisant les trois différents arbres de scénarios générés (facile, médian et dur). Nous mentionnons aussi les résultats de la méthode (que l'on appellera mixte) consistant à cumuler les deux modifications proposées par les deux approches VaR_{CA} et VaR_{benef}.

Sortie	Nominale	VaR_{CA}	VaR_{benef}	Mixte
Moyenne	467 529 291	488 271 561	459 515 733	459 991 109
Écart type	47 864 238	46 609 492	31 750 330	30 597 481
VaR 1%	671 599 214	672 311 733	557 798 070	558 719 609
VaR 5%	543 786 128	574 245 712	517 836 912	518 856 055

TAB. 4.2 – Caractéristiques centrales et de dispersion de la distribution empirique des coûts simulés sur l'arbre facile (horizon de gestion d'une année)

Sortie	Nominale	VaR_{CA}	VaR_{benef}	Mixte
Moyenne	466 498 435	486 587 235	462 184 762	462 334 138
Écart type	46 540 234	47 075 591	29 154 061	28 857 445
VaR 1%	689 406 053	679 536 989	557 838 702	554 297 141
VaR 5%	548 912 439	573 426 666	516 533 109	515 210 039

TAB. 4.3 – Caractéristiques centrales et de dispersion de la distribution empirique des coûts simulés sur l'arbre médian (horizon de gestion d'une année)

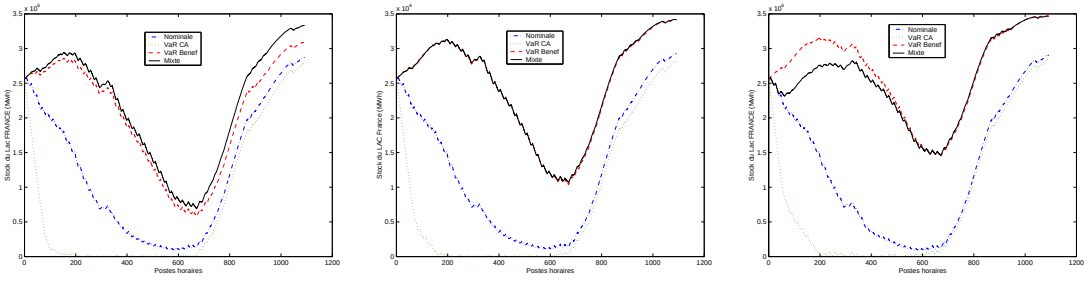
Sortie	Nominale	VaR_{CA}	VaR_{benef}	Mixte
Moyenne	464 712 495	479 773 359	465 886 909	465 097 476
Écart type	44 615 708	47 172 746	28 146 821	28 272 953
VaR 1%	667 480 170	693 735 355	556 573 568	556 464 708
VaR 5%	543 728 341	562 040 194	517 110 690	518 142 699

TAB. 4.4 – Caractéristiques centrales et de dispersion de la distribution empirique des coûts simulés sur l'arbre dur (horizon de gestion d'une année)

Nous remarquons essentiellement les points suivants.

- Les coûts de gestion moyens sur l'ensemble des scénarios et pour chacun des trois arbres sont assez proches pour chacune des 4 méthodes considérées. Pour la méthode VaR_{CA}, les coûts moyens sont entre 3.2% et 4.4% supérieurs à ceux de la méthode nominale. Pour les méthodes VaR_{benef} et Mixte, les coûts peuvent être supérieurs ou inférieurs à ceux du modèle nominal. Néanmoins, ces coûts varient seulement entre 0.08% et 1.7% par rapport aux coûts du modèle nominal.
- La méthode VaR_{CA} ne va pas vraiment dans le bon sens puisque l'écart type des coûts ainsi que les VaR d'ordre 1% et 5% sont plus grandes que les mêmes quantités calculées pour le modèle nominal. En ce qui concerne les méthodes VaR_{benef} et Mixte, elles conduisent dans tous les cas à une réduction de l'écart type des coûts (jusqu'à 38% de réduction sur l'arbre difficile) et des VaR à 1% et 5%.

Étude des trajectoires du plus gros réservoir.



Arbre facile Arbre médian Arbre difficile

FIG. 4.8 – Évolution du stock moyen (en MWh et sur l'ensemble des scénarios) du plus gros réservoir au cours de l'année pour toutes les méthodes et en utilisant les arbres faciles, médians et durs comme support du problème d'optimisation. Pour chaque jour, on repère le niveau du réservoir à 3 instants de la journée.

Pour une méthode donnée, la stratégie de gestion du plus gros réservoir ne varie que légèrement lorsqu'on change d'arbre de scénarios. Les méthodes nominales et VaR_{CA} tendent à vider plus le réservoir dont le niveau remonte en fin d'année. Les méthodes VaR_{benef} et Mixte n'utilisent pas ou que très peu le réservoir au début de l'année. Globalement, le réservoir a un niveau plus élevé avec ces méthodes et est presque plein (son stock maximal est 3500 GWh) en fin d'année. Dans ce qui suit, nous dirons que le plus gros réservoir at-

teint un niveau bas s'il contient au plus 5% de son stock maximal. Nous dirons par ailleurs qu'un réservoir atteint un niveau élevé si son volume est supérieur ou égal à son volume au début de l'année moins 5% de son stock maximal. Le tableau ci-après permet de préciser des tendances déjà observées sur les courbes ci-dessus.

# semaines	Réservoir au niveau haut				Réservoir au niveau bas			
	Nominal	VaR_{CA}	VaR_{benef}	Mixte	Nominal	VaR_{CA}	VaR_{benef}	Mixte
1	437	405	456	456	426	456	5	9
2	421	387	456	456	423	456	4	4
3	408	377	456	456	423	456	3	3
4	390	344	456	456	417	456	3	2
5	375	312	456	456	412	456	3	2
10	201	64	456	449	365	456	0	0
15	69	0	449	410	256	456	0	0
20	29	0	438	373	92	453	0	0
25	10	0	414	329	14	257	0	0
30	5	0	371	268	0	25	0	0

TAB. 4.5 – Nombre de scénarios parmi 456 pour lesquels le gros réservoir est au moins X semaines au niveau haut ou au niveau bas (l'optimiseur ayant été lancé sur l'arbre difficile et avec toutes les méthodes)

Comparaison de la distribution des coûts.

Les densités empiriques des coûts de gestion (représentées sur la figure 4.9 ci-après) pour les méthodes nominales et VaR_{CA} ont des queues de distribution plus grosse que celles des méthodes VaR_{benef} et Mixte. Quelques scénarios sont de coût très élevé pour les méthodes nominale et VaR_{CA}. Par contre, la dispersion des coûts est inférieure pour les modèles VaR_{benef} et Mixte. On peut illustrer ces propos par quelques chiffres :

- Le scénario de coût le plus élevé correspond à un coût de 7.448×10^8, 8.232×10^8, 5.738×10^8 et 5.753×10^8 pour respectivement les méthodes nominale, VaR_{CA}, VaR_{benef} et Mixte.
- Les scénarios de coût minimum pour chacun des modèles ont des coûts voisins qui s'élèvent à 3.822×10^8, 3.868×10^8, 3.997×10^8 et 3.991×10^8 pour respectivement les méthodes nominale, VaR_{CA}, VaR_{benef} et Mixte.

L'allure de la fonction de répartition empirique des coûts de gestion (représentée sur la figure 4.10 ci-après) confirme cette tendance. La méthode nominale a un plus grand nombre de scénarios dont le coût est inférieur à 4.7×10^8. En outre, un nombre non négligeable de scénarios est de coût élevé pour cette méthode. Pour la méthode VaR_{CA}, c'est encore pire puisque le nombre de scénarios dont le coût est inférieur à 4.7×10^8 est faible.

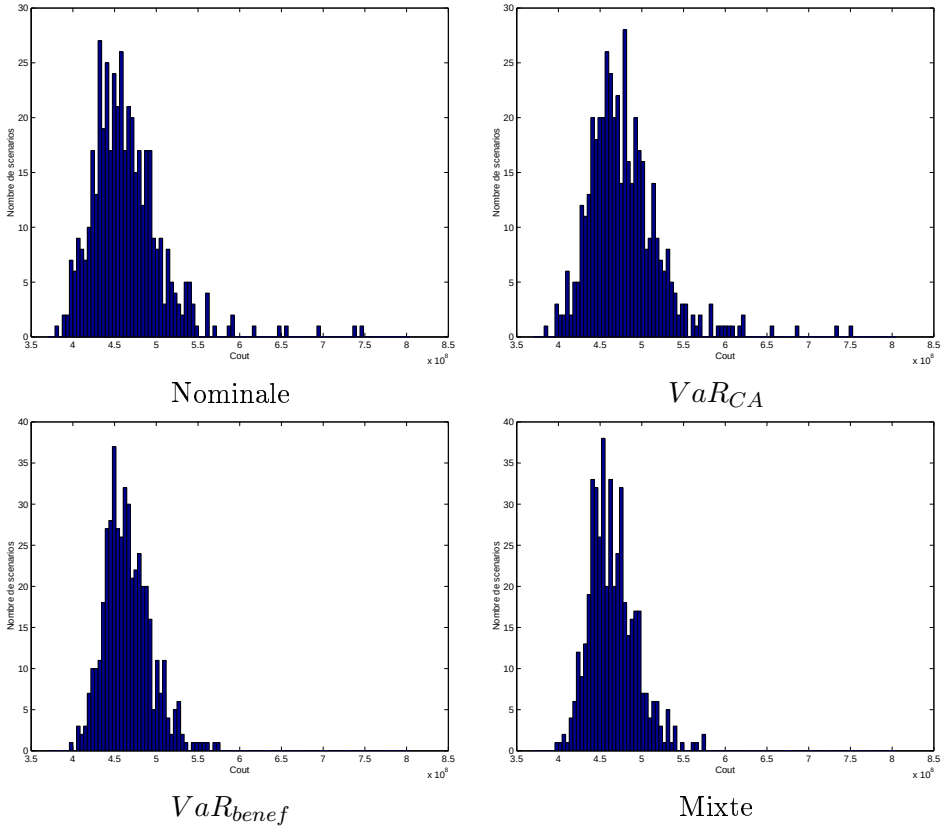


FIG. 4.9 – Densités empiriques des coûts de gestion sur l'arbre dur et pour chaque méthode (horizon de gestion : 1 an).

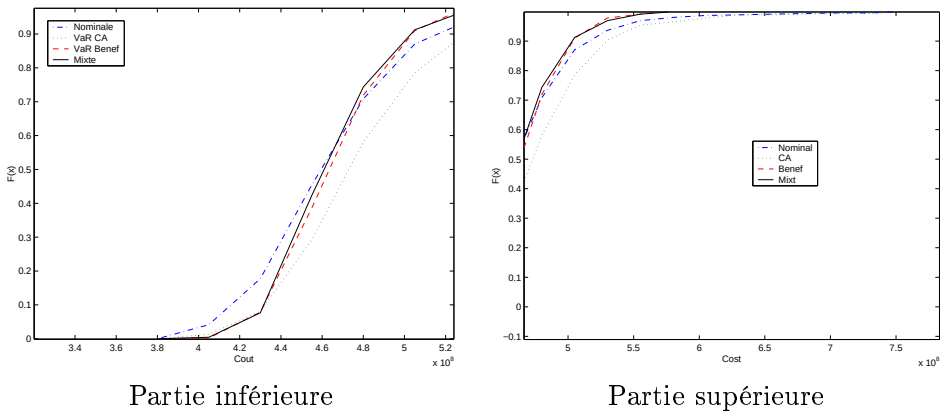


FIG. 4.10 – Fonction de répartition empirique des coûts sur l'arbre dur pour les quatre méthodes.

Dans ce qui suit, nous testons les autres méthodes de gestion introduites dans ce chapitre et les 456 scénarios sont agrégés de manière à fournir 456 scénarios de taille 12 (on

détermine alors des plans de production mensuels) ou 24 (on détermine dans ce cas des plans de production bimensuels). Ceci permettra en particulier de fournir des CRA et des CRAA de taille raisonnable.

4.7.1.2 Régularisation de la fonction duale

Les scénarios sont ici agrégés sous la forme de 456 scénarios de taille 24 pour tester la méthode de régularisation de la fonction duale présentée en section 4.4.1. Nous supposons ici que les scénarios d'apports sont connus. Deux calibrations de la matrice Γ sont essayées. Une première (notée C1 dans le tableau ci-dessous), consiste à estimer Γ par la matrice de covariance empirique de la demande obtenue à partir des 456 scénarios de demande disponibles. Une seconde (notée C2 dans le tableau ci-dessous), consiste à choisir une matrice diagonale dont la diagonale est celle de la matrice de covariance empirique (cette estimation étant inversible permet d'éviter tout problème numérique de résolution). On choisit naturellement pour $\bar{d}$ la moyenne empirique des demandes sur les 456 scénarios. La moyenne et l'écart type des coûts de gestion simulés obtenus en choisissant $\kappa(\varepsilon) = 1$ et $\kappa(\varepsilon) = 2$ sont alors donnés dans le tableau ci-dessous.

Calibration	C1 - $\kappa(\varepsilon) = 1$	C2 - $\kappa(\varepsilon) = 1$	C1 - $\kappa(\varepsilon) = 2$	C2 - $\kappa(\varepsilon) = 2$
Moyenne	4.268×10^8	4.198×10^8	4.237×10^8	4.219×10^8
Écart type	12 239 392	13 207 628	12 081 557	12 603 941

TAB. 4.6 – Moyenne et écart type des coûts de gestion simulés en ayant régularisé la fonction duale comme en section 4.4.1.

La calibration 2 a tendance à fournir un coût moyen un peu inférieur mais un écart type un peu supérieur par rapport à la calibration 1. On observe néanmoins, quelle que soit la calibration de Γ utilisée, et pour les valeurs de κ testées, une réduction significative (par rapport aux méthodes VaR_{CA} et VaR_{benef}) du coût moyen de gestion et une réduction très importante de l'écart type des coûts de gestion. Même si la réduction de variance peut en partie dépendre de la réduction du nombre de pas temps utilisé pour résoudre le problème, cette méthode s'avère très efficace. En effet, les coûts moyens obtenus avec cette méthode sur ces simulations sont entre 1.9% et 3.6% du coût moyen optimal qui est de 4.12×10^8. D'un point de vue pratique, la calibration de la matrice Γ devra soigneusement être étudiée.

