

Université de Montréal

**Modélisation des modèles autorégressifs vectoriels avec  
variables exogènes et sélection d'indices**

par

**Mylène Oscar**

Département de mathématiques et de statistique  
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de  
Maître ès sciences (M.Sc.)  
en statistique

7 mai 2022



# Université de Montréal

Faculté des arts et des sciences

---

Ce mémoire intitulé

## **Modélisation des modèles autorégressifs vectoriels avec variables exogènes et sélection d'indices**

présenté par

**Mylène Oscar**

a été évalué par un jury composé des personnes suivantes :

*Martin Bilodeau*

---

(président-rapporteur)

*Pierre Duchesne*

---

(directeur de recherche)

*Alejandro Murua*

---

(membre du jury)



# Sommaire

---

Ce mémoire porte sur l'étude des modèles autorégressifs avec variables exogènes et sélection d'indices. La littérature classique regorge de textes concernant la sélection d'indices dans les modèles autorégressifs. Ces modèles sont particulièrement utiles pour des données macroéconomiques mesurées sur des périodes de temps modérées à longues. Effectivement, la lourde paramétrisation des modèles complets peut souvent être allégée en utilisant la sélection d'indices aboutissant ainsi à des modèles plus parcimonieux. Les modèles à variables exogènes sont très intéressants dans le contexte où il est connu que les variables à l'étude sont affectées par d'autres variables, jouant le rôle de variables explicatives, que l'analyste ne veut pas forcément modéliser. Ce mémoire se propose donc d'étudier les modèles autorégressifs vectoriels avec variables exogènes et sélection d'indices. Ces modèles ont été explorés, entre autres, par Lütkepohl (2005), qui se contente cependant d'esquisser les développements mathématiques. Nous concentrons notre étude sur l'inférence statistique sous des conditions précises, la modélisation ainsi que les prévisions. Notre objectif est de comparer les modèles avec sélection d'indices aux modèles autorégressifs avec variables exogènes complets classiques. Nous désirons déterminer si l'utilisation des modèles avec sélection d'indices est marquée par une différence favorable au niveau du biais et de l'écart-type des estimateurs ainsi qu'au niveau des prévisions de valeurs futures. Nous souhaitons également comparer l'efficacité de la sélection d'indices dans les modèles autorégressifs ayant des variables exogènes à celle dans les modèles autorégressifs. Il est à noter qu'une motivation première dans ce mémoire est l'estimation dans les modèles autorégressifs avec variables exogènes à sous-ensemble d'indices.

Dans le premier chapitre, nous présentons les séries temporelles ainsi que les diverses notions qui y sont rattachées. De plus, nous présentons les modèles linéaires classiques multivariés, les modèles à variables exogènes puis des modèles avec sélection d'indices. Dans le deuxième chapitre, nous exposons le cadre théorique de l'estimation des moindres carrés dans les modèles autorégressifs à sous-ensemble d'indices ainsi que le comportement asymptotique de l'estimateur. Ensuite, nous développons la théorie pour l'estimation des moindres carrés (LS) ainsi que la loi asymptotique des estimateurs pour les modèles

autorégressifs avec sélection d'indices (SVAR) puis nous faisons de même pour les modèles autorégressifs avec variables exogènes et tenant compte de la sélection des indices (SVARX). Spécifiquement, nous établissons la convergence ainsi que la distribution asymptotique pour l'estimateur des moindres carrés d'un processus autorégressif vectoriel à sous-ensemble d'indices et avec variables exogènes. Dans le troisième chapitre, nous appliquons la théorie spécifiée précédemment lors de simulations de Monte Carlo. Nous évaluons de manière empirique les biais et les écarts-types des coefficients trouvés lors de l'estimation ainsi que la proportion de fois que le modèle ajusté correspond au vrai modèle pour différents critères de sélection, tailles échantillonales et processus générateurs des données. Dans le quatrième chapitre, nous appliquons la théorie élaborée aux chapitres 1 et 2 à un vrai jeu de données provenant du système canadien d'information socioéconomique (CANSIM), constitué de la production mensuelle de fromage mozzarella, cheddar et ricotta au Canada, expliquée par les prix mensuels du lait de bovin non transformé dans les provinces de Québec, d'Ontario et de la Colombie-Britannique pour la période allant de janvier 2003 à juillet 2021. Nous ajustons ces données à un modèle autorégressif avec variables exogènes complet puis à un modèle autorégressif avec variables exogènes et sélection d'indices. Nous comparons ensuite les résultats obtenus avec le modèle complet à ceux obtenus avec le modèle restreint.

**Mots-clés :** Processus autorégressif à sous-ensemble d'indices, variables exogènes, estimation des moindres carrés, sélection de modèle, séries chronologiques multivariées, processus stochastiques, séries chronologiques.

# Summary

---

This Master's Thesis focuses on the study of subset autoregressive models with exogenous variables. Many texts from the classical literature deal with the selection of indexes in autoregressive models. These models are particularly useful for macroeconomic data measured over moderate to long periods of time. Indeed, the heavy parameterization of full models can often be simplified by using the selection of indexes, thus resulting in more parsimonious models. Models with exogenous variables are very interesting in the context where it is known that the variables under study are affected by other variables, playing the role of explanatory variables, not necessarily modeled by the analyst. This Master's Thesis therefore proposes to study vector subset autoregressive models with exogenous variables. These models have been explored, among others, by Lütkepohl (2005), who merely sketches proofs of the statistical properties. We focus our study on statistical inference under precise conditions, modeling and forecasting for these models. Our goal is to compare restricted models to full classical autoregressive models with exogenous variables. We want to determine whether the use of restricted models is marked by a favorable difference in the bias and standard deviation properties of the estimators as well as in forecasting future values. We also compare the efficiency of index selection in autoregressive models with exogenous variables to that in autoregressive models. It should be noted that a primary motivation in this Master's Thesis is the estimation in subset autoregressive models with exogenous variables.

In the first chapter, we present time series as well as the various concepts which are attached to them. In addition, we present the classical multivariate linear models, models with exogenous variables and then we present subset models. In the second chapter, we present the theoretical framework for least squares estimation in subset autoregressive models as well as the asymptotic behavior of the estimator. Then, we develop the theory for the estimation of least squares (LS) as well as the asymptotic distribution of the estimators for the subset autoregressive models (SVAR), and we do the same for the subset autoregressive models with exogenous variables (SVARX). Specifically, we establish the convergence as well as the asymptotic distribution for the least squares estimator of a

subset autoregressive process with exogenous variables. In the third chapter, we apply the theory specified above in Monte Carlo simulations. We evaluate empirically the biases and the standard deviations of the coefficients found during the estimation as well as the proportion of times that the adjusted model matches the true model for different selection criteria, sample size and data generating processes. In the fourth chapter, we apply the theory developed in chapters 1 and 2 to a real dataset from the Canadian Socio-Economic Information System (CANSIM) consisting of the monthly production of mozzarella, cheddar and ricotta cheese in Canada, explained by the monthly prices of unprocessed bovine milk in the provinces of Quebec, Ontario and British Columbia from January 2003 to July 2021. We fit these data with a full autoregressive model with exogenous variables and then to a subset autoregressive model with exogenous variables. Afterwards, we compare the results obtained with the complete model to those obtained with the subset model.

**Keywords :** Subset autoregressive process, exogenous variables, least squares estimation, model selection, multivariate time series, stochastic process, time series.

# Table des matières

---

Sommaire .....	v
Summary .....	vii
Liste des tableaux .....	xi
Table des figures .....	xiii
Remerciements .....	xv
Introduction .....	1
<b>Chapitre 1. Préliminaires .....</b>	<b>5</b>
1.1. Généralités .....	5
1.2. Modèles multivariés classiques .....	9
1.3. Modèles linéaires multivariés avec sélection des indices .....	14
1.3.1. Modèles SVAR .....	14
1.3.2. Modèles SVARX .....	15
<b>Chapitre 2. Inférence statistique dans les modèles à sélection d'indices, avec           statistiques de type diagnostic et calcul des prévisions .....</b>	<b>17</b>
2.1. Inférence statistique dans le modèle SVAR .....	17
2.1.1. Estimation des moindres carrés dans le modèle SVAR .....	18
2.1.2. Propriétés asymptotiques du modèle SVAR .....	20
2.2. Inférence statistique et vérification diagnostique dans le modèle SVARX .....	22
2.2.1. Estimation des moindres carrés dans le modèle SVARX .....	22
2.2.2. Propriétés asymptotiques du modèle SVARX .....	25
2.2.3. Vérification diagnostique des modèles SVARX .....	37
2.3. Calcul des prévisions de valeurs futures .....	38
2.3.1. Problème de la prévision dans le modèle SVAR .....	38
2.3.2. Construction d'un système de prévision dans les modèles VARX .....	40

2.3.3.	Construction d'un système de prévision dans les modèles SVARX .....	42
<b>Chapitre 3.</b>	<b>Étude de simulations : Estimation et prévision dans les modèles à sous-ensemble d'indices .....</b>	<b>45</b>
3.1.	Modèles utilisés dans les simulations .....	46
3.2.	Modèles ajustés .....	48
3.3.	Critères de sélection .....	49
3.4.	Choix des modèles .....	51
3.5.	Estimation des paramètres .....	52
3.6.	Prévisions .....	55
<b>Chapitre 4.</b>	<b>Analyse de données réelles .....</b>	<b>71</b>
4.1.	Description du jeu de données .....	71
4.2.	Ajustement du modèle .....	73
4.2.1.	Sélection des ordres maximaux $p$ et $s$ .....	74
4.2.2.	Ajustement d'un SVARX( $p,s$ ) .....	75
4.2.3.	Estimation des paramètres .....	76
4.2.4.	Évaluation du modèle .....	77
4.3.	Prévisions de valeurs futures .....	78
<b>Conclusion</b>	.....	<b>81</b>
<b>Références</b>	.....	<b>83</b>

## Liste des tableaux

---

3.1	Table des paramètres .....	47
3.2	Proportion de fois où le bon modèle est choisi .....	52
3.3	Proportion de fois où la bonne décision est prise pour la matrice .....	53
3.4	Fréquence de sélection des modèles pour les VAR .....	54
3.5	Fréquence de sélection des modèles pour les VARX.....	57
3.6	Biais moyen des coefficients obtenus pour le DGP <sub>1</sub> .....	58
3.7	Écart-type moyen des coefficients obtenus pour le DGP <sub>1</sub> .....	59
3.8	Biais moyen des coefficients obtenus pour le DGP <sub>2</sub> .....	60
3.9	Écart-type moyen des coefficients obtenus pour le DGP <sub>2</sub> .....	61
3.10	Biais moyen des coefficients obtenus pour le DGP <sub>3</sub> .....	62
3.11	Écart-type moyen des coefficients obtenus pour le DGP <sub>3</sub> .....	63
3.12	Biais moyen des coefficients obtenus pour le DGP <sub>4</sub> .....	64
3.13	Écart-type moyen des coefficients obtenus pour le DGP <sub>4</sub> .....	65
3.14	Biais moyen des coefficients obtenus pour le DGP <sub>5</sub> .....	66
3.15	Écart-type moyen des coefficients obtenus pour le DGP <sub>5</sub> .....	67
3.16	Biais moyen des coefficients obtenus pour le DGP <sub>6</sub> .....	68
3.17	Écart-type moyen des coefficients obtenus pour le DGP <sub>6</sub> .....	69
3.18	Erreur quadratique moyenne de la prévision d’horizon un .....	70
4.1	Résultats de l’estimation du modèle SVARX( $\{1,2,3,4,5,7,9,12\},\{0,3\}$ ) pour la première différence des logarithmes, pour la production mensuelle de fromage expliquée par les prix du lait de bovin non transformé au Québec, en Ontario et en Colombie-Britannique. Les écarts-types sont indiquées entre parenthèses. ....	76
4.2	Valeurs de la statistique de test portemanteau basée sur le noyau $T_n$ lorsque $\mathcal{K}$ est le noyau uniforme tronqué, le noyau Bartlett et le noyau Daniell pour $P_n \in \{2,3,5,10,15\}$ . .....	78

4.3	Biais moyen et mesure d'EQM des prévisions ponctuelles pour les prévisions d'horizon un de la production mensuelle de fromage mozzarella, cheddar et ricotta au Canada transformée .....	79
-----	--	----

## Table des figures

---

4.1	Production mensuelle de fromage mozzarella, cheddar et ricotta au Canada et prix du lait de bovin non transformé au Québec, en Ontario et en Colombie-Britannique (janvier 2003 à juillet 2021). . . . .	72
4.2	Première différence des logarithmes, pour la production mensuelle de fromage mozzarella, cheddar et ricotta au Canada, et pour les prix du lait de bovin non transformé au Québec, en Ontario et en Colombie-Britannique (janvier 2003 à juillet 2021). . . . .	74
4.3	Autocorrélations et corrélations croisées des résidus, modèle restreint . . . . .	77



# Remerciements

---

Tout d'abord, je tiens à remercier mon directeur de recherche Monsieur Pierre Duchesne, qui m'a accompagnée durant mon projet de recherche. Je lui suis grandement reconnaissante pour le soutien financier mais également pour les précieux conseils prodigués ainsi que les nombreux encouragements qu'il a su proférer et qui m'ont permis de compléter mon mémoire. Merci pour votre patience tout au long du projet. Merci pour les indications, les corrections, et les relectures.

Je veux ensuite remercier ma famille, particulièrement ma mère, ma soeur Agnès et mon frère Augustin qui m'ont supportée dans ma décision de poursuivre mes études. Merci énormément à vous pour votre appui et pour votre soutien.

Finalement, je souhaite remercier mes amis proches qui ont su m'encourager et m'épauler tout au long de ma maîtrise.



# Introduction

---

Les processus autorégressifs à variables exogènes et leurs versions multivariées, abrégés par les acronymes ARX et VARX, respectivement, constituent une classe de modèles vaste et flexible pour les séries temporelles stationnaires. Ils sont efficaces pour modéliser la dynamique conjointe des séries temporelles et ont l'avantage d'intégrer l'impact des variables exogènes stochastiques non modélisées. La classe des modèles VARX comprend les processus autorégressifs vectoriels (VAR) et généralise des modèles de régression linéaire multivariée dont les variables explicatives incluent des variables exogènes ainsi que les valeurs retardées des variables dépendantes et exogènes. Les applications des modèles ARX et VARX sont diverses en économie, plus particulièrement en macroéconomie (voir, par exemple, Lütkepohl (2005) ou Nicholson, Matteson et Bien (2017)) ainsi qu'en ingénierie (voir, par exemple, da Silva, da Fonseca Neto et Nagem (2009), Diversi, Guidorzi et Soverini (2007) ou Sanandaji et al. (2011)). Les propriétés théoriques de ces modèles sont discutées et étudiées dans Hannan et Deistler (1988) et Lütkepohl (2005). Voir également Ocampo et Rodríguez (2012), qui présentent une revue de l'estimation des modèles VARX réduits et structurels avec des applications utiles.

Sous les conditions générales, les propriétés asymptotiques des estimateurs des moindres carrés (LS) dans les modèles VARX ont été étudiées dans Hannan, Dunsmuir et Deistler (1980) et Hannan et Deistler (1988). Lütkepohl (1993, Section 10.3) présente les grandes lignes de la preuve conduisant aux propriétés de convergence en probabilité, ainsi qu'à la normalité asymptotique des estimateurs des moindres carrés généralisés (GLS) et estimateurs GLS estimés (EGLS). La structure de covariance asymptotique des estimateurs du modèle VARX y est également donnée. Cela est un résultat utile pour l'inférence statistique. D'autres références présentant les propriétés statistiques des modèles VARX incluent Guo, Huang et Hannan (1990), Penm, Penm et Terrell (1993), Boutahar et Deniau (1995) et Duchesne (2005).

Un inconvénient majeur des modèles VARX est leur lourde paramétrisation. Étant donné que les variables indépendantes et exogènes sont en général des vecteurs, les paramètres

libres sont inclus dans les matrices donc l'espace des paramètres croît quadratiquement avec le nombre de séries temporelles incluses dans le modèle. Conséquemment, diverses tentatives ont exploré des méthodes pour réduire la taille de l'espace des paramètres. Voir Nicholson, Matteson et Bien (2017) pour une discussion intéressante. Dans Lütkepohl (1993, Section 10.3), des contraintes sur les paramètres sont permises. Une autre stratégie, particulièrement intéressante du point de vue de la sélection des modèles, repose sur les modèles VARX de sous-ensemble, où pour certains délais sélectionnés, les matrices complètes de paramètres dans le modèle plein sont contraintes à zéro. La question est de choisir quels délais expliquent le mieux la variable réponse. C'est la voie empruntée dans ce mémoire.

Les modèles autorégressifs à sous-ensemble d'indices ont une longue histoire dans l'analyse des séries chronologiques. Naturellement, le choix optimal de l'ordre du délai maximal dans les modèles autorégressifs est d'importance fondamentale. Cependant, dans un ouvrage fondateur, McClave (1975) note qu'en prenant l'ordre autorégressif optimal, la variance résiduelle estimée de certains modèles autorégressifs intermédiaires peut être plus petite lorsque certains retards intermédiaires sont contraints à avoir des coefficients nuls. Une représentation plus parcimonieuse des modèles autorégressifs est obtenue avec des autorégressions à sous-ensemble d'indices, ce qui peut être particulièrement important dans le contexte des modèles multivariés. Afin de trouver le meilleur modèle, McClave (1975) a présenté un algorithme s'appuyant sur les estimateurs de Yule-Walker. Des critères tels que le critère d'information d'Akaike (AIC) ou le critère d'information bayésien (BIC) ont été utilisés. En utilisant les estimateurs LS, un autre algorithme facile à mettre en oeuvre a été proposé par Haggan et Oyetunji (1984). Des méthodes bayésiennes pour l'autorégression à sous-ensemble d'indices ont été proposées par Chen (1999) et Unnikrishnan (2004). Dans McLeod et Zhang (2006), des modèles d'autorégression de sous-ensembles sont étudiés en utilisant les récursions de Durbin-Levinson ainsi qu'une paramétrisation alternative.

Dans les modèles autorégressifs vectoriels à sous-ensemble d'indices, Brüggemann et Lütkepohl (2001) discutent de quatre stratégies pour la spécification de modèles VAR à sous-ensemble d'indices, où aucune restriction n'est imposée sur certains des coefficients. Des approches univariées, c'est-à-dire des approches d'équations individuelles, et multivariées sont discutées. Les variables et les délais à éliminer sont basés sur des critères de sélection de modèles tels que le critère d'information d'Akaike (AIC), voir Akaike (1974). Voir également la monographie de Brüggemann (2004). Dans Brüggemann (2002), des expériences de Monte-Carlo supplémentaires comparent les stratégies de modélisation à sous-ensemble d'indices avec la méthode de réduction automatisée, de générale à spécifique, proposée par Krolzig et Hendry (2001). Dans Lütkepohl (1993, 2005), des dérivations partielles de la distribution asymptotique des estimateurs LS complets, en permettant des contraintes linéaires

sur les paramètres, sont exposées. Ces résultats incluent la possibilité de contraindre à zéro des paramètres. Dans la littérature de sélection de modèles, on parle également de pénaliser des paramètres à zéro. En séries chronologiques, par exemple dans Lütkepohl (2005), il est discuté de comment contraindre des paramètres à zéro. Dans Brockwell, Davis et Alexandre Trindade (2004), en utilisant la théorie des martingales et sous des conditions précises, des théorèmes centraux limites pour les estimateurs des moindres carrés avec sélection d'indices, et pour les estimateurs de type Yule-Walker et de Burg, toujours dans ce même contexte d'estimateurs des paramètres d'un sous-ensemble d'un modèle VAR, sont établis. Par sélection d'indices nous signifions le choix des variables retardées à inclure dans le modèle. Il a été montré qu'asymptotiquement tous les estimateurs partagent la même distribution limite normale multivariée.

Ce mémoire est composé de quatre chapitres. Dans le premier chapitre, nous élaborons le cadre théorique pour l'étude des séries temporelles, incluant des définitions de bruit blanc, de stationnarité, d'inversibilité, d'autocorrélation et de corrélation croisée. Nous présentons les modèles linéaires classiques multivariés, soit les modèles VMA( $q$ ), VAR( $p$ ), et VARMA( $p,q$ ), suivis des modèles à variables exogènes VARX( $p,s$ ), VARMAX( $p,s,q$ ) puis des modèles à sous-ensemble d'indices soit les SVAR( $p$ ) et les SVARX( $p,s$ ).

Dans le deuxième chapitre, nous voyons le cadre théorique pour l'estimation des moindres carrés dans les modèles SVAR et nous étudions le comportement asymptotique de l'estimateur. Ensuite, nous développerons la théorie pour l'estimation des moindres carrés ainsi que la loi asymptotique des estimateurs pour les modèles VARX puis nous développerons le problème de la prévision pour ces modèles. Nous verrons que pour les deux modèles à sélection d'indices, l'estimateur des moindres carrés converge en probabilité vers les vrais paramètres et nous verrons que leur distribution limite est une distribution normale multivariée.

Le troisième chapitre est consacré à des simulations Monte Carlo pour les modèles SVAR( $p$ ) et SVARX( $p,s$ ). Notre étude empirique s'attarde à plusieurs modèles SVAR(3) ainsi qu'à des modèles SVARX(3,0) de dimensions variées. Nous simulons ainsi des séries puis nous estimons les paramètres de ces modèles et discutons des difficultés d'estimation présentes selon la taille échantillonnale des séries, selon les valeurs des coefficients des processus générés, selon le modèle généré et selon le critère de sélection utilisé. Nous calculons le biais des coefficients trouvés lors de l'estimation ainsi que l'écart-type. Ensuite, nous calculons la proportion de fois où le vrai modèle est choisi ainsi que la proportion de fois où la bonne décision pour chaque matrice de coefficients. Nous considérons qu'une bonne décision est prise lorsque l'estimateur d'une matrice nulle dans le vrai modèle est également nul ou lorsque l'estimateur d'une matrice non nulle dans le vrai modèle est également non nul. Finalement, nous discutons des résultats particulièrement des différences obtenues entre les modèles SVAR et SVARX. Nous observons ainsi que les résultats varient grandement en

fonction des matrices composant les processus générant les données et de la taille échantillonnale.

Le quatrième chapitre comporte une application de la théorie des modèles VARX élaborée dans les chapitres 1 et 2 à un vrai jeu de données constitué de la production mensuelle de fromage mozzarella, cheddar et ricotta au Canada expliqués par les prix mensuels du lait de bovin non transformé dans les provinces de Québec, d'Ontario et de la Colombie-Britannique pour la période de janvier 2003 à juillet 2021. Plus spécifiquement, les variables concernant la production des divers fromages sont utilisées comme variables endogènes alors que les variables concernant les prix du lait sont utilisées comme variables exogènes. Après avoir déterminé les ordres maximaux  $(p,s)$ , nous ajustons ces données à un modèle VARX $(p,s)$  plein puis nous considérons tous les modèles SVARX possibles et choisissons celui qui minimise le critère de sélection AIC. Par la suite, nous calculons les prévisions ponctuelles obtenues lors de l'ajustement des données au modèle complet VARX et les comparons à celles obtenues lors de l'ajustement des données au modèle SVARX choisi précédemment. Nous discutons des résultats obtenus, à savoir si le modèle SVARX mène à une réelle différence au niveau des prévisions dans le cadre de données réelles. Nous concluons que, pour notre jeu de données, en considérant le critère de l'erreur quadratique moyenne, le modèle SVARX mène à de meilleures prévisions pour deux des trois variables du modèle.

# Chapitre 1

---

## Préliminaires

Dans ce chapitre, nous présentons d'abord des notions générales liées aux séries chronologiques multivariées. Par la suite, nous présentons les modèles linéaires multivariés classiques. Ces modèles incluent les modèles vectoriels autorégressifs  $\text{VAR}(p)$ , les modèles vectoriels à moyenne mobile  $\text{VMA}(q)$  ainsi que les modèles vectoriels autorégressifs à moyenne mobile  $\text{VARMA}(p,q)$ . Les modèles multivariés classiques comprennent également les modèles ayant des variables exogènes. Parmi ceux-ci, nous présentons les modèles vectoriels autorégressifs avec variables exogènes  $\text{VARX}(p,s)$  et les modèles vectoriels autorégressifs à moyenne mobile avec variables exogènes  $\text{VARMAX}(p,s,q)$ . Pour finir, nous présentons certains modèles avec sélection d'indices, soient les modèles autorégressifs  $\text{SVAR}(\mathcal{I})$ , où  $\mathcal{I}$  contient les indices retenus de la partie autorégressive. De plus, nous introduirons les modèles autorégressifs avec variables exogènes avec sélection d'indices  $\text{SVARX}(\mathcal{I},\mathcal{J})$ , où  $\mathcal{I}$  et  $\mathcal{J}$  contiennent les indices retenus de la partie autorégressive et exogène, respectivement ; le modèle  $\text{SVARX}(\mathcal{I},\mathcal{J})$  est le principal sujet du présent mémoire.