4.7.1.3 Contrepartie robuste ajustée

Les scénarios sont toujours agrégés sous la forme de 456 scénarios de taille 24. Nous mettons en place la CRA du problème (4.12) (présentée en section 4.4.2) avec 24 pas de temps et en choisissant pour ensemble d'incertitude l'enveloppe convexe des 456 scénarios ou un sous ensemble de ces scénarios.
Commençons par illustrer, sur un exemple, que la CRA et la CR ne sont pas nécessairement équivalentes. On choisit comme ensemble d'incertitude l'enveloppe convexe des 20 premiers scénarios. La valeur optimale de la CR est alors 4.67×10^8 contre 4.38×10^8 pour la CRA et dans cet exemple, la CRA et la CR ne sont donc pas équivalentes.
Nous montrons à présent, que plus le nombre de scénarios augmente et plus la solution de

la CRA devient intéressante par rapport à la solution de la CR. Dans ce but, nous calculons les solutions de la CR et de la CRA obtenues en choisissant pour ensemble d'incertitude l'enveloppe convexe des x premiers scénarios (x prenant successivement les valeurs 5, 200 et 456). Le nombre de variables de décision de la CRA augmente linéairement avec le nombre de scénarios et s'élève à 74 615 lorsqu'on considère 200 scénarios et à 170 103 lorsqu'on considère 456 scénarios. La CRA fournit, pour chaque scénario ξ_k, une solution (u, v_k) adaptée à ce scénario. Dans le tableau ci-dessous, 'CoutCRA' désigne le coût moyen (sur l'ensemble des x premiers scénarios) obtenu en utilisant ces solutions adaptatives et 'etCRA' désigne l'écart type de ces coûts. Nous indiquons aussi les coûts moyens 'CoutR', 'CoutS', 'CoutOpt' et les écarts types 'etR', 'etS' et 'etOpt', liés respectivement à l'utilisation de la solution de la CR, la solution du problème où chaque paramètre incertain est remplacé par sa moyenne sur l'ensemble des x scénarios et enfin la solution optimale.

x	CoutCRA	etCRA	CoutS	etS	CoutR	etR	CoutOpt	etOpt
5	4.23×10^8	1.11×10^6	4.17×10^8	1.41×10^6	4.54×10^8	2.65×10^6	4.15×10^8	5.53×10^6
200	4.40×10^8	1.17×10^6	4.14×10^8	1.70×10^6	7.02×10^8	3.4×10^6	4.12×10^8	12.1×10^6
456	4.54×10^8	1.01×10^6	4.14×10^8	1.80×10^6	8.14×10^8	3.5×10^6	4.12×10^8	12.7×10^6

TAB. 4.7 – Coût moyen et écart type des coûts sur un ensemble de scénarios pour les paramètres incertains en utilisant la technique de la contrepartie robuste ajustée.

Les coûts obtenus en remplaçant chaque paramètre incertain par une estimation de sa moyenne sur l'ensemble des scénarios sont très proches des coûts optimaux. Néanmoins, cette politique ne peut pas être utilisée en pratique car pour un bon nombre de pas de temps, la contrainte de satisfaction de la demande (qui est une contrainte dure) n'est pas satisfaite. Les coûts moyens de gestion obtenus en utilisant la CRA sont, comme on s'y attendait, inférieurs à ceux obtenus en utilisant la CR. Lorsque le nombre de scénarios augmente, l'utilisation d'une solution robuste conduit naturellement à une solution très éloignée de la solution optimale (pratiquement deux fois plus coûteuse lorsqu'on utilise 456 scénarios).
La solution robuste utilisée est la même quelle que soit le scénario choisi. Si les apports étaient fixés, l'écart type des coûts de gestion obtenus en utilisant la CR seraient nuls. Cependant, les différences entre les apports des réservoirs d'un scénario à l'autre font que, même en utilisant la solution de la CR, les niveaux des réserves en fin de période de gestion ne sont pas les mêmes sur tous les scénarios. Ainsi, il est bien normal d'obtenir un écart type des coûts de gestion strictement positif en utilisant la CR. On remarque même ici que les écarts types des coûts de gestion obtenus en utilisant la CR sont supérieurs à ceux obtenus en utilisant la CRA. La CRA satisfait donc doublement nos objectifs : réduction du coût moyen et réduction de l'écart type des coûts par rapport à l'utilisation de la CR. En ayant choisi comme seules variables non ajustables les commandes du premier pas de temps, on aurait néanmoins pu s'attendre à ce que le coût 'CoutCRA' soit plus proche du coût optimal 'CoutOpt'. Le grand nombre de scénarios et la présence de certains scénarios durs sont à l'origine de cet écart.
Ces résultats sur l'utilisation de la CRA restent malgré tout théoriques. En pratique, afin de déterminer les commandes au pas de temps t, on utilisera les commandes non ajustables de la CRA correspondant au problème de gestion de production électrique associé aux $T-t+1$ derniers pas de temps. En utilisant cette approche et comme ensemble d'incertitude les 5

premiers scénarios, le coût moyen est de 4.30×10^8 et l'écart type des coûts de 1.75×10^6. On se trouve donc, déjà pour 5 scénarios, assez éloigné du coût moyen optimal.

4.7.1.4 Contrepartie robuste ajustée affinement

Nous illustrons à présent les résultats obtenus en utilisant la CRAA du problème de gestion de production électrique. Afin de réduire la taille des CRAA, les 456 scénarios sont ici agrégés sous la forme de 456 scénarios de taille 12 (on détermine donc maintenant des plans de production mensuels).

Ensemble d'incertitude polytopique

Commençons par mettre en place la CRAA (donnée par (4.54) en annexe du chapitre 4) obtenue en choisissant pour ensemble d'incertitude l'enveloppe convexe d'un sous ensemble des 456 scénarios. Tout d'abord, remarquons que la CRAA et la CR ne sont pas forcément équivalentes. Par exemple, la valeur optimale de la CRAA obtenue en choisissant les schémas de dépendance (4.19) (pour les commandes thermiques et EJP) et (4.20) pour les productions des centrales hydrauliques ; et pour ensemble d'incertitude l'enveloppe convexe des 6 premiers scénarios est de 4.254×10^8 contre 4.57×10^8 pour la CR obtenue avec le même ensemble d'incertitude.
La CRAA obtenue en utilisant les 456 scénarios est de trop grande taille pour pouvoir être résolue par les moyens dont on dispose. Nous nous contentons donc dans cette section de choisir un nombre restreint de scénarios (6 et 20) afin de montrer l'influence du choix de l'intervalle I_t et des schémas de dépendance des variables ajustables en fonction des paramètres incertains sur la qualité des solutions de la CRAA.
Dans un premier temps, on suppose que l'ensemble d'incertitude est l'enveloppe convexe des 6 premiers scénarios et que les variables ajustables sont données par (4.19). Pour chaque pas de temps, l'utilisation de la solution de la CRAA et des équations (4.19) fournit les commandes préconisées par la CRAA à ce pas de temps. On fait alors varier la taille de l'intervalle $I_t = \{1, \ldots, t-m\}$, et l'on mesure, pour chaque choix de l'intervalle I_t, le coût moyen et l'écart type des coûts sur l'ensemble des 6 scénarios.

Valeur de m	$m=1$	$m=2$	$m=4$	$m=6$	$m=12$
Moyenne	4.26×10^8	4.261×10^8	4.288×10^8	4.33×10^8	4.57×10^8
Écart type	1.04×10^5	3.95×10^4	7.66×10^4	1.71×10^5	4.05×10^6

TAB. 4.8 – Influence de la quantité d'information utilisée sur les performances de la CRAA.

Comme prévu, les coûts moyens sont une fonction croissante de m. En effet, plus m augmente et moins on utilise d'information pour expliquer les commandes ajustables et donc plus l'ensemble de faisabilité est petit. Lorsque m vaut 12 (le nombre de pas de temps), on retrouve la solution de la CR.
Nous fixons maintenant $I_t = \{1, \ldots, t-1\}$, et on observe l'influence du choix des variables explicatives sur les performances de la CRAA. On envisage quatre schémas de dépendance pour les variables ajustables en fonction des paramètres incertains :

- Pour la première méthode, dite méthode 1, les commandes thermiques, hydrauliques et EJP sont ajustables et données par (4.19). Pour la deuxième méthode, dite mé-

thode 2, les commandes thermiques et hydrauliques sont ajustables et données par (4.19). Les commandes EJP ne sont pas ajustables.

– Pour la troisième méthode, dite méthode 3, toutes les commandes sont ajustables ; les commandes thermiques et EJP étant données par (4.19) et les commandes hydrauliques étant données par (4.20). Pour la dernière méthode, dite méthode 4, les commandes thermiques et hydrauliques sont ajustables et données par respectivement (4.19) et (4.20). Les commandes EJP ne sont pas ajustables.

En choisissant comme ensemble d'incertitude l'enveloppe convexe des 6 puis des 20 premiers scénarios nous obtenons les résultats suivants.

Méthode	Méthode 1	Méthode 2	Méthode 3
Moyenne	4.26×10^8	4.26×10^8	4.254×10^8
Écart type	1.04×10^5	9.00×10^4	6.87×10^4

TAB. 4.9 – Influence du choix des variables explicatives sur les performances de la CRAA. 6 premiers scénarios.

Méthode	Méthode 1	Méthode 2	Méthode 3	Méthode 4
Moyenne	4.471×10^8	4.472×10^8	4.456×10^8	4.456×10^8
Écart type	2.391×10^6	2.387×10^6	6.223×10^4	6.394×10^4

TAB. 4.10 – Influence du choix des variables explicatives sur les performances de la CRAA. 20 premiers scénarios.

Comme prévu, plus l'information utilisée est riche et plus le coût moyen diminue. Il semble intéressant d'utiliser à la fois les consommations électriques et les apports afin d'expliquer les commandes hydrauliques. Par contre, il n'apparaît pas nécessaire, au vu de ces simulations, de considérer que les variables EJP sont ajustables. Enfin, l'évolution des résultats lorsqu'on passe de 6 à 20 scénarios laisse penser que le coût moyen que l'on obtiendrait sur les 456 scénarios avec la solution de la CRAA doit être assez élevé.

Ensemble d'incertitude défini par des contraintes de boîte

Nous terminons cette section en choisissant un ensemble d'incertitude défini par des contraintes de boîte du type (4.21). On obtient alors la CRAA (4.52) donnée en annexe de ce chapitre. On calibre tout simplement $d^{\min}$ et $d^{\max}$ en estimant pour tout pas de temps t, $d_t^{\min}$ et $d_t^{\max}$ par les valeurs minimales et maximales des consommations au pas de temps t sur l'ensemble des scénarios. Les vecteurs $a^{\ell,\min}, a^{\ell,\max}$, $\tau^{\ell,\min}$ et $\tau^{\ell,\max}$ sont calibrés de façon similaire. On choisit pour I_t, l'intervalle $I_t = \{1, \ldots, t-1\}$. Deux CRAA sont mises en place. Les variables ajustables sont des fonctions des paramètres incertains données par (4.19) pour la première (notée $CRAA_1$). Pour la seconde (notée $CRAA_2$), les commandes thermiques et EJP sont données par (4.19) et les commandes hydrauliques par (4.20). On notera $CRAA_2'$ la méthode $CRAA_2$ glissante. Les coûts moyens et l'écart type des coûts sur l'ensemble des scénarios sont donnés, pour ces différentes $CRAA$, dans le tableau 4.11 ci-après.

	Coût optimal	$CRAA_1$	$CRAA_2$	$CRAA'_2$
Moyenne	4.12×10^8	5.01×10^8	4.70×10^8	4.75×10^8
Écart type	1.27×10^7	2.4×10^6	3.57×10^6	1.2×10^7

TAB. 4.11 – Performances des CRAA. Ensemble d'incertitude défini par des contraintes de boîte.

L'utilisation des apports pour expliquer les variables de commandes hydrauliques s'avère ici particulièrement intéressant. Les coûts moyens de $CRAA_2$ et $CRAA'_2$ sont meilleurs que le coût moyen obtenu avec la méthode VaR_{CA} mais moins bons que ceux obtenus avec les autres méthodes de gestion.
Notons $\bar{d} = \frac{1}{2}(d^{\min}+d^{\max})$, le centre de l'ensemble d'incertitude sur les consommations électriques. Afin de souligner l'importance de la finesse de la calibration de l'ensemble d'incertitude, précisons que si on avait utilisé l'ensemble d'incertitude $[\bar{d}-\frac{1}{8}(\bar{d}-d^{\min}), \bar{d}+\frac{1}{8}(d^{\max}-\bar{d})]$ pour les consommations électriques, et en remplaçant les scénarios de demande par des scénarios tirés uniformément dans $[\bar{d}-\frac{1}{8}(\bar{d}-d^{\min}), \bar{d}+\frac{1}{8}(d^{\max}-\bar{d})]$, on aurait vu le coût moyen lié à l'utilisation de la CRAA $CRAA_2$ diminuer fortement, passant de 4.7×10^8 à 4.51×10^8.

4.7.1.5 Tableau comparatif des méthodes de gestion

Nous récapitulons, dans le tableau ci-dessous, le coût moyen et l'écart type des coûts de gestion simulés pour les différentes méthodes de gestion. Pour chaque méthode, on choisit la calibration des paramètres de la méthode conduisant à la valeur la plus faible de la somme moyenne plus écart type des coûts. On notera 'Arbre' la méthode décrite en section 4.3.2 et 'Stab. Duale' la méthode de stabilisation de la fonction duale présentée en section 4.4.1. Enfin $CRAA_2$ et $CRAA'_2$ sont les méthodes définies dans le paragraphe précédent. Les méthodes sont classées en fonction de la valeur de la somme moyenne plus écart type.

	Stab. Duale	CRA	$CRAA_2$	$CRAA'_2$	VaR_{benef}	Arbre	VaR_{CA}
Moyenne	4.22×10^8	4.54×10^8	4.70×10^8	4.75×10^8	4.60×10^8	4.65×10^8	4.80×10^8
Écart type	1.04×10^5	1.01×10^6	3.57×10^6	1.2×10^7	3.18×10^7	4.46×10^7	4.72×10^7
Rang	1	2	3	4	5	6	7

TAB. 4.12 – Tableau comparatif des méthodes de gestion.

4.7.2 Simulations sur des données réelles

Nous terminons cette section consacrée à la comparaison des méthodes de gestion en les testant sur des données réelles. À partir de l'historique des consommations d'électricité en France de 1996 à 2003 inclus (étudié dans la section 4.6), l'objectif est d'utiliser les x premières années de cet historique (pour $2 \leq x \leq 7$) afin de déterminer un plan de production pour l'année $x+1$. Nous utilisons le même parc de production que dans la section précédente. Les apports des deux réservoirs sont fixés et on suppose que tous les groupes thermiques de toutes les centrales thermiques sont disponibles pour tous les pas de temps. Nous testerons 3 CRAA et 3 CR obtenues en prenant des ensembles d'incertitude définis par des contraintes de boîte et calibrés dans la section 4.6.