### 1.1. Généralités

Avant tout, nous présentons le concept de série temporelle univariée. Une série temporelle univariée, également appelée série chronologique univariée, est une réalisation finie d'un processus stochastique  $\{Y_t, t \in \mathbb{Z}\}$  de taille  $n$ . Un processus stochastique repose sur une collection infinie de variables aléatoires. Par réalisation finie, on veut dire que l'on observe qu'un nombre fixé à  $n$  de variables qui constitueront notre série chronologique. Dans notre contexte, l'intervalle temporel entre deux observations d'un processus stochastique est fixe de sorte que l'on suppose disposer des informations de la série chronologique aux temps  $t = 1, 2, \dots, n$ . Ainsi, la série chronologique peut être observée sur une base annuelle, mensuelle ou même selon une échelle plus fine (hebdomadaire ou horaire), mais pas de manière irrégulièrement espacée (comme dans certaines applications financières, par exemple lors de l'observation du prix des actions observée en temps réel).

La généralisation au cas multivarié est aisée et directe. Un processus stochastique multivarié  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  est constitué de vecteurs aléatoires de dimension  $k \times 1$ , tel que  $\mathbf{Y}_t = (Y_t(1), Y_t(2), \dots, Y_t(k))^\top$ , où  $\mathbf{A}^\top$  dénote la transposée du vecteur ou de la matrice  $\mathbf{A}$ . On supposera aussi observer une réalisation finie de taille  $n$  des vecteurs  $\mathbf{Y}_t, t = 1, \dots, n$ . Considérons  $k$  séries chronologiques univariées  $\{Y_t(1)\}, \dots, \{Y_t(k)\}$ . Le vecteur aléatoire  $\mathbf{Y}_t$  de dimension  $k \times 1$  où la  $i^e$  ligne est donc  $Y_t(i)$  définit une série chronologique multivariée. Autrement dit, à chaque temps  $t$ , on a  $\mathbf{Y}_t = (Y_t(1), \dots, Y_t(k))^\top$ . Le choix des séries temporelles univariées  $\{Y_t(i)\}$  à inclure dans le vecteur  $\mathbf{Y}_t$  dépend du sujet, de la compréhension du système à l'étude ainsi que de la disponibilité des données. En général, il est cependant implicite que ces séries seront interdépendantes à travers le temps et à travers les délais. Par ailleurs, la modélisation des relations entre les séries univariées est l'intérêt principal des séries chronologiques multivariées.

Le concept de stationnarité est d'une importance capitale dans le contexte des séries chronologiques. En effet, lorsque satisfaite, cette propriété permet d'obtenir des résultats d'inférence à des fins de modélisation utiles à partir d'une réalisation finie d'un processus. Deux types de stationnarité sont traditionnellement étudiés. Il s'agit de la stationnarité au sens strict ainsi que de la stationnarité au sens large. La stationnarité au sens strict est un concept probabiliste qui porte sur la distribution de probabilité du processus, alors que la stationnarité au sens large pose des conditions sur les deux premiers moments du processus (ainsi sur la fonction moyenne et la fonction d'autocovariance). Nous définissons formellement ces concepts. Les définitions proviennent de la monographie de Reinsel (1997).

**Définition 1.1.1.** *Un processus  $\{\mathbf{Y}_t\}$  est dit stationnaire au sens strict si la distribution de probabilité des vecteurs  $(\mathbf{Y}_{t_1}, \mathbf{Y}_{t_2}, \dots, \mathbf{Y}_{t_n})$  et  $(\mathbf{Y}_{t_1+l}, \mathbf{Y}_{t_2+l}, \dots, \mathbf{Y}_{t_n+l})$  est la même pour tout  $t_1, t_2, \dots, t_n$ , pour tout  $n$  et pour tout délai  $l = 0, \pm 1, \pm 2, \dots$*

**Définition 1.1.2.** *En supposant l'existence des moments d'ordre un et d'ordre deux, un processus  $\{\mathbf{Y}_t\}$  est dit stationnaire au sens large si :*

- (1) *La fonction moyenne satisfait  $E(\mathbf{Y}_t) = \boldsymbol{\mu}$  pour tout  $t \in \mathbb{Z}$  où  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)^\top$  est le vecteur de la moyenne du processus,*
- (2) *La matrice de covariance des  $\mathbf{Y}_t$  dénotée par  $\boldsymbol{\Gamma}_{\mathbf{Y}} \equiv \boldsymbol{\Gamma}(0) = E\{(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})^\top\}$  existe et est constante, indépendamment de  $t$ , et ce pour tout  $t$ ,*
- (3) *La fonction d'autocovariance peut être définie, de sorte que  $\gamma_{ij}(l) = \text{cov}(Y_{t+l}(i), Y_t(j)) = E\{(Y_{t+l}(i) - \mu_i)(Y_t(j) - \mu_j)\}$  ne dépend que du délai  $l$  et est indépendant du temps  $t$  pour  $i, j = 1, \dots, k$ .*

La matrice  $k \times k$  contenant les covariances pour les délais  $l = 0, \pm 1, \pm 2, \dots$  est dénotée  $\mathbf{\Gamma}_{\mathbf{Y}}(l)$ . Plus précisément, on peut introduire :

$$\mathbf{\Gamma}_{\mathbf{Y}}(l) = E \left\{ (\mathbf{Y}_{t+l} - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})^\top \right\} = \begin{bmatrix} \gamma_{11}(l) & \gamma_{12}(l) & \dots & \gamma_{1k}(l) \\ \gamma_{21}(l) & \gamma_{22}(l) & \dots & \gamma_{2k}(l) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k1}(l) & \gamma_{k2}(l) & \dots & \gamma_{kk}(l) \end{bmatrix}.$$

De plus,  $\rho_{ij}(l) = \text{corr}(Y_{t+l}(i), Y_t(j)) = \gamma_{ij}(l) / \{\gamma_{ii}(0)\gamma_{jj}(0)\}^{1/2}$  est un élément de la matrice  $k \times k$  contenant les corrélations pour les délais  $l = 0, \pm 1, \pm 2, \dots$ , et est dénotée  $\boldsymbol{\rho}_{\mathbf{Y}}(l)$  et est telle que :

$$\boldsymbol{\rho}_{\mathbf{Y}}(l) = \mathbf{V}^{-1/2} \mathbf{\Gamma}_{\mathbf{Y}}(l) \mathbf{V}^{-1/2} = \begin{bmatrix} \rho_{11}(l) & \rho_{12}(l) & \dots & \rho_{1k}(l) \\ \rho_{21}(l) & \rho_{22}(l) & \dots & \rho_{2k}(l) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1}(l) & \rho_{k2}(l) & \dots & \rho_{kk}(l) \end{bmatrix},$$

où  $\mathbf{V}^{-1/2} = \text{diag} \left\{ \gamma_{11}^{-1/2}(0), \dots, \gamma_{kk}^{-1/2}(0) \right\}$ .

Un processus strictement stationnaire avec une matrice des variances et covariances  $\mathbf{\Gamma}_{\mathbf{Y}}(0)$  qui existe sera stationnaire au sens large mais l'inverse n'est pas forcément vrai. En pratique, il est difficile de vérifier qu'un processus est strictement stationnaire alors que c'est plus aisé dans le cas de réalisations de processus vérifiant la condition de stationnarité au sens large. Il s'agit de la principale raison pour laquelle la stationnarité au sens strict est un concept peu utilisé dans les applications, mais qui est un outil théorique puissant dans l'établissement des résultats théoriques, facilitant les preuves et développements mathématiques. Dans le cadre de ce mémoire, lorsque la stationnarité sera mentionnée, il s'agira de la stationnarité au sens large. De plus, nous ne nous attarderons uniquement qu'à des processus qui sont stationnaires. Bien entendu, dans la section avec des données réelles, nous utiliserons la méthodologie usuelle afin de rendre stationnaire un processus qui violerait les conditions de stationnarité au sens large.

Un exemple élémentaire de processus stationnaire au sens large est le bruit blanc. Le bruit blanc est couramment utilisé comme terme d'erreur dans la modélisation statistique. Nous le définissons de façon explicite.

**Définition 1.1.3.** *Un bruit blanc multivarié de dimension  $k \times 1$  est un processus  $\{\boldsymbol{\epsilon}_t\}$  tel que*

- (1) La fonction moyenne est centrée en zéro :  $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$ ,
- (2) La matrice des variances et covariances  $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top) = \boldsymbol{\Sigma}_\boldsymbol{\epsilon}$  est une matrice de dimension  $k \times k$  définie positive,
- (3) Propriété d'absence de corrélation, c'est-à-dire  $E(\boldsymbol{\epsilon}_{t+l} \boldsymbol{\epsilon}_t^\top) = \mathbf{0}$  pour  $l \neq 0$ .

Un bruit blanc (multivarié) est donc constitué de vecteurs aléatoires non corrélés, d'espérance nulle et d'une matrice des variances et covariances finie et constante. Soit le processus  $\{\boldsymbol{\epsilon}_t\}$ . Dans le cas où les vecteurs de ce processus sont indépendants et identiquement distribués, on notera  $\{\boldsymbol{\epsilon}_t\} \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_\boldsymbol{\epsilon})$ . Dans une telle situation, le bruit blanc est parfois qualifié de bruit blanc fort. Dans le cas contraire, sous la seule hypothèse d'absence de corrélation, il est appelé bruit blanc faible. Lorsque  $\{\boldsymbol{\epsilon}_t\}$  est composé de vecteurs aléatoires de distributions normales et indépendantes, le bruit blanc est dit Gaussien. En outre, un processus  $\mathbf{Y}_t$  de moyenne  $\boldsymbol{\mu}$  est linéaire si on peut écrire :  $\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\epsilon}_{t-j}$  avec  $\sum_{j=0}^{\infty} \|\boldsymbol{\Psi}_j\| < \infty$ , où  $\|\boldsymbol{\Psi}_j\| = \text{tr}\{\boldsymbol{\Psi}_j \boldsymbol{\Psi}_j^\top\}^{1/2}$ . Il est à noter que les poids  $\boldsymbol{\Psi}_j$  sont des matrices et que  $\text{tr}(\cdot)$  dénote la trace d'une matrice. Un processus linéaire peut donc s'écrire comme une combinaison linéaire de bruits blancs. Il peut être montré qu'un tel processus est stationnaire.

En principe, le concept de stationnarité dans le cas multivarié englobe la stationnarité conjointe entre deux processus, univariés ou multivariés. Cependant, compte tenu que nous considérerons des modèles impliquant deux processus multivariés (un endogène, l'autre exogène), il est utile d'avoir des notations particulières pour la stationnarité conjointe entre deux processus (multivariés).

**Définition 1.1.4.** Soient  $\{\mathbf{X}_t\}$  un processus multivarié de dimension  $k_x \times 1$  et  $\{\mathbf{Y}_t\}$  un processus multivarié de dimension  $k \times 1$ . Les deux processus  $\{\mathbf{X}_t\}$  et  $\{\mathbf{Y}_t\}$  sont dits conjointement stationnaires si les conditions suivantes sont satisfaites :

- (1) Le processus  $\{\mathbf{X}_t\}$  est stationnaire au sens large et de même  $\{\mathbf{Y}_t\}$  est stationnaire au sens large,
- (2) La fonction de covariance

$$\gamma_{X_i, Y_j}(l) = \text{cov}(X_{t+l}(i), Y_t(j)) = E \left\{ (X_{t+l}(i) - \mu_{x_i})(Y_t(j) - \mu_{y_j}) \right\}$$

est une fonction du délai  $l$  et est indépendante du temps  $t$  pour  $i = 1, \dots, k_x$  et  $j = 1, \dots, k$ .

Les covariances entre  $\{\mathbf{X}_t\}$  et  $\{\mathbf{Y}_t\}$  sont appelées covariances croisées, et sont définies pour les délais  $l = 0, \pm 1, \pm 2, \dots$

La matrice  $k_x \times k$  contenant ces covariances croisées est dénotée  $\mathbf{\Gamma}_{\mathbf{XY}}(l)$  et elle est définie par la matrice suivante :

$$\mathbf{\Gamma}_{\mathbf{XY}}(l) = E \left\{ (\mathbf{X}_{t+l} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{Y}_t - \boldsymbol{\mu}_{\mathbf{Y}})^\top \right\} = \begin{bmatrix} \gamma_{X_1, Y_1}(l) & \gamma_{X_1, Y_2}(l) & \dots & \gamma_{X_1, Y_k}(l) \\ \gamma_{X_2, Y_1}(l) & \gamma_{X_2, Y_2}(l) & \dots & \gamma_{X_2, Y_k}(l) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{X_{k_x}, Y_1}(l) & \gamma_{X_{k_x}, Y_2}(l) & \dots & \gamma_{X_{k_x}, Y_k}(l) \end{bmatrix}.$$

De plus, la fonction de corrélation croisée est donnée par  $\rho_{X_i, Y_j}(l) = \text{corr}(X_{it}; Y_{j, t+l}) = \gamma_{X_i, Y_j}(l) / \{\gamma_X(0)\gamma_Y(0)\}^{1/2}$  et la matrice  $k_x \times k$  contenant les corrélations croisées pour les délais  $l = 0, \pm 1, \pm 2, \dots$  est notée  $\boldsymbol{\rho}_{\mathbf{XY}}(l)$  et est telle que :

$$\boldsymbol{\rho}_{\mathbf{XY}}(l) = \begin{bmatrix} \rho_{X_1, Y_1}(l) & \rho_{X_1, Y_2}(l) & \dots & \rho_{X_1, Y_k}(l) \\ \rho_{X_2, Y_1}(l) & \rho_{X_2, Y_2}(l) & \dots & \rho_{X_2, Y_k}(l) \\ \vdots & \vdots & \dots & \vdots \\ \rho_{X_{k_x}, Y_1}(l) & \rho_{X_{k_x}, Y_2}(l) & \dots & \rho_{X_{k_x}, Y_k}(l) \end{bmatrix}.$$

Un second concept d'intérêt dans le contexte des séries chronologiques est celui d'inversibilité. Nous définissons la notion d'inversibilité.

**Définition 1.1.5.** *Un processus  $\{\mathbf{Y}_t\}$  de moyenne  $\boldsymbol{\mu}$  est inversible si on peut écrire le processus sous la forme suivante :  $\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{j=1}^{\infty} \mathbf{\Pi}_j(\mathbf{Y}_{t-j} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t$  où les poids  $\mathbf{\Pi}_j$  sont des matrices telles que  $\sum_{j=1}^{\infty} \|\mathbf{\Pi}_j\| < \infty$ .*

Un processus est donc inversible s'il peut être exprimé en fonction de ses valeurs retardées plus un terme d'erreur aléatoire. Les notions de stationnarité et d'inversibilité sont essentielles dans le contexte des séries chronologiques car elles sont nécessaires dans l'élaboration du calcul des prévisions. Effectivement, l'hypothèse de stationnarité simplifie le calcul de la variance des erreurs de prévision alors que l'inversibilité facilite le calcul des prévisions elles-mêmes.

Nous avons exposé les notions élémentaires liés aux séries chronologiques. La prochaine section présente les modèles classiques.

## 1.2. Modèles multivariés classiques

Divers modèles de séries chronologiques sont présentés dans la littérature. Parmi les modèles univariés classiques se trouvent les processus autorégressifs d'ordre  $p$ , notés  $\text{AR}(p)$ , les processus à moyennes mobile d'ordre  $q$ , notés  $\text{MA}(q)$ , ainsi que les processus autorégressifs à moyenne mobile d'ordre  $(p, q)$ , notés  $\text{ARMA}(p, q)$ . Nous ne rappellerons pas ces modèles

univariés, qui sont présentés dans la plupart des ouvrages de base portant sur les séries chronologiques, voir par exemple Shumway et Stoffer (2017). De plus, ils sont des cas particuliers des modèles multivariés, qui eux seront présentés.

En effet, présenter les versions multivariées des modèles univariés s'avère beaucoup plus pertinent dans le cadre de ce mémoire. Ainsi, on retrouve les autorégressifs vectoriels VAR( $p$ ), les moyennes mobiles vectoriels VMA( $q$ ) et plus généralement les modèles ARMA vectoriels, qui sont notés VARMA( $p, q$ ). Ces modèles sont pertinents lors de l'ajustement de données se présentant comme des vecteurs. La littérature économétrique a consacré beaucoup d'attention sur les modèles de type VAR, compte tenu que ces modèles sont similaires aux modèles de régression multivariés. Il est bien connu que les VARMA font face à des problèmes d'identification, et même si au niveau de l'écriture les généralisations du cas AR au VAR, ou encore ARMA au VARMA sont immédiates, l'ajustement des VARMA semble moins populaire que les VAR. Voir Brockwell et Davis (2016, Section 8.4).

Des processus multivariés exogènes, constitués de vecteurs aléatoires jouant le rôle de variables explicatives, peuvent également être intégrés aux modèles multivariés classiques. Plus précisément, en ajoutant un processus exogène aux modèles VAR( $p$ ) et VARMA( $p, q$ ), de nouveaux modèles sont obtenus, notés VARX( $p, s$ ) et VARMAX( $p, s, q$ ). On note que  $s$  est l'ordre exogène, et lorsque  $s = 0$ , une variable explicative contemporaine est incluse dans le modèle, et lorsque  $s > 0$ , des retards de la variable exogène sont également inclus dans le modèle. Autrement dit, un VARX( $p, 0$ ) ne se réduit pas à un VAR( $p$ ). Plus de détails sont donnés dans ce qui suit.

Puisqu'ils sont particulièrement utiles lors de l'ajustement d'un système de données influencées par d'autres variables, ces modèles sont aussi présentés dans cette section. Nous commençons par définir formellement le modèle VAR( $p$ ).

**Définition 1.2.1.** *Un modèle vectoriel autorégressif d'ordre  $p$ , noté VAR( $p$ ), satisfait l'équation aux différences stochastique :*

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{Y}_{t-j} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t, \quad (1.2.1)$$

où  $\mathbf{Y}_t$  est un vecteur aléatoire de dimension  $k \times 1$ ;  $\boldsymbol{\mu}$  est un vecteur de constantes de dimension  $k \times 1$ ;  $\boldsymbol{\Phi}_j$  est une matrice de paramètres de dimension  $k \times k$  pour  $j = 1, \dots, p$ ,  $\boldsymbol{\Phi}_p \neq \mathbf{0}$  et  $\{\boldsymbol{\epsilon}_t\}$  est un bruit blanc de dimension  $k \times 1$ .

En utilisant l'opérateur retard,  $B\mathbf{Y}_t = \mathbf{Y}_{t-1}$ , le processus peut être réécrit de la façon suivante :

$$\boldsymbol{\Phi}(B)(\mathbf{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\epsilon}_t, \quad (1.2.2)$$

avec  $\Phi(B) = \mathbf{I} - \Phi_1 B - \dots - \Phi_p B^p$ .

Pour que le processus soit stationnaire, toutes les racines de  $\det\{\Phi(B)\}=0$  doivent être à l'extérieur du disque unitaire dans les complexes (les racines pouvant être réelles ou des nombres complexes, pouvant donc être représentées dans le plan complexe), où  $\det(\mathbf{A})$  représente le déterminant de la matrice  $\mathbf{A}$ . Ce résultat est démontré dans Brockwell et Davis (1991). De par la définition précédente, il devrait être assez évident qu'un processus VAR( $p$ ) stationnaire sera toujours inversible. Nous définissons à présent le modèle VMA( $q$ ), la généralisation multivariée d'un modèle moyenne mobile.

**Définition 1.2.2.** *Un modèle vectoriel de type moyenne mobile d'ordre  $q$ , noté VMA( $q$ ), satisfait l'équation aux différences stochastiques :*

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \sum_{j=1}^q \boldsymbol{\Theta}_j \boldsymbol{\epsilon}_{t-j}, \quad (1.2.3)$$

où  $\mathbf{Y}_t$  est un vecteur aléatoire de dimension  $k \times 1$ ;  $\boldsymbol{\mu}$  est un vecteur de constantes de dimension  $k \times 1$ ;  $\boldsymbol{\Theta}_j$  est une matrice de paramètres de dimension  $k \times k$  pour  $j = 1, \dots, q$ ,  $\boldsymbol{\Theta}_q \neq \mathbf{0}$  et  $\{\boldsymbol{\epsilon}_t\}$  est un bruit blanc de dimension  $k \times 1$ .

En invoquant encore l'opérateur retard, le processus peut être réécrit de la façon compacte suivante :

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Theta}(B)\boldsymbol{\epsilon}_t, \quad (1.2.4)$$

avec  $\boldsymbol{\Theta}(B) = \mathbf{I} + \boldsymbol{\Theta}_1 B + \dots + \boldsymbol{\Theta}_q B^q$ .

Par la définition précédente, un processus VMA( $q$ ) est toujours stationnaire. Cependant, pour que le processus soit inversible, les racines de  $\det\{\boldsymbol{\Theta}(B)\}=0$  doivent être supérieures à un en module, où  $\det(\mathbf{A})$  est le déterminant de la matrice  $\mathbf{A}$ . Plus de détails se trouvent dans l'ouvrage Brockwell et Davis (1991).

Finalement, le modèle VARMA( $p, q$ ) est une combinaison des processus VAR( $p$ ) et VMA( $q$ ). Nous le définissons formellement.

**Définition 1.2.3.** *Un modèle vectoriel autorégressif à moyenne mobile d'ordre ( $p, q$ ), noté VARMA( $p, q$ ), satisfait l'équation aux différences stochastique :*

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{j=1}^p \Phi_j (\mathbf{Y}_{t-j} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t + \sum_{j=1}^q \boldsymbol{\Theta}_j \boldsymbol{\epsilon}_{t-j}, \quad (1.2.5)$$

où  $\mathbf{Y}_t$  est un vecteur aléatoire de dimension  $k \times 1$ ;  $\boldsymbol{\mu}$  est un vecteur de constantes de dimension  $k \times 1$ ;  $\boldsymbol{\Phi}_j$  est une matrice de paramètres de dimension  $k \times k$  pour  $j = 1, \dots, p$ ,  $\boldsymbol{\Phi}_p \neq \mathbf{0}$ ;  $\boldsymbol{\Theta}_j$  est une matrice de paramètres de dimension  $k \times k$  pour  $j = 1, \dots, q$ ,  $\boldsymbol{\Theta}_q \neq \mathbf{0}$  et  $\{\boldsymbol{\epsilon}_t\}$  est un bruit blanc de dimension  $k \times 1$ .

À nouveau, le processus peut être réécrit en faisant usage de l'opérateur retard :

$$\boldsymbol{\Phi}(B)(\mathbf{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(B)\boldsymbol{\epsilon}_t, \quad (1.2.6)$$

avec  $\boldsymbol{\Phi}(B) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \dots - \boldsymbol{\Phi}_p B^p$  et  $\boldsymbol{\Theta}(B) = \mathbf{I} + \boldsymbol{\Theta}_1 B + \dots + \boldsymbol{\Theta}_q B^q$ .

Un processus VARMA( $p, q$ ) est stationnaire si les racines de  $\det\{\boldsymbol{\Phi}(B)\}=0$  sont supérieures à un en module et est inversible si les racines de  $\det\{\boldsymbol{\Theta}(B)\}=0$  sont supérieures à un en module. On note que lorsque  $k = 1$  on retrouve en cas particulier les modèles AR, MA et ARMA. Les modèles venant d'être présentés ne tiennent pas compte de l'influence potentielle de variables observées à l'extérieur du système à l'étude. De telles variables sont dites exogènes alors que les variables déterminées à l'intérieur du système sont dites endogènes.

L'exogénéité est un concept fort utilisé dans les applications économétriques. Il existe plusieurs types d'exogénéité. En effet, on peut parler d'exogénéité faible, forte, stricte et même de super-exogénéité. Une condition technique particulièrement commode est l'exogénéité stricte selon laquelle les processus  $\{\mathbf{X}_t\}$  et  $\{\boldsymbol{\epsilon}_t\}$  sont supposés indépendants. Selon cette définition,  $\mathbf{X}_s$  et  $\boldsymbol{\epsilon}_t$  sont des vecteurs aléatoires indépendants pour tout  $s, t$ . Ainsi, lorsque nous ferons référence à l'exogénéité dans la présent mémoire, il s'agira de l'exogénéité stricte.

Nous présentons donc quelques modèles avec variables exogènes, à savoir les modèles VARX et VARMAX, qui constituent une grande classe de modèles de processus multivariés qui sous certaines conditions seront stationnaires. Ils reposent sur une série de variables de sortie (en anglais, on parle de variables d'*outputs*), et la variable  $\mathbf{Y}_t = (Y_t(1), Y_t(2), \dots, Y_t(k))^T$  est dite endogène, le processus endogène étant  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ , avec fonction moyenne  $E(\mathbf{Y}_t) = \boldsymbol{\mu}$  indépendante du temps  $t$  et  $\text{cov}(\mathbf{Y}_{t+l}, \mathbf{Y}_t) = E\{(\mathbf{Y}_{t+l} - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})^T\} = \boldsymbol{\Gamma}_Y(l)$  finie et indépendante du temps  $t$ . La matrice  $\boldsymbol{\Gamma}_Y(l)$  est l'autocovariance du processus  $\{\mathbf{Y}_t\}$  au délai  $l$ . Ce qui distingue les modèles VARX et VARMAX des modèles VAR et VARMA est l'utilisation de variables explicatives également appelées variables exogènes  $\mathbf{X}_t = (X_t(1), X_t(2), \dots, X_t(k_x))^T$  qui sont des variables d'entrées (ou des variables d'*input*). Ces séries sont présumées de moyennes nulles, c'est-à-dire que  $E(\mathbf{X}_t) = \mathbf{0}$  et de fonction d'autocovariance  $E(\mathbf{X}_{t+l}\mathbf{X}_t^T) = \boldsymbol{\Gamma}_X(l)$ , finie et indépendante de  $t$ .

**Définition 1.2.4.** *Un modèle vectoriel autorégressif à variables exogènes d'ordre  $(p, s)$ , noté VARX( $p, s$ ) satisfait l'équation suivante :*

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{Y}_{t-j} - \boldsymbol{\mu}) + \sum_{j=0}^s \mathbf{B}_j \mathbf{X}_{t-j} + \boldsymbol{\epsilon}_t, \quad (1.2.7)$$

où  $\mathbf{Y}_t$  est un vecteur aléatoire de dimension  $k \times 1$ ;  $\mathbf{X}_t$  est un vecteur aléatoire exogène de dimension  $k_x \times 1$ ,  $\boldsymbol{\mu}$  est un vecteur de constantes de dimension  $k \times 1$ ;  $\boldsymbol{\Phi}_j$  est une matrice de paramètres de dimension  $k \times k$  pour  $j = 1, \dots, p$ ,  $\boldsymbol{\Phi}_p \neq \mathbf{0}$ ;  $\mathbf{B}_j$  est une matrice de paramètres de dimension  $k \times k_x$  pour  $j = 0, \dots, s$ ,  $\mathbf{B}_s \neq \mathbf{0}$  et  $\{\boldsymbol{\epsilon}_t\}$  est un bruit blanc de dimension  $k \times 1$ .

Une fois de plus, le processus peut être réécrit en faisant usage de l'opérateur retard :

$$\boldsymbol{\Phi}(B)(\mathbf{Y}_t - \boldsymbol{\mu}) - \mathbf{B}(B)\mathbf{X}_t = \boldsymbol{\epsilon}_t, \quad (1.2.8)$$

avec  $\boldsymbol{\Phi}(B) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \dots - \boldsymbol{\Phi}_p B^p$  et  $\mathbf{B}(B) = \mathbf{B}_0 + \mathbf{B}_1 B + \dots + \mathbf{B}_s B^s$ .

Supposons que le processus exogène  $\{\mathbf{X}_t\}$  soit un processus stationnaire. Dans ce cas, le VARX( $p,s$ ) sera stationnaire si la portion autorégressive du processus est stationnaire. Le processus VARX sera donc stationnaire si les racines de  $\det \{\boldsymbol{\Phi}(B)\} = 0$  sont supérieures à un en module.

De façon analogue, un vecteur exogène peut être ajouté au modèle VARMA( $p,q$ ) afin d'obtenir le modèle VARMAX( $p,s,q$ ).

**Définition 1.2.5.** Un modèle vectoriel autorégressif à moyenne mobile à variables exogènes d'ordre ( $p,s,q$ ), noté VARMAX( $p,s,q$ ), satisfait l'équation suivante :

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{Y}_{t-j} - \boldsymbol{\mu}) + \sum_{j=0}^s \mathbf{B}_j \mathbf{X}_{t-j} + \boldsymbol{\epsilon}_t + \sum_{j=1}^q \boldsymbol{\Theta}_j \boldsymbol{\epsilon}_{t-j}, \quad (1.2.9)$$

où  $\{\mathbf{Y}_t\}$  est un vecteur de dimension  $k \times 1$ ;  $\{\mathbf{X}_t\}$  est un vecteur exogène de dimension  $k_x \times 1$ ,  $\boldsymbol{\mu}$  est un vecteur de constantes de dimension  $k \times 1$ ;  $\boldsymbol{\Phi}_j$  est une matrice de paramètres de dimension  $k \times k$  pour  $j = 1, \dots, p$ ,  $\boldsymbol{\Phi}_p \neq \mathbf{0}$ ;  $\mathbf{B}_j$  est une matrice de paramètres de dimension  $k \times k_x$  pour  $j = 0, \dots, s$ ,  $\mathbf{B}_s \neq \mathbf{0}$ ;  $\boldsymbol{\Theta}_j$  est une matrice de paramètres de dimension  $k \times k$  pour  $j = 0, \dots, q$ ,  $\boldsymbol{\Theta}_q \neq \mathbf{0}$  et  $\{\boldsymbol{\epsilon}_t\}$  est un bruit blanc de dimension  $k \times 1$ .

Le processus peut également être noté comme suit :

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(B)(\mathbf{Y}_{t-j} - \boldsymbol{\mu}) + \mathbf{B}(B)\mathbf{X}_{t-j} + \boldsymbol{\Theta}(B)\boldsymbol{\epsilon}_{t-j}, \quad (1.2.10)$$

avec  $\boldsymbol{\Phi}(B) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \dots - \boldsymbol{\Phi}_p B^p$ ,  $\mathbf{B}(B) = \mathbf{B}_0 + \mathbf{B}_1 B + \dots + \mathbf{B}_s B^s$  et  $\boldsymbol{\Theta}(B) = \mathbf{I} + \boldsymbol{\Theta}_1 B + \dots + \boldsymbol{\Theta}_q B^q$ .

Un inconvénient majeur des modèles classiques est le nombre élevé de paramètres à estimer, qui augmente rapidement avec les ordres  $p$ ,  $s$  et  $q$ . Ceci est particulièrement vrai dans le contexte multivarié et dans le cas où des vecteurs exogènes sont considérés. Dans la section suivante, nous présentons donc des modèles qui pallient à ce problème, à savoir les modèles avec sélection d'indices.

### 1.3. Modèles linéaires multivariés avec sélection des indices

Dans la modélisation des modèles multivariés, le nombre de paramètres à estimer augmente rapidement. Par exemple, pour un VAR( $p$ ) de dimension  $k$  et de moyenne nulle, il y a  $p$  matrices de paramètres de dimensions  $k \times k$ , donnant donc un total de  $k^2p$  paramètres à estimer. En présence de variables exogènes, le nombre de paramètres s'accroît davantage compte tenu des matrices associées aux variables exogènes contemporaines et retardées. Effectivement, pour un VARX( $p,s$ ) avec un vecteur endogène de dimension  $k$  et un vecteur exogène de dimension  $k_x$ , le nombre de paramètres à estimer devient  $k(kp + k_x(s + 1))$ .

Une première stratégie, dans la modélisation des modèles VAR, est de considérer une sous-sélection d'indices, expliquant le mieux le mieux le phénomène d'intérêt. Nous débutons avec les modèles VAR avec sélection d'indices avant d'introduire des variables exogènes qui s'ajoutent dans le modèle.

#### 1.3.1. Modèles SVAR

Plusieurs méthodes ont été développées dans le but de réduire la quantité de paramètres à estimer dans les modèles classiques vus précédemment, particulièrement dans les VAR( $p$ ). Parmi celles-ci se trouvent l'imposition de contraintes linéaires sur les coefficients. Plus récemment, la régression LASSO offre une approche alternative à la sélection des coefficients (voir, e.g., Tibshirani (1996) et Nardi et Rinaldo (2011)), ainsi que la sélection d'indices bayésienne. La première méthode est naturelle et consiste à poser certains coefficients comme nuls lors de l'estimation des paramètres. Par souci de parcimonie, il est raisonnable de vouloir contraindre à zéro des coefficients non significatifs. La recherche d'un meilleur modèle est étroitement liée à la réduction du nombre de paramètres. Dans la littérature portant sur les méthodes de type LASSO, ceci amène à pénaliser à zéro des coefficients jugés inutiles ou non significatifs. Dans un cadre de prévisions, des variables inutilement incluses dans le modèle engendrent potentiellement un bruit additionnel qui aura pour effet d'augmenter la variance de l'erreur de prévision. Ces restrictions peuvent être imputées de différentes manières. Certainement, une stratégie est d'examiner toutes les combinaisons possibles, d'ajuster tous les modèles, et de calculer des critères de sélection de modèles pour chaque modèle ajusté. Bien

que possible, ceci risque d'être coûteux en temps de calcul car il y aurait  $2^{k^2 p}$  modèles possibles en incluant le modèle complet. Par exemple, si un modèle VAR inclut  $k = 3$  variables, et s'il affiche un comportement saisonnier mensuel (menant à considérer  $p = 12$ ), alors un total de  $2^{9 \times 12} \approx 3 \times 10^{32}$  modèles doit être examiné.

Une approche alternative est l'élimination de matrices complètes, ce qui réduirait le nombre de modèles à examiner à  $2^p$ , chacune des matrices étant incluse ou non. Suivant cette approche, le meilleur modèle choisi est déterminé en se basant sur une autorégression basée sur un sous-ensemble d'indices. Par la suite, si désirée, des coefficients non-significatifs pourraient être retirés, menant à une parcimonie accrue.

Ces modèles sont particulièrement attrayants lorsqu'il est question de modéliser des phénomènes périodiques ou saisonniers. Des travaux sur les séries chronologiques multivariées saisonnières vont dans cette direction (Ursu et Duchesne (2009a, 2009b, 2009c), Duchesne et Lafaye de Micheaux (2013)). La sélection du meilleur sous-ensemble a généré beaucoup de travaux dans les modèles univariés, notamment McLeod et Zhang (2006, 2008). Nous présentons maintenant formellement le modèle SVAR( $\mathcal{I}$ ), où  $\mathcal{I}$  contient les indices retenus dans la structure autorégressive.

**Définition 1.3.1.** *Considérons  $\mathcal{I}_0 = \{1, \dots, p\}$ . Un modèle vectoriel autorégressif à sous-ensemble d'indices, noté SVAR( $\mathcal{I}$ ), satisfait l'équation aux différences stochastique :*

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{i \in \mathcal{I}} \boldsymbol{\Phi}_i (\mathbf{Y}_{t-i} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t, \quad (1.3.1)$$

où  $\mathcal{I} = \{i_1, \dots, i_a\}$  est un sous-ensemble de  $\mathcal{I}_0 = \{1, 2, \dots, p\}$ , avec  $i_1 < \dots < i_a$ , tel que  $\boldsymbol{\Phi}_i = \mathbf{0}$  pour  $i \notin \mathcal{I}$  et  $\boldsymbol{\Phi}_i \neq \mathbf{0}$  pour  $i \in \mathcal{I}$ .

De façon naturelle, on peut introduire les modèles SVAR avec variables exogènes, notés SVARX, qui sont introduits dans la prochaine section.

### 1.3.2. Modèles SVARX

De la même manière, les modèles SVARX sont obtenus en posant certaines matrices comme nulles. Nous présentons donc le modèle VARX à indices restreints.

**Définition 1.3.2.** *Un modèle vectoriel autorégressif avec variables exogènes à sous-ensemble d'indices, noté SVARX( $\mathcal{I}, \mathcal{J}$ ) satisfait l'équation :*

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{i \in \mathcal{I}} \boldsymbol{\Phi}_i (\mathbf{Y}_{t-i} - \boldsymbol{\mu}) + \sum_{j \in \mathcal{J}} \mathbf{B}_j \mathbf{X}_{t-j} + \boldsymbol{\epsilon}_t, \quad (1.3.2)$$

où  $\mathcal{I} = \{i_1, \dots, i_a\}$  est un sous-ensemble de  $\mathcal{I}_0 = \{1, 2, \dots, p\}$ ,  $i_1 < \dots < i_a$ , tel que  $\boldsymbol{\Phi}_i = \mathbf{0}$  pour  $i \notin \mathcal{I}$  et  $\boldsymbol{\Phi}_i \neq \mathbf{0}$  pour  $i \in \mathcal{I}$  et  $\mathcal{J} = \{j_1, \dots, j_b\}$  est un sous-ensemble de

$\mathcal{J}_0 = \{0, 1, \dots, s\}$ ,  $j_1 < \dots < j_b$ , tel que  $\mathbf{B}_j = \mathbf{0}$  pour  $j \notin \mathcal{J}$  et  $\mathbf{B}_j \neq \mathbf{0}$  pour  $j \in \mathcal{J}$ .

Bien entendu, les modèles à indices restreints sont présentés sans détailler à ce stade comment choisir les indices dans le modèle. Les techniques usuelles de sélection de modèles peuvent être utilisées, ce qui sera abordé plus loin dans ce mémoire.

Les concepts généraux concernant les séries chronologiques et les modèles à sous-ensemble d'indices étant maintenant présentés, nous procédons à la mise en oeuvre de l'estimation des moindres carrés pour ces modèles.

Selon nous et au mieux de notre connaissance, les résultats théoriques les plus satisfaisants sont ceux de Brockwell, Davis et Alexandre Trindade (2004), qui ont déjà exposé l'estimation sous des conditions précises pour les modèles SVAR. Nous exposerons ces résultats au chapitre suivant. Nous proposerons aussi une approche largement similaire, et nous nous inspirerons des développements théoriques de Brockwell, Davis et Alexandre Trindade (2004), pour effectuer l'estimation dans le contexte des modèles SVARX. Cette contribution technique constitue une contribution originale de ce mémoire, puisque les développements mathématiques semblaient absents de la littérature, toujours au mieux de notre connaissance.

## Chapitre 2

---

# Inférence statistique dans les modèles à sélection d'indices, avec statistiques de type diagnostic et calcul des prévisions

Dans ce chapitre, nous présentons l'estimation des moindres carrés pour les modèles autorégressifs vectoriels avec sélection d'indices, ainsi que des versions de ces modèles qui incluent des variables exogènes. Nous présentons d'abord l'estimateur des moindres carrés ainsi que sa loi asymptotique pour les modèles SVAR suivant l'approche rigoureuse décrite dans Brockwell, Davis et Alexandre Trindade (2004).

Par la suite, nous adaptons les arguments de Brockwell, Davis et Alexandre Trindade (2004) dans le cas des VAR avec variables exogènes et sélection d'indices. Plus précisément, nous obtenons les résultats de convergence en probabilité vers la vraie valeur ainsi que la normalité asymptotique de l'estimateur des moindres carrés dans le cadre des modèles SVARX sous des conditions précises. Afin d'établir la distribution asymptotique normale, un théorème central limite pour différences de martingales est invoqué.

Dans les applications, diagnostiquer l'ajustement d'un modèle de travail aux données disponibles représente une étape importante. Dans les modèles linéaires classiques, l'analyse des résidus fait partie d'une analyse complète. Dans un contexte de séries chronologiques, les autocorrélations résiduelles peuvent être étudiées et le calcul des statistiques de type portemanteau est courant. Suivant les résultats de Duchesne et Roy (2004), les tests portemanteaux généralisés reposant sur la méthode du noyau sont discutés pour diagnostiquer des modèles SVARX. Finalement, nous abordons le calcul des prévisions dans le modèle SVARX ainsi que les problèmes qui y sont associés.

### 2.1. Inférence statistique dans le modèle SVAR

Dans cette section, nous définissons tout d'abord l'estimateur des moindres carrés pour les modèles SVAR puis nous obtenons les propriétés asymptotiques de cet estimateur.

### 2.1.1. Estimation des moindres carrés dans le modèle SVAR

Brockwell, Davis et Alexandre Trindade (2004) ont développé les estimateurs des moindres carrés, de Burg et de Yule-Walker pour les modèles vectoriels autorégressifs à sélection d'indices. Les estimateurs des moindres carrés et de Yule-Walker sont bien connus dans les modèles univariés de séries chronologiques. En fait, l'estimateur de Yule-Walker est essentiellement associé à la méthode des moments, voir par exemple Shumway et Stoffer (2017). De plus, cette méthode permet d'obtenir les coefficients de la meilleure prévision linéaire de l'observation au temps  $p + 1$ , en fonction des observations aux temps  $1, \dots, p$ . La méthode de Burg est discutée dans Brockwell et Davis (2016), et repose sur une compréhension des autocorrélations partielles et de leurs propriétés. Puisque Brockwell, Davis et Alexandre Trindade (2004), utilisant un théorème limite central pour différences de martingales, ont démontré que ces trois estimateurs possédaient la même distribution asymptotique, nous privilégions dans ce mémoire uniquement l'étude de l'estimateur des moindres carrés.

Dans un contexte où l'on se concentre sur les prévisions, des approches alternatives pourraient faire appel à des techniques comme la régression PLS, et l'utilisation de la validation croisée serait différente des critères AIC et BIC. Cependant, dans un contexte de séries chronologiques, focaliser sur les délais est utile et c'est ce qui explique l'approche préconisée dans ce mémoire.

Dans cette section, nous rappelons les propriétés asymptotiques de cet estimateur dans le cadre des modèles SVAR. Ainsi, il n'y a pas de nouveaux résultats et nous nous proposons d'exposer les résultats principaux obtenus par Brockwell, Davis et Alexandre Trindade (2004). Nous supposons que le processus étudié est de moyenne nulle. Des résultats généraux stipulent que d'estimer la moyenne par la moyenne échantillonnale mène à un estimateur asymptotiquement efficace du paramètre moyenne et estimer la moyenne du processus et les paramètres du modèle sont deux considérations qui peuvent être faites séparément. Ainsi, peu est à gagner (au moins asymptotiquement) à estimer conjointement la moyenne et les autres paramètres par la méthode des moindres carrés. Pour ces raisons, on suppose la moyenne du processus connue et nulle. En pratique, on estime la moyenne du processus (ce dernier étant présumé stationnaire)  $E(\mathbf{Y}_t) = \boldsymbol{\mu}$  par la moyenne échantillonnale et on considère les observations centrées  $\mathbf{Y}_t - \bar{\mathbf{Y}}_n$ ,  $t = 1, \dots, n$ , avec  $\bar{\mathbf{Y}}_n = n^{-1} \sum_{t=1}^n \mathbf{Y}_t$ . La compréhension de cette section demeure importante pour la généralisation au cas des modèles SVARX.

Nous supposons comme hypothèse générale que le processus  $\{\mathbf{Y}_t\}$  est strictement stationnaire et ergodique. La notion d'ergodicité est un concept probabiliste qui veut dire essentiellement que des lois de type *Loi forte des grands nombres* s'appliquent. Francq et Zakoïan

(2010) expliquent le concept d'ergodicité et son utilisation en statistique. Ainsi on présume que  $\{\mathbf{Y}_t\}$  est l'unique solution stationnaire d'un SVAR( $\mathcal{I}$ ) :

$$\mathbf{Y}_t = \sum_{i \in \mathcal{I}} \Phi_i \mathbf{Y}_{t-i} + \boldsymbol{\epsilon}_t, \quad (2.1.1)$$

où  $\mathcal{I} = \{i_1, \dots, i_a\}$ ,  $i_1 < \dots < i_a$ , est un sous-ensemble de  $\{1, \dots, p\}$ , avec terme d'erreur  $\{\boldsymbol{\epsilon}_t\} \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_\boldsymbol{\epsilon})$ , avec  $\boldsymbol{\Sigma}_\boldsymbol{\epsilon}$  matrice définie positive. Puisque le processus  $\{\mathbf{Y}_t\}$  est de moyenne nulle, la fonction d'autocovariance satisfait et sera notée comme suit :  $\boldsymbol{\Gamma}(l) = E(\mathbf{Y}_{t+l} \mathbf{Y}_t^\top)$ ,  $l = 0, \pm 1, \dots$ . Comme Brockwell, Davis et Alexandre Trindade (2004), nous introduisons quelques matrices auxiliaires. La matrice  $\Phi_{\mathcal{I}} = (\Phi_{\mathcal{I}i_1}, \dots, \Phi_{\mathcal{I}i_a})$  est une matrice bloc de dimension  $k \times ka$  contenant les matrices de coefficients, alors que  $\boldsymbol{\Gamma}_{\mathcal{I}} = (\boldsymbol{\Gamma}(i_1), \dots, \boldsymbol{\Gamma}(i_a))$  est une matrice bloc de dimension  $k \times ka$  contenant les matrices des covariances. On pose également  $\mathbf{R}_{\mathcal{I}}(k, j) = \boldsymbol{\Gamma}(i_{j-1} - i_{k-1})$ ,  $k \leq j, k, j = 1, \dots, s+1$  avec  $i_0 = 0$ . Finalement,  $\mathbf{G}_{\mathcal{I}}$  est une matrice de dimension  $ka \times ka$  obtenue de  $\mathbf{R}_{\mathcal{I}}$  en supprimant la première ligne et la première colonne de chaque matrice  $k \times k$ .

Le modèle en (2.1.1) peut être réécrit sous forme matricielle de la façon suivante :

$$\mathbf{Y}_t = \begin{bmatrix} \Phi_{i_1} & \dots & \Phi_{i_a} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-i_1} \\ \vdots \\ \mathbf{Y}_{t-i_a} \end{bmatrix} + \boldsymbol{\epsilon}_t.$$

Afin de rendre l'estimateur des moindres carrés calculable, nous devons négliger les premières  $i_a$  observations. Ainsi, pour  $\mathbf{Y} = (\mathbf{Y}_{i_a+1}, \dots, \mathbf{Y}_n)$ , on peut écrire le modèle de manière compacte sous la forme d'un modèle de régression linéaire multivarié :

$$\mathbf{Y} = \Phi_{\mathcal{I}} \mathbf{M}_{\mathcal{I}} + \mathbf{E}, \quad (2.1.2)$$

avec  $\mathbf{E} = (\boldsymbol{\epsilon}_{i_a+1}, \dots, \boldsymbol{\epsilon}_n)$ ,  $\Phi_{\mathcal{I}} = (\Phi_{i_1} \dots \Phi_{i_a})$  et  $\mathbf{M}_{\mathcal{I}} = \left( \begin{bmatrix} \mathbf{Y}_{i_a+1-i_1} \\ \vdots \\ \mathbf{Y}_1 \end{bmatrix}, \dots, \begin{bmatrix} \mathbf{Y}_{n-i_1} \\ \vdots \\ \mathbf{Y}_{n-i_a} \end{bmatrix} \right)$ . On peut

ensuite appliquer l'opérateur  $\text{vec}(\cdot)$  de chaque côté de l'égalité (2.1.2). Des vecteurs utiles sont définis, à savoir  $\mathbf{y} = \text{vec}(\mathbf{Y})$ ,  $\boldsymbol{\alpha}_{\mathcal{I}} = \text{vec}(\Phi_{\mathcal{I}})$  et  $\mathbf{e} = \text{vec}(\mathbf{E})$ . On obtient ainsi :

$$\mathbf{y} = (\mathbf{M}_{\mathcal{I}}^\top \otimes \mathbf{I}_k) \boldsymbol{\alpha}_{\mathcal{I}} + \mathbf{e},$$

où  $\otimes$  est le produit de Kronecker. Dénotons  $\boldsymbol{\Sigma}_{\mathbf{E}} = \text{var}(\mathbf{e})$ . On aura donc :

$$\boldsymbol{\Sigma}_{\mathbf{E}} = \begin{bmatrix} \boldsymbol{\Sigma}_\boldsymbol{\epsilon} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\Sigma}_\boldsymbol{\epsilon} \end{bmatrix} = \mathbf{I}_{n-i_a} \otimes \boldsymbol{\Sigma}_\boldsymbol{\epsilon}.$$

L'estimateur des moindres carrés est obtenu en minimisant la quantité suivante :

$$S(\boldsymbol{\alpha}_{\mathcal{I}}) = \mathbf{e}^\top \boldsymbol{\Sigma}_{\mathbf{E}}^{-1} \mathbf{e}.$$

Ainsi, par dérivation on obtient les équations normales  $(\mathbf{M}_{\mathcal{I}}\mathbf{M}_{\mathcal{I}}^{\top} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1})\hat{\boldsymbol{\alpha}}_{\mathcal{I}} = (\mathbf{M}_{\mathcal{I}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1})\mathbf{y}$  dont la solution est :

$$\hat{\boldsymbol{\alpha}}_{\mathcal{I}} = \text{vec}\{\mathbf{Y}\mathbf{M}_{\mathcal{I}}^{\top}(\mathbf{M}_{\mathcal{I}}\mathbf{M}_{\mathcal{I}}^{\top})^{-1}\}.$$

Puisque  $\hat{\boldsymbol{\alpha}}_{\mathcal{I}} = \text{vec}(\hat{\boldsymbol{\Phi}}_{\mathcal{I}})$ , on déduit que l'estimateur des moindres carrés satisfait :

$$\begin{aligned} \hat{\boldsymbol{\Phi}}_{\mathcal{I}} &= \mathbf{Y}\mathbf{M}_{\mathcal{I}}^{\top}(\mathbf{M}_{\mathcal{I}}\mathbf{M}_{\mathcal{I}}^{\top})^{-1}, \\ &= (\boldsymbol{\Phi}_{\mathcal{I}}\mathbf{M}_{\mathcal{I}} + \mathbf{E})\mathbf{M}_{\mathcal{I}}^{\top}(\mathbf{M}_{\mathcal{I}}\mathbf{M}_{\mathcal{I}}^{\top})^{-1}, \\ &= \boldsymbol{\Phi}_{\mathcal{I}} + \mathbf{E}\mathbf{M}_{\mathcal{I}}^{\top}(\mathbf{M}_{\mathcal{I}}\mathbf{M}_{\mathcal{I}}^{\top})^{-1}. \end{aligned} \tag{2.1.3}$$

La prochaine section contient les propriétés asymptotiques de l'estimateur des moindres carrés pour les modèles SVAR.