- On notera $CRAA_1$ et CR_1 les CRAA et CR obtenues en prenant comme ensemble d'incertitude sur les consommations l'ensemble déterminé en utilisant la méthode détaillée en section 4.6.1. On utilise donc pour expliquer la consommation l'année $x+1$, les températures et nébulosités de cette même année. Le niveau de confiance sur l'intervalle de prédiction de chaque consommation est 95%.
- On notera $CRAA_2$ et CR_2 les CRAA et CR obtenues lorsque l'ensemble d'incertitude sur les consommations est déterminé en utilisant la méthode détaillée en section 4.6.2 (méthode SARIMA).
- On notera enfin $CRAA_3$ et CR_3 les CRAA et CR obtenues lorsque l'ensemble d'incertitude $[d^{\min}, d^{\max}]$ sur les consommations est tel que $d_t^{\min}$ est la plus petite des consommations au pas de temps t sur les x premières années et $d_t^{\max}$ la plus grande des consommations au pas de temps t sur les x premières années.

On notera respectivement CS_1, CS_2, CS_3, les méthodes consistant à remplacer les consommations incertaines par leurs prévisions en utilisant respectivement les méthodes de prévision détaillées aux sections 4.6.1, 4.6.2 et les consommations moyennes sur les x premières années. Pour la mise en place des CRAA, on choisit au pas de temps t, $I_t = \{1, \ldots, t-1\}$; et on suppose que les variables ajustables sont des fonctions des consommations données par (4.19). Les résultats sont les suivants (la première ligne du tableau donne les coûts optimaux).

x	3	4	5	6	7	8	m
Opt	3.80×10^8	3.86×10^8	3.98×10^8	4.12×10^8	4.27×10^8	4.42×10^8	4.07×10^8
$CRAA_1$	3.98×10^8	4.13×10^8	4.17×10^8	4.18×10^8	4.44×10^8	4.49×10^8	4.23×10^8
$CRAA_2$	-	4.50×10^8	4.53×10^8	4.64×10^8	4.77×10^8	4.89×10^8	4.67×10^8
$CRAA_3$	4.05×10^8	4.12×10^8	4.25×10^8	4.39×10^8	4.47×10^8	-	4.26×10^8
CR_1	4.32×10^8	4.39×10^8	4.52×10^8	4.67×10^8	4.75×10^8	-	4.53×10^8
CR_2	-	4.75×10^8	4.78×10^8	4.83×10^8	5.03×10^8	5.15×10^8	4.91×10^8
CR_3	4.25×10^8	4.39×10^8	4.43×10^8	4.55×10^8	4.69×10^8	4.74×10^8	4.51×10^8
CS_1	4.69×10^8	4.54×10^8	4.56×10^8	4.62×10^8	4.69×10^8	4.71×10^8	4.63×10^8
CS_2	-	4.70×10^8	4.79×10^8	4.89×10^8	5.08×10^8	5.19×10^8	4.93×10^8
CS_3	4.65×10^8	4.72×10^8	4.84×10^8	5.00×10^8	5.07×10^8	-	4.86×10^8

TAB. 4.13 – Utilisation de 8 ans d'historiques des consommations d'électricité en France de 1996 à 2003 inclus. Coûts des CRAA et CR du problème de gestion de production électrique (4.12) mises en place les années 3, 4, 5, 6, 7 et 8 en utilisant ces historiques.

Les CRAA donnent ici les résultats les plus satisfaisants. Pour $CRAA_1$, en particulier, le coût moyen est à un peu moins de 4.5% du coût moyen optimal. Les résultats de la CRAA utilisant un ensemble d'incertitude polytopique sont très mauvais en moyenne et ne sont pas présentés ici. Ceci peut s'expliquer par le fait que cet ensemble d'incertitude ne reflète pas du tout la dynamique de la demande (notamment sa tendance à la croissance). Par conséquent, le vecteur des consommations électriques l'année $x+1$ n'appartient pas forcément à l'enveloppe convexe des scénarios des x premières années. Dans ces conditions, aucune garantie n'est donnée sur la qualité de la solution de la CRAA ou de la CR.

4.8 Conclusion sur les problèmes de gestion de production électrique

Dans ce chapitre, nous avons présenté des méthodes stables ou robustes pour le problème de gestion de production électrique que nous avons comparées sur des données réelles et simulées. Notre but était d'améliorer le modèle décrit en section 4.3.2 consistant à résoudre le problème par programmation stochastique multi-étapes sur un arbre de scénarios. En effet, une telle approche conduit à un écart type des coûts de gestion simulés assez élevé et a tendance à trop vider le plus gros des réservoirs.
Trois modèles visant à régulariser la fonction duale ont tout d'abord été proposés afin de pallier ces défauts. Si le premier modèle VaR_{CA} ne s'est pas révélé concluant en pratique, les deux autres modèles se sont avérés adaptés à nos objectifs. Concernant le modèle VaR_{benef}, d'une part ce modèle conduit à une diminution de l'écart type des coûts de gestion simulés et d'autre part il tend à vider moins le plus gros des réservoirs. La dernière méthode de stabilisation de la fonction duale est celle introduite dans la section 4.4.1. L'intérêt de cette approche est de tenir à la fois compte de l'incertitude sur la demande et sur les taux de disponibilité. Parmi toutes les méthodes testées, cette méthode est celle qui fournit les meilleurs résultats en moyenne (son coût moyen pouvant être [en fonction de la calibration de Γ et de κ] à environ 0.08% du coût moyen optimal). D'autre part, l'écart type des coûts de gestion simulés en utilisant cette méthode est environ 4 fois inférieur à celui obtenu en utilisant la méthode VaR_{benef}. Enfin, cette méthode conserve le schéma de décomposition par les prix. Aucun arbre de scénario n'est utilisé et la taille des sous-problèmes duaux est de l'ordre du nombre de pas de temps. Par conséquent, cette méthode permet de résoudre des problèmes de gestion de production comportant un très grand nombre de pas de temps et d'unités de production. Apportons un bémol aux résultats encourageant de cette méthode. Les simulations ont été conduites pour cette méthode sur 24 pas de temps au lieu de 364 comme pour les méthodes VaR_{CA} et VaR_{benef} et on supposait connaître les apports. Il conviendrait de tester les résultats de cette méthode pour diverses agrégations des pas de temps afin d'observer si la qualité de la solution dépend beaucoup ou non de cette politique d'agrégation des pas de temps.

Dans un deuxième temps, nous avons proposé des CR, CRA et CRAA du problème de gestion de production électrique. Plusieurs problèmes ont alors été abordés et notamment :

1. le choix des variables ajustables ;
2. le choix de l'intervalle I_t et des schémas de dépendance des variables ajustables en fonction des paramètres incertains ;
3. la calibration des ensembles d'incertitude.

Pour ces CR, CRA et CRAA, l'ensemble d'incertitude le plus simple auquel on peut penser est l'enveloppe convexe des scénarios disponibles. Cependant, cette approche conduit à des résultats assez conservatifs, comme on a pu le voir sur les CRAA avec ensemble d'incertitude polytopique qui ont été testées. D'un point de vue statistique, des ensembles d'incertitude définis par des contraintes de boîte semblent les plus aisés à calibrer finement. Il s'agit alors d'étudier les processus stochastiques mis en jeu et c'est ce qui a été fait dans la section 4.6 afin d'étudier la série chronologique des consommations électriques pour tester les méthodes de gestion en utilisant cette série. On a pu voir, en utilisant ces données réelles, l'intérêt de la CRAA par rapport à la CR.

Les tests effectués sur des données simulées ont montré que les CRAA offrent un meilleur compromis coût moyen/écart type des coûts que la méthode VaR_{benef} mais un coût moyen supérieur. Cependant, il faut noter que les ensembles d'incertitude retenus pour mettre en place les CRAA sur ces données simulées étaient assez conservatifs.
Quant à la CRA, elle fournit en théorie de meilleurs résultats que la CRAA. Cependant, le schéma de dépendance des variables ajustables en fonction des paramètres incertains étant inconnu, une adaptation de la CRA a été proposée mais elle ne s'est pas avérée concluante.

4.9 Le problème de gestion du transport et de la distribution du gaz

L'objet de cette section est d'étudier le problème de la gestion du transport et de la distribution du gaz en France. Dans un premier temps le problème est présenté puis exprimé sous la forme d'un problème d'optimisation linéaire stochastique. Cette section est ensuite l'occasion de revenir sur le concept de CRA. Afin de tenir compte de l'incertitude sur certains paramètres du problème, nous proposons deux manières pour robustifier ce problème, par CRA puis par CRAA. En particulier, l'utilisation des CRA et CRAA prend ici tout son sens puisque la CR du problème est infaisable.

4.9.1 Présentation du problème

Afin de satisfaire la demande en gaz en un certain nombre de villes ou plus généralement de points d'alimentation $i(i = 1, \ldots, n)$, et pour des pas de temps $t(t = 1, \ldots, T)$, la France dispose d'un certain nombre de contrats (disons C) d'approvisionnement en gaz. Au pas de temps t, un certain nombre de contrats sont en activité (les contrats u tels que $t_{\min}(u) \leq t \leq t_{\max}(u)$). Chaque contrat permet de faire venir, via des gazoducs haute pression (70 bars) du gaz d'un site de production gazier vers le réseau de gazoducs français. Le transport du gaz dans les gazoducs haute pression est assuré par des stations de compression régulièrement espacées (tous les 100 ou 200 kilomètres) sur les gazoducs. Les gazoducs convergent d'une part vers des points d'alimentation espacés sur tout le territoire et d'autre part vers un réseau de gazoducs moyenne pression. Le gaz est fourni aux utilisateurs soit directement à partir de gazoducs moyenne pression (si les utilisateurs ont un régulateur) ou à partir d'un réseau basse pression relié au réseau moyenne pression.

Concernant les coûts d'approvisionnement, plus le site de production est éloigné et plus les coûts sont élevés. Le coût dépend aussi de la quantité de gaz acheté au fournisseur. Le prix du gaz sera d'autant plus faible qu'une grande quantité de gaz est achetée. Un autre élément influençant le prix du gaz est la capacité du client à commander des quantités constantes et régulièrement au cours de l'année. Si le client adopte une telle politique (ce qui est le cas de la France), il doit mettre en place des moyens de stockage du gaz dans des réservoirs afin, ensuite, de faire face aux pics de consommation hivernaux. La construction de ces réservoirs a un coût dépendant de la taille du réservoir et de la configuration du terrain. Le coût de stockage dépend ensuite de la quantité de gaz injecté ou soutiré de la réserve. Les coûts de transport et de distribution du gaz dépendent quant à eux de la distance parcourue par le gaz dans les pipelines et des types de pipelines empruntés.

4.9.2 Modélisation

L'objectif est de maximiser le bénéfice sur l'horizon de gestion (du pas de temps $t = 1$ au pas de temps $t = T$) tout en satisfaisant la demande aux différents points d'approvisionnement. Le problème peut être modélisé sous la forme d'un graphe E dont les arêtes matérialisent le réseau de transport et de distribution et dont les sommets sont de quatre types :

- Les sommets sources. Chaque sommet source est associé à un site de production. S_o est l'ensemble des sommets sources.
- Les sommets puits (sans arcs sortant) en lesquels une certaine quantité de gaz doit être transportée à chaque pas de temps. P est l'ensemble des puits du graphe.
- Des sommets neutres, correspondant à des points de connection de différents pipelines. N est l'ensemble des sommets neutres.
- Des sommets correspondant à des lieux de stockage. S est l'ensemble de ces sommets.

Nous utilisons les notations suivantes :

- x_u^t est la quantité de gaz fournie par le contrat u au pas de temps t et pour le prix unitaire p_u^t (qui inclue le coût d'achat et de transport). Les prix $(p_u^t)_{u,t}$ ne sont pas fixés à l'avance et sont donc des paramètres incertains du système au début de l'horizon de gestion.
- $y_{i,j}^t$ est la quantité de gaz allant du sommet i au sommet j au pas de temps t (i n'est pas un sommet source). Le coût de transport (ou de distribution) est linéaire, le coût unitaire de transport entre les sommets i et j au pas de temps t étant $q_{i,j}^t > 0$.
- d_i^t est la demande au puits i et au pas de temps t. Toutes les demandes doivent être satisfaites. Ces demandes sont inconnues au début de la période de gestion. Au pas de temps t, et en un puits i, le gaz est vendu au prix unitaire λ_i^t. L'effacement η_i^t correspond à la demande non satisfaite au puits i, et au pas de temps t. Par ailleurs, en un puits i, on ne peut pas effacer plus d'une quantité P_i et l'effacement doit être limité en chaque puits et pour chaque pas de temps, à savoir $0 \leq \eta_i^t \leq \varepsilon_i d_i^t$, où ε_i est une constante fixée.
- s_i^t est le stock de gaz au noeud de stockage i, au début du pas de temps $t+1$. Le stock s_i^0 de la réserve i au début de la période de gestion est connu. Par ailleurs, le coût de stockage unitaire est r_i^t pour le stock i au pas de temps t.