### 2.1.2. Propriétés asymptotiques du modèle SVAR

Nous venons de montrer en (2.1.3) que l'estimateur des moindres carrés peut être exprimé en fonction des vraies matrices de paramètres. Ceci sera utile afin de déterminer les propriétés asymptotiques de cet estimateur. Plus particulièrement, nous montrerons la normalité asymptotique de cet estimateur pour les modèles vectoriels autorégressifs avec sélection d'indices. Nous débutons par énoncer deux lemmes importants.

**Lemme 2.1.1.** *Pour un processus  $\{\mathbf{Y}_t\}$  satisfaisant l'équation (2.1.1) où le bruit  $\{\boldsymbol{\epsilon}_t\} \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$ , on a le résultat de convergence en probabilité suivant :*

$$n^{-1}\mathbf{M}_{\mathcal{I}}\mathbf{M}_{\mathcal{I}}^{\top} \xrightarrow{p} \mathbf{G}_{\mathcal{I}}.$$

Le normalité asymptotique de l'estimateur des moindres carrés provient de l'étude de  $n^{-1/2}\text{vec}(\mathbf{E}\mathbf{M}_{\mathcal{I}}^{\top})$ . Ceci est rendu plus précis dans le lemme suivant.

**Lemme 2.1.2.** *Pour un processus  $\{\mathbf{Y}_t\}$  satisfaisant l'équation (2.1.1) où le bruit  $\{\boldsymbol{\epsilon}_t\} \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$ , le vecteur  $n^{-1/2}\text{vec}(\mathbf{E}\mathbf{M}_{\mathcal{I}}^{\top})$  est asymptotiquement normal :*

$$n^{-1/2}\text{vec}(\mathbf{E}\mathbf{M}_{\mathcal{I}}^{\top}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{G}_{\mathcal{I}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}).$$

Afin de montrer la normalité asymptotique de l'estimateur des moindres carrés, Brockwell, Davis et Alexandre Trindade (2004) font appel à l'outil de Cramér-Wold, qui consiste à montrer la normalité asymptotique pour une combinaison linéaire arbitraire en introduisant un certain vecteur fixe et arbitraire  $\boldsymbol{\lambda}$ . Pour une description de la technique de Cramér-Wold,

voir Serfling (1980). Dans notre contexte, dénotons

$$\mathbf{U}_t = \text{vec} \left( \boldsymbol{\epsilon}_t \begin{bmatrix} \mathbf{Y}_{t-i_1}^\top & \cdots & \mathbf{Y}_{t-i_a}^\top \end{bmatrix} \right) = \begin{bmatrix} \mathbf{Y}_{t-i_1} \\ \vdots \\ \mathbf{Y}_{t-i_a} \end{bmatrix} \otimes \boldsymbol{\epsilon}_t.$$

Nous introduisons le vecteur fixe  $\boldsymbol{\lambda} \in \mathbb{R}^{k^2 a}$  et nous considérons

$$\mathbf{W}_{n,t} = n^{-1/2} \boldsymbol{\lambda}^\top \mathbf{U}_t.$$

On remarque que  $\{\mathbf{W}_{n,t}\}$  satisfait la condition de différence de martingales, puisque

$$E(\mathbf{W}_{n,t} \mid \mathcal{F}_{t-1}) = \mathbf{0},$$

où  $\mathcal{F}_t$  dénote l'information disponible au temps  $t$  (on note que pour les besoins de la cause l'interprétation usuelle est suffisante, mais une définition probabiliste plus précise de  $\mathcal{F}_t$  utilisant les  $\sigma$ -algèbres est possible, mais pas nécessaire dans notre contexte). On peut consulter Lütkepohl (2005, p. 689) pour plus de détails sur les différences de martingales. En utilisant les propriétés du produit de Kronecker, on aura alors que :

$$\begin{aligned} E(\mathbf{U}_t \mathbf{U}_t^\top) &= E \left( \begin{bmatrix} \mathbf{Y}_{t-i_1} \\ \vdots \\ \mathbf{Y}_{t-i_a} \end{bmatrix} \otimes \boldsymbol{\epsilon}_t \right) \left( \begin{bmatrix} \mathbf{Y}_{t-i_1}^\top & \cdots & \mathbf{Y}_{t-i_a}^\top \end{bmatrix} \otimes \boldsymbol{\epsilon}_t^\top \right), \\ &= E \left( \begin{bmatrix} \mathbf{Y}_{t-i_1} \\ \vdots \\ \mathbf{Y}_{t-i_a} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-i_1}^\top & \cdots & \mathbf{Y}_{t-i_a}^\top \end{bmatrix} \otimes \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \right), \\ &= E \left( \begin{bmatrix} \mathbf{Y}_{t-i_1} \mathbf{Y}_{t-i_1}^\top & \cdots & \mathbf{Y}_{t-i_1} \mathbf{Y}_{t-i_a}^\top \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_{t-i_a} \mathbf{Y}_{t-i_1}^\top & \cdots & \mathbf{Y}_{t-i_a} \mathbf{Y}_{t-i_a}^\top \end{bmatrix} \right) \otimes E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top) = \mathbf{G}_I \otimes \boldsymbol{\Sigma}_\epsilon. \end{aligned}$$

On obtient également que

$$\text{var}(\mathbf{W}_{n,t}) = n^{-1} \boldsymbol{\lambda}^\top E(\mathbf{U}_t \mathbf{U}_t^\top) \boldsymbol{\lambda} = n^{-1} \boldsymbol{\lambda}^\top (\mathbf{G}_I \otimes \boldsymbol{\Sigma}_\epsilon) \boldsymbol{\lambda}. \quad (2.1.4)$$

Selon Brockwell, Davis et Alexandre Trindade (2004),  $\{\mathbf{W}_{n,t}\}$  satisfait les conditions suivantes :

- (1)  $\sum_{t=1}^n E \left[ \mathbf{W}_{n,t}^2 \mid \mathcal{F}_{t-1} \right] \xrightarrow{p} \boldsymbol{\lambda}^\top (\mathbf{G}_I \otimes \boldsymbol{\Sigma}_\epsilon) \boldsymbol{\lambda}$ ,
- (2)  $\max_{1 \leq t \leq n} |\mathbf{W}_{n,t}| \xrightarrow{p} \mathbf{0}$ ,
- (3)  $E(\max_{1 \leq t \leq n} \mathbf{W}_{n,t}^2)$  est borné en  $n$ .

Ces conditions sont suffisantes pour obtenir la normalité asymptotique. En effet, selon Hall and Heyde (1980, Théorème 3.2, Corollaire 3.1), Brockwell, Davis et Alexandre Trindade

(2004) montrent le résultat suivant :

$$n^{-1/2} \boldsymbol{\lambda}^\top \text{vec}(\mathbf{E} \mathbf{M}_{\mathcal{I}}^\top) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\lambda}^\top (\mathbf{G}_{\mathcal{I}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}) \boldsymbol{\lambda}).$$

On peut donc établir la convergence en probabilité de l'estimateur des moindres carrés pour les modèles autorégressifs avec sélection d'indices ainsi qu'un théorème limite central. Ceci est résumé dans le théorème suivant.

**Théorème 2.1.3.** *Considérons un processus  $\{\mathbf{Y}_t\}$  satisfaisant l'équation (2.1.1) où le bruit  $\{\boldsymbol{\epsilon}_t\} \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$ . On estime les paramètres du modèle par moindres carrés. L'estimateur des moindres carrés est convergent en probabilité et sa distribution asymptotique est normale. Plus précisément :*

- (a)  $\hat{\boldsymbol{\Phi}}_I \xrightarrow{p} \boldsymbol{\Phi}_I,$
- (b)  $n^{1/2} (\hat{\boldsymbol{\alpha}}_I - \boldsymbol{\alpha}_I) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{G}_{\mathcal{I}}^{-1} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}).$

La convergence en probabilité et la normalité asymptotique de l'estimateur des moindres carrés étant obtenues pour le modèle SVAR, nous allons établir des résultats similaires dans le contexte des modèles SVARX, sous des conditions comparables, mais pas exactement identiques (compte tenu de la présence de variables exogènes stochastiques), à ce qui est fait dans Brockwell, Davis et Alexandre Trindade (2004). Nous présentons notre démarche dans la prochaine section.

## 2.2. Inférence statistique et vérification diagnostique dans le modèle SVARX

Afin de parvenir à démontrer la convergence en probabilité et la normalité asymptotique de l'estimateur des moindres carrés pour les modèles SVARX, nous commençons par obtenir une expression explicite de cet estimateur.

### 2.2.1. Estimation des moindres carrés dans le modèle SVARX

Afin d'étudier l'estimateur des moindres carrés dans le contexte des modèles SVARX, nos points de départ sont Brockwell, Davis et Alexandre Trindade (2004) pour les SVAR, ainsi que la preuve esquissée dans Lütkepohl (1993) pour les modèles VARX complets. Notre but est de justifier rigoureusement l'approche heuristique de Lütkepohl (1993), dans le contexte des modèles autorégressifs à variables exogènes et choix des indices. Nous emprunterons une notation semblable à celle de la section précédente. Nous supposons, sans pertes de généralités, que la moyenne des processus endogène et exogène est nulle. Le modèle VARX complet est :

$$\mathbf{Y}_t = \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{Y}_{t-j} + \sum_{j=0}^s \mathbf{B}_j \mathbf{X}_{t-j} + \boldsymbol{\epsilon}_t, \quad (2.2.1)$$

tandis que le modèle à sous-ensemble d'indices est :

$$\mathbf{Y}_t = \Phi_{\mathcal{I}_{i_1}} \mathbf{Y}_{t-i_1} + \dots + \Phi_{\mathcal{I}_{i_a}} \mathbf{Y}_{t-i_a} + \mathbf{B}_{\mathcal{J}_{j_1}} \mathbf{X}_{t-j_1} + \dots + \mathbf{B}_{\mathcal{J}_{j_b}} \mathbf{X}_{t-j_b} + \epsilon_t. \quad (2.2.2)$$

Définissons les ensembles d'indices  $\mathcal{I} = \{i_1, i_2, \dots, i_a\}$ ,  $i_1 < \dots < i_a$  et  $\mathcal{J} = \{j_1, j_2, \dots, j_b\}$ ,  $j_1 < \dots < j_b$ . On suppose que  $\mathcal{I} \subset \mathcal{I}_0$ ,  $\mathcal{J} \subset \mathcal{J}_0$ , avec  $\mathcal{I}_0 = \{1, 2, \dots, p\}$  et  $\mathcal{J}_0 = \{0, 1, \dots, s\}$ . Ainsi,  $a$  est le nombre de termes de la partie autorégressive et  $b$  est le nombre de terme de la partie exogène du modèle SVARX. Nous introduisons également  $\Phi_{\mathcal{I}} = (\Phi_{i_1} \dots \Phi_{i_a})$  de dimension  $k \times ka$  et  $\mathbf{B}_{\mathcal{J}} = (\mathbf{B}_{j_1} \dots \mathbf{B}_{j_b})$  de dimension  $k \times k_x b$ . Le modèle restreint (2.2.2) peut être réécrit sous forme matricielle de la façon suivante

$$\mathbf{Y}_t = \begin{bmatrix} \Phi_{i_1} & \dots & \Phi_{i_a} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-i_1} \\ \vdots \\ \mathbf{Y}_{t-i_a} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_{j_1} & \dots & \mathbf{B}_{j_b} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-j_1} \\ \vdots \\ \mathbf{X}_{t-j_b} \end{bmatrix} + \epsilon_t. \quad (2.2.3)$$

Posons  $d = \max(i_a, j_b) + 1$ . En outre, supposons que les erreurs  $\{\epsilon_t\}$  sont indépendantes et identiquement distribuées telles que  $\{\epsilon_t\} \sim IID(\mathbf{0}, \Sigma_{\epsilon})$ . Le modèle SVARX sous forme matricielle peut alors être reformulé comme un modèle de régression multivariée. Effectivement, nous commençons par définir :

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} \mathbf{Y}_{d+1} & \dots & \mathbf{Y}_n \end{bmatrix}, \\ \mathbf{U}_{\mathcal{I}} &= \begin{bmatrix} \mathbf{Y}_{d+1-i_1} & \dots & \mathbf{Y}_{n-i_1} \\ \vdots & & \vdots \\ \mathbf{Y}_{d+1-i_a} & \dots & \mathbf{Y}_{n-i_a} \end{bmatrix}, \\ \mathbf{V}_{\mathcal{J}} &= \begin{bmatrix} \mathbf{X}_{d+1-j_1} & \dots & \mathbf{X}_{n-j_1} \\ \vdots & & \vdots \\ \mathbf{X}_{d+1-j_b} & \dots & \mathbf{X}_{n-j_b} \end{bmatrix}, \\ \mathbf{E} &= \begin{bmatrix} \epsilon_{d+1} & \dots & \epsilon_n \end{bmatrix}. \end{aligned}$$

Les matrices  $\mathbf{Y}$ ,  $\mathbf{U}_{\mathcal{I}}$ ,  $\mathbf{V}_{\mathcal{J}}$  et  $\mathbf{E}$  sont de dimension  $k \times (n-d)$ ,  $ka \times (n-d)$ ,  $k_x b \times (n-d)$  et  $k \times (n-d)$ , respectivement. Avec ces matrices, le modèle peut être réécrit sous forme d'un modèle de régression multivariée, à savoir :

$$\mathbf{Y} = \Phi_{\mathcal{I}} \mathbf{U}_{\mathcal{I}} + \mathbf{B}_{\mathcal{J}} \mathbf{V}_{\mathcal{J}} + \mathbf{E}. \quad (2.2.4)$$

De façon encore plus succincte, nous pouvons dire :

$$\mathbf{Y} = \begin{bmatrix} \Phi_{\mathcal{I}} & \mathbf{B}_{\mathcal{J}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\mathcal{I}} \\ \mathbf{V}_{\mathcal{J}} \end{bmatrix} + \mathbf{E}, \quad (2.2.5)$$

$$= \beta_{\mathcal{I}\mathcal{J}} \mathbf{Z} + \mathbf{E}, \quad (2.2.6)$$

avec  $\beta_{\mathcal{IJ}} = [\Phi_{\mathcal{I}} \ \mathbf{B}_{\mathcal{J}}]$  et  $\mathbf{Z} = \begin{bmatrix} \mathbf{U}_{\mathcal{I}} \\ \mathbf{V}_{\mathcal{J}} \end{bmatrix}$ . Avant de continuer, nous rappelons une propriété pratique de l'opérateur  $\text{vec}(\cdot)$ . Considérons des matrices  $\mathbf{A}$ ,  $\mathbf{B}$  et  $\mathbf{C}$ , on aura alors que  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^{\top} \otimes \mathbf{A})\text{vec}(\mathbf{B})$ . Ainsi, nous pouvons appliquer l'opérateur vecteur de chaque côté de (2.2.6) et nous obtenons :

$$\mathbf{y} = (\mathbf{Z}^{\top} \otimes \mathbf{I}_k)\text{vec}(\beta_{\mathcal{IJ}}) + \mathbf{e}. \quad (2.2.7)$$

Il est utile de noter les dimensions des matrices de (2.2.7). On a  $\mathbf{y} = \text{vec}(\mathbf{Y})$  qui est de dimension  $k(n-d) \times 1$  puisque la matrice  $\mathbf{Y}$  est de dimension  $k \times (n-d)$ . On a ensuite  $\text{vec}(\beta_{\mathcal{IJ}}\mathbf{Z}) = \text{vec}(\mathbf{I}_k\beta_{\mathcal{IJ}}\mathbf{Z})$ , où  $\mathbf{I}_k$  est la matrice identité de dimension  $k \times k$ . On rappelle que  $\beta_{\mathcal{IJ}} = [\Phi_{\mathcal{I}} \ \mathbf{B}_{\mathcal{J}}]$  est de dimension  $k \times (ka + k_xb)$  faisant en sorte que  $\text{vec}(\beta_{\mathcal{IJ}}\mathbf{Z})$  est de dimension  $k(ka + k_xb) \times 1$ . Pour finir, on note que  $\mathbf{e} = \text{vec}(\mathbf{E})$  est de dimension  $k(n-d) \times 1$ . L'estimateur des moindres carrés calculé dans le modèle avec sélection d'indices est un cas particulier de la démarche générale trouvée dans Lütkepohl (2005) pour les modèles VARX avec contraintes linéaires sur les paramètres. On écrit

$$\begin{aligned} \mathbf{y} &= (\mathbf{Z}^{\top} \otimes \mathbf{I}_k)\text{vec}(\beta_{\mathcal{IJ}}) + \mathbf{e}, \\ &= (\mathbf{Z}^{\top} \otimes \mathbf{I}_k)\mathbf{b}_{\mathcal{IJ}} + \mathbf{e}, \end{aligned}$$

en notant  $\mathbf{b}_{\mathcal{IJ}} = \text{vec}(\beta_{\mathcal{IJ}})$ . On peut donc calculer l'estimateur des moindres carrés de  $\mathbf{b}_{\mathcal{IJ}}$  :

$$\begin{aligned} \hat{\mathbf{b}}_{\mathcal{IJ}} &= (\mathbf{ZZ}^{\top} \otimes \mathbf{I}_k)^{-1}(\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{y}, \\ &= \{(\mathbf{ZZ}^{\top})^{-1} \otimes \mathbf{I}_k\} (\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{y}, \\ &= \{(\mathbf{ZZ}^{\top})^{-1}\mathbf{Z} \otimes \mathbf{I}_k\} \mathbf{y}. \end{aligned}$$

Cet estimateur utilise la structure particulière des contraintes se trouvant dans Lütkepohl (2005) et se trouve essentiellement dans Brockwell, Davis et Alexandre Trindade (2004) dont nous reprenons les arguments.

$$\begin{aligned} \hat{\mathbf{b}}_{\mathcal{IJ}} &= \{(\mathbf{ZZ}^{\top})^{-1}\mathbf{Z} \otimes \mathbf{I}_k\} \mathbf{y}, \\ &= \{(\mathbf{ZZ}^{\top})^{-1}\mathbf{Z} \otimes \mathbf{I}_k\} (\mathbf{Z}^{\top} \otimes \mathbf{I}_k\mathbf{b}_{\mathcal{IJ}} + \mathbf{e}), \\ &= \mathbf{b}_{\mathcal{IJ}} + \{(\mathbf{ZZ}^{\top})^{-1}\mathbf{Z} \otimes \mathbf{I}_k\} \mathbf{e}. \end{aligned}$$

Aussi, puisque  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^{\top} \otimes \mathbf{A})\text{vec}(\mathbf{B})$ , on peut écrire

$$\begin{aligned} \hat{\mathbf{b}}_{\mathcal{IJ}} &= \{(\mathbf{ZZ}^{\top})^{-1}\mathbf{Z} \otimes \mathbf{I}_k\} \mathbf{y}, \\ &= \text{vec} \{ \mathbf{YZ}^{\top}(\mathbf{ZZ}^{\top})^{-1} \}. \end{aligned}$$

Donc

$$\begin{aligned}\hat{\mathbf{b}}_{\mathcal{IJ}} &= (\mathbf{ZZ}^\top \otimes \mathbf{I}_k)^{-1} (\mathbf{Z} \otimes \mathbf{I}_k) \mathbf{y}, \\ &= \text{vec} \left\{ \mathbf{YZ}^\top (\mathbf{ZZ}^\top)^{-1} \right\}.\end{aligned}$$

La dernière relation peut être computationnellement intéressante car elle ne repose pas sur le calcul de produits de Kronecker mais seulement sur la vectorisation d'un produit de matrices, ce qui peut être fait efficacement avec les logiciels disponibles. On trouve alors un résultat analogue aux formules (12) et (13) de Brockwell, Davis et Alexandre Trindade (2004). Plus précisément :

$$\begin{aligned}\hat{\beta}_{\mathcal{IJ}} &= \mathbf{YZ}^\top (\mathbf{ZZ}^\top)^{-1}, \\ &= (\beta_{\mathcal{IJ}} \mathbf{Z} + \mathbf{E}) \mathbf{Z}^\top (\mathbf{ZZ}^\top)^{-1}, \\ &= \beta_{\mathcal{IJ}} + \mathbf{EZ}^\top (\mathbf{ZZ}^\top)^{-1}.\end{aligned}\tag{2.2.8}$$

L'estimateur des moindres carrés pour les modèles SVARX étant maintenant obtenu, il devient possible d'étudier les propriétés asymptotiques de cet estimateur. C'est ce qui sera abordé dans la prochaine section.

## 2.2.2. Propriétés asymptotiques du modèle SVARX

Nous nous intéressons particulièrement à la convergence en probabilité ainsi qu'à la loi asymptotique de l'estimateur des moindres carrés pour le modèle SVARX. Effectivement, comme dans le cas des modèles SVAR, nous voulons montrer que l'estimateur des moindres carrés tend en probabilité vers les vrais paramètres et nous voulons également montrer qu'il tend en loi vers une distribution normale. Un intérêt de la distribution limite multinormale est la matrice des variances et covariances, qui représente un résultat important afin d'effectuer de l'inférence statistique.

Pour se faire, nous débutons avec l'expression (2.2.8). Il est utile de rappeler que  $\mathbf{E}$  est une matrice aléatoire telle que  $\mathbf{E} = (\epsilon_{d+1} \dots \epsilon_n)$  de dimension  $k \times (n - d)$  et  $\epsilon_i \sim \mathcal{L}(\mathbf{0}, \Sigma_\epsilon)$ , c'est-à-dire que  $\epsilon_i$  admet une certaine loi dont les deux premiers moments sont  $\mathbf{0}$  et  $\Sigma_\epsilon$ . Il est également utile de noter que  $\mathbf{Z} = \begin{bmatrix} \mathbf{U}_{\mathcal{I}} \\ \mathbf{V}_{\mathcal{J}} \end{bmatrix}$  de dimension  $k(k_a + k_x b) \times (n - d)$ .

Notons tout d'abord que

$$\begin{aligned}\hat{\beta}_{\mathcal{IJ}} &= \mathbf{YZ}^\top (\mathbf{ZZ}^\top)^{-1}, \\ &= \beta_{\mathcal{IJ}} + \mathbf{EZ}^\top (\mathbf{ZZ}^\top)^{-1}, \\ &= \beta_{\mathcal{IJ}} + n^{-1} \mathbf{EZ}^\top (n^{-1} \mathbf{ZZ}^\top)^{-1}.\end{aligned}$$

On aura donc

$$\hat{\beta}_{\mathcal{I}\mathcal{J}} - \beta_{\mathcal{I}\mathcal{J}} = n^{-1}\mathbf{E}\mathbf{Z}^\top \left( n^{-1}\mathbf{Z}\mathbf{Z}^\top \right)^{-1},$$

et

$$n^{1/2} \left( \hat{\beta}_{\mathcal{I}\mathcal{J}} - \beta_{\mathcal{I}\mathcal{J}} \right) = n^{-1/2}\mathbf{E}\mathbf{Z}^\top \left( n^{-1}\mathbf{Z}\mathbf{Z}^\top \right)^{-1}.$$

Afin d'étudier le comportement asymptotique de  $n^{1/2} \left( \hat{\beta}_{\mathcal{I}\mathcal{J}} - \beta_{\mathcal{I}\mathcal{J}} \right)$  nous devons donc examiner les termes  $n^{-1/2}\mathbf{E}\mathbf{Z}^\top$  et  $\left( n^{-1}\mathbf{Z}\mathbf{Z}^\top \right)$  lorsque  $n \rightarrow \infty$ . Ceci est essentiellement ce qui a été fait dans les lemmes 2.1.1 et 2.1.2.

Nous commençons notre étude par le terme  $n^{-1}\mathbf{Z}\mathbf{Z}^\top$ . On avait  $\mathbf{Z} = \begin{bmatrix} \mathbf{U}_{\mathcal{I}} \\ \mathbf{V}_{\mathcal{J}} \end{bmatrix}$ .

En effectuant le produit matriciel, nous avons donc

$$\mathbf{Z}\mathbf{Z}^\top = \begin{bmatrix} \mathbf{U}_{\mathcal{I}} \\ \mathbf{V}_{\mathcal{J}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\mathcal{I}}^\top & \mathbf{V}_{\mathcal{J}}^\top \end{bmatrix}, \quad (2.2.9)$$

$$= \begin{bmatrix} \mathbf{U}_{\mathcal{I}}\mathbf{U}_{\mathcal{I}}^\top & \mathbf{U}_{\mathcal{I}}\mathbf{V}_{\mathcal{J}}^\top \\ \mathbf{V}_{\mathcal{J}}\mathbf{U}_{\mathcal{I}}^\top & \mathbf{V}_{\mathcal{J}}\mathbf{V}_{\mathcal{J}}^\top \end{bmatrix}. \quad (2.2.10)$$

Par souci de clarté, désignons les éléments de la matrice  $n^{-1}\mathbf{Z}\mathbf{Z}^\top$  :

$$n^{-1}\mathbf{Z}\mathbf{Z}^\top = \begin{bmatrix} A_{11n} & A_{12n} \\ A_{12n}^\top & A_{22n} \end{bmatrix},$$

avec  $A_{11n} = n^{-1}\mathbf{U}_{\mathcal{I}}\mathbf{U}_{\mathcal{I}}^\top$ ,  $A_{12n} = n^{-1}\mathbf{U}_{\mathcal{I}}\mathbf{V}_{\mathcal{J}}^\top$  et  $A_{22n} = n^{-1}\mathbf{V}_{\mathcal{J}}\mathbf{V}_{\mathcal{J}}^\top$ . Nous allons analyser les composantes de la matrice  $n^{-1}\mathbf{Z}\mathbf{Z}^\top$  en débutant par  $A_{11n}$ . On a

$$\begin{aligned} \mathbf{U}_{\mathcal{I}} &= \begin{bmatrix} \mathbf{Y}_{d+1-i_1} & \cdots & \mathbf{Y}_{n-i_1} \\ \vdots & & \vdots \\ \mathbf{Y}_{d+1-i_a} & \cdots & \mathbf{Y}_{n-i_a} \end{bmatrix}, \\ &= \begin{bmatrix} \mathbf{u}_{\mathcal{I},d+1} & \cdots & \mathbf{u}_{\mathcal{I},n} \end{bmatrix}, \end{aligned}$$

que l'on obtient en posant  $\mathbf{u}_{\mathcal{I},i} = \begin{bmatrix} \mathbf{Y}_{i-i_1} \\ \vdots \\ \mathbf{Y}_{i-i_a} \end{bmatrix}$  pour  $i = d+1, \dots, n$ . On obtient ainsi

$$\begin{aligned} A_{11n} &= n^{-1}\mathbf{U}_{\mathcal{I}}\mathbf{U}_{\mathcal{I}}^\top, \\ &= n^{-1} \begin{bmatrix} \mathbf{u}_{\mathcal{I},d+1} & \cdots & \mathbf{u}_{\mathcal{I},n} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{\mathcal{I},d+1}^\top \\ \vdots \\ \mathbf{u}_{\mathcal{I},n}^\top \end{bmatrix}, \\ &= n^{-1} \sum_{i=d+1}^n \mathbf{u}_{\mathcal{I},i}\mathbf{u}_{\mathcal{I},i}^\top, \end{aligned}$$

où on rappelle que  $\mathcal{I} = \{i_1, i_2, \dots, i_a\}$  est un l'ensemble des indices temporels de la partie autorégressive inclus dans le modèle SVARX et est de cardinalité  $a$ . Sous les mêmes hypothèses que dans Brockwell, Davis et Alexandre Trindade (2004), nous avons que :

$$\begin{aligned} n^{-1}\mathbf{U}_{\mathcal{I}}\mathbf{U}_{\mathcal{I}}^{\top} &\xrightarrow{p} E(\mathbf{u}_{\mathcal{I},1}\mathbf{u}_{\mathcal{I},1}^{\top}) = \begin{bmatrix} E(\mathbf{Y}_{1-i_1}\mathbf{Y}_{1-i_1}^{\top}) & \dots & E(\mathbf{Y}_{1-i_1}\mathbf{Y}_{1-i_a}^{\top}) \\ \vdots & & \vdots \\ E(\mathbf{Y}_{1-i_a}\mathbf{Y}_{1-i_1}^{\top}) & \dots & E(\mathbf{Y}_0\mathbf{Y}_0^{\top}) \end{bmatrix}, \\ &= [\mathbf{\Gamma}_{\mathbf{Y}}(i_d - i_e)], \end{aligned}$$

avec  $d, e = 1, \dots, i_a$ . Étudions à présent l'élément  $A_{22n}$  de la matrice  $n^{-1}\mathbf{Z}\mathbf{Z}^{\top}$ . On aait

$$\begin{aligned} \mathbf{V}_{\mathcal{J}} &= \begin{bmatrix} \mathbf{X}_{d+1-j_1} & \dots & \mathbf{X}_{n-j_1} \\ \vdots & & \vdots \\ \mathbf{X}_{d+1-j_b} & \dots & \mathbf{X}_{n-j_b} \end{bmatrix}, \\ &= [\mathbf{x}_{\mathcal{J},d+1} \quad \dots \quad \mathbf{x}_{\mathcal{J},n}], \end{aligned}$$

que l'on obtient en posant  $\mathbf{x}_{\mathcal{J},j} = \begin{bmatrix} \mathbf{X}_{j-j_1} \\ \vdots \\ \mathbf{X}_{j-j_a} \end{bmatrix}$  pour  $j = d+1, \dots, n$ . Cela mène à

$$\begin{aligned} A_{22n} &= n^{-1}\mathbf{V}_{\mathcal{J}}\mathbf{V}_{\mathcal{J}}^{\top}, \\ &= n^{-1} [\mathbf{x}_{\mathcal{J},d+1} \quad \dots \quad \mathbf{x}_{\mathcal{J},n}] \begin{bmatrix} \mathbf{x}_{\mathcal{J},d+1}^{\top} \\ \vdots \\ \mathbf{x}_{\mathcal{J},n}^{\top} \end{bmatrix}, \\ &= n^{-1} \sum_{j=d+1}^n \mathbf{x}_{\mathcal{J},j}\mathbf{x}_{\mathcal{J},j}^{\top}, \end{aligned}$$

où l'on rappelle que  $\mathcal{J} = \{j_1, j_2, \dots, j_b\}$  est l'ensemble des indices temporels de la portion exogène inclus dans le modèle SVARX et est de cardinalité  $b$ . De la même façon que pour l'étude de  $A_{11n}$  et sous les mêmes hypothèses, nous parvenons au résultat suivant :

$$n^{-1}\mathbf{V}_{\mathcal{J}}\mathbf{V}_{\mathcal{J}}^{\top} \xrightarrow{p} E(\mathbf{x}_{\mathcal{J},1}\mathbf{x}_{\mathcal{J},1}^{\top}) = [\mathbf{\Gamma}_{\mathbf{X}}(j_d - j_e)],$$

avec  $d, e = 1, \dots, j_b$ .

Finalement, nous étudions l'expression  $A_{12n}$  de la matrice  $n^{-1}\mathbf{Z}\mathbf{Z}^{\top}$ . Nous avons que  $\mathbf{U}_{\mathcal{I}} = [\mathbf{u}_{\mathcal{I},d+1} \quad \dots \quad \mathbf{u}_{\mathcal{I},n}]$  et  $\mathbf{V}_{\mathcal{J}} = [\mathbf{x}_{\mathcal{J},d+1} \quad \dots \quad \mathbf{x}_{\mathcal{J},n}]$ . Ainsi, par multiplication directe,

nous obtenons :

$$\begin{aligned} A_{12n} &= n^{-1} \mathbf{U}_{\mathcal{I}} \mathbf{V}_{\mathcal{J}}^{\top}, \\ &= n^{-1} \sum_{i=d+1}^n \mathbf{u}_{\mathcal{I},i} \mathbf{x}_{\mathcal{J},i}. \end{aligned}$$

Et donc,

$$\begin{aligned} E(\mathbf{u}_{\mathcal{I}1} \mathbf{x}_{\mathcal{J}1}^{\top}) &= E \left[ \begin{pmatrix} \mathbf{Y}_{1-i_1} \\ \vdots \\ \mathbf{Y}_{1-i_a} \end{pmatrix} \left( \mathbf{X}_{1-j_1}^{\top} \quad \cdots \quad \mathbf{X}_{1-j_b}^{\top} \right) \right], \\ &= [\boldsymbol{\Gamma}_{\mathbf{YX}}(j_d - i_e)], \end{aligned}$$

avec  $d = 1, \dots, a$  et  $e = 1, \dots, b$ . Invoquant des arguments en tout points similaires aux cas précédents, nous obtenons le résultat.

En nous inspirant des notations usuelles, nous posons

$$\begin{aligned} \boldsymbol{\Gamma}_{\mathcal{I}\mathcal{I},\mathbf{Y}\mathbf{Y}} &= [\boldsymbol{\Gamma}_{\mathbf{Y}\mathbf{Y}}(i_d - i_e)], \text{ où } d, e = 1, \dots, a, \\ \boldsymbol{\Gamma}_{\mathcal{J}\mathcal{J},\mathbf{X}\mathbf{X}} &= [\boldsymbol{\Gamma}_{\mathbf{X}\mathbf{X}}(j_d - j_e)], \text{ où } d, e = 1, \dots, b, \\ \boldsymbol{\Gamma}_{\mathcal{I}\mathcal{J},\mathbf{Y}\mathbf{X}} &= [\boldsymbol{\Gamma}_{\mathbf{Y}\mathbf{X}}(j_d - i_e)], \text{ où } e = 1, \dots, a, \text{ et } d = 1, \dots, b. \end{aligned}$$

En regroupant les trois résultats précédents, nous obtenons que la matrice  $n^{-1} \mathbf{Z}\mathbf{Z}^{\top}$  tend en probabilité vers une matrice que l'on note  $\boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}}$ . En effet, on a  $n^{-1} \mathbf{Z}\mathbf{Z}^{\top} \xrightarrow{p} \boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}}$  avec

$$\boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} = \begin{bmatrix} \boldsymbol{\Gamma}_{\mathcal{I}\mathcal{I},\mathbf{Y}\mathbf{Y}} & \boldsymbol{\Gamma}_{\mathcal{I}\mathcal{J},\mathbf{Y}\mathbf{X}} \\ \boldsymbol{\Gamma}_{\mathcal{I}\mathcal{J},\mathbf{Y}\mathbf{X}}^{\top} & \boldsymbol{\Gamma}_{\mathcal{J}\mathcal{J},\mathbf{X}\mathbf{X}} \end{bmatrix}.$$

Passons maintenant à l'étude du comportement asymptotique du terme restant de  $n^{1/2} (\hat{\boldsymbol{\beta}}_{\mathcal{I}\mathcal{J}} - \boldsymbol{\beta}_{\mathcal{I}\mathcal{J}})$ , soit  $n^{-1/2} \mathbf{E}\mathbf{Z}^{\top}$ . On a  $\mathbf{E} = [\boldsymbol{\epsilon}_{d+1} \quad \cdots \quad \boldsymbol{\epsilon}_n]$  qui est un vecteur aléatoire de dimension  $k \times (n - d)$  et on a :

$$\begin{aligned} \mathbf{Z} &= \begin{bmatrix} \mathbf{U}_{\mathcal{I}} \\ \mathbf{V}_{\mathcal{J}} \end{bmatrix}, \\ &= \begin{bmatrix} \mathbf{u}_{\mathcal{I},d+1} & \cdots & \mathbf{u}_{\mathcal{I},n} \\ \mathbf{x}_{\mathcal{J},d+1} & \cdots & \mathbf{x}_{\mathcal{J},n} \end{bmatrix}. \end{aligned}$$

Ainsi, puisque  $n^{-1/2} \mathbf{E}\mathbf{Z}^{\top}$  est une matrice aléatoire, on utilise à nouveau l'opérateur  $\text{vec}(\cdot)$ . Le résultat attendu, compte tenu de la première partie et compte tenu des formules (10.3.7) et (10.3.8) de Lütkepohl (2005), est

$$n^{-1/2} \text{vec}(\mathbf{E}\mathbf{Z}^{\top}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}),$$

avec  $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$  et  $\text{var}(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ . On peut également noter  $\mathbf{Z} = [\mathbf{w}_{\mathcal{I}\mathcal{J},d+1} \ \dots \ \mathbf{w}_{\mathcal{I}\mathcal{J},n}]$ , où  $\mathbf{w}_{\mathcal{I}\mathcal{J},i} = \begin{bmatrix} \mathbf{u}_{\mathcal{I},i} \\ \mathbf{x}_{\mathcal{J},i} \end{bmatrix}$ . On note ainsi que

$$\begin{aligned} \mathbf{E}\mathbf{Z}^\top &= \begin{bmatrix} \boldsymbol{\epsilon}_{d+1} & \dots & \boldsymbol{\epsilon}_n \end{bmatrix} \begin{bmatrix} \mathbf{w}_{\mathcal{I}\mathcal{J},d+1}^\top \\ \vdots \\ \mathbf{w}_{\mathcal{I}\mathcal{J},n}^\top \end{bmatrix}, \\ &= \sum_{i=d+1}^n \boldsymbol{\epsilon}_i \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top. \end{aligned}$$

Utilisant la linéarité de l'opérateur  $\text{vec}(\cdot)$ , et la relation  $\text{vec}(\mathbf{a}\mathbf{b}^\top) = \mathbf{b} \otimes \mathbf{a}$ , nous obtenons les égalités suivantes :

$$\begin{aligned} \text{vec}(\mathbf{E}\mathbf{Z}^\top) &= \text{vec} \left( \sum_{i=d+1}^n \boldsymbol{\epsilon}_i \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top \right), \\ &= \sum_{i=d+1}^n \text{vec} \left( \boldsymbol{\epsilon}_i \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top \right), \\ &= \sum_{i=d+1}^n \mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i. \end{aligned}$$

On note que  $d$  est fixe et finie, c'est-à-dire  $d < \infty$ , constante fixée par rapport à  $n$ . Ainsi,  $\sum_{i=1}^d \mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i = \mathcal{O}_p(1)$  est une somme finie de variables aléatoires, et cette somme est une variable aléatoire bornée en probabilité. Pour faciliter les calculs théoriques, on peut considérer la somme allant de  $1, \dots, n$  et étudier :

$$\begin{aligned} \text{vec}(\mathbf{E}\mathbf{Z}^\top) &= \sum_{i=1}^n \mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i + \mathcal{O}_p(1), \\ &= \sum_{i=1}^n \mathbf{F}_i + \mathcal{O}_p(1), \end{aligned}$$

où  $\mathbf{F}_i = \mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i$ . Nous allons utiliser l'outil de Cramér-Wold afin d'établir la normalité asymptotique. Considérons donc un vecteur  $\boldsymbol{\lambda}$ . On a alors que

$$n^{-1/2} \boldsymbol{\lambda}^\top \text{vec}(\mathbf{E}\mathbf{Z}^\top) = n^{-1/2} \sum_{i=1}^n \boldsymbol{\lambda}^\top \mathbf{F}_i + \mathcal{O}_p(n^{-1/2}).$$

Posons

$$f_{n,i} = n^{-1/2} \boldsymbol{\lambda}^\top \mathbf{F}_i,$$

une variable aléatoire univariée. On a donc

$$n^{-1/2} \boldsymbol{\lambda}^\top \text{vec}(\mathbf{E}\mathbf{Z}^\top) = \sum_{i=1}^n f_{n,i} + \mathcal{O}_p(n^{-1/2}).$$

Remarquons que la formule  $f_{n,i}$  est divisée par  $n^{1/2}$ . Dans le cas des VAR, on avait  $\mathbf{X}_t = \sum_{i \in \mathcal{I}} \Phi_i \mathbf{X}_{t-i} + \epsilon_t$  où  $\epsilon_t$  est indépendant du passé. On aura en fait que  $\mathbf{X}_t = \sum_{j \in \mathcal{K}} \Psi_j \epsilon_{t-j}$  avec des poids  $\Psi$  qui sont des matrices et des fonctions des indices  $\mathcal{K} = \{k_1, k_2, \dots, k_m\}$  choisis. Ceci permet de dire que la causalité du modèle implique que  $\epsilon_i$  est indépendant de

$\{\mathbf{X}_{t-k_1}, \mathbf{X}_{t-k_2}, \dots, \mathbf{X}_{t-k_m}\}$ . Rappelons la définition de  $\mathbf{U}_t$ .  $\mathbf{U}_t = \begin{bmatrix} \mathbf{X}_{t-k_1} \\ \vdots \\ \mathbf{X}_{t-k_m} \end{bmatrix} \otimes \mathbf{Z}_t$ , et est une

différence de matringales vectorielles :

$$\begin{aligned} E(\mathbf{U}_t | \mathcal{F}_{t-1}) &= \left( \begin{bmatrix} \mathbf{X}_{t-k_1} \\ \vdots \\ \mathbf{X}_{t-k_m} \end{bmatrix} \otimes \mathbf{Z}_t \mid \mathcal{F}_{t-1} \right), \\ &= \begin{bmatrix} \mathbf{X}_{t-k_1} \\ \vdots \\ \mathbf{X}_{t-k_m} \end{bmatrix} \otimes E(\mathbf{Z}_t | \mathcal{F}_{t-1}), \\ &= \begin{bmatrix} \mathbf{X}_{t-k_1} \\ \vdots \\ \mathbf{X}_{t-k_m} \end{bmatrix} \otimes E(\mathbf{Z}_t), \\ &= \mathbf{0}. \end{aligned}$$

On considère  $n^{-1/2} \sum_{i=1}^n \boldsymbol{\lambda}^\top \mathbf{F}_i$

$$\text{où } \mathbf{F}_i = \mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \epsilon_i \text{ et } \mathbf{w}_{\mathcal{I}\mathcal{J},i} = \begin{bmatrix} \mathbf{u}_{\mathcal{I},i} \\ \mathbf{x}_{\mathcal{J},i} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{i-i_1} \\ \vdots \\ \mathbf{Y}_{i-i_a} \\ \mathbf{X}_{i-j_1} \\ \vdots \\ \mathbf{X}_{i-j_b} \end{bmatrix},$$

$i_1, \dots, i_a \in \{1, \dots, p\}$  et  $j_1, \dots, j_b \in \{1, \dots, s\}$ . Si  $j_k = 0$  pour  $k \in \{1, \dots, b\}$ , on aura que  $\mathbf{X}_i$  aura le même indice que  $\epsilon_i$ .

Soit  $\mathcal{F}_t = \sigma \{\mathbf{X}_t, \mathbf{Y}_s, s \leq t\}$ . On sait que  $\{\mathbf{X}_t\}$  et  $\{\epsilon_t\}$  sont indépendants par exogénéité stricte.

$$\begin{aligned}
E(\mathbf{F}_i \mid \mathcal{F}_{i-1}) &= E(\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i \mid \mathcal{F}_{i-1}), \\
&= E \left( \begin{array}{c} \left[ \begin{array}{c} \mathbf{Y}_{i-i_1} \otimes \boldsymbol{\epsilon}_i \\ \vdots \\ \mathbf{Y}_{i-i_a} \otimes \boldsymbol{\epsilon}_i \\ \mathbf{X}_{i-j_1} \otimes \boldsymbol{\epsilon}_i \\ \vdots \\ \mathbf{X}_{i-j_b} \otimes \boldsymbol{\epsilon}_i \end{array} \right] \\ \left. \vphantom{\begin{array}{c} \left[ \begin{array}{c} \mathbf{Y}_{i-i_1} \otimes \boldsymbol{\epsilon}_i \\ \vdots \\ \mathbf{Y}_{i-i_a} \otimes \boldsymbol{\epsilon}_i \\ \mathbf{X}_{i-j_1} \otimes \boldsymbol{\epsilon}_i \\ \vdots \\ \mathbf{X}_{i-j_b} \otimes \boldsymbol{\epsilon}_i \end{array} \right]} \right| \mathcal{F}_{i-1} \end{array} \right), \\
&= \mathbf{0}.
\end{aligned}$$

On note que si  $j_i = 0$  pour un certain  $i \in \{1, \dots, b\}$ , on peut utiliser le résultat suivant pour trouver  $E(\mathbf{X}_i \otimes \boldsymbol{\epsilon}_i \mid \mathcal{F}_{i-1})$ . Si  $(\mathbf{X}, \mathbf{Z})$  est indépendant de  $\mathbf{Y}$  alors,  $P(\mathbf{X} \leq \mathbf{x}, \mathbf{Y} \leq \mathbf{y}, \mathbf{Z} \leq \mathbf{z}) = P(\mathbf{X} \leq \mathbf{x}, \mathbf{Z} \leq \mathbf{z})P(\mathbf{Y} \leq \mathbf{y})$ . Or, si  $(\mathbf{X}, \mathbf{Z})$  est indépendant de  $\mathbf{Y}$ , on aura que  $\mathbf{Z}$  est indépendant de  $\mathbf{Y}$  et donc  $P(\mathbf{Y} \leq \mathbf{y}) = P(\mathbf{Y} \leq \mathbf{y} \mid \mathbf{Z} \leq \mathbf{z})$ . On aura alors  $P(\mathbf{X} \leq \mathbf{x}, \mathbf{Y} \leq \mathbf{y}, \mathbf{Z} \leq \mathbf{z}) = P(\mathbf{X} \leq \mathbf{x}, \mathbf{Z} \leq \mathbf{z})P(\mathbf{Y} \leq \mathbf{y} \mid \mathbf{Z} \leq \mathbf{z})$ . Considérons maintenant les lois de  $\mathbf{X} \mid \mathbf{Z}$  et de  $\mathbf{Y} \mid \mathbf{Z}$ .

**Théorème 2.2.1.** *Soient  $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$  des vecteurs aléatoires admettant une densité. On suppose que  $\mathbf{X}$  et  $\mathbf{Z}$  sont conjointement indépendant de  $\mathbf{Y}$ . Alors,  $\mathbf{X}$  et  $\mathbf{Y}$  sont conditionnellement indépendantes étant donné  $\mathbf{Z}$ .*

**Démonstration.** Puisque l'on présume l'existence des densités, on aura :

$$\begin{aligned}
f_{\mathbf{X}\mathbf{Y}\mathbf{Z}}(\mathbf{x}, \mathbf{y}, \mathbf{z}) &= f_{\mathbf{X}\mathbf{Z}}(\mathbf{x}, \mathbf{z})f_{\mathbf{Y}}(\mathbf{y}), \\
&= f_{\mathbf{X}\mathbf{Z}}(\mathbf{x}, \mathbf{z})f_{\mathbf{Y} \mid \mathbf{Z}=\mathbf{z}}(\mathbf{y}).
\end{aligned} \tag{2.2.11}$$

Dans les deux égalités précédentes, la première provient de l'hypothèse d'indépendance de  $(\mathbf{X}, \mathbf{Z})$  avec  $\mathbf{Y}$ , alors que la deuxième repose sur le fait que si  $(\mathbf{X}, \mathbf{Z})$  est indépendant du vecteur aléatoire  $\mathbf{Y}$ , alors  $\mathbf{Z}$  et  $\mathbf{Y}$  sont indépendants, et donc la distribution conditionnelle de  $\mathbf{Y} \mid \mathbf{Z}$  est la même que la densité de  $\mathbf{Y}$ . On divise la relation (2.2.11) par la densité  $f_{\mathbf{Z}}(\mathbf{z})$ , qui est supposée strictement positive. On obtient alors :

$$\frac{f_{\mathbf{X}\mathbf{Y}\mathbf{Z}}(\mathbf{x}, \mathbf{y}, \mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})} = \frac{f_{\mathbf{X}\mathbf{Z}}(\mathbf{x}, \mathbf{z})f_{\mathbf{Y} \mid \mathbf{Z}=\mathbf{z}}(\mathbf{y})}{f_{\mathbf{Z}}(\mathbf{z})}.$$

Cependant, par définition même d'une distribution conditionnelle, on aura

$$\begin{aligned}
\frac{f_{\mathbf{X}\mathbf{Y}\mathbf{Z}}(\mathbf{x}, \mathbf{y}, \mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})} &= f_{\mathbf{X}\mathbf{Y} \mid \mathbf{Z}=\mathbf{z}}(\mathbf{x}, \mathbf{y}), \\
\frac{f_{\mathbf{X}\mathbf{Z}}(\mathbf{x}, \mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})} &= f_{\mathbf{X} \mid \mathbf{Z}=\mathbf{z}}(\mathbf{x}).
\end{aligned}$$

Finalement, on trouve donc que :

$$f_{\mathbf{XY} | \mathbf{Z}=\mathbf{z}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X} | \mathbf{Z}=\mathbf{z}}(\mathbf{x}) f_{\mathbf{Y} | \mathbf{Z}=\mathbf{z}}(\mathbf{y}).$$

En conclusion,  $\mathbf{X} | \mathbf{Z}$  et  $\mathbf{Y} | \mathbf{Z}$  sont indépendants. Or  $\mathbf{Y}$  et  $\mathbf{Z}$  sont indépendants et donc  $f_{\mathbf{Y} | \mathbf{Z}=\mathbf{z}}(\mathbf{y}) = f_{\mathbf{Y}}(\mathbf{y})$ . Donc on obtient que  $\mathbf{X}$  et  $\mathbf{Y}$  sont conditionnellement indépendantes étant donné  $\mathbf{Z}$ , ce qui montre le résultat annoncé.  $\square$

Utilisant l'hypothèse d'exogénéité stricte et la causalité du processus  $\{\mathbf{Y}_t\}$ , on aura que  $\mathbf{X}_i$  et l'information contenue dans  $\mathcal{F}_{i-1}$  sont indépendants du vecteur aléatoire  $\boldsymbol{\epsilon}_i$ . Ainsi, pour les distributions conditionnelles, une génération de l'argument menant au théorème 2.2.1 permet d'affirmer que :

$$\begin{aligned} E(\mathbf{X}_i \otimes \boldsymbol{\epsilon}_i | \mathcal{F}_{i-1}) &= E(\mathbf{X}_i | \mathcal{F}_{i-1}) \otimes E(\boldsymbol{\epsilon}_i | \mathcal{F}_{i-1}), \\ &= E(\mathbf{X}_i | \mathcal{F}_{i-1}) \otimes E(\boldsymbol{\epsilon}_i), \\ &= \mathbf{0}. \end{aligned}$$

On généralise (2.1.4) aux modèles SVARX. On a

$$\begin{aligned} \text{vec}(\mathbf{E}\mathbf{Z}^\top) &= \sum_{i=1}^n \mathbf{F}_i + \mathcal{O}_p(1), \\ &= \sum_{i=1}^n (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i) + \mathcal{O}_p(1), \end{aligned}$$

où l'on rappelle que :

$$\mathbf{w}_{\mathcal{I}\mathcal{J},i} = \begin{bmatrix} \mathbf{U}_{\mathcal{I},i} \\ \mathbf{x}_{\mathcal{J},i} \end{bmatrix}.$$

Or,  $E\{(\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i)(\mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top \otimes \boldsymbol{\epsilon}_i^\top)\} = E\{\mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top \otimes \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top\}$ , car  $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$ , où il est sous-entendu que les dimensions sont compatibles.