Les sommets neutres correspondent soit à des points du réseau ou le coût unitaire change (typiquement la taille du gazoduc change en ce point du réseau) ou à un point de rencontre entre au moins trois gazoducs. On représente, sur la figure 4.11 ci-après, un exemple de graphe E d'un réseau simple de transport et de distribution du gaz.
Soit $F(u)$ le sommet du graphe alimenté par le contrat u et soit A_i^t l'ensemble des contrats fournissant du gaz au sommet i au pas de temps t :

$$A_i^t = \{u \mid 1 \leq u \leq C,\ \ t_{\min(u)} \leq t \leq t_{\max}(u),\ \ F(u) = i\}.$$

Les équations de conservation du flot s'écrivent :

$$\sum_{u \in A_i^t} x_u^t + \sum_{j|(j,i)\in E} y_{j,i}^t = \sum_{j|(i,j)\in E} y_{i,j}^t,\ \ \forall\, t = 1, \ldots, T,\ i \in N, \tag{4.46}$$

pour les noeuds neutres et

$$s_i^{t-1} + \sum_{u \in A_i^t} x_u^t + \sum_{j|(j,i)\in E} y_{j,i}^t = s_i^t + \sum_{j|(i,j)\in E} y_{i,j}^t,\ \ \forall\, t = 1, \ldots, T,\ \ i \in S, \tag{4.47}$$

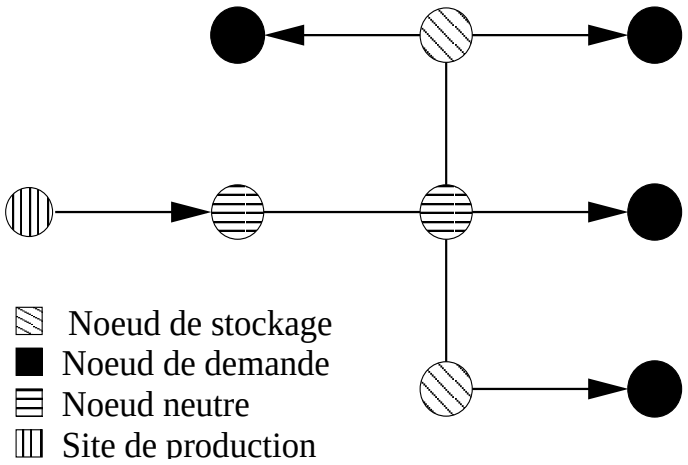


FIG. 4.11 – Réseau de transport et de distribution du gaz.

pour les noeuds de stockage. Le chiffre d'affaire des ventes de gaz s'exprime au pas de temps t par : $\sum_{i\in P} \lambda_i^t \; d_i^t 1(\eta_i^t = 0) + \sum_{i\in P} \lambda_i^t \; (d_i^t - \eta_i^t) 1(\eta_i^t > 0)$, ou encore $\sum_{i\in P} \lambda_i^t \; (d_i^t - \eta_i^t)$. Il va s'agir alors, pour tous les pas de temps, de déterminer les quantités de gaz à acheter sur les différents contrats (en respectant les quantités minimales et maximales à acheter qu'exigent ces contrats) ainsi que les quantités de gaz à faire transiter sur le réseau, de façon à maximiser les bénéfices. Ainsi, si tous les paramètres du problème étaient connus à l'avance, la maximisation, sur l'horizon de gestion, du bénéfice lié au problème de transport et de distribution du gaz, conduirait au problème :

$$
\left\{
\begin{array}{l}
\min \; \displaystyle\sum_{u=1}^{C} \sum_{t_{\min}(u)\le t\le t_{\max}(u)} p_u^t x_u^t + \sum_{1\le t\le T} \left(\sum_{(i,j)\in E} q_{i,j}^t \; y_{i,j}^t + \sum_{i\in S} r_i^t \; s_i^t + \sum_{i\in P} \lambda_i^t \; \eta_i^t \right) \\
\displaystyle\sum_{u\in A_i^t} x_u^t + \sum_{j|(j,i)\in E} y_{j,i}^t = \sum_{j|(i,j)\in E} y_{i,j}^t, \;\; \forall\, t = 1,\ldots,T, \; i \in N, \\
\displaystyle s_i^{t-1} + \sum_{u\in A_i^t} x_u^t + \sum_{j|(j,i)\in E} y_{j,i}^t = s_i^t + \sum_{j|(i,j)\in E} y_{i,j}^t, \;\; \forall\, t = 1,\ldots,T, \;\; i \in S, \\
\displaystyle\sum_{k|(k,i)\in E} y_{k,i}^t + \eta_i^t = d_i^t, \;\; \forall\, i \in P, \forall\, t = 1,\ldots,T, \\
\displaystyle\sum_{t=1}^{T} \eta_i^t \le P_i, \;\; 0 \le \eta_i^t \le \varepsilon_i \, d_i^t, \;\; \forall\, i \in P, \; \forall\, t = 1,\ldots,T, \\
0 \le s_i^t \le s_i^{\max}, \;\; \forall\, i \in S, \;\; t = 1,\ldots,T, \\
x_u^{t,\min} \le x_u^t \le x_u^{t,\max}, \;\; \forall\, u = 1,\ldots,C, \; t = 1,\ldots,T, \\
0 \le y_{i,j}^t \le y_{i,j}^{\max}, \;\; \forall\, t = 1,\ldots,T, \; i \in N \bigcup S.
\end{array}
\right. \tag{4.48}
$$

On montre facilement que pour chaque unité de temps, dans un pipeline donné, le gaz ne circulera que dans une seule direction.

Lemme 4.9.1 *Soit* $(\bar{x}, \bar{y}, \bar{s}, \bar{\eta})$, *une solution optimale du problème d'optimisation* (4.48). *Pour tout* $t = 1,\ldots,T$, *et pour tout sommet* i *et* j *du graphe, si* $\bar{y}_{ij}^t > 0$, *alors* $\bar{y}_{ji}^t = 0$.

Preuve Soit $1 \leq t \leq T$, et soient i et j deux sommets du graphe. Supposons que $\bar{y}_{ij}^t > 0$ et $\bar{y}_{ji}^t > 0$. Définissons $\tilde{y}$ par $\tilde{y}_{mn}^t = \bar{y}_{mn}^t$ pour tout sommet m et n du graphe différent de i et j et :

$$\begin{array}{ll} \tilde{y}_{ij}^t = \bar{y}_{ij}^t - \bar{y}_{ji}^t, \;\; \tilde{y}_{ji}^t = 0, & \text{si } \bar{y}_{ij}^t > \bar{y}_{ji}^t, \\ \tilde{y}_{ij}^t = 0, \;\; \tilde{y}_{ji}^t = \bar{y}_{ji}^t - \bar{y}_{ij}^t, & \text{sinon.} \end{array} \tag{4.49}$$

On montre ensuite que le quadruplet $(\tilde{x}, \tilde{y}, \tilde{s}, \tilde{\eta})$, avec $\tilde{x} = \bar{x}, \tilde{s} = \bar{s}, \tilde{\eta} = \bar{\eta}$ est faisable et conduit à une meilleure solution que $(\bar{x}, \bar{y}, \bar{s}, \bar{\eta})$, ce qui est absurde.
On vérifie tout d'abord la faisabilité de $(\tilde{x}, \tilde{y}, \tilde{s}, \tilde{\eta})$. On a bien $0 \leq \tilde{y}_{ij}^t \leq \bar{y}_{ij}^t \leq y_{ij}^{\max}$, et $0 \leq \tilde{y}_{ji}^t \leq y_{ji}^{\max}$. Il suffit ensuite de vérifier la conservation du flot aux sommets i et j (qui ne peuvent être des puits). Nous avons, pour tout pas de temps t et au sommet i :

$$\begin{aligned} \sum_{(k,i)\in E} \tilde{y}_{ki}^t - \sum_{(i,k)\in E} \tilde{y}_{ik}^t &= \sum_{(k,i)\in E,\, k\neq i} \bar{y}_{ki}^t - \sum_{(i,k)\in E,\, k\neq i} \bar{y}_{ik}^t + \bar{y}_{ji}^t - \bar{y}_{ij}^t \\ &= \sum_{(k,i)\in E} \bar{y}_{ki}^t - \sum_{(i,k)\in E} \bar{y}_{ik}^t. \end{aligned}$$

On montre de même la conservation du flot au sommet j. De plus, les coûts des politiques $(\bar{x}, \bar{y}, \bar{s}, \bar{\eta})$ et $(\tilde{x}, \tilde{y}, \tilde{s}, \tilde{\eta})$ diffèrent seulement par les coûts de transport entre les sommets i et j. Si $\bar{y}_{ij}^t > \bar{y}_{ji}^t$, pour le choix des quantités transportées $\tilde{y}$, ce coût est $q_{ij}^t(\bar{y}_{ij}^t - \bar{y}_{ji}^t)$ qui est strictement inférieur (puisque $\bar{y}_{ji}^t > 0$) au coût de transport $q_{ij}^t(\bar{y}_{ij}^t + \bar{y}_{ji}^t)$ du gaz entre i et j pour la politique $\bar{y}$. L'autre cas est symétrique.

□

4.9.3 Contrepartie robuste ajustée

Afin de robustifier le problème (4.48), diverses voies peuvent être empruntées. On peut pénaliser l'objectif par un terme du type $-\sqrt{x^T P x}$, avec x le vecteur donnant les quantités achetées sur chaque contrat et P une matrice définie positive. Nous allons montrer comment robustifier ce problème de gestion en utilisant la méthodologie de la contrepartie robuste ajustée présentée dans le chapitre 1. Remarquons tout de suite que la contrepartie robuste du problème (4.48) est infaisable puisque les contraintes de satisfaction de la demande (laquelle est incertaine) sont des contraintes d'égalité. Pour déterminer une contrepartie robuste ajustée, il convient dans un premier temps de déterminer les variables non ajustables. On peut chercher à fixer les quantités de gaz à acheter à l'avance et ajuster ensuite les quantités de gaz à faire transiter sur le réseau et gérer les stocks afin de satisfaire la demande. Dans cette modélisation, les variables non ajustables sont donc les variables $(x_u^t)_{u,t}$, les variables restantes (y, s, η) étant ajustables. Cette politique permettra de pouvoir avoir une meilleure idée des prix du gaz (le fournisseur connaissant les quantités exactes qui seront prises). Néanmoins, ces prix ne sont pas connus de façon exacte au début de la période de gestion. De même, les demandes en gaz d_i^t pour les puits i et les différents pas de temps t ne sont pas connues à l'avance. Nous supposons en outre que les prix de vente du gaz λ_i^t sont quant à eux fixés sur la période de gestion (c'est le cas en pratique) et que la demande et les prix d'achat du gaz appartiennent à l'enveloppe convexe des S' scénarios $((p_u^t(s), d_i^t(s))_{u,i,t})_{1\leq s\leq S'}$. En utilisant, le théorème 2.2 de [11] et en introduisant des variables de décision $(y_{ij}^t(s), s_i^t(s), \eta_i^t(s))_{s=1}^{S'}$ associées aux différents scénarios,

la contrepartie robuste ajustée du problème (4.48) est alors le problème linéaire suivant :

$$\left\{\begin{array}{l}
\min \quad F \\
\displaystyle\sum_{u=1}^{C} \sum_{t_{\min}(u)\le t\le t_{\max}(u)} p_u^t(k) x_u^t + \sum_{1\le t\le T} \left(\sum_{(i,j)\in E} q_{i,j}^t\, y_{i,j}^t(k) + \sum_{i\in S} r_i^t\, s_i^t(k) + \sum_{i\in P} \lambda_i^t\, \eta_i^t(k) \right) \le F,\ \forall\, k = 1,\dots,S', \\
\displaystyle\sum_{u\in A_i^t} x_u^t + \sum_{j|(j,i)\in E} y_{j,i}^t(k) = \sum_{j|(i,j)\in E} y_{i,j}^t(k),\ \ \forall\, t = 1,\dots,T,\ i\in N,\ k = 1,\dots,S', \\
\displaystyle s_i^{t-1}(k) + \sum_{u\in A_i^t} x_u^t + \sum_{j|(j,i)\in E} y_{j,i}^t(k) = s_i^t(k) + \sum_{j|(i,j)\in E} y_{i,j}^t(k),\ \ \forall\, t = 1,\dots,T,\ \ i\in S,\ k = 1,\dots,S', \\
\displaystyle\sum_{k|(k,i)\in E} y_{k,i}^t(k) + \eta_i^t(k) = d_i^t(k),\ \ \forall\, i\in P,\ \forall\, t = 1,\dots,T,\ \forall k = 1,\dots,S', \\
\displaystyle\sum_{t=1}^{T} \eta_i^t(k) \le P_i,\ \ 0 \le \eta_i^t(k) \le \varepsilon_i\, d_i^t(k),\ \ \forall\, i\in P,\ \forall\, t = 1,\dots,T,\ \forall k = 1,\dots,S', \\
0 \le s_i^t(k) \le s_i^{\max},\ \ \forall\, i\in S,\ \ t = 1,\dots,T,\ k = 1,\dots,S', \\
x_u^{t,\min} \le x_u^t \le x_u^{t,\max},\ \ \forall\, u = 1,\dots,C,\ t = 1,\dots,T, \\
0 \le y_{i,j}^t(k) \le y_{i,j}^{\max},\ \ \forall\, t = 1,\dots,T,\ i\in N\bigcup S,\ k = 1,\dots,S'.
\end{array}\right. \tag{4.50}$$

Afin d'illustrer les résultats de cette méthode, nous considérons un réseau de petite taille comportant 6 puits, deux sites de stockage (de capacités maximales 5000) et dont les contenances initiales sont de 50. On dispose de deux contrats permettant d'acheter entre 0 et 4000 à chaque pas de temps, le coût unitaire de stockage étant de 2, le prix de vente 5, et la demande et les prix sont dans l'enveloppe convexe des scénarios $(p,d) = ((5\mathbf{e},5\mathbf{e}),(10\mathbf{e},10\mathbf{e}),(15\mathbf{e},20\mathbf{e}),(20\mathbf{e},50\mathbf{e}),(40\mathbf{e},100\mathbf{e}))$. On choisit $\varepsilon_i = 1$ pour tout puits et 12 pas de temps. Le coût moyen sur l'ensemble des scénarios est de $1.5.10^4$ en utilisant la solution fournie par la contrepartie robuste ajustée alors que le coût optimal sur l'ensemble de ces scénarios est $1.43.10^4$. Cet exemple montre que la contrepartie robuste ajustée peut fournir une solution dont le coût moyen est proche du coût moyen optimal.

On peut aussi proposer une contrepartie robuste ajustée affinement. On considère que les seuls paramètres incertains sont les prix d'achat p_u^t du gaz. La modélisation est ici complètement différente de l'approche précédente quant au choix des variables ajustables et non ajustables. Il semble ici plus simple d'essayer d'expliquer les achats x_u^t de gaz sur le contrat u au pas de temps t par les demandes passées des puits du sous-graphe du graphe E dont la racine est le sommet alimenté par le contrat u. On peut alors proposer le schéma de dépendance suivant pour les commandes x_u^t :

$$x_u^t = \sum_{r\in I_t} \alpha_u^{t,r} D_{F(u)}^r \text{ avec } D_j^r = \sum_{k\in S_j} d_k^r, \tag{4.51}$$

où S_j est l'ensemble des puits du sous-graphe du graphe E dont la racine est le sommet j et $I_t = \{1,\dots,t-1\}$. Les termes en x dans la première contrainte de (4.48) peuvent alors s'écrire sous la forme : $\displaystyle\sum_{r=1}^{T}\sum_{k\in P} \beta_k^r\, d_k^r$, avec

$$\beta_k^r = \sum_{u\ |\ k\in S_{F(u)}} \sum_{t\in B_r^u} p_u^t\ \alpha_u^{t,r},$$

où

$$B_r^u = \{t\ |\ \ r\in I_t,\ \ t_{\min}(u) \le t \le t_{\max}(u)\}.$$

En utilisant le théorème 4.2 donné dans [11], on peut alors résoudre un problème semidéfini positif qui donnera la solution optimale de cette CRAA correspondant à un ensemble d'incertitude ellipsoïdal pour les prix p_u^t.

4.10 CRA et CRAA du problème de gestion de production

Nous donnons dans cette dernière section, les CRA présentées dans ce chapitre (les notations utiles à la présentation de ces CRA sont données à la fin de cette section).