L'exogénéité stricte, la causalité et le fait que le processus  $\{\boldsymbol{\epsilon}_t\}$  soit constitué de vecteurs aléatoires iid impliquent que  $\mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top$  et  $\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top$  sont indépendants. Or, si  $\mathbf{X}$  et  $\mathbf{Y}$  sont indépendants, alors  $g(\mathbf{X})$  et  $h(\mathbf{Y})$  sont indépendants et on a  $E\{g(\mathbf{X})h(\mathbf{Y})\} = E\{g(\mathbf{X})\} E\{h(\mathbf{Y})\}$ . Donc

$$\begin{aligned} E\{(\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i)(\mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top \otimes \boldsymbol{\epsilon}_i^\top)\} &= E\left(\mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top \otimes \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top\right), \\ &= E\left(\mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top\right) E\left(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top\right), \\ &= \boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}. \end{aligned}$$

Ainsi, on fait comme en (2.1.4) et on trouve

$$\text{var}(f_{n,i}) = \text{var}(n^{-1/2} \boldsymbol{\lambda}^\top \mathbf{F}_i) = n^{-1} \boldsymbol{\lambda}^\top \text{var}(\mathbf{F}_i) \boldsymbol{\lambda},$$

où

$$\text{var}(\mathbf{F}_i) = E(\mathbf{F}_i \mathbf{F}_i^\top) = \boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} \otimes \boldsymbol{\Sigma}_\epsilon.$$

Donc

$$\text{var}(f_{n,i}) = n^{-1} \boldsymbol{\lambda}^\top (\boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} \otimes \boldsymbol{\Sigma}_\epsilon) \boldsymbol{\lambda}.$$

Si on procède comme Brockwell, Davis et Alexandre Trindade (2004), il faut montrer qu'en posant  $\mathcal{F}_{t-1} = \sigma \{ \mathbf{X}_s, \boldsymbol{\epsilon}_s, s \leq t-1 \}$  alors

$$(1) \sum_{t=1}^n E \left( f_{n,t}^2 \mid \mathcal{F}_t - 1 \right) \xrightarrow{p} \boldsymbol{\lambda}^\top (\boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} \otimes \boldsymbol{\Sigma}_\epsilon) \boldsymbol{\lambda},$$

$$(2) \max_{1 \leq t \leq n} |f_{n,t}| \xrightarrow{p} 0,$$

$$(3) E \left\{ \max_{1 \leq t \leq n} f_{n,t}^2 \right\} = \mathcal{O}(n).$$

Alternativement, on applique directement le théorème limite central des martingales cité dans Brown (1971). Ceci aurait l'avantage de simplifier l'argumentation. On a

$$n^{-1/2} \text{vec}(\mathbf{E}\mathbf{Z}^\top) = n^{-1/2} \sum_{i=1}^n (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i) + \mathcal{O}_p(n^{-1/2}).$$

On pose

$$U_n = n^{-1/2} \boldsymbol{\lambda}^\top \sum_{i=1}^n (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i) = n^{-1} \sum_{i=1}^n U_{in},$$

où

$$U_{in} = n^{1/2} \boldsymbol{\lambda}^\top \{ \mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i \}.$$

On pose  $\sigma^2(n) = E(U_n^2)$ . On aura la normalité asymptotique, à savoir :

$$U_n \sigma^{-1}(n) \xrightarrow{d} \mathcal{N}(0,1),$$

si les conditions suffisantes suivantes sont respectées :

$$a) \quad n^{-2} \sigma^{-2}(n) \sum_{i=1}^n E \left\{ U_{in}^2 \mathbb{I}(|U_{in}| > \epsilon n \sigma(n)) \right\} \longrightarrow 0, \forall \epsilon > 0,$$

$$b) \quad n^{-2} \sigma^{-2}(n) = U_{in}^2 \xrightarrow{p} 1,$$

où  $U_{in}^2 = E(U_{in}^2 \mid \mathcal{F}_{i-1})$  et  $\mathcal{F}_{i-1} = \sigma \{ (\mathbf{X}_s, \boldsymbol{\epsilon}_s), s \leq i-1 \}$ . Commençons par montrer la première condition. Le condition dite de Lyapounov est suffisante pour établir a) :

$$\frac{1}{n^4} \frac{1}{\sigma^4(n)} \sum_{i=1}^n E(U_{in}^4) \longrightarrow 0.$$

On procède au calcul explicite de  $\sigma^2(n) = E(U_n^2)$ , où  $U_n = \boldsymbol{\lambda}^\top n^{-1/2} \sum_{i=1}^n (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i)$ . On a

$$\begin{aligned} U_n^2 &= U_n U_n = U_n U_n^\top, \\ &= n^{-1} \boldsymbol{\lambda}^\top \sum_{i=1}^n \sum_{j=1}^n (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i) (\mathbf{w}_{\mathcal{I}\mathcal{J},j}^\top \otimes \boldsymbol{\epsilon}_j^\top) \boldsymbol{\lambda}. \end{aligned}$$

Pour calculer  $E(U_n^2)$ , nous séparons la double somme  $U_n^2$  en trois cas, pour les indices  $i < j$ ,  $i = j$  et finalement  $i > j$ . Plus précisément :

$$\begin{aligned} E(U_n^2) &= n^{-1} \boldsymbol{\lambda}^\top \sum_{i < j} E \left( \mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},j}^\top \otimes \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^\top \right) \boldsymbol{\lambda} + \\ &\quad n^{-1} \boldsymbol{\lambda}^\top \sum_{i=1}^n E \left( \mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top \otimes \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top \right) \boldsymbol{\lambda} + \\ &\quad n^{-1} \boldsymbol{\lambda}^\top \sum_{i > j} E \left( \mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},j}^\top \otimes \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^\top \right) \boldsymbol{\lambda}. \end{aligned}$$

En reprennant l'argumentation faite précédemment, on constate que les premier et troisième termes sont nuls. On trouve que

$$E(U_n^2) = \boldsymbol{\lambda}^\top (\boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}) \boldsymbol{\lambda},$$

et que

$$U_{in} = n^{1/2} \boldsymbol{\lambda}^\top (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i).$$

Ainsi,

$$|U_{in}|^4 = U_{in}^4 = \|U_{in}\|_E^4,$$

où  $\|\mathbf{x}\|_E = \sqrt{\sum_i^n x_i^2}$ ,  $\|\mathbf{x}\|_E^2 = \sum_i^n x_i^2$  et  $\|\mathbf{x}\|_E^4 = (\sum_i^n x_i^2)^2$ . Or, on sait par l'inégalité de Cauchy-Schwarz que  $\mathbf{x}^\top \mathbf{y} \leq \|\mathbf{x}\|_E \|\mathbf{y}\|_E$  donc  $|\mathbf{x}^\top \mathbf{y}| \leq \|\mathbf{x}\|_E \|\mathbf{y}\|_E$ . Autrement dit,

$$\begin{aligned} |U_{in}| &= n^{1/2} |\boldsymbol{\lambda}^\top (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i)|, \\ &\leq n^{1/2} \|\boldsymbol{\lambda}^\top (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i)\|_E, \end{aligned}$$

et

$$U_{in}^4 = n^2 \|\boldsymbol{\lambda}^\top\|_E^4 \|(\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i)\|_E^4.$$

Ainsi,

$$n^{-4} \sum_{i=1}^n E(U_{in}^4) \leq n^{-2} \|\boldsymbol{\lambda}\|_E^4 \sum_{i=1}^n E(\|\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i\|_E^4).$$

**Lemme 2.2.2.** Soit  $\mathbf{A}$  un vecteur de dimension  $p \times q$  et  $\mathbf{B}$  un vecteur de dimension  $r \times s$ . Alors, la norme euclidienne du produit de Kronecker de  $\mathbf{A}$  et  $\mathbf{B}$  équivaut au produit de la norme euclidienne des deux vecteurs c'est-à-dire  $\|\mathbf{A} \otimes \mathbf{B}\|_E = \|\mathbf{A}\|_E \|\mathbf{B}\|_E$ .

**Démonstration.** En effet,

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij} \mathbf{B}).$$

Donc,

$$\begin{aligned} \|\mathbf{A} \otimes \mathbf{B}\|_E^2 &= \sum_{i=1}^p \sum_{j=1}^q \left\{ a_{ij}^2 \left( \sum_{k=1}^r \sum_{l=1}^s b_{kl}^2 \right) \right\}, \\ &= \|\mathbf{A}\|_E^2 \|\mathbf{B}\|_E^2. \end{aligned}$$

□

On aura alors que

$$\|\mathbf{w}_{\mathcal{I}\mathcal{J},i} \otimes \boldsymbol{\epsilon}_i\|_E = \|\mathbf{w}_{\mathcal{I}\mathcal{J},i}\|_E \|\boldsymbol{\epsilon}_i\|_E.$$

Cela entraîne que

$$\begin{aligned} n^{-4} \sum_{i=1}^n E(U_{in}^4) &\leq n^{-2} \|\boldsymbol{\lambda}\|_E^4 \sum_{i=1}^n E \left( \|\mathbf{w}_{\mathcal{I}\mathcal{J},i} \times \boldsymbol{\epsilon}_i\|_E^4 \right), \\ &= n^{-2} \|\boldsymbol{\lambda}\|_E^4 \sum_{i=1}^n E \left( \|\mathbf{w}_{\mathcal{I}\mathcal{J},i}\|_E^4 \|\boldsymbol{\epsilon}_i\|_E^4 \right), \\ &= n^{-2} \|\boldsymbol{\lambda}\|_E^4 \sum_{i=1}^n E \left( \|\mathbf{w}_{\mathcal{I}\mathcal{J},i}\|_E^4 \right) E \left( \|\boldsymbol{\epsilon}_i\|_E^4 \right), \\ &= \mathcal{O}(n^{-1}). \end{aligned}$$

Pour montrer la seconde condition b), on pose  $\ddot{U}_n^2 = n^{-2} \sum_{i=1}^n \ddot{U}_{in}^2$ . On remarque que par la définition de  $\ddot{U}_{in}^2$  :

$$\begin{aligned} \ddot{U}_{in}^2 &= E(U_{in}^2 | \mathcal{F}_{i-1}), \\ &= n \boldsymbol{\lambda}^\top E \left\{ (\mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top \otimes \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top) | \mathcal{F}_{i-1} \right\} \boldsymbol{\lambda}, \\ &= n \boldsymbol{\lambda}^\top \left\{ E(\mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top | \mathcal{F}_{i-1}) \otimes \boldsymbol{\Sigma}_\boldsymbol{\epsilon} \right\} \boldsymbol{\lambda}. \end{aligned}$$

Ainsi :

$$\ddot{U}_n^2 = n^{-1} \sum_{i=1}^n \boldsymbol{\lambda}^\top \left\{ E(\mathbf{w}_{\mathcal{I}\mathcal{J},i} \mathbf{w}_{\mathcal{I}\mathcal{J},i}^\top | \mathcal{F}_{i-1}) \otimes \boldsymbol{\Sigma}_\boldsymbol{\epsilon} \right\} \boldsymbol{\lambda}$$

Une application du théorème ergodique donne  $\ddot{U}_n^2 \xrightarrow{p} \sigma^2(n) = \boldsymbol{\lambda}^\top (\boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} \otimes \boldsymbol{\Sigma}_\boldsymbol{\epsilon}) \boldsymbol{\lambda}$ . Ceci montre la seconde condition. Ayant montré les deux conditions, nous obtenons le résultat de normalité asymptotique :

$$n^{-1/2} \text{vec}(\mathbf{E}\mathbf{Z}^\top) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}} \otimes \boldsymbol{\Sigma}_\boldsymbol{\epsilon}).$$

Ce résultat est l'équivalent du lemme 2.1.2 pour les VARX. On se propose de généraliser aux VARX le théorème 2.1.3.

**Théorème 2.2.3.** *Considérons un processus  $\{\mathbf{Y}_t\}$  satisfaisant l'équation (2.2.2) où le bruit  $\{\boldsymbol{\epsilon}_t\} \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_\boldsymbol{\epsilon})$ . On estime les paramètres du modèle par moindres carrés. L'estimateur des moindres carrés est convergent en probabilité et sa distribution asymptotique est normale. Plus précisément :*

$$\begin{aligned} \text{a)} \quad &\hat{\boldsymbol{\beta}}_{\mathcal{I}\mathcal{J}} \xrightarrow{p} \boldsymbol{\beta}_{\mathcal{I}\mathcal{J}}, \\ \text{b)} \quad &n^{-1/2} (\hat{\mathbf{b}}_{\mathcal{I}\mathcal{J}} - \mathbf{b}_{\mathcal{I}\mathcal{J}}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\mathcal{I}\mathcal{J}}^{-1} \otimes \boldsymbol{\Sigma}_\boldsymbol{\epsilon}), \end{aligned}$$

où l'estimateur  $\hat{\boldsymbol{\beta}}_{\mathcal{I}\mathcal{J}}$  est défini par (2.2.8),  $\mathbf{b}_{\mathcal{I}\mathcal{J}} = \text{vec}(\boldsymbol{\beta}_{\mathcal{I}\mathcal{J}})$  et  $\hat{\mathbf{b}}_{\mathcal{I}\mathcal{J}} = \text{vec}(\hat{\boldsymbol{\beta}}_{\mathcal{I}\mathcal{J}})$ .

**Démonstration.** Pour (a), on reprend l'argument :

$$\hat{\beta}_{IJ} - \beta_{IJ} = n^{-1} \mathbf{E} \mathbf{Z}^\top (n^{-1} \mathbf{Z} \mathbf{Z}^\top).$$

On a montré que

$$n^{-1} \mathbf{Z} \mathbf{Z}^\top \xrightarrow{p} \Omega_{IJ},$$

et que

$$n^{-1/2} \mathbf{E} \mathbf{Z}^\top = \mathcal{O}_p(1),$$

car  $n^{-1/2} \text{vec}(\mathbf{E} \mathbf{Z}^\top) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega_{IJ} \otimes \Sigma_\epsilon)$ . Donc

$$n^{-1} \mathbf{E} \mathbf{Z}^\top = \mathcal{O}_p(n^{-1/2}),$$

et

$$n^{-1} \mathbf{E} \mathbf{Z}^\top \xrightarrow{p} \mathbf{0}.$$

Ceci montre que

$$\hat{\beta}_{IJ} - \beta_{IJ} \xrightarrow{p} \mathbf{0}.$$

Pour (b), l'argument est également standard.

$$\begin{aligned} \hat{\beta}_{IJ} - \beta_{IJ} &= n^{-1} \mathbf{E} \mathbf{Z}^\top (n^{-1} \mathbf{Z} \mathbf{Z}^\top)^{-1}, \\ \text{vec}(\hat{\beta}_{IJ} - \beta_{IJ}) &= \text{vec} \left\{ n^{-1} \mathbf{E} \mathbf{Z}^\top (n^{-1} \mathbf{Z} \mathbf{Z}^\top)^{-1} \right\}, \\ n^{-1/2} (\hat{\mathbf{b}}_{IJ} - \mathbf{b}_{IJ}) &= \left\{ (n^{-1} \mathbf{Z} \mathbf{Z}^\top)^{-1} \otimes \mathbf{I} \right\} n^{-1/2} \text{vec}(\mathbf{E} \mathbf{Z}^\top), \end{aligned}$$

car on sait que  $\text{vec}(\mathbf{A} \mathbf{B} \mathbf{C}) = (\mathbf{C}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ .

Or

$$n^{-1/2} \text{vec}(\mathbf{E} \mathbf{Z}^\top) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega_{IJ} \otimes \Sigma_\epsilon),$$

et

$$(n^{-1} \mathbf{Z} \mathbf{Z}^\top)^{-1} \otimes \mathbf{I} \xrightarrow{p} \Omega_{IJ}^{-1} \otimes \mathbf{I}.$$

Donc

$$n^{1/2} (\hat{\mathbf{b}}_{IJ} - \mathbf{b}_{IJ}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega_{IJ} \otimes \Sigma_\epsilon).$$

Or, si  $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{B})$  alors  $\mathbf{A} \mathbf{X} \sim \mathcal{N}_p(\mathbf{A} \boldsymbol{\mu}, \mathbf{A} \mathbf{B} \mathbf{A}^\top)$ . Autrement formulé,  $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{B}) \equiv \mathcal{N}_p(\mathbf{A} \boldsymbol{\mu}, \mathbf{A} \mathbf{B} \mathbf{A}^\top)$ . On aura donc

$$(\Omega_{IJ}^{-1} \otimes \mathbf{I}) (\Omega_{IJ} \otimes \Sigma_\epsilon) (\Omega_{IJ}^{-1} \otimes \mathbf{I}) = (\Omega_{IJ}^{-1} \otimes \Sigma_\epsilon),$$

et

$$n^{1/2} (\hat{\mathbf{b}}_{IJ} - \mathbf{b}_{IJ}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega_{IJ}^{-1} \otimes \Sigma_\epsilon).$$

□

Nous venons de démontrer deux résultats utiles concernant l'estimateur des moindres carrés pour les modèles SVARX. Dans la prochaine section, nous allons présenter des outils servant à la vérification diagnostic pour ces modèles.

### 2.2.3. Vérification diagnostique des modèles SVARX

La vérification diagnostique des modèles de séries chronologiques représente une question pratique importante. En utilisant une approche de densité spectrale basée sur le noyau ainsi que les résultats de Duchesne et Roy (2004), de puissantes statistiques de test de type portemanteau peuvent être proposées. Les statistiques portemanteaux sont généralement performantes (c'est-à-dire affichant une puissance non triviale supérieure au niveau du test) pour des hypothèses alternatives qui peuvent être assez générales. Par exemple, dans notre situation, le terme d'erreur sera non corrélé sous l'hypothèse nulle versus une hypothèse alternative de corrélation de structure inconnue et non spécifiée. Plus précisément, afin de vérifier un modèle SVARX( $\mathcal{I}, \mathcal{J}$ ) tel que défini par (1.3.2), l'hypothèse d'intérêt est que le processus d'erreur de (1.3.2) est un bruit blanc processus contre l'alternative de la corrélation sérielle de forme arbitraire. De manière plus formelle :

$$H_0 : \mathbf{\Gamma}_{\boldsymbol{\epsilon}}(h) = \mathbf{0}, \forall h \neq 0, \text{ contre}$$

$$H_1 : \mathbf{\Gamma}_{\boldsymbol{\epsilon}}(h) \neq \mathbf{0}, \text{ pour au moins un } h \neq 0.$$

Définissons  $\mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h) = n^{-1} \sum \hat{\boldsymbol{\epsilon}}_t \hat{\boldsymbol{\epsilon}}_{t-h}^\top$ ,  $h = 0, 1, \dots, n$ ,  $M_n(\mathcal{K}) = \sum_{h=1}^{n-1} (1 - h/n) \mathcal{K}^2(h/P_n)$  et  $V_n(\mathcal{K}) = \sum_{h=1}^{n-2} (1 - h/n)(1 - (h+1)/n) \mathcal{K}^4(h/P_n)$ , avec  $P_n \rightarrow \infty$  tel que  $P_n/n \rightarrow 0$  et  $\mathcal{K}(\cdot)$  correspond à un noyau. Une statistique de test justifiée avec un estimateur de densité spectrale à noyau est donnée par :

$$T_n = \frac{n \sum_{h=1}^{n-1} \mathcal{K}^2(h/P_n) \text{tr} \left\{ \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^\top(h) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^{-1}(0) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^{-1}(0) \right\} - k^2 M_n(\mathcal{K})}{\{2k^2 V_n(\mathcal{K})\}^{1/2}}. \quad (2.2.12)$$

L'utilisation du noyau uniforme tronqué représente essentiellement une statistique test de Box-Pierce-Ljung généralisée. Cependant, les noyaux  $\mathcal{K}(\cdot)$  de Daniell ou de Bartlett s'avèrent plus puissants dans de nombreuses situations. Voir Duchesne et Roy (2004) pour plus de détails. Étant donné le taux de convergence des estimateurs des moindres carrés étudiés à la section précédente, c'est-à-dire  $\hat{\boldsymbol{\beta}}_{\mathcal{I}\mathcal{J}} - \boldsymbol{\beta}_{\mathcal{I}\mathcal{J}} = \mathcal{O}_p(n^{-1/2})$ , et le taux de convergence plus lent de l'estimateur de densité spectrale basé sur le noyau de  $\{\boldsymbol{\epsilon}_t\}$ , l'estimation n'a pas d'impact sur la distribution asymptotique de  $T_n$  sous  $H_0$ . Il peut être montré que les arguments de Duchesne et Roy (2004, Théorème 1), particulièrement leurs résultats A.5 et A.6, sont toujours valables pour les modèles SVARX et

$$T_n \xrightarrow{d} \mathcal{N}(0,1)$$

sous l'hypothèse nulle d'adéquation. Ainsi, des statistiques de test faciles à mettre en œuvre et puissantes sont disponibles afin vérifier les modèles SVARX. Elles seront utilisées au chapitre 4.

La section suivante s'intéresse aux prévisions dans les modèles SVAR et SVARX.

## 2.3. Calcul des prévisions de valeurs futures

La prévision de valeurs futures représente certainement une application particulièrement fondamentale des séries chronologiques. Les prévisions sont fréquemment utilisées dans les domaines de l'économie, de la finance ou même en hydrologie. L'ajustement d'un modèle vectoriel autorégressif à des séries temporelles permet de modéliser la structure des données. Disposant d'un modèle, ce dernier a pour effet de faciliter la prévision de valeurs futures pour des séries multivariées.

Supposons que nous disposions des observations de plusieurs séries jusqu'au temps  $n$ , c'est-à-dire que nous connaissons les valeurs de la série chronologique  $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$ . Il devient alors intéressant de prévoir les valeurs que prendront ces séries dans le futur, c'est-à-dire les valeurs que prendront ces séries aux temps  $n + 1, n + 2, \dots, n + l$ , où  $l$  est appelé l'horizon, satisfaisant  $l > 0$ .

Deux types de prévisions sont traditionnellement considérées dans la littérature portant sur les séries chronologiques. Le premier type étant la recherche de la meilleure prévision linéaire au sens de l'erreur quadratique moyenne. Cette prévision est une fonction linéaire des valeurs du processus jusqu'au temps  $n$ . Le second type de prévision vise à déterminer la meilleure prévision (parmi toutes les prévisions, linéaires ou non) au sens de l'erreur quadratique moyenne (EQM). Sous des conditions générales, il peut être montré que la meilleure prévision au sens de l'erreur quadratique moyenne est l'espérance conditionnelle, qui peut être plus compliquée à calculer que la meilleure prévision linéaire. Dans les modèles autorégressifs, compte tenu de la structure du modèle (qui est linéaire), il est bien connu que la meilleure prévision linéaire et la meilleure prévision au sens de l'EQM coïncident, en autant que le terme d'erreur soit un bruit blanc fort et que le nombre de données soit supérieur à l'ordre autorégressif.

Dans les prochaines sections, nous présenterons ces deux prévisions pour les modèles SVAR et SVARX.