CRAA-Contraintes de boîte

La CRAA introduite en section 4.4.3.2 est le problème linéaire suivant :

$$
\left\{
\begin{array}{l}
\min \ F \\
\sum_{\ell\in\mathcal{L}_T}\sum_{t=1}^{T} c_\ell q_t^{\ell,0} + \sum_{t=1}^{T-m}(\nu_t^1 d_t^{\max} + \nu_t^2 d_t^{\min}) - \sum_{\ell\in\mathcal{L}_H} a_\ell - \sum_{\ell\in\mathcal{L}_J} b_\ell \le F, \\
\nu_k^1 \ge 0,\ \nu_k^2 \le 0,\ \nu_k^1 \ge \xi_k^{'},\ \nu_k^2 \le \xi_k^{'},\ \nu_k^1 + \nu_k^2 = \xi_k^{'},\ k = 1,\dots,T-m, \\
\alpha_{t,1}^{\ell,k} \ge 0,\ \beta_{t,1}^{\ell,k} \le 0,\ \alpha_{t,1}^{\ell,k} \ge \gamma_{t,1}^{\ell,k},\ \beta_{t,1}^{\ell,k} \le \gamma_{t,1}^{\ell,k},\ \alpha_{t,1}^{\ell,k} + \beta_{t,1}^{\ell,k} = \gamma_{t,1}^{\ell,k},\ t \ge m+1,\ k \in I_t, \\
\alpha_{t,2}^{\ell,k} \ge 0,\ \beta_{t,2}^{\ell,k} \le 0,\ \alpha_{t,2}^{\ell,k} \ge \gamma_{t,2}^{\ell,k} - 1,\ \beta_{t,2}^{\ell,k} \le \gamma_{t,2}^{\ell,k} - 1,\ \alpha_{t,2}^{\ell,k} + \beta_{t,2}^{\ell,k} = \gamma_{t,2}^{\ell,k} - 1,\ t \ge m+1,\ k \in I_t, \\
a_\ell + c_{H,k}^{\ell}\left(\sum_{t=1}^{T}(dev_t^\ell + r_t^{\ell,0}) + \sum_{t=1}^{T-m}(\alpha_{T,1}^{\ell,t} d_t^{\max} + \beta_{T,1}^{\ell,t} d_t^{\min} + \alpha_{T,2}^{\ell,t} a_t^{\ell,\max} + \beta_{T,2}^{\ell,t} a_t^{\ell,\min})\right) \\
\quad \le c_{H,k}^{\ell}(x_1^\ell + \sum_{t=T-m+1}^{T} a_t^{\ell,\min}) + d_{H,k}^{\ell},\ \ \ell \in \mathcal{L}_H,\ k \in M_H^\ell, \\
x_1^\ell + \sum_{j=1}^{t} a_j^{\ell,\max} - x_{\max}^\ell \le \sum_{j=1}^{t} dev_j^\ell + r_j^{\ell,0} \le x_1^\ell + \sum_{j=1}^{t} a_j^{\ell,\min} - x_{\min}^\ell,\ \ell \in \mathcal{L}_H,\ 1 \le t \le m, \\
x_1^\ell + \sum_{j=t-m+1}^{t} a_j^{\ell,\max} - x_{\max}^\ell \le \sum_{j=1}^{t} dev_j^\ell + r_j^{\ell,0} + \sum_{k=1}^{t-m} \alpha_{t,1}^{\ell,k} d_k^{\min} + \beta_{t,1}^{\ell,k} d_k^{\max} + \alpha_{t,2}^{\ell,k} a_k^{\ell,\min} + \beta_{t,2}^{\ell,k} a_k^{\ell,\max},\ \ell \in \mathcal{L}_H, t > m, \\
\sum_{j=1}^{t} dev_j^\ell + r_j^{\ell,0} + \sum_{k=1}^{t-m} \alpha_{t,1}^{\ell,k} d_k^{\max} + \beta_{t,1}^{\ell,k} d_k^{\min} + \alpha_{t,2}^{\ell,k} a_k^{\ell,\max} + \beta_{t,2}^{\ell,k} a_k^{\ell,\min} \le x_1^\ell + \sum_{j=t-m+1}^{t} a_j^{\ell,\min} - x_{\min}^\ell,\ \ell \in \mathcal{L}_H, t > m, \\
\delta_{t,1}^{\ell,k} + \hat{\delta}_{t,1}^{\ell,k} = r_{t,1}^{\ell,k},\ \delta_{t,1}^{\ell,k} \ge 0,\ \delta_{t,1}^{\ell,k} \ge r_{t,1}^{\ell,k},\ \hat{\delta}_{t,1}^{\ell,k} \le 0,\ \hat{\delta}_{t,1}^{\ell,k} \le r_{t,1}^{\ell,k},\ t \ge m+1,\ k \in I_t,\ \ell \in \mathcal{L}_H, \\
\delta_{t,2}^{\ell,k} + \hat{\delta}_{t,2}^{\ell,k} = r_{t,2}^{\ell,k},\ \delta_{t,2}^{\ell,k} \ge 0,\ \delta_{t,2}^{\ell,k} \ge r_{t,2}^{\ell,k},\ \hat{\delta}_{t,2}^{\ell,k} \le 0,\ \hat{\delta}_{t,2}^{\ell,k} \le r_{t,2}^{\ell,k},\ t \ge m+1,\ k \in I_t,\ \ell \in \mathcal{L}_H, \\
0 \le r_t^{\ell,0} + \sum_{k\in I_t}(\delta_{t,1}^{\ell,k} d_k^{\min} + \hat{\delta}_{t,1}^{\ell,k} d_k^{\max} + \delta_{t,2}^{\ell,k} a_k^{\ell,\min} + \hat{\delta}_{t,2}^{\ell,k} a_k^{\ell,\max}),\ \ 1 \le t \le T,\ \ell \in \mathcal{L}_H, \\
r_t^{\ell,0} + \sum_{k\in I_t}(\delta_{t,1}^{\ell,k} d_k^{\max} + \hat{\delta}_{t,1}^{\ell,k} d_k^{\min} + \delta_{t,2}^{\ell,k} a_k^{\ell,\max} + \hat{\delta}_{t,2}^{\ell,k} a_k^{\ell,\min}) \le Duree(t) P_{\max}^\ell,\ \ 1 \le t \le T,\ \ell \in \mathcal{L}_H, \\
\rho_{t,1}^{\ell,k} + \rho_{t,2}^{\ell,k} = q_t^{\ell,k},\ \rho_{t,1}^{\ell,k} \ge 0,\ \rho_{t,2}^{\ell,k} \le q_t^{\ell,k},\ \rho_{t,2}^{\ell,k} \le 0,\ \rho_{t,1}^{\ell,k} \ge q_t^{\ell,k},\ t \ge m+1,\ k \in I_t,\ \ell \in \mathcal{L}_T, \\
0 \le q_t^{\ell,0} + \sum_{k\in I_t}(\rho_{t,1}^{\ell,k} d_k^{\min} + \rho_{t,2}^{\ell,k} d_k^{\max}),\ \ \ell \in \mathcal{L}_T, 1 \le t \le T, \\
q_t^{\ell,0} + \sum_{k\in I_t}(\rho_{t,1}^{\ell,k} d_k^{\max} + \rho_{t,2}^{\ell,k} d_k^{\min}) \le Duree(t)\, \tau_t^{\ell,\min} P_{\max}^\ell,\ \ \ell \in \mathcal{L}_T, 1 \le t \le T, \\
\nu_{t,1}^k + \nu_{t,2}^k = \xi_t^k,\ \nu_{t,1}^k \ge 0,\ \nu_{t,1}^k \ge \xi_t^k,\ \nu_{t,2}^k \le 0,\ \nu_{t,2}^k \le \xi_t^k,\ t \ge m+1,\ k \in I_t, \\
b_\ell + c_J^{k,\ell}(\sum_{t=1}^{T} p_t^{\ell,0} + \sum_{t=1}^{T-m} g_{t,1}^\ell\, d_t^{\max} + g_{t,2}^\ell\, d_t^{\min}) \le c_{J,k}^\ell x_1^\ell + d_{J,k}^\ell,\ \ell \in \mathcal{L}_J, k \in M_J^\ell, \\
g_{k,1}^\ell \ge 0,\ g_{k,1}^\ell \ge f_k^\ell,\ g_{k,2}^\ell \le 0,\ g_{k,2}^\ell \le f_k^\ell,\ g_{k,1}^\ell + g_{k,2}^\ell = f_k^\ell,\ \ \ell \in \mathcal{L}_J, 1 \le k \le T-m, \\
\sum_{t=1}^{T} p_t^{\ell,0} + \sum_{t=1}^{T-m} g_{t,1}^\ell\, d_t^{\max} + g_{t,2}^\ell\, d_t^{\min} \le x_1^\ell,\ \ell \in \mathcal{L}_J, \\
h_{t,1}^{\ell,k} \ge 0,\ h_{t,1}^{\ell,k} \ge p_t^{\ell,k},\ h_{t,2}^{\ell,k} \le 0,\ h_{t,2}^{\ell,k} \le p_t^{\ell,k},\ h_{t,1}^{\ell,k} + h_{t,2}^{\ell,k} = p_t^{\ell,k},\quad \ell \in \mathcal{L}_J,\ t \ge m+1,\ k \in I_t, \\
0 \le p_t^{\ell,0} + \sum_{k\in I_t} h_{t,1}^{\ell,k} d_k^{\min} + h_{t,2}^{\ell,k} d_k^{\max},\ \ \ell \in \mathcal{L}_J,\ 1 \le t \le T, \\
p_t^{\ell,0} + \sum_{k\in I_t} h_{t,1}^{\ell,k} d_k^{\max} + h_{t,2}^{\ell,k} d_k^{\min} \le Duree(t) P_{\max}^\ell,\ \ell \in \mathcal{L}_J,\ \ 1 \le t \le T, \\
d_t^{\max} \le \sum_{\ell\in\mathcal{L}_J} p_t^{\ell,0} + \sum_{\ell\in\mathcal{L}_T} q_t^{\ell,0} + \sum_{\ell\in\mathcal{L}_H} r_t^{\ell,0} + \sum_{k\in I_t} \nu_{t,1}^k\, d_k^{\min} + \nu_{t,2}^k\, d_k^{\max} + \sum_{\ell\in\mathcal{L}_H}\sum_{k\in I_t} \delta_{t,2}^{\ell,k}\, a_k^{\ell,\min} + \hat{\delta}_{t,2}^{\ell,k}\, a_k^{\ell,\max},\ \ 1 \le t \le T, \\
0 \le dev_t^\ell,\ \ \ell \in \mathcal{L}_H,\ 1 \le t \le T.
\end{array}
\right. \tag{4.52}
$$

CRA et CRAA. Incertitude polytopique

La CRA introduite en section 4.4.2 est le problème linéaire suivant :

$$
\left\{
\begin{array}{l}
\min \ F \\
\displaystyle\sum_{\ell\in\mathcal{L}_T} c_\ell\,(u_1^\ell + \sum_{t=2}^{T} u_t^\ell(s)) - \sum_{\ell\in\mathcal{L}_H} a_\ell(s) - \sum_{\ell\in\mathcal{L}_J} b_\ell(s) \le F,\ s\in S', \\
\displaystyle a_\ell(s) + c_{H,k}^\ell\,(v_1^\ell + dev_1^\ell(s) + \sum_{t=2}^{T}(v_t^\ell(s) + dev_t^\ell(s)) \le c_{H,k}^\ell\,(x_1^\ell + \sum_{t=1}^{T} a_t^\ell(s)) + d_{H,k}^\ell,\ \ \ell\in\mathcal{L}_H, 0\le k\le m_H^\ell - 1,\ s\in S', \\
\displaystyle b_\ell(s) + c_{J,k}^\ell\,(u_1^\ell + \sum_{t=2}^{T} u_t^\ell(s)) \le c_{J,k}^\ell x_1^\ell + d_{J,k}^\ell,\ \ \ell\in\mathcal{L}_J, 0\le k\le m_J^\ell - 1,\ s\in S', \\
\displaystyle x_1^\ell + \sum_{k=1}^{t} a_k^\ell(s) - x_{\max}^\ell \le v_1^\ell + dev_1^\ell(s) + \sum_{k=2}^{t}(v_k^\ell(s) + dev_k^\ell(s)) \le x_1^\ell + \sum_{k=1}^{t} a_k^\ell(s) - x_{\min}^\ell,\ \ \ell\in\mathcal{L}_H, 1\le t\le T,\ s\in S', \\
0 \le v_1^\ell \le Duree(1)\,P_{\max}^\ell,\ \ 0\le v_t^\ell(s)\le Duree(t)\,P_{\max}^\ell,\ \ \ell\in\mathcal{L}_H, 2\le t\le T,\ s\in S', \\
0 \le dev_t^\ell(s),\ \ \ell\in\mathcal{L}_H, 1\le t\le T, s\in S', \\
0 \le u_1^\ell \le Duree(1)\,\tau_1^\ell\,P_{\max}^\ell,\ 0\le u_t^\ell(s)\le Duree(t)\,\tau_t^\ell\,P_{\max}^\ell,\ \ \ell\in\mathcal{L}_T, 2\le t\le T,\ s\in S', \\
\displaystyle 0 \le x_1^\ell - u_1^\ell - \sum_{k=2}^{t} u_k^\ell(s),\ \ \ell\in\mathcal{L}_J, 2\le t\le T, s\in S', \\
\displaystyle d_t(s) \le \sum_{\ell\in\mathcal{L}_T} u_t^\ell(s) + \sum_{\ell\in\mathcal{L}_H} v_t^\ell(s) + \sum_{\ell\in\mathcal{L}_J} u_t^\ell(s),\ \ 2\le t\le T,\ s\in S', \\
\displaystyle \max_{s=1,\dots,S} d_1(s) \le \sum_{\ell\in\mathcal{L}_T} u_1^\ell + \sum_{\ell\in\mathcal{L}_H} v_1^\ell + \sum_{\ell\in\mathcal{L}_J} u_1^\ell.
\end{array}
\right. \tag{4.53}
$$