### 2.3.1. Problème de la prévision dans le modèle SVAR

Les prévisions pour les modèles VAR avec sélection d'indices ont été développées dans Brockwell, Davis et Alexandre Trindade (2004). Cette section présente leurs résultats les plus utiles et importants pour la suite quand nous introduirons des variables exogènes. Soit

$\mathbf{Y}_t = (Y_t(1), Y_t(2), \dots, Y_t(k))^\top$  un SVAR( $p$ ) stationnaire dont l'équation est la suivante :

$$\mathbf{Y}_t = \sum_{i \in \mathcal{I}} \Phi_{I_i} \mathbf{Y}_{t-i} + \epsilon_t. \quad (2.3.1)$$

La meilleure prévision linéaire au sens de l'EQM de la valeur  $\mathbf{Y}_t$  en fonction de  $\{\mathbf{Y}_{t-i}, i \in \mathcal{I}\}$  sera de la forme :

$$\hat{\mathbf{Y}}_t^{(f)} = \sum_{i \in \mathcal{I}} \Phi_i^{(f)} \mathbf{Y}_{t-i}, \quad (2.3.2)$$

et la matrice des variances et covariances des erreurs de prévision est donnée par l'expression générale :

$$\mathbf{U}_{\mathcal{I}}^{(f)} \equiv E \left[ (\mathbf{Y}_t - \hat{\mathbf{Y}}_t^{(f)}) (\mathbf{Y}_t - \hat{\mathbf{Y}}_t^{(f)})^\top \right].$$

Afin de déterminer les coefficients  $\Phi_i^{(f)}$ , il suffit de résoudre les équations empiriques de Yule-Walker qui font appel à la fonction d'autocovariance du processus :

$$\Gamma_{\mathbf{Y}}(i) = \sum_{k \in \mathcal{I}} \Phi_i^{(f)} \Gamma_{\mathbf{Y}}(i-k), i \in \mathcal{I}.$$

Il peut être montré que la matrice des variances et covariances des erreurs de prévision devient :

$$\mathbf{U}_{\mathcal{I}}^{(f)} = \sum_{i \in \mathcal{I}} \Phi_i^{(f)} \Gamma_{\mathbf{Y}}(i)^\top.$$

Une hypothèse faible est de supposer que  $\epsilon_t$  et  $\{\mathbf{Y}_s, s < t\}$  sont non corrélés, et ce pour tout  $t$ . Dans ce cas, on a  $\Phi_i^{(f)} = \Phi_i, i \in \mathcal{I}$  et  $\mathbf{U}_{\mathcal{I}}^{(f)} = \Sigma$ .

Dans l'éventualité où la fonction d'autocovariance est inconnue, ce qui, il faut bien l'avouer, est largement le cas en pratique, on utilise la fonction d'autocovariance échantillonnale pour estimer la fonction d'autocovariance théorique, qui est définie ainsi (dans le cas d'un processus de moyenne nulle) :

$$\hat{\Gamma}_{\mathbf{Y}}(l) = \begin{cases} n^{-1} \sum_{t=1}^{n-l} \mathbf{Y}_{t+l} \mathbf{Y}_t^\top, & \text{pour } l = 0, \dots, n-1, \\ \hat{\Gamma}_{\mathbf{Y}}(-l), & \text{pour } l = -n+1, \dots, -1. \end{cases} \quad (2.3.3)$$

La meilleure prévision linéaire telle que définie ci-haut utilise les observations passées. On dit qu'il s'agit de la meilleure prévision vers l'avant. Une prévision utilisant les valeurs futures des observations peut également être obtenue. On dit qu'il s'agit de la meilleure prévision linéaire vers l'arrière et cette prévision est de la forme :

$$\hat{\mathbf{Y}}_t^{(b)} = \sum_{i \in \mathcal{I}} \Psi_i^{(b)} \mathbf{Y}_{t+i}, \quad (2.3.4)$$

et la matrice de covariance des erreurs de prévision est donnée par  $\mathbf{V}_{\mathcal{I}}^{(b)} \equiv E \left[ (\mathbf{Y}_t - \hat{\mathbf{Y}}_t^{(b)}) (\mathbf{Y}_t - \hat{\mathbf{Y}}_t^{(b)})^\top \right]$ . Les coefficients  $\Psi_i^{(b)}$ , peuvent être trouvés en résolvant les équations de Yule-Walker vers l'arrière :

$$\Gamma_{\mathbf{Y}}(i)^\top = \sum_{k \in \mathcal{I}} \Psi_i^{(b)} \Gamma_{\mathbf{Y}}(i-k), i \in \mathcal{I}, \quad (2.3.5)$$

$$\mathbf{V}_{\mathcal{I}}^{(b)} = \mathbf{\Gamma}_{\mathbf{Y}}(0) - \sum_{i \in \mathcal{I}} \mathbf{\Psi}_i^{(b)} \mathbf{\Gamma}_{\mathbf{Y}}(i)^\top. \quad (2.3.6)$$

On discute maintenant le calcul de la meilleure prévision à l'horizon un au sens de l'erreur quadratique moyenne, sans se restreindre à la recherche de la meilleure prévision linéaire. Comme on l'a vu, une hypothèse faible consiste à ne supposer que l'absence de corrélation entre  $\boldsymbol{\epsilon}_t$  et  $\{\mathbf{Y}_s, s < t\}$ , et ce pour tout  $t$ . En supposant plus, à savoir l'indépendance entre  $\boldsymbol{\epsilon}_t$  et  $\{\mathbf{Y}_s, s < t\}$ , on peut calculer aisément l'espérance conditionnelle dans le modèle. En effet, on aura alors que la meilleure prévision de  $\mathbf{Y}_t$  est :

$$\begin{aligned} E(\mathbf{Y}_t | \mathbf{Y}_{t-i}, i \in \mathcal{I}) &= \sum_{i \in \mathcal{I}} \mathbf{\Phi}_{I_i} \mathbf{Y}_{t-i} + E(\boldsymbol{\epsilon}_t | \mathbf{Y}_{t-i}, i \in \mathcal{I}), \\ &= \sum_{i \in \mathcal{I}} \mathbf{\Phi}_{I_i} \mathbf{Y}_{t-i}. \end{aligned} \quad (2.3.7)$$

Si on suppose l'indépendance entre  $\boldsymbol{\epsilon}_t$  et  $\{\mathbf{Y}_s, s < t\}$ , on aura en particulier l'indépendance entre  $\boldsymbol{\epsilon}_t$  et  $\{\mathbf{Y}_{t-i}, i \in \mathcal{I}\}$ , et on aura finalement :

$$E(\boldsymbol{\epsilon}_t | \mathbf{Y}_{t-i}, i \in \mathcal{I}) = E(\boldsymbol{\epsilon}_t) = \mathbf{0}.$$

On constate alors que la meilleure prévision obtenue en (2.3.7) est alors la même que la meilleure prévision linéaire obtenue en (2.3.4) : les deux prévisions coïncident.

Le segment suivant développe les arguments permettant d'obtenir des prévisions pour les modèles SVARX.

### 2.3.2. Construction d'un système de prévision dans les modèles VARX

Considérons les variables aléatoires  $\{\mathbf{Y}_{t-i}, i \in \mathcal{I}, \mathbf{X}_{t-j}, j \in \mathcal{J}\}$ . Nous nous intéressons à trouver la meilleure prévision linéaire de  $\{\mathbf{Y}_t\}$  dans le sens de l'erreur quadratique moyenne de prévision lorsque  $\{\mathbf{Y}_t\}$  suit un modèle autorégressif à variables exogènes à sous-ensemble d'indices. On cherche donc la meilleure prévision linéaire de  $\{\mathbf{Y}_t\}$  en fonction des valeurs passées, par exemple  $\{\mathbf{Y}_{t-i}, i \in \mathcal{I}, \mathbf{X}_{t-j}, j \in \mathcal{J}\}$ . On verra que la présence des valeurs exogènes suscite des questions sur la disponibilité de l'information exogène.

On procède comme dans Brockwell et Davis (1991). On considère que  $\mathbf{Y} = (Y_1, \dots, Y_k)^\top$  et  $\mathbf{X} = (X_1, \dots, X_{k_x})^\top$  sont des vecteurs aléatoires dont les seconds moments existent. Autrement dit, on a  $E(Y_j^2) < \infty$  pour  $j = 1, \dots, k$  et  $E(X_j^2) < \infty$  pour  $j = 1, \dots, k_x$ .

Soient des vecteurs  $\mathbf{U}$  de dimension  $n \times 1$  et  $\mathbf{V}$  de dimension  $m \times 1$ . On dira que  $\mathbf{U}$  et  $\mathbf{V}$  sont orthogonaux si  $E(\mathbf{UV}^\top) = \mathbf{0}_{m \times n}$ .

On introduit les espaces suivants :

$$\begin{aligned} S_n^X &= \text{span} \{X_{jt}, t = 1, \dots, n, j = 1, \dots, k_x\}. \\ S_n^Y &= \text{span} \{Y_{jt}, t = 1, \dots, n, j = 1, \dots, k\}. \\ S_n &= \text{span} \{X_{jt}, t = 1, \dots, n, j = 1, \dots, k_x, Y_{jt}, j = 1, \dots, k\}. \end{aligned}$$

Par  $\text{span} \{X_{jt}, t = 1, \dots, n, j = 1, \dots, k_x\}$ , nous considérons l'espace engendré par  $X_{jt}, t = 1, \dots, n, j = 1, \dots, k_x$ . On pourra dire comme dans Brockwell et Davis (1991) que  $\mathbf{U} \in S_n$  si chaque composante de  $\mathbf{U} \in S_n$ . Si on cherche à prédire  $\mathbf{U}$  par l'information disponible, il faudra trouver  $P(\mathbf{U} \mid \mathbf{Y}_1, \dots, \mathbf{Y}_n, \mathbf{X}_1, \dots, \mathbf{X}_n)$ , c'est-à-dire la meilleure prévision linéaire de  $\mathbf{U}$  exprimée en fonction des variables aléatoires dans l'ensemble  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n, \mathbf{X}_1, \dots, \mathbf{X}_n\}$ .

De façon plus spécifique dans un processus stochastique de type VARX (on discutera des modèles avec sélection d'indices SVARX un plus tard), on pourrait également s'intéresser à la prévision suivante :

$$P(\mathbf{Y}_{n+1} \mid \mathbf{Y}_1, \dots, \mathbf{Y}_n, \mathbf{X}_1, \dots, \mathbf{X}_{n+1}) \quad (2.3.8)$$

afin de prévoir  $\mathbf{Y}_{n+1}$ . On note que dans la formule (2.3.8), l'information disponible inclut l'information disponible au temps  $n + 1$  pour la variable exogène  $\mathbf{X}_t$ , en plus des valeurs passées (on note que dans un modèle SVARX il faudra tenir compte des indices retenus dans le modèle). Autrement dit, on va présumer disposer des valeurs passées de la variable endogène, et pour prévoir  $\mathbf{Y}_{n+1}$ , on supposera disposer des valeurs passées de la variable exogène mais également de la valeur actuelle au temps  $n + 1$ .

La disponibilité de l'information exogène est une particularité des modèles VARX. Un exemple réaliste concerne les prévisions des débits d'eau journaliers en hydrologie statistique. Afin de prévoir le débit au temps  $n + 1$ , des variables exogènes possibles pourraient être reliées à la température. Or, les prévisions de température par Environnement Canada peuvent être considérées assez fiables et précises. En fait, la prévision d'horizon  $l$ , que l'on peut noter  $\hat{\mathbf{Y}}_{n+l}$  pourrait reposer sur  $\{\mathbf{Y}_t, t = 1, \dots, n, \mathbf{X}_t, t = 1, \dots, n\}$ , ou encore sur l'ensemble plus grand  $\{\mathbf{Y}_t, t = 1, \dots, n, \mathbf{X}_t, t = 1, \dots, n + l\}$ . Ceci pourrait être utile lors des prévisions pour des horizons supérieurs à un. On considère alors la recherche de la meilleure prévision linéaire reposant sur  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n, \mathbf{X}_1, \dots, \mathbf{X}_{n+l}\}$ . On cherche à prévoir  $\mathbf{Y}_{n+l}$  pour  $l$  fixé,  $l \geq 1$  un entier. Dans l'exemple d'hydrologie statistique, la température prévue par Environnement Canada est reconnue assez fiable dans des horizons assez courts, disons une semaine.

Afin de simplifier les notations, nous notons :

$$P(\mathbf{Y}_{n+l} \mid \mathbf{Y}_i, i = 1, \dots, n, \mathbf{X}_j, j = 1, \dots, n + l) = P(\mathbf{Y}_{n+l} \mid \mathbf{Y}_{(1:n)}, \mathbf{X}_{(1:(n+l))}),$$

où  $P(\mathbf{Y} \mid \cdot)$  désigne la meilleure prévision linéaire. On aura alors que  $P(\mathbf{Y}_{n+l} \mid \mathbf{Y}_{(1:n)}, \mathbf{X}_{(1:(n+l))}) \in \text{span} \{\mathbf{X}_1, \dots, \mathbf{X}_{n+l}, \mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ , que  $\mathbf{Y}_{n+l} - P(\mathbf{Y}_{n+l} \mid \mathbf{Y}_{(1:n)}, \mathbf{X}_{(1:(n+l))}) \perp \mathbf{Y}_i, i = 1, \dots, n$ , c'est-à-dire que les erreurs de prévisions sont orthogonales aux variables  $\mathbf{Y}_i, i = 1, \dots, n$ , mais

également nous aurons les relations d'orthogonalité  $\mathbf{Y}_{n+l} - P(\mathbf{Y}_{n+l} | \mathbf{Y}_{(1:n)}, \mathbf{X}_{(1:(n+l))}) \perp \mathbf{X}_i, i = 1, \dots, n+l$ .

Concernant les relations d'orthogonalité, pour deux processus de moyennes nulles nous utilisons la notation  $\mathbf{U} \perp \mathbf{V}$  si  $E(\mathbf{UV}^\top) = \mathbf{0}_{m \times n}$  où  $\mathbf{U}$  est de dimension  $n \times 1$  et  $\mathbf{V}$  est de dimension  $m \times 1$ .

Pour  $n \geq 1$ , il existe des matrices  $\Phi_{n1}, \dots, \Phi_{nn}$  ainsi que des matrices  $\mathbf{B}_{n1}, \dots, \mathbf{B}_{n,n+l}$  telles que

$$\hat{\mathbf{Y}}_{n+l} = \Phi_{n1} \mathbf{Y}_n + \dots + \Phi_{nn} \mathbf{Y}_1 + \mathbf{B}_{n1} \mathbf{X}_{n+l} + \dots + \mathbf{B}_{n,n+l} \mathbf{X}_1. \quad (2.3.9)$$

On aura alors les équations de prévision suivantes :

$$E \left\{ \left( \sum_{i=1}^n \Phi_{ni} \mathbf{Y}_{n-i+1} + \sum_{i=1}^{n+l} \mathbf{B}_{ni} \mathbf{X}_{n+l-i+1} \right) \mathbf{Y}_{n+l-i+1}^\top \right\} = \Gamma_{\mathbf{Y}}(j+l-i), j = 1, \dots, n, \quad (2.3.10)$$

$$E \left\{ \left( \sum_{i=1}^n \Phi_{ni} \mathbf{Y}_{n-i+1} + \sum_{i=1}^{n+l} \mathbf{B}_{ni} \mathbf{X}_{n+l-i+1} \right) \mathbf{X}_{n+l-j+1}^\top \right\} = \Gamma_{\mathbf{YX}}(j+l-1), j = 1, \dots, n+l. \quad (2.3.11)$$

Autrement dit, avec  $\{\mathbf{X}_t\}$  et  $\{\mathbf{Y}_t\}$  conjointement stationnaires au sens large :

$$\sum_{i=1}^n \Phi_{ni} \Gamma_{\mathbf{Y}}(j-i) + \sum_{i=1}^{n+l} \mathbf{B}_{ni} \Gamma_{\mathbf{XY}}(l+j-i) = \Gamma_{\mathbf{Y}}(j+l-i), j = 1, \dots, n, \quad (2.3.12)$$

$$\sum_{i=1}^n \Phi_{ni} \Gamma_{\mathbf{YX}}(l+j-i) + \sum_{i=1}^{n+l} \mathbf{B}_{ni} \Gamma_{\mathbf{X}}(j-i) = \Gamma_{\mathbf{YX}}(j+l-i), j = 1, \dots, n+l. \quad (2.3.13)$$

Jusqu'à présent, la discussion n'est encore que générale et porte sur les VARX complets. La prochaine section aborde les modèles SVARX avec sous-sélection d'indices.

### 2.3.3. Construction d'un système de prévision dans les modèles SVARX

Dans cette section nous discutons maintenant des VARX avec sélection d'indices. Posons  $\mathcal{I} = \{i_1, i_2, \dots, i_a\} \subset \{1, \dots, p\}$ ,  $i_1 < \dots < i_a$ , et  $\mathcal{J} = \{j_1, j_2, \dots, j_b\} \subset \{0, 1, \dots, s\}$ ,  $j_1 < \dots < j_b$ . On note que l'ensemble  $\mathcal{I}$  contient des indices strictement positifs, et donc  $\mathbf{Y}_t$  est naturellement prévu par des variables passées. Cependant, dans l'ensemble  $\mathcal{J}$ , l'indice  $j_1$  pourrait être 0, et donc  $\mathbf{Y}_t$  pourrait être expliqué par  $\mathbf{X}_t$ . Il faut donc supposer la disponibilité de l'information exogène contemporaine. Or, la vraie valeur du processus au temps  $t$  est

$$\mathbf{Y}_t = \sum_{i \in \mathcal{I}} \Phi_i \mathbf{Y}_{t-i} + \sum_{j \in \mathcal{J}} \mathbf{B}_j \mathbf{X}_{t-j} + \epsilon_t. \quad (2.3.14)$$

La meilleure prévision linéaire reposant sur  $\{\mathbf{Y}_{t-i}, i \in \mathcal{I}\}$  et sur  $\{\mathbf{X}_{t-j}, j \in \mathcal{J}\}$ , est donc de la forme :

$$\hat{\mathbf{Y}}_t(\mathcal{I}, \mathcal{J}) = \sum_{i \in \mathcal{I}} \Phi_{I_i}^{(f)} \mathbf{Y}_{t-i} + \sum_{j \in \mathcal{J}} \mathbf{B}_{J_j}^{(f)} \mathbf{X}_{t-j}, \quad (2.3.15)$$

puisque la meilleure prévision doit être dans l'espace suivant :

$$\text{span} \{ \mathbf{Y}_{t-i}, i \in \mathcal{I}, \mathbf{X}_{t-j}, j \in \mathcal{J} \}.$$

En suivant la même démarche que pour les modèles vectoriels autorégressifs avec sélection d'indices, on trouve que les coefficients sont obtenus en résolvant les équations suivantes :

$$\sum_{i \in \mathcal{I}} \Phi_{I_i}^{(f)} \Gamma_{\mathbf{Y}}(k-i) + \sum_{j \in \mathcal{J}} \mathbf{B}_{J_j}^{(f)} \Gamma_{\mathbf{X}\mathbf{Y}}(k-i+1) = \Gamma_{\mathbf{Y}}(k), \quad k \in \mathcal{I}, \quad (2.3.16)$$

$$\sum_{i \in \mathcal{I}} \Phi_{I_i}^{(f)} \Gamma_{\mathbf{Y}\mathbf{X}}(k-i+1) + \sum_{j \in \mathcal{J}} \mathbf{B}_{J_j}^{(f)} \Gamma_{\mathbf{X}}(k-i) = \Gamma_{\mathbf{Y}\mathbf{X}}(k), \quad k \in \mathcal{J}. \quad (2.3.17)$$

Dans les cas où les fonctions d'autocovariance ne sont pas connues, on utilise la version échantillonnale vue en (2.3.3). De même, dans le cas où les fonctions de covariances croisées réelles ne sont pas connues, il est possible d'utiliser la version échantillonnale donnée par :

$$\hat{\Gamma}_{\mathbf{X}\mathbf{Y}}(l) = \begin{cases} n^{-1} \sum_{t=1}^{n-l} \mathbf{X}_{t+l} \mathbf{Y}_t^\top, & \text{pour } l = 0, \dots, n-1 \\ \hat{\Gamma}_{\mathbf{X}\mathbf{Y}}(-l), & \text{pour } l = -n+1, \dots, -1. \end{cases} \quad (2.3.18)$$

La meilleure prévision linéaire venant d'être définie est la meilleure prévision linéaire avant. On peut aussi définir une meilleure prévision linéaire arrière de forme :

$$\hat{\mathbf{Y}}_t^{(b)} = \sum_{i \in \mathcal{I}} \Psi_i^{(b)} \mathbf{Y}_{t+i} + \sum_{j \in \mathcal{J}} \eta_j^{(b)} \mathbf{X}_{t+j}, \quad (2.3.19)$$

et la matrice de covariance des erreurs de prévision est donnée par  $\mathbf{V}_{\mathcal{I}}^{(b)} \equiv E [(\mathbf{Y}_t - \hat{\mathbf{Y}}_t^{(b)})(\mathbf{Y}_t - \hat{\mathbf{Y}}_t^{(b)})^\top]$ . Les coefficients  $\Psi_i^{(b)}$  et  $\eta_j^{(b)}$ , peuvent être trouvés en résolvant les équations de Yule-Walker vers l'arrière :

$$\Gamma_{\mathbf{Y}}(i)^\top = \sum_{k \in \mathcal{I}} \Psi_k^{(b)} \Gamma_{\mathbf{Y}}(i-k) + \sum_{k \in \mathcal{J}} \eta_k^{(b)} \Gamma_{\mathbf{Y}\mathbf{X}}(j-k), \quad i \in \mathcal{I}, j \in \mathcal{J}, \quad (2.3.20)$$

$$\mathbf{V}_{\mathcal{I}}^{(b)} = \Gamma_{\mathbf{Y}}(0) - \sum_{i \in \mathcal{I}} \Psi_i^{(b)} \Gamma_{\mathbf{Y}}(i)^\top - \sum_{j \in \mathcal{J}} \eta_j^{(b)} \Gamma_{\mathbf{X}\mathbf{Y}}(j). \quad (2.3.21)$$

Comme précédemment, en supposant que  $\boldsymbol{\epsilon}_t$  et  $\{\mathbf{Y}_s, s < t\}$  sont indépendants pour tout  $t$  et que  $\boldsymbol{\epsilon}_t$  et  $\{\mathbf{X}_s, s < t\}$  sont également indépendants pour tout  $t$  alors, il découle que la meilleure prévision, sans restriction de linéarité, est l'espérance conditionnelle. Ainsi, si l'on cherche à prévoir  $\mathbf{Y}_t$  reposant sur  $\{\mathbf{Y}_{t-i}, i \in \mathcal{I}, \mathbf{X}_{t-j}, j \in \mathcal{J}\}$ , on calcule l'espérance

conditionnelle.

$$\begin{aligned}
E(\mathbf{Y}_t \mid \mathbf{Y}_{t-i}, i \in \mathcal{I}, \mathbf{X}_{t-j}, j \in \mathcal{J}) &= \sum_{i \in \mathcal{I}} \Phi_i \mathbf{Y}_{t-i} \\
&\quad + \sum_{j \in \mathcal{J}} \mathbf{B}_j \mathbf{X}_{t-j} + E(\boldsymbol{\epsilon}_t \mid \mathbf{Y}_{t-i}, i \in \mathcal{I}, \mathbf{X}_{t-j}, j \in \mathcal{J}), \\
&= \sum_{i \in \mathcal{I}} \Phi_i \mathbf{Y}_{t-i} + \sum_{j \in \mathcal{J}} \mathbf{B}_{\mathcal{I}_j} \mathbf{X}_{t-j}.
\end{aligned}$$

En effet, si on suppose l'indépendance entre  $\boldsymbol{\epsilon}_t$  et  $\mathbf{Y}_{t-i}$ ,  $i \in \mathcal{I}$ ,  $\mathbf{X}_{t-j}$ ,  $j \in \mathcal{J}$ , on aura alors que :

$$E(\boldsymbol{\epsilon}_t \mid \mathbf{Y}_{t-i}, i \in \mathcal{I}, \mathbf{X}_{t-j}, j \in \mathcal{J}) = E(\boldsymbol{\epsilon}_t) = \mathbf{0}.$$

Ainsi, on constate que la meilleure prévision coïncide avec la meilleure prévision linéaire dans le cadre des modèles SVARX également.

Dans ce chapitre, nous avons vu la théorie concernant l'estimateur des moindres carrés dans les modèles SVAR et SVARX. Dans le chapitre suivant, nous allons examiner l'aspect pratique en effectuant des simulations.

## Chapitre 3

---

# Étude de simulations : Estimation et prévision dans les modèles à sous-ensemble d'indices

Au chapitre 2, les estimateurs des moindres carrés dans les modèles SVAR et SVARX ainsi que leurs propriétés asymptotiques ont été établies et discutées. Dans ce chapitre, des simulations seront effectuées afin d'apprécier certaines propriétés empiriques de ces estimateurs. En particulier, les biais et écarts-types empiriques de ces estimateurs seront évalués. Pour ce faire, six modèles de séries chronologiques multivariées de taille  $n$  ont été choisis. Ces tailles sont  $n = 50, 100, 200$  et  $n = 400$ . On peut affirmer qu'une taille de  $n = 50$  est petite dans un contexte de séries chronologiques, et  $n = 400$  est relativement grande. Avec des simulations de Monte Carlo, les estimateurs ont été comparés selon les modèles choisis, les diverses tailles échantillonnales et des critères de sélection de modèles couramment utilisés dans la littérature des séries chronologiques.

Les modèles générant les données sont des modèles de type SVAR et SVARX. Les données ont ensuite ajustées à différents modèles. Le modèle choisi est celui qui minimisera le critère de sélection. Les critères de sélection employés afin de choisir le modèle sont le critère d'information d'Akaike (AIC), le critère d'Hannan et Quinn (HQ) ainsi que le critère d'information bayésien (BIC). Ces critères sont décrits dans Brockwell et Davis (2016) et Shumway et Stoffer (2017). Pour chaque critère de sélection de modèles, la proportion de fois que le modèle choisi correspond effectivement au modèle ayant généré les données a été calculée. De plus, pour chaque matrice du processus générateur des données, la proportion de fois où le modèle choisi prend la bonne décision est également calculée. Nous considérons qu'une bonne décision est prise lorsqu'une matrice non nulle dans le processus générant les données est également non nulle dans le modèle choisi ou encore lorsqu'une matrice nulle dans le processus générant les données est également nulle dans le modèle choisi.