La CRAA introduite en section 4.4.3.1 est le problème linéaire suivant :

$$
\left\{
\begin{array}{l}
\min \ F \\
\displaystyle\sum_{t=1}^{T}\sum_{\ell\in\mathcal{L}_T} c_\ell\,q_t^{\ell,0} + \sum_{t=1}^{T}\sum_{\ell\in\mathcal{L}_T}\sum_{j\in I_t} c_\ell\,q_t^{\ell,j}\,d_j(s) - \sum_{\ell\in\mathcal{L}_H} a_\ell - \sum_{\ell\in\mathcal{L}_J} b_\ell \le F,\ s\in S', \\
\displaystyle a_\ell + c_{H,k}^\ell\left(\sum_{t=1}^{T}(r_t^{\ell,0} + dev_t^\ell + \sum_{j\in I_t}(r_{t,1}^{\ell,j}\,d_j(s) + r_{t,2}^{\ell,j}\,a_j^\ell(s)))\right) \le c_{H,k}^\ell\,(x_1^\ell + \sum_{t=1}^{T} a_t^\ell(s)) + d_{H,k}^\ell,\ \ell\in\mathcal{L}_H,\ k\in M_H^\ell,\ s\in S', \\
\displaystyle b_\ell + c_{J,k}^\ell\left(\sum_{t=1}^{T} p_t^{\ell,0} + \sum_{t=1}^{T-m}(\sum_{j\,|\,t\in I_j} p_j^{\ell,t})\,d_t(s)\right) \le c_{J,k}^\ell x_1^\ell + d_{J,k}^\ell,\ \ell\in\mathcal{L}_J,\ k\in M_J^\ell,\ s\in S', \\
\displaystyle x_1^\ell + \sum_{j=1}^{t} a_j^\ell(s) - x_{\max}^\ell \le \sum_{j=1}^{t}(r_j^{\ell,0} + dev_j^\ell) + \sum_{j=1}^{t-m}\sum_{k\le t\,|\,j\in I_k}(r_{k,1}^{\ell,j}\,d_j(s) + r_{k,2}^{\ell,j}\,a_j^\ell(s)),\ \ell\in\mathcal{L}_H,\ 1\le t\le T,\ s\in S', \\
\displaystyle \sum_{j=1}^{t}(r_j^{\ell,0} + dev_j^\ell) + \sum_{j=1}^{t-m}\sum_{k\le t\,|\,j\in I_k}(r_{k,1}^{\ell,j}\,d_j(s) + r_{k,2}^{\ell,j}\,a_j^\ell(s)) \le x_1^\ell + \sum_{j=1}^{t} a_j^\ell(s) - x_{\min}^\ell,\ \ell\in\mathcal{L}_H,\ 1\le t\le T,\ s\in S', \\
\displaystyle 0 \le p_t^{\ell,0} + \sum_{k\in I_t} p_t^{\ell,k}\,d_k(s) \le Duree(t)\,P_{\max}^\ell,\ \ell\in\mathcal{L}_J,\ 1\le t\le T,\ s\in S', \\
\displaystyle 0 \le q_t^{\ell,0} + \sum_{k\in I_t} q_t^{\ell,k}\,d_k(s) \le Duree(t)\,P_{\max}^\ell\,\tau_t^\ell(s),\ \ell\in\mathcal{L}_T,\ 1\le t\le T,\ s\in S', \\
\displaystyle 0 \le r_t^{\ell,0} + \sum_{k\in I_t}(r_{t,1}^{\ell,k}\,d_k(s) + r_{t,2}^{\ell,k}\,a_k^\ell(s)) \le Duree(t)\,P_{\max}^\ell,\ \ \ell\in\mathcal{L}_H,\ 1\le t\le T,\ s\in S', \\
\displaystyle \sum_{t=1}^{T} p_t^{\ell,0} + \sum_{t=1}^{T-m}\sum_{k\,|\,t\in I_k} p_k^{\ell,t}\,d_t(s) \le x_1^\ell, \ell\in\mathcal{L}_J, s\in S', \\
\displaystyle 0 \le dev_t^\ell,\ \ \ell\in\mathcal{L}_H, 1\le t\le T, \\
\displaystyle d_t(s) \le \sum_{\ell\in\mathcal{L}_J} p_t^{\ell,0} + \sum_{\ell\in\mathcal{L}_T} q_t^{\ell,0} + \sum_{\ell\in\mathcal{L}_H} r_t^{\ell,0} + \sum_{k\in I_t}\left((\sum_{\ell\in\mathcal{L}_J} p_t^{\ell,k} + \sum_{\ell\in\mathcal{L}_T} q_t^{\ell,k} + \sum_{\ell\in\mathcal{L}_H} r_{t,1}^{\ell,k})d_k(s) + (\sum_{\ell\in\mathcal{L}_H} r_{t,2}^{\ell,k})a_k^\ell(s)\right),\ \forall t,s.
\end{array}
\right. \tag{4.54}
$$

CRAA-Incertitude ellipsoïdale

La CRAA introduite en section 4.4.3.3 est le problème conique suivant :

$$
\left\{
\begin{array}{l}
\min \ F \\
\sum_{\ell\in\mathcal{L}_T}\sum_{t=1}^{T} c_\ell q_t^{\ell,0} + \sum_{t=1}^{T-m} \bar{\xi}_t \bar{d}_t + \kappa\sqrt{\bar{\xi}^T\Gamma_{T-m}\bar{\xi}} - \sum_{\ell\in\mathcal{L}_H} a_\ell - \sum_{\ell\in\mathcal{L}_J} b_\ell \le F, \\
\bar{\xi}_k = \sum_{\ell\in\mathcal{L}_T}\sum_{t|k\in I_t} c_\ell \ q_t^{\ell,k}, \ \ k=1,\ldots,T-m, \\
\alpha_{t,2}^{\ell,k}\ge 0,\ \beta_{t,2}^{\ell,k}\le 0,\ \alpha_{t,2}^{\ell,k}\ge\gamma_{t,2}^{\ell,k}-1,\ \beta_{t,2}^{\ell,k}\le\gamma_{t,2}^{\ell,k}-1,\ \alpha_{t,2}^{\ell,k}+\beta_{t,2}^{\ell,k}=\gamma_{t,2}^{\ell,k}-1,\ t>m,\ k\in I_t, \ell\in\mathcal{L}_H, \\
a_\ell + c_{H,k}^\ell\left(\sum_{t=1}^{T}(dev_t^\ell + r_t^{\ell,0}) + \sum_{t=1}^{T-m}(\gamma_{T,1}^{\ell,t}\bar{d}_t + \alpha_{T,2}^{\ell,t}a_t^{\ell,\max} + \beta_{T,2}^{\ell,t}a_t^{\ell,\min})\right) \\
\quad +\kappa\sqrt{(\gamma_{T,1}^\ell)^T\Gamma_{T-m}\gamma_{T,1}^\ell}\le c_{H,k}^\ell(x_1^\ell + \sum_{t=T-m+1}^{T} a_t^{\ell,\min}) + d_{H,k}^\ell, \ \ \ell\in\mathcal{L}_H,\ k\in M_H^\ell, \\
x_1^\ell + \sum_{j=t-m+1}^{t} a_j^{\ell,\max} - x_{\max}^\ell \le \sum_{j=1}^{t} dev_j^\ell + r_j^{\ell,0} + \sum_{k=1}^{t-m}\gamma_{t,1}^{\ell,k}\bar{d}_k + \alpha_{t,2}^{\ell,k}a_k^{\ell,\min} + \beta_{t,2}^{\ell,k}a_k^{\ell,\max} \\
\quad -\kappa\sqrt{(\gamma_{t,1}^\ell)^T\Gamma_{t-m}\gamma_{t,1}^\ell},\ \ell\in\mathcal{L}_H, t\ge m+1, \\
\sum_{j=1}^{t} dev_j^\ell + r_j^{\ell,0} + \sum_{k=1}^{t-m}\gamma_{t,1}^{\ell,k}\bar{d}_k + \alpha_{t,2}^{\ell,k}a_k^{\ell,\max} + \beta_{t,2}^{\ell,k}a_k^{\ell,\min} + \kappa\sqrt{(\gamma_{t,1}^\ell)^T\Gamma_{t-m}\gamma_{t,1}^\ell} \\
\quad \le x_1^\ell + \sum_{j=t-m+1}^{t} a_j^{\ell,\min} - x_{\min}^\ell,\ \ell\in\mathcal{L}_H, t\ge m+1, \\
x_1^\ell - x_{\max}^\ell + \sum_{j=1}^{t} a_j^{\ell,\max} \le \sum_{j=1}^{t} r_j^{\ell,0} + dev_j^\ell \le x_1^\ell - x_{\min}^\ell + \sum_{j=1}^{t} a_j^{\ell,\min}, \ \ \ell\in\mathcal{L}_H, 1\le t\le m, \\
\delta_{t,2}^{\ell,k} + \hat{\delta}_{t,2}^{\ell,k} = r_{t,2}^{\ell,k},\ \delta_{t,2}^{\ell,k}\ge 0,\ \delta_{t,2}^{\ell,k}\ge r_{t,2}^{\ell,k},\ \hat{\delta}_{t,2}^{\ell,k}\le 0,\ \hat{\delta}_{t,2}^{\ell,k}\le r_{t,2}^{\ell,k},\ t\ge m+1,\ k\in I_t,\ \ell\in\mathcal{L}_H, \\
0\le r_t^{\ell,0} + \sum_{k\in I_t} r_{t,1}^{\ell,k}\bar{d}_k - \kappa\sqrt{(r_{t,1}^\ell)^T\Gamma_{t-m}r_{t,1}^\ell} + \sum_{k\in I_t}\delta_{t,2}^{\ell,k}a_k^{\ell,\min} + \hat{\delta}_{t,2}^{\ell,k}a_k^{\ell,\max},\ \ell\in\mathcal{L}_H, m<t, \\
r_t^{\ell,0} + \sum_{k\in I_t} r_{t,1}^{\ell,k}\bar{d}_k + \kappa\sqrt{(r_{t,1}^\ell)^T\Gamma_{t-m}r_{t,1}^\ell} + \sum_{k\in I_t}\delta_{t,2}^{\ell,k}a_k^{\ell,\max} + \hat{\delta}_{t,2}^{\ell,k}a_k^{\ell,\min} \\
\quad \le Duree(t)P_{\max}^\ell,\ \ell\in\mathcal{L}_H,\ m+1\le t, \\
0\le r_t^{\ell,0}\le Duree(t)P_{\max}^\ell, \ \ \ell\in\mathcal{L}_H,\ 1\le t\le m, \\
0\le q_t^{\ell,0} + \sum_{k\in I_t} q_t^{\ell,k}\bar{d}_k - \kappa\sqrt{(q_t^\ell)^T\Gamma_{t-m}q_t^\ell},\ \ell\in\mathcal{L}_T,\ t>m, \\
q_t^{\ell,0} + \sum_{k\in I_t} q_t^{\ell,k}\bar{d}_k + \kappa\sqrt{(q_t^\ell)^T\Gamma_{t-m}q_t^\ell} \le Duree(t)\tau_t^{\ell,\min}P_{\max}^\ell,\ \ell\in\mathcal{L}_T,\ t>m, \\
0\le q_t^{\ell,0}\le Duree(t)\,\tau_t^{\ell,\min}P_{\max}^\ell, \ \ \ell\in\mathcal{L}_T,\ 1\le t\le m, \\
b_\ell + c_{J,k}^\ell\sum_{t=1}^{T} u_t^\ell \le cx_1^\ell + d_{H,k}^\ell, \ \ \ell\in\mathcal{L}_J,\ k\in M_J^\ell, \\
\forall\ell\in\mathcal{L}_J,\ 0\le u_t^\ell\le Duree(t)P_{\max}^\ell,\ t=1,\ldots,T, \quad \sum_{t=1}^{T} u_t^\ell \le x_1^\ell, \\
\forall m+1\le t\le T, \ \ x_t(k)=\xi_t^k \text{ si } k\in I_t,\ x_t(t)=-1, \text{ et } x_t(k)=0, \text{ si } k<t \text{ et } k\notin I_t, \\
\bar{d}_t \le \sum_{\ell\in\mathcal{L}_J} p_t^{\ell,0} + \sum_{\ell\in\mathcal{L}_T} q_t^{\ell,0} + \sum_{\ell\in\mathcal{L}_H} r_t^{\ell,0} + \sum_{k\in I_t}\xi_t^k\bar{d}_k - \kappa\sqrt{x_t^T\Gamma_t x_t} \\
\quad + \sum_{\ell\in\mathcal{L}_H}\sum_{k\in I_t}\delta_{t,2}^{\ell,k}a_k^{\ell,\min} + \hat{\delta}_{t,2}^{\ell,k}a_k^{\ell,\max},\ t\ge m+1, \\
\bar{d}_t \le \sum_{\ell\in\mathcal{L}_J} p_t^{\ell,0} + \sum_{\ell\in\mathcal{L}_T} q_t^{\ell,0} + \sum_{\ell\in\mathcal{L}_H} r_t^{\ell,0}, t\le m.
\end{array}
\right. \tag{4.55}
$$

Les notations suivantes sont utilisées pour présenter les CRA et CRAA de ce chapitre. On notera $M_H^\ell = \{0, \ldots, m_H^\ell - 1\}$ et $M_J^\ell = \{0, \ldots, m_J^\ell - 1\}$. Nous introduisons, par ailleurs pour tout $t \geq m+1$ et pour tout $k \in I_t$, les quantités suivantes :

$$\begin{array}{ll} \xi'_j = \sum_{t \,|\, j \in I_t} \sum_{\ell \in \mathcal{L}_T} c_\ell \, q_t^{\ell,j}, 1 \leq j \leq T-m, & \xi_t^k = \sum_{\ell \in \mathcal{L}_J} p_t^{\ell,k} + \sum_{\ell \in \mathcal{L}_T} q_t^{\ell,k} + \sum_{\ell \in \mathcal{L}_H} r_{t,1}^{\ell,k}, \\ \gamma_{t,1}^{\ell,k} = \sum_{j \leq t \,|\, k \in I_j} r_{j,1}^{\ell,k}, \ \ell \in \mathcal{L}_H, & \gamma_{t,2}^{\ell,k} = \sum_{j \leq t \,|\, k \in I_j} r_{j,2}^{\ell,k}, \ \ell \in \mathcal{L}_H; \end{array} \tag{4.56}$$

et enfin $f_k^\ell = \sum_{j \,|\, k \in I_j} p_j^{\ell,k}$, pour $\ell \in \mathcal{L}_J$ et $1 \leq k \leq T-m$.

On note $\gamma_{t,1}^\ell$ (resp. $\gamma_{t,2}^\ell$) le vecteur dont les composantes sont $\gamma_{t,1}^{\ell,r}$, $r \in I_t$ (resp. $\gamma_{t,2}^{\ell,r}$, $r \in I_t$) et q_t^ℓ (resp. $r_{t,1}^\ell$) le vecteur dont les composantes sont $q_t^{\ell,r}$, $r \in I_t$ (resp. $r_{t,1}^{\ell,j}$, $j \in I_t$). Enfin, pour $1 \leq t \leq T$, la sous-matrice $\Gamma(1:t, 1:t)$ de la matrice Γ est notée Γ_t.

On peut remarquer que pour le problème (4.52), les variables $h_{t,1}^{\ell,k}$, $h_{t,2}^{\ell,k}$, $g_{k,1}^\ell$, $g_{k,2}^\ell$, $\rho_{t,1}^{\ell,k}$, $\rho_{t,2}^{\ell,k}$, $\alpha_{t,1}^{\ell,k}$, $\alpha_{t,2}^{\ell,k}$, $\beta_{t,1}^{\ell,k}$, $\beta_{t,2}^{\ell,k}$, $\delta_{t,1}^{\ell,k}$, $\delta_{t,2}^{\ell,k}$, $\hat{\delta}_{t,1}^{\ell,k}$, $\hat{\delta}_{t,2}^{\ell,k}$ sont des variables slack et on aura, pour la solution optimale :

$$\begin{array}{llll} h_{t,1}^{\ell,k} = \max(p_t^{\ell,k}, 0), & h_{t,2}^{\ell,k} = \min(p_t^{\ell,k}, 0), & g_{k,1}^\ell = \max(f_k^\ell, 0), & g_{k,2}^\ell = \min(f_k^\ell, 0), \\ \rho_{t,1}^{\ell,k} = \max(q_t^{\ell,k}, 0), & \rho_{t,2}^{\ell,k} = \min(q_t^{\ell,k}, 0), & \alpha_{t,1}^{\ell,k} = \max(\gamma_{t,1}^{\ell,k}, 0), & \alpha_{t,2}^{\ell,k} = \max(\gamma_{t,2}^{\ell,k} - 1, 0), \\ \beta_{t,1}^{\ell,k} = \min(\gamma_{t,1}^{\ell,k}, 0), & \beta_{t,2}^{\ell,k} = \min(\gamma_{t,2}^{\ell,k} - 1, 0), & \delta_{t,1}^{\ell,k} = \max(r_{t,1}^{\ell,k}, 0), & \delta_{t,2}^{\ell,k} = \max(r_{t,2}^{\ell,k}, 0), \\ \hat{\delta}_{t,1}^{\ell,k} = \min(r_{t,1}^{\ell,k}, 0), & \hat{\delta}_{t,2}^{\ell,k} = \min(r_{t,2}^{\ell,k}, 0). & & \end{array}$$

Conclusion générale et perspectives

Cette thèse est consacrée à l'étude de problèmes d'optimisation stochastique. Dans ce mémoire, nous avons examiné plusieurs problèmes liés à la modélisation, l'analyse et la résolution de problèmes d'optimisation stochastique. Les contributions dans chaque domaine sont les suivantes.

Dans le premier chapitre, nous avons étendu certains résultats de [11] pour déterminer les contreparties robustes ajustées de deux classes de problèmes.

Dans le deuxième chapitre, nous avons proposé un algorithme adaptatif afin de construire des estimateurs de la moyenne et de la matrice de covariance d'un processus à un instant donné sous l'hypothèse d'homogénéité temporelle locale. Les qualités théoriques de ces estimateurs ont été étudiées. D'autre part, cet algorithme adaptatif a été utilisé pour un modèle à saut (dans la moyenne ou la matrice de covariance) afin de déterminer un estimateur de l'instant de la rupture. Dans ce contexte, des bornes ont été données sur les erreurs de première et de deuxième espèce pour tester si un intervalle donné est d'homogénéité ou non.

Dans un modèle à saut, afin de déterminer l'instant de la rupture, un autre algorithme non paramétrique a été étudié d'un point de vue théorique et numérique dans la section 2.4.

Les estimations adaptatives des paramètres des modèles VaR et de Markowitz sont ensuite utilisées pour proposer des mises en oeuvre de ces modèles dont on contrôle la qualité. Les qualités des estimations adaptatives et des mises en oeuvre des modèles proposées sont mesurées en utilisant des données réelles et simulées.

La chapitre 3 propose une analyse de sensibilité du modèle de Markowitz et d'une classe particulière de problèmes d'optimisation quadratiques. Cette étude, combinée aux résultats du chapitre 2, motive deux calibrations de la matrice de covariance que nous proposons dans la section 3.4 dont le but est de stabiliser la composition des portefeuilles de Markowitz par rapport à des perturbations des paramètres du modèle. L'une de ces calibrations est testée en détail en utilisant des données réelles et simulées. Nous proposons également dans le chapitre 3 des contreparties robustes simples du modèle de Markowitz utilisant les résultats du chapitre 2.

Dans la deuxième partie de cette thèse (le chapitre 4), nous étudions deux problèmes de gestion de production : le problème de gestion de production électrique et le problème du transport et de la distribution du gaz. Pour ce dernier problème, nous appliquons la méthodologie de la CRA et de la CRAA. Pour le problème de gestion de production élec-

trique, divers aspects ont été examinés.

Après avoir présenté la modélisation des unités de production, on explique comment le problème est en général traité par programmation stochastique multi-étapes.

Du point de vue de la modélisation, diverses méthodes sont ensuite proposées afin d'assurer un meilleur compromis coût moyen/écart type des coûts sur un ensemble de scénarios tests. Un premier groupe de méthodes s'attache à régulariser la fonction duale obtenue en dualisant les contraintes couplantes du problème. Un deuxième groupe de méthodes repose sur l'utilisation des CRA et des CRAA.

D'un point de vue algorithmique, nous montrons ensuite comment le problème de gestion de production électrique peut dans certains cas se résoudre directement par programmation dynamique. Cette étude est utilisée pour montrer comment on peut résoudre les problèmes hydrauliques et EJP duaux de façon efficace.

Les différentes méthodes de gestion sont ensuite testées sur des données simulées puis sur des données réelles. À cette fin, on présente dans la section 4.6 une étude détaillée de la série des consommations électriques en France de 1996 à 2003.

Ces travaux offrent un certain nombre de perspectives.

Le premier chapitre présente un certain nombre d'équivalents déterministes de certaines contreparties robustes ajustées et ajustées affinement. Il serait intéressant d'étudier les contreparties robustes ajustées de problèmes d'optimisation coniques et semidéfinis positifs.

Nous sommes en train d'utiliser et généraliser les résultats du chapitre 2 dans deux articles en préparation ([56] et [57]). Il s'agit de présenter une méthode de traitement de l'incertitude pour résoudre un problème d'optimisation stochastique du type $\min_{x\in X} f_0(x,\theta)$ sous les contraintes $f_j(x,\theta)\leq 0$. Le paramètre θ, incertain, a une forme particulière et les fonctions f_j satisfont $|f_j(x,\theta)-f_j(x,\theta')|\leq C_0\|\theta-\theta'\|_\infty^{\alpha_0}\,\|x\|_1^{p_0}$, avec $0<\alpha_0\leq 2$, $0<p_0\leq 2$ et $C_0<\infty$. Dans [57], on traite le cas ou l'incertitude intervient seulement dans la fonction objectif. Le cas général sera traité dans [56].

Le chapitre 2 permet de déterminer des estimateurs adaptatifs de la matrice de covariance et de la moyenne dont on contrôle la qualité dans un cadre multidimensionnel et non paramétrique. Le même travail pourrait être fait dans un cadre paramétrique. On pourrait alors sûrement obtenir des bornes plus fines. Dans le contexte financier, un cadre paramétrique d'intérêt pour la modélisation des rentabilités est le modèle log-normal.

Dans le chapitre 4, les CRA et CRAA glissantes testées sur des données simulées se sont révélées moins performantes que leur homologue non glissant. Ces méthodes devraient cependant fournir de meilleurs résultats en pratique, en utilisant des données réelles. En effet, les scénarios que nous avons utilisés pour tester les méthodes non glissantes sont ceux ayant servi à calibrer ces méthodes. Ainsi, les CRA et CRAA non glissantes ont été

mises en place au mieux en tenant compte de ces scénarios et sont performantes sur ces scénarios. D'un coté, les commandes utilisées par les CRA glissantes sont les commandes non ajustables de la CRA utilisant des horizons de plus en plus courts et on n'utilise pas les variables ajustables. D'un autre coté, d'un point de vue pratique, lorsque les méthodes glissantes sont utilisées, pour déterminer les commandes d'un pas de temps, on tient compte (contrairement au cas où les CRA et CRAA sont utilisées) de l'état réel du système. D'autre part, plus on avance dans l'année, moins on a d'intervalles de prévision à calculer et plus les informations sur les processus mis en jeu s'enrichissent.

La modélisation présentée en section 4.5.3 pourrait être robustifiée en remplaçant les fonctions $\phi_j(\tau)$ par :

$$\tilde{\phi}_j(\tau) = \min_{\alpha} [\alpha\tau + \max_{\nu \in \mathcal{P}} E_\nu[\psi_{M_j}(d_j - \alpha)] \ : \ 0 \leq \alpha \leq V_j],$$

où $\mathcal{P}$ est un ensemble de distributions de probabilité.

La recherche de stratégies de gestion assurant un bon compromis coût moyen/écart type des coûts reste à faire dans le cas plus général où l'on autorise la mise en route et l'arrêt de centrales sur la période de gestion.

Pour les problèmes d'optimisation stochastique, afin de tenir compte du fait que les prévisions que l'on peut faire pour les paramètres incertains sont d'autant plus précises et les ensembles d'incertitude sur ces prévisions (pour un niveau de confiance fixé) d'autant plus petits que l'horizon de prévision est faible, on peut choisir une discrétisation non homogène de l'horizon de gestion. On pourrait choisir une résolution fine (grain journalier) pour les k_1 premières variables, une résolution plus grossière pour les k_2 variables suivantes et ainsi de suite.
Nous pourrions en particulier appliquer cette technique au problème de gestion de production électrique étudié dans le chapitre 4. Cette technique permettrait de réduire de façon significative la dimension du problème de gestion de production électrique (4.12) correspondant et les CRA et CRAA présentées dans les sections 4.4.2 et 4.4.3 deviennent alors traitables en temps raisonnable même dans le cas où la contrepartie robuste ajustée affinement comprend des contraintes coniques. Une autre propriété intéressante de cette technique d'agrégation multi-résolution des paramètres incertains est la réduction de la taille des ensembles d'incertitudes. En effet, en supposant les demandes d_i aléatoires et de même variance, l'écart type de la demande agrégée $\sum_{i=t}^{i=t+k-1} d_i$ augmente seulement comme $\sqrt{k}$, (au lieu de k pour l'écart type de la demande d_i) alors que sa valeur moyenne croît comme k. Les variations de la demande agrégée autour de sa moyenne étant plus faible, les contreparties robustes ajustées sont moins conservatives.

Le travail restant à faire pour rendre l'approche par contrepartie robuste affinement ajustable avec agrégation d'état opérationnelle, est d'une part de définir des méthodes pour choisir la granularité de la partition et d'autre part d'estimer les ensembles d'incertitude des paramètres agrégés. Si les pas de temps de l'horizon de gestion sont partitionnés en $P_1 \cup \ldots \cup P_m = \{1, \ldots, T\}$, les variables de commande au pas temps agrégé k du problème d'optimisation obtenu après agrégation des états sont les sommes des variables de commandes pour les pas de temps élémentaires de l'ensemble P_k. D'autre part, au lieu d'avoir à étudier

le processus aléatoire $\xi = (d_t, (a_t^\ell)_{\ell \in \mathcal{L}_H}, (\tau_t^\ell)_{\ell \in \mathcal{L}_T})_{t=1}^T$ pour déterminer les ensembles d'incertitude, on doit étudier le processus $(\sum_{t \in P_k} d_t, ((\sum_{t \in P_k} a_t^\ell)_{\ell \in \mathcal{L}_H}, ((\sum_{t \in P_k} Duree(t)\, \tau_t^\ell)_{\ell \in \mathcal{L}_T})_{k=1}^m$. Les ensembles d'incertitude ne doivent pas être trop conservatifs mais pour autant permettre de déterminer des commandes pour chaque pas de temps élémentaire contenu dans les partitions.

Enfin, les méthodes que nous avons utilisées dans cette thèse pour résoudre le problème de gestion de production électrique (en particulier la régularisation de la fonction duale et les CRA glissantes) pourraient être utilement appliquées à d'autres problèmes d'optimisation stochastique.

Contributions logicielles

Les contributions logicielles suivantes ont été apportées par cette thèse.

- Un premier module en finance a été développé en Matlab (en utilisant les librairies mosek et SDPPack pour résoudre les problèmes d'optimisation). Ce module fournit une implémentation de l'algorithme adaptatif, les calibrations de la matrice de covariance présentées dans les sections 3.4.3 et 3.5 et diverses fonctions de tests nécessaires pour les simulations proposées dans les chapitres 2 et 3.
- Un module de gestion de production électrique codé en Scilab, C et Fortran permettant de résoudre le problème de gestion de production électrique sur un arbre de scénarios et d'effectuer une simulation de la stratégie de gestion qui en résulte sur un ensemble de scénarios de tests. Ce module propose aussi une implémentation des méthodes VaR_{CA} et VaR_{benef}. Une documentation technique détaillée a été écrite sur ce module. Ce module utilise mosek pour résoudre les problèmes linéaires et la méthode de faisceaux proxmyqn (en Fortran) [[65]].
- Un module, codé en C, utilisant l'API de mosek et la méthode de faisceaux proxmyqn [[65]], implémentant la méthode de stabilisation de la fonction duale présentée dans la section 4.4.1.
- Un dernier module de gestion de production électrique en Matlab et R, utilisant la librairie mosek (pour résoudre les problèmes linéaires et coniques) permettant d'une part d'étudier la série des consommations électriques en France de 1996 à 2003 et d'autre part de fournir des CRA et CRAA du problème de gestion de production électrique. Les CRAA proposées se servent d'ensembles d'incertitude polytopiques, d'ensembles définis par des contraintes de boîte ou d'ensembles d'incertitude ellipsoïdaux pour la demande. Diverses variantes de ces CRAA, utilisant différents schémas de dépendance des variables ajustables en fonction des paramètres incertains ont été envisagées.
- Un module en Matlab (utilisant la librairie mosek) pour résoudre le problème de gestion du transport et de la distribution du gaz ainsi que sa CRA.

Bibliographie

[1] Bacaud L., Lemaréchal C., Renaud A., Sagastizábal C. – Bundle methods in stochastic optimal power management :A disaggregated approach using preconditioners. *Computational Optimization and Applications 20, 227-244*, 2001.

[2] Batut J. et Renaud A. – Daily generation scheduling with transmission constraints :A new class of algorithms. *IEEE Transactions on Power Systems*, 1992.

[3] Beale E.M.L. – On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society, 17B, p.173-184*, 1955.

[4] Bellman R.R. – *Dynamic Programming.* – Princeton University Press, 1957.

[5] Ben-Tal A. et Nemirovski A. – Robust Optimization - Methodology and Applications. 1996.

[6] Ben-Tal A. et Nemirovski A. – Robust solutions to uncertain linear programs via Convex Programming. *Operations Research Letters*, 1996.

[7] Ben-Tal A. et Nemirovski A. – Robust Convex Optimization. *Mathematics of Operations Research*, 1997.

[8] Ben-Tal A. et Nemirovski A. – Robust Solutions of Linear Programming problems contamined with uncertain data. *Mathematical Programming*, 1997.

[9] Ben-Tal A. et Nemirovski A. – Robust Truss Topology Design via Semidefinite Programming. *SIAM J. Optim. 7 991-1016*, 1997.

[10] Ben-Tal A., Margelit T. et Nemirovski A. – Robust Modeling of Multi-Stage Portfolio Problems. *In [45], p.303-328*, 2000.

[11] Ben-Tal A., Nemirovski A. – Adjustable robust counterpart of uncertain linear programs. 2003.

[12] Bertsekas D.P. – *Dynamic Programming and Optimal Control, Volumes 1 et 2.* – Athena Scientific, Belmont, Massachusetts, 1995.

[13] Bertsimas D. et Popescu I. – Optimal inequalities in probability theory : a convex optimization approach, MIT/INSEAD working paper, available at http ://faculty.insead.edu/popescu/ionna. *Submitted to Operations Research*, 2000.

[14] Best M.J. et Grauer R.R. – On the Sensitivity of Mean-Variance-Efficient Portfolios to changes in asset means : some analytical and computational results. *Management Science*, 1991.

[15] Best M.J. et Grauer R.R. – Sensitivity Analysis for Mean-Variance portfolio problems. *Management Science*, 1991.

[16] Bevan E. et Winkelmann K. – Using the Black-Litterman Global Asset Allocation Model : Three Years of Practical Experience. *Fixed Income Research*, 1998.

[17] Birge John R. et Louveaux F. – *Introduction to Stochastic Programming.* – Springer Series in Operations Research, 1997.

[18] Bonnans J.F. et Shapiro A. – *Perturbation Analysis of Optimization Problems.* – Springer, 2000, *Springer Series in Operations Research.*

[19] Bonnans J.F., Gilbert J.C., Lemaréchal C. et Sagastizábal C. – *Optimisation Numérique, aspects théoriques et pratiques.* – Springer, 1997, *Mathématiques & Applications.*

[20] Box, G.E.P, Jenkins G.M et Reinsel G.C. – *Time Series Analysis, Forecasting and Control.* – Holden-Day, San Francisco, 1976.

[21] Brignol S. et Renaud A. – A new model for stochastic optimization of weekly generation schedule. *APSCOM-97 Proc,Hong-Kong*, 1997.

[22] Brignol S. et Rippault G. – Risk management applied to weekly generation scheduling. *IEEE Winter Meeting Proceedings*, 1999.

[23] Broens D.-F. et Haneveld W. K. – Investment evaluation based on the commercial scope : The production of natural gas. *Annals of Operations Research*, 1995.

[24] Calafiore, G. – Approximation of n-dimensional Data Using Spherical and Ellipsoidal Primitives. *IEEE Transactions on Systems, Man, and Cybernetics, vol. 32, n 2, pp. 269-278*, 2002.

[25] Caroe C.C, Schultz R. – A Two-Stage Stochastic Program for Unit Commitment Under Uncertainty in a Hydro-Thermal Power System. *Preprint 98-13*, 1998.

[26] Carpentier P., Cohen G. et Culioli J.C. – Stochastic optimal control and decomposition-coordination methods. Part I : theory. *Seventh French-German Conference on Optimization, Dijon*, 1994.

[27] Charnes A. et Cooper W. – Chance-Constrained Programming. *Management Science*, 1959.

[28] Charnes A. et Cooper, W. – Chance-Constraints et Normal Derivatives. *J. Am. Statist. Assoc*, 1962.

[29] Charnes A. et Cooper W. – Deterministic Equivalents for Optimizing and Satisficing under Chance-Constraints. *Operation Research*, 1963.

[30] Chiralaksanakul A. – Monte Carlo Methods for Multi-stage Stochastic Programs-Thèse. 2003.

[31] Conejo A.J. et Redondo N.J. – Short-term hydro-thermal coordination by Lagrangian relaxation :solution of the dual problem. *IEEE Transactions on Power Systems*, 1998.

[32] Daniel J. W. – Stability of the solutions of definite quadratic programs. *Mathematical Programming*, 1973.

[33] Dantzig G.B. – Linear programming under uncertainty. *Management science,1,p.197-206*, 1955.

[34] Dantzig G.B. et Infanger G. – Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research,Vol 45,59-76*, 1993.

[35] Dentcheva D. et Römisch W. – Optimal power generation under uncertainty via stochastic programming. *Economics and Mathematical Systems*, 1998.

[36] Dupačová J. – Stochastic programming : Approximation via scenarios. 1996.

[37] Dupačová J., Gröwe-Kuska N., Römisch W. – Scenario reduction in stochastic programming : An approach using probability metrics. *"Preprint 00-9, Institut für Mathematik, Humboldt-Universitat Berlin"*, 2000.

[38] El Ghaoui L. et Ben-Tal A. et Nemirovski A. – Robust Semidefinite Programming. *In [82]*, 2000.

[39] El Ghaoui L., Oustry F. et Oks M. – Worst-Case Value-at-Risk and Robust Asset Allocation : a Semidefinite Programming Approach. *Operations Research*, 2004.

[40] El-Ghaoui L.,Oustry F.,Lebret H. – Robust solutions to uncertain semidefinite programs. *SIAM J. on Optimization 9,33-52*, 1998.

[41] Ermoliev Y. et Wets R. J.-B. – *Numerical techniques for Stochastic Optmization Problems.* – Springer, Berlin, 1988.

[42] Escudero L. F. et al. – Production planning via scenario modeling. *Annals of Operations Research. 42, 311-336*, 1993.

[43] Falk J. E. – Exact Solutions of Inexact Linear Programs. 1975.

[44] Frauendorfer K. – Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming,75 :277-293*, 1996.

[45] Frenk H., Roos C., Terlaky T., Zhang S. – *High Performance Optimization.* – Kluwer Academic Publishers, 2000.

[46] Gold et Iyengar. – Robust portfolio selection problems. *Mathematics of Operations Research*, 2001.

[47] Gourieroux C. et Monfort A. – *Séries temporelles et modèles dynamiques.* – Economica, 1995.

[48] Gröwe-Kuska N. et Römisch W. – Stochastic unit commitment in hydro-thermal power production planning. 1999.

[49] Gröwe-Kuska N., Kiwiel K.C., Nowak M.P., Römisch W., Wegner I. – Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. *Decision Making under Uncertainty : Energy and Power (Greengard, C., Ruszczyński, A. eds.), IMA Volumes in Mathematics and its Applications Vol. 128, Springer-Verlag, New York 2002, 39-70.*

[50] Gröwe N. et Römisch W. – A stochastic programming model for optimal power dispatch : Stability and numerical treatment. *In Stochastic Optimization- Numerical Methods and Technical Applications (K.Marti, ed.), LN in Economics and Math. Systems Vol. 379, Springer, Berlin, p. 111-139*, 1992.

[51] Gröwe N., Römisch W. et Schultz R. – A simple recourse model for power dispatch under uncertain demand. *Annals of Operations Research, 59,135-164*, 1995.

[52] Guigues V. – New modelings and solution methods for the electric generation management problem. *Working paper.*

[53] Guigues V. – Robust electricity power management. *Working paper.*

[54] Guigues V. – Sensitivity analysis and stable and robust calibration of covariance matrices. *Working paper.*

[55] Guigues V., Aïd R., Ndiaye P., Oustry F. et Romanet F. – A Value-at-Risk approach for robust electricity power management. *Soumis à IEEE Transaction on Automatic and Control*, 2004.

[56] Guigues V. and Judistki A. – Statistical inference for stochastic optimization. *Working paper*.

[57] Guigues V. and Juditski A. – Treatment of the uncertainty for a class of stochastic optimization problems. *Working paper*.

[58] Heitsch H., Römisch W. – Scenario reduction algorithms in stochastic programming. *Preprint 01-8, Institute of Mathematics, Humboldt-University Berlin*, 2001.

[59] Hoffman A.J. – On approximate solutions of systems of linear inequlities. *Journal of Research of National Bureau of Standards*, 1952.

[60] Hoylet K. , Kaut M. et Wallace S. W. – A Heuristic for Generating Scenario Trees for Multistage Decision Problems. 2001.

[61] Judin D. B. – Mathematical Methods of Management under Uncertainty. *Soviet Radio*, 1974.

[62] Kataoka S. – A Stochastic Programming Model. *Econometrica*, 1963.

[63] Laloux L., Cizeau P.,Bouchaud J.-B. et Potters M. – Noise Dressing of Financial Correlation Matrices. *Physical Review Letters*, 1998.

[64] Lemaréchal C. et Hiriart-Urruty J.-B. – *Convex Analysis and Minimization Algorithms I et II.* – Springer-Verlag, 1993.

[65] Lemaréchal C. et Sagastizábal C. – Variable metric bundle methods : from conceptual to implementable forms. *Mathematical Programming,volume 76*, 1997.

[66] Litterman R. et Guangliang H. – The Intuition behind Black Litterman Model Portfolios. 1999.

[67] Litterman R. et Winkelmann K. – Estimating Covariance Matrices. 1998.

[68] Louveaux F. V. et Smeers Y. – A stochastic model for electricity generation. *Preprint*, 1980.

[69] Markowitz HM. – Portfolio Selection. *Journal of Finance*, 1952.

[70] Mercurio D. and Spokoiny V. – Statistical inference for time-inhomogeneous volatility models. *Annals of statistics*, 2004.

[71] Miller B. L. et Wagner H. M. – Chance-Constrained Programming with Joint Probability Constraints. *Operation Research*, 1965.

[72] Moore M. – On the estimation of a convex set. *Annals of Statistics. 12,1090-1099*, 1984.

[73] Nesterov Yu. – Rounding of convex sets and efficient gradient methods for linear programming problems. *CORE discussion paper*, 2004.

[74] Nielsen S. S. et Zenios S.A. – Solving multistage stochastic network programs. *Math. Progr. 73,227-250*, 1996.

[75] Nowak. – Stochastic Lagrangian Relaxation in Power Scheduling of a Hydro-Thermal System under Uncertainty. *Ph.D. thesis, Insitute of Mathematics, Humboldt-University Berlin (http ://dochost.rz.hu-brlin.de/dissertationen/*, 2000.

[76] Oustry F., NDiaye P. – Kato Covariance Mmatrix correction. *Raise partner internal report*, 2002.

[77] Perira M.V.F. et Pinto L.M.V.G. – Multi-stage stochastic Optimization applied to energy planning. *Mathematical Programming*, 1991.

[78] Pflug G.Ch. – Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming, Ser. B 89*, 2001.

[79] Philpott A.B., Craddock M., Waterer H. – Hydro-electric unit commitment subject to uncertain demand. *European Journal of Operational Research*, 1999.

[80] Ripley B.B et Rasson J.P. – Finding the edge of a Poisson forest. *J. Appl. Prob. 14,483-491*, 1977.

[81] Roos C., Bai Y. et Chaerani D. – Conic Optimization Models for Robust Truss Topology Design with Examples. 2003.

[82] Saigal R.,Wolkowitcz H., Vetenberghe L. – *Handbook on Semidefinite Programming.* – Kluwer Academic Publishers, 2000.

[83] Shapiro J.F. – Stochastic programming models for dedicated portfolio selection. *In Mathematical Models for Decision Support (B. Mitra, ed.), NATO ASI Series, Vol. F48, Springer, Berlin, 1988, 587-611.*, 1988.

[84] Smith J. – Generalized Chebyshev inequalities : theory and application in decision analysis. *Operations Research*, 1995.

[85] Somlyódy L. et Wets R. J.-B. – Stochastic optimization models for lake eurotrophication management. *Operations Research. 36, 660-681*, 1988.

[86] Soyster A. L. – Convex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming. *Operations Research. 21, 1154-1157*, 1973.

[87] Soyster A. L. – A Duality Theory for Convex Programming with Set-Inclusive Constraints. *Operations Research. 22, 892-898*, 1974.

[88] Soyster A. L. – Erratum. *Operations Research. 22, 1279-1280*, 1974.

[89] Takriti S., Krasenbrink B., Wu L.S.-Y. – Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. *Operations Research 48, 268-280*, 2000.

[90] Valenca M. et Ludermir T. – Monthly streamflow forecasting using a Neural Fuzzy Network Model. 2000.

[91] Van Slyke R.M et Wets R.J.-B. – L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics,17 :638-663*, 1969.

[92] Wallace S. W. . – A two-stage stochastic facility-location problem with time dependant supply. *In [41], 489-514.*

Résumé : L'objet de cette thèse est de modéliser et analyser des problèmes d'optimisation stochastique et de proposer des méthodes de résolution pour ces problèmes.
Dans une première partie, on considère des problèmes d'allocation d'actifs se formulant comme des problèmes d'optimisation convexes. La fonction coût et les contraintes dépendent d'un paramètre multidimensionnel inconnu. On montre, sous des hypothèses assez générales sur le processus des rendements et notamment s'il y a homogénéité temporelle locale, que l'on peut construire des approximations du problème original se servant d'une estimation adaptative du paramètre inconnu. La précision du problème approché est $O\left(\sqrt{\frac{\max(\ln n, \ln N)}{N}}\right)$, où n est la dimension du problème et N le nombre d'observations disponibles. Cette méthode a été appliquée sur les problème VaR et de Markowitz et l'on présente les résultats de simulations numériques sur des données réelles et simulées. On propose ensuite une analyse de sensibilité pour une classe de problèmes quadratiques dont on déduit une analyse de sensibilité du problème de Markowitz. Pour ce problème, on propose alors une calibration stable de la matrice de covariance et des contreparties robustes.
La deuxième partie porte sur l'analyse de problèmes de gestion de production. On s'intéresse essentiellement au problème de gestion de production électrique. Nous proposons de nouvelles modélisations pour ce problème et des moyens pour les mettre en oeuvre. L'un des modèles conduit à une résolution par décomposition par les prix. Dans ce cas, on montre comment calculer la fonction duale par programmation dynamique. On explique enfin comment dans chaque cas, une stratégie de gestion est mise en place. Les différentes méthodes de gestion sont comparées sur des données réelles et simulées.

Mots clés : Optimisation stochastique et robuste, analyse de sensibilité, calibration de matrice de covariance, optimisation semidéfinie, estimation adaptative, séries chronologiques.

Abstract : In this thesis, stochastic optimization problems are modelized and analyzed and we propose ways to solve these problems.
In the first part, we consider asset allocation problems formulated as convex optimization problems. The cost function and the constraints depend on an unknwon multi-dimensional parameter. We show, under quite general assumptions on the return process, and in particular under local time homogeneity, that we can construct a data-driven approximation of the original problem using an adaptive estimation of the unknown parameter. The accuracy of the approximate problem is $O\left(\sqrt{\frac{\max(\ln n, \ln N)}{N}}\right)$, where n is the problem dimension and N the number of available observations. This method has been applied on the VaR and Markowitz problems and we present the results of numerical experiments with simulated and real-world data. We then propose a sensitivity analysis for a class of quadratic problems which is used to propose a sensitivity analysis of the Markowitz problem. For this problem, we finally propose a stable calibration of the covariance matrix and robust counterparts.
In the second part, we analyse production management problems. We mainly deal with the problem of electricity production generation. We propose new modelings for this problem and ways of implementing them. One of the models we propose is solved using price decomposition. In this case, we show how to efficiently compute, by dynammic programming, the dual function. We finally explain, in each case, how a management strategy is determined. The different management methods are compared using real and simulated data.

Keywords : Stochastic optimization, robust optimization, sensitivity analysis, covariance matrix calibration, semidefinite optimization, adaptive estimation, time series.