Pour finir, puisque les prévisions représentent un aspect primordial des séries chronologiques, nous évaluerons les prévisions d'horizon un en utilisant l'erreur quadratique moyenne

selon les processus générant les données, les tailles échantillonales ainsi que les critères de sélection.

Dans la prochaine section nous présentons les processus générateurs de données utilisés pour les simulations.

### 3.1. Modèles utilisés dans les simulations

Les processus générateurs des données (DGP) suivants ont été utilisés pour les simulations :

- i)  $DGP_1 : \mathbf{Y}_t = \phi_{01} + \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-3} + \epsilon_{1t},$
- ii)  $DGP_2 : \mathbf{Y}_t = \phi_{01} + \Phi_3 \mathbf{Y}_{t-1} + \Phi_4 \mathbf{Y}_{t-3} + \epsilon_{1t},$
- iii)  $DGP_3 : \mathbf{Y}_t = \phi_{01} + \Phi_5 \mathbf{Y}_{t-1} + \Phi_6 \mathbf{Y}_{t-3} + \epsilon_{1t},$
- iv)  $DGP_4 : \mathbf{Y}_t = \phi_{02} + \Phi_3 \mathbf{Y}_{t-1} + \Phi_4 \mathbf{Y}_{t-3} + \mathbf{B}_0 \mathbf{X}_t + \epsilon_{1t},$
- v)  $DGP_5 : \mathbf{Y}_t = \phi_{02} + \Phi_3 \mathbf{Y}_{t-1} + \Phi_4 \mathbf{Y}_{t-3} + \mathbf{B}_0 \mathbf{X}_t + \epsilon_{1t},$
- vi)  $DGP_6 : \mathbf{Y}_t = \phi_{02} + \Phi_5 \mathbf{Y}_{t-1} + \Phi_6 \mathbf{Y}_{t-3} + \mathbf{B}_1 \mathbf{X}_t + \epsilon_{1t}.$

Les modèles (i) et (ii) sont des SVAR(3) bivariés alors que (iii) est un SVAR(3) trivarié. Les modèles (iv) et (v) sont des SVARX(3,0) dont le vecteur endogène est bivarié. On note qu'un modèle SVARX(3,0) contient seulement une variable exogène contemporaine. Finalement, le modèle (vi) est un SVARX(3,0) dont le vecteur endogène est trivarié. Pour les trois modèles de type SVARX, le vecteur exogène est univarié. Il s'agit d'un  $AR(1)$  de moyenne nulle et ayant un paramètre autorégressif  $\phi = 0,4$ . Pour chacun des modèles, les bruits blancs sont indépendants avec distribution gaussienne et de variance égale à un dans le cas univarié, et de matrice identité dans le cas multivarié. Les paramètres des modèles sont détaillés dans le tableau 3.1.

Pour les simulations, nous avons fait usage du *package MTS* en *R*. Ce *package* pour les séries chronologiques multivariées a été créé par l'économètre et statisticien Ruey S. Tsay, voir Tsay (2014). Les processus SVAR(3) ont été simulés à l'aide de la fonction *VARMAsim* du *package MTS* tandis qu'une fonction pour générer des modèles de types VARX a été créée afin de générer les trois derniers modèles de simulation.

Tel que mentionné précédemment, une propriété primordiale des séries chronologiques est l'hypothèse de stationnarité. La première étape, avant même de faire les simulations, consiste donc à vérifier que les processus choisis sont bien stationnaires. Pour s'assurer qu'un processus SVAR( $p$ ) est stationnaire, il faut utiliser le fait qu'un VAR( $p$ ) de dimension  $k$  peut être réécrit sous forme d'un VAR(1) de dimension  $kp$ . Effectivement, on a  $\mathbf{Y}_t = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_{t-p+1}^\top)^\top$ . On peut écrire  $\mathbf{Y}_t = \Phi \mathbf{Y}_{t-1} + \mathbf{e}_t$ , avec  $\mathbf{e}_t = (\epsilon_t^\top, \mathbf{0}^\top, \dots, \mathbf{0}^\top)^\top$  et  $\Phi$  une matrice de dimension  $kp \times kp$  associée à l'opérateur  $\Phi(B)$  du processus VAR( $p$ ). Plus spécifiquement, la matrice

**Tableau 3.1.** Table des paramètres

$$\begin{aligned}
 \phi_{01} &= \begin{bmatrix} 0,020 \\ 0,030 \end{bmatrix}, & \Phi_1 &= \begin{bmatrix} 0,500 & 0,100 \\ 0,400 & 0,500 \end{bmatrix}, & \Phi_2 &= \begin{bmatrix} 0,000 & 0,000 \\ 0,250 & 0,000 \end{bmatrix}, \\
 \mathbf{B}_0 &= \begin{bmatrix} 1,200 \\ 0,800 \end{bmatrix}, & \Phi_3 &= \begin{bmatrix} 0,500 & 0,150 \\ -0,400 & 0,500 \end{bmatrix}, & \Phi_4 &= \begin{bmatrix} 0,200 & 0,100 \\ -0,200 & 0,250 \end{bmatrix}, \\
 \phi_{02} &= \begin{bmatrix} 0,020 \\ 0,030 \\ 0,040 \end{bmatrix}, & \Phi_5 &= \begin{bmatrix} 0,320 & 0,146 & 0,996 \\ 0,044 & -0,153 & 0,289 \\ -0,002 & 0,225 & 0,264 \end{bmatrix}, & \Phi_6 &= \begin{bmatrix} 0,119 & 0,353 & -0,408 \\ -0,009 & -0,071 & 0,120 \\ -0,001 & -0,098 & 0,091 \end{bmatrix}, \\
 & & \mathbf{B}_1 &= \begin{bmatrix} 1,200 \\ 0,800 \\ 0,400 \end{bmatrix}.
 \end{aligned}$$

autorégressive associée au VAR(1) est :

$$\Phi = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdot & \cdot & \cdot & \Phi_p \\ \mathbf{I} & \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdot & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdot & \cdot & \mathbf{I} & \mathbf{0} \end{bmatrix}.$$

Si l'on dispose d'un VAR(1)  $(\mathbf{I} - \Phi B)\mathbf{Y}_t = \epsilon_t$ , on sait que le processus est stationnaire si et seulement si les racines de  $\det\{\mathbf{I} - \Phi B\} = 0$  sont supérieures à un en module. Or,  $\det\{\mathbf{I} - \Phi B\} = 0$  si et seulement si  $\det\{\lambda \mathbf{I} - \Phi\} = 0$  avec  $\lambda = 1/B$ . La condition de stationnarité pour les VAR(1) est donc équivalente à dire que les racines de  $\det\{\lambda \mathbf{I} - \Phi\} = 0$  doivent être inférieures à un en valeur absolue. Cela revient donc à dire que les valeurs propres de  $\Phi$  doivent être strictement inférieures à un en module.

Ainsi, la condition de stationnarité pour un VAR(1) peut également être utilisée pour un VAR( $p$ ), en reformulant le VAR( $p$ ) comme un VAR(1). En conséquence, un VAR( $p$ ) sera

stationnaire pourvu que les valeurs propres de la matrice  $\Phi$  soient inférieures à un en module. Selon ce concept et à l'aide de la fonction *eigenvalue* en  $R$ , la stationnarité des trois premiers DGP, qui sont des modèles VAR, a été vérifiée. En ce qui à trait aux trois derniers DGP qui sont de type VARX, on sait qu'ils seront stationnaires si la partie autorégressive et la partie exogène sont toutes deux stationnaires. On rappelle que la partie exogène consiste en un processus AR(1) avec  $\phi = 0,4$ . Ce processus est définitivement stationnaire car la racine de  $1 - 0,4z = 0$  est 2,5 qui est supérieure à un. La partie autorégressive est la même que celle des trois premiers DGP et donc les modèles VARX générés sont tous stationnaires.

Ainsi, des données de taille échantillonale  $n + 1$  ont été générées selon les six modèles énumérés ci-haut. Les  $n$  observations ont été utilisées pour l'estimation alors que la dernière observation a été utilisée pour l'évaluation des prévisions.

Une fois les données générées, elles ont été ajustées à différents modèles afin de déterminer le modèle minimisant le critère de sélection.

Maintenant que les DGP sont présentés, nous allons, à la section suivante, exposer les modèles utilisés pour l'ajustement des données générées.

## 3.2. Modèles ajustés

Les modèles générant les données sont des modèles de type SVAR et SVARX. Les données des DGP de type VAR ont été ajustées aux modèles VAR suivants :

- (1)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_1 Y_{t-1} + \hat{A}_2 Y_{t-2} + \hat{A}_3 Y_{t-3}$ ,
- (2)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_2 Y_{t-2} + \hat{A}_3 Y_{t-3}$ ,
- (3)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_1 Y_{t-1} + \hat{A}_3 Y_{t-3}$ ,
- (4)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_1 Y_{t-1} + \hat{A}_2 Y_{t-2}$ ,
- (5)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_3 Y_{t-3}$ ,
- (6)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_1 Y_{t-1}$ ,
- (7)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_2 Y_{t-2}$ ,
- (8)  $\hat{Y}_t = \hat{A}_0$ .

On note que le premier modèle est le modèle VAR complet, le troisième modèle correspond au modèle générant les données et le huitième modèle est celui qui ne contient que la constante. Les données des DGP de type VARX ont été ajustées aux modèles VARX et VAR suivants :

- (1)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_1 Y_{t-1} + \hat{A}_2 Y_{t-2} + \hat{A}_3 Y_{t-3} + \hat{B}_0 X_t$ ,
- (2)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_2 Y_{t-2} + \hat{A}_3 Y_{t-3} + \hat{B}_0 X_t$ ,
- (3)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_1 Y_{t-1} + \hat{A}_3 Y_{t-3} + \hat{B}_0 X_t$ ,
- (4)  $\hat{Y}_t = \hat{A}_0 + \hat{A}_1 Y_{t-1} + \hat{A}_2 Y_{t-2} + \hat{B}_0 X_t$ ,

- (5)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_3 \mathbf{Y}_{t-3} + \hat{\mathbf{B}}_0 X_t,$
- (6)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_1 \mathbf{Y}_{t-1} + \hat{\mathbf{B}}_0 X_t,$
- (7)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_2 \mathbf{Y}_{t-2} + \hat{\mathbf{B}}_0 X_t,$
- (8)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{B}}_0 X_t,$
- (9)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_1 \mathbf{Y}_{t-1} + \hat{\mathbf{A}}_2 \mathbf{Y}_{t-2} + \hat{\mathbf{A}}_3 \mathbf{Y}_{t-3},$
- (10)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_2 \mathbf{Y}_{t-2} + \hat{\mathbf{A}}_3 \mathbf{Y}_{t-3},$
- (11)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_1 \mathbf{Y}_{t-1} + \hat{\mathbf{A}}_3 \mathbf{Y}_{t-3},$
- (12)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_1 \mathbf{Y}_{t-1} + \hat{\mathbf{A}}_2 \mathbf{Y}_{t-2},$
- (13)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_3 \mathbf{Y}_{t-3},$
- (14)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_1 \mathbf{Y}_{t-1},$
- (15)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0 + \hat{\mathbf{A}}_2 \mathbf{Y}_{t-2},$
- (16)  $\hat{\mathbf{Y}}_t = \hat{\mathbf{A}}_0.$

Les huit premiers modèles sont de type VARX. Parmi ceux-ci, le premier modèle est le modèle VARX(3,0) complet et le troisième correspond au modèle ayant généré les données. Les huit derniers modèles sont de type VAR.

Les fonctions utilisées sont disponibles dans la librairie *MTS*, en particulier les fonctions *VAR* et *VARX*. Le meilleur modèle ajusté est déterminé en utilisant des critères de sélection. En effet, le modèle ajusté choisi par un critère est celui minimisant le dit critère. La prochaine section présente les critères de sélection de modèles les plus courants en séries chronologiques.

### 3.3. Critères de sélection

Nous utiliserons trois critères de sélection dans nos simulations. Effectivement, nous utiliserons les critères AIC, HQ et BIC, que nous définissons maintenant. Plus précisément, ces critères de sélection de modèle sont de la forme suivante :

$$\chi = \log\{\det(\hat{\Sigma}_{\epsilon})\} + \text{nombre de paramètres libres} \times \frac{C_{n'}}{n'},$$

où  $\hat{\Sigma}_{\epsilon}$  est l'estimateur de  $\Sigma_{\epsilon}$  donné par l'expression  $\sum \hat{\epsilon}\hat{\epsilon}^{\top}/n'$ , et  $C_{n'}$  représente la pénalité associée au critère de sélection. Différentes valeurs ont été considérées dans la littérature pour la pénalité, que nous discutons maintenant.

En effet, le critère AIC a un terme de pénalité valant  $C_{n'} = 2$ , alors que le critère HQ a un terme de pénalité qui est quant à lui  $C_{n'} = 2 \log \log n'$ . Finalement, le critère BIC a un terme de pénalité  $C_{n'} = \log n'$ . Ainsi, pour un VAR( $p$ ) complet, les formules les critères de

sélection AIC, HQ et BIC sont respectivement :

$$\begin{aligned} AIC &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{2}{n'}k(kp + 1), \\ HQ &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{2 \log \log n'}{n'}k(kp + 1), \\ BIC &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{\log n'}{n'}k(kp + 1), \end{aligned}$$

avec  $n' = n - p$ .

Dans le cas des modèles à sous-ensembles d'indices, certaines matrices du modèle sont posées comme étant nulles. Cela fait en sorte que le nombre de paramètres libres diminue. Pour les modèles SVAR( $p$ ), on obtient donc la formule suivante pour nos critères de sélection :

$$\begin{aligned} AIC &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{2}{n'}k(ka + 1), \\ HQ &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{2 \log \log n'}{n'}k(ka + 1), \\ BIC &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{\log n'}{n'}k(ka + 1), \end{aligned}$$

avec  $n' = n - p$  et où l'on rappelle que  $a$  est la cardinalité de l'ensemble  $\mathcal{I} = \{i_1, i_2, \dots, p\}$ .

On a également les formules pour les modèles VARX complets :

$$\begin{aligned} AIC &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{2}{n'}k\{kp + k_x(s + 1)\}, \\ HQ &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{2 \log \log n'}{n'}k\{kp + k_x(s + 1)\}, \\ BIC &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{\log n'}{n'}k\{kp + k_x(s + 1)\}, \end{aligned}$$

avec  $n' = n - \max(p, s)$ .

De même, pour les modèles SVARX avec sélection d'indices, on obtient les formules suivantes pour les critères de sélection :

$$\begin{aligned} AIC &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{2}{n'}k(ka + k_x b), \\ HQ &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{2 \log \log n'}{n'}k(ka + k_x b), \\ BIC &= \log\{\det(\hat{\Sigma}_{\epsilon})\} + \frac{\log n'}{n'}k(ka + k_x b), \end{aligned}$$

avec  $n' = n - \max(p, s)$  et où l'on rappelle que  $a$  est la cardinalité de l'ensemble  $\mathcal{I} = \{i_1, i_2, \dots, i_a\}$ ,  $i_1 < \dots < i_a$  et que  $b$  est la cardinalité de l'ensemble  $\mathcal{J} = \{j_1, j_2, \dots, j_b\}$ ,  $j_1 < \dots < j_b$ .

Les sections subséquentes présentent les résultats des simulations. Nous commençons par exposer les résultats concernant les modèles choisis.

### 3.4. Choix des modèles

Les tables 3.2 et 3.3 contiennent la proportion de fois que le modèle ajusté choisi est le même que celui du DGP ainsi que la proportion de fois où la bonne décision est prise pour chaque matrice du processus pour 10 000 simulations de Monte Carlo. Une bonne décision peut être prise de deux manières : Premièrement, une matrice nulle dans le DGP est également nulle dans le modèle minimisant le critère de sélection. Deuxièmement, une matrice non nulle dans le DGP est également non nulle dans le modèle minimisant le critère de sélection. Les tables 3.4 et 3.5 reportent le nombre de fois que chaque modèle a été choisi pour les DGP de type VAR et pour les DGP de type VARX respectivement.

Tout d'abord, nous remarquons que la proportion de fois où le bon modèle est choisi augmente à mesure que la taille échantillonnale augmente. Effectivement, les taux sont relativement faible pour des séries chronologiques assez courtes, c'est-à-dire  $n = 50$ , alors qu'ils s'approchent de un pour des séries plus longues, correspondant au cas  $n = 400$ . Ensuite, nous constatons que, généralement, le bon modèle est choisi plus fréquemment pour les données générées par des modèles SVAR, soient les DGP<sub>1</sub> à DGP<sub>3</sub>, que pour celles générées par des SVARX, soient les DGP<sub>4</sub> à DGP<sub>6</sub>. Les modèles SVARX demeurent plus complexes dans la description de la dynamique d'une variable réponse expliquée par des retards et une variable exogène stochastique. De plus, les proportions sont meilleures pour certains DGP que pour d'autres. En effet, les taux du DGP<sub>1</sub> sont inférieurs à ceux du DGP<sub>2</sub> et que ceux du DGP<sub>4</sub> sont inférieurs à ceux du DGP<sub>5</sub>. Cela est dû aux matrices composant les processus. Les DGP<sub>1</sub> et DGP<sub>4</sub> incluent la matrice  $\Phi_2$  dont la majorité des éléments sont nuls. Cela a pour effet de diminuer la capacité à sélectionner le bon modèle particulièrement pour le critère BIC. Puisque le critère BIC pénalise plus fortement pour le nombre de paramètres, il choisit de supprimer la matrice et donc prend la mauvaise décision pour cette matrice. D'ailleurs, le tableau 3.4 montre que pour le DGP<sub>1</sub>, les trois critères choisissent le modèle 6, où seule la matrice  $\hat{A}_1$  est différente de zéro, beaucoup plus souvent que pour les DGP<sub>2</sub> et DGP<sub>3</sub>. En outre, les modèles 2,5,7 et 8 ne sont que très peu sélectionnés. En particulier, ces modèles ne sont jamais sélectionnés pour le DGP<sub>3</sub> et ne sont sélectionnés que pour les petites tailles échantillonales pour les DGP<sub>1</sub> et DGP<sub>2</sub>. Notons que ces modèles sont ceux qui posent la matrice  $\hat{A}_1$  comme nulle. Le tableau 3.4 montre également que parmi les trois critères, l'AIC choisit plus souvent le modèle complet (modèle 1) et le BIC choisit plus souvent le modèle qui ne contient le premier délai (modèle 6). Ces comportements sont bien évidemment atténués à mesure que la taille échantillonnale augmente.

Analysant les résultats que l'on trouve dans le tableau 3.5, nous constatons que, similairement, pour les DGP de type SVARX, l'AIC est le critère sélectionnant le plus souvent le modèle entier et que le critère BIC est celui choisissant le plus souvent le modèle 14 où seule la matrice  $\hat{A}_1$  est différente de zéro. Par ailleurs, ce modèle est plus régulièrement élu pour le DGP<sub>4</sub>,

dont la matrice  $\hat{\mathbf{A}}_3$  est très près de la matrice nulle, que pour les DGP<sub>5</sub> et DGP<sub>6</sub>. Nous observons également que, pour des tailles échantillonales plus faibles, le critère BIC a tendance à sélectionner les modèles qui excluent la variable exogène. Comme dans le cas des DGP de type SVAR, les modèles où la matrice  $\hat{\mathbf{A}}_1$  est nulle sont très peu sélectionnés. À mesure que la taille échantillonnale augmente, la sélection des modèles se joue entre le vrai modèle SVARX et le SVAR correspondant.

Nous avons présenté les résultats concernant le choix modèles selon les DGP, les critères et les tailles échantillonales. La prochaine section expose les résultats des simulations obtenus quant à l'estimation des paramètres.

**Tableau 3.2.** Proportion de fois où le bon modèle est choisi

	Critère	$n = 50$	$n = 100$	$n = 200$	$n = 400$
DGP <sub>1</sub>	AIC	0,3996	0,5525	0,7722	0,8936
	HQ	0,2816	0,3724	0,6144	0,9210
	BIC	0,1181	0,1344	0,2651	0,6479
DGP <sub>2</sub>	AIC	0,6325	0,8393	0,8977	0,8984
	HQ	0,5720	0,8566	0,9844	0,9922
	BIC	0,3693	0,6798	0,9699	1,0000
DGP <sub>3</sub>	AIC	0,6785	0,9124	0,9124	0,9566
	HQ	0,5846	0,8963	0,8963	1,0000
	BIC	0,2422	0,5492	0,5492	1,0000
DGP <sub>4</sub>	AIC	0,1039	0,3388	0,7400	0,8941
	HQ	0,0443	0,1350	0,5019	0,9302
	BIC	0,0053	0,0169	0,1193	0,6274
DGP <sub>5</sub>	AIC	0,1770	0,5202	0,8433	0,9036
	HQ	0,0888	0,3094	0,7597	0,9852
	BIC	0,0164	0,0841	0,3959	0,9005
DGP <sub>6</sub>	AIC	0,0970	0,4022	0,8377	0,9609
	HQ	0,0340	0,1468	0,5513	0,9678
	BIC	0,0024	0,0110	0,1387	0,6704

### 3.5. Estimation des paramètres

Pour chacune des 10 000 simulations, pour chaque DGP et pour chaque taille échantillonnale, les paramètres du modèle choisi par le critère de sélection ont été estimés selon les

**Tableau 3.3.** Proportion de fois où la bonne décision est prise pour la matrice

Critère	$n = 50$				$n = 100$				$n = 200$				$n = 400$				
	$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{B}_0$													
DGP <sub>1</sub>	AIC	0,9980	0,8328	0,4761	-	1,0000	0,8630	0,6232	-	1,0000	0,8870	0,8542	-	1,0000	0,8995	0,9897	-
	HQ	0,9941	0,9284	0,3032	-	1,0000	0,9627	0,3803	-	1,0000	0,9786	0,6215	-	1,0000	0,9875	0,9275	-
	BIC	0,9777	0,9816	0,1270	-	0,9999	0,9940	0,1351	-	1,0000	0,9971	0,2652	-	1,0000	0,9987	0,6479	-
DGP <sub>2</sub>	AIC	0,9987	0,8209	0,7471	-	1,0000	0,8689	0,9515	-	1,0000	0,8977	0,9997	-	1,0000	0,8984	1,0000	-
	HQ	0,9953	0,9134	0,6103	-	1,0000	0,9634	0,8765	-	1,0000	0,9871	0,9958	-	1,0000	0,9922	1,0000	-
	BIC	0,9818	0,9725	0,3773	-	1,0000	0,9905	0,6810	-	1,0000	0,9980	0,9703	-	1,0000	1,0000	1,0000	-
DGP <sub>3</sub>	AIC	0,9997	0,8034	0,8034	-	1,0000	0,9215	0,9802	-	1,0000	0,9506	0,9999	-	1,0000	0,9566	1,0000	-
	HQ	0,9955	0,9315	0,6050	-	1,0000	0,9883	0,8994	-	1,0000	0,9988	0,9986	-	1,0000	1,0000	1,0000	-
	BIC	0,9204	0,9859	0,2456	-	1,0000	0,9980	0,5492	-	1,0000	0,9998	0,9655	-	1,0000	1,0000	1,0000	-
DGP <sub>4</sub>	AIC	0,9982	0,8305	0,4650	0,3003	1,0000	0,8574	0,6221	0,6261	1,0000	0,8832	0,8683	0,9444	1,0000	0,8996	0,9897	0,9997
	HQ	0,9959	0,9258	0,3019	0,1642	1,0000	0,9590	0,3798	0,3572	1,0000	0,9759	0,6475	0,7676	1,0000	0,9896	0,9435	0,9916
	BIC	0,9872	0,9796	0,1265	0,0515	1,0000	0,9941	0,1337	0,1137	1,0000	0,9979	0,2791	0,3894	1,0000	0,9988	0,6932	0,8963
DGP <sub>5</sub>	AIC	0,9989	0,8209	0,7424	0,2826	1,0000	0,8714	0,9543	0,6116	1,0000	0,8941	0,9993	0,9436	1,0000	0,9038	1,0000	0,9998
	HQ	0,9970	0,9153	0,6053	0,1520	1,0000	0,9662	0,881	0,3497	1,0000	0,9870	0,9979	0,7696	1,0000	0,9932	1,0000	0,9920
	BIC	0,9840	0,9732	0,3801	0,0458	1,0000	0,9905	0,6892	0,1114	1,0000	0,9988	0,9772	0,4034	1,0000	1,0000	1,0000	0,9005
DGP <sub>6</sub>	AIC	0,9999	0,8054	0,7914	0,1600	1,0000	0,9217	0,9785	0,4418	1,0000	0,9475	1,0000	0,8820	1,0000	0,9621	1,0000	0,9987
	HQ	0,9941	0,9332	0,5817	0,0652	1,0000	0,9873	0,8920	0,1665	1,0000	0,9993	0,9993	0,5519	1,0000	0,9998	1,0000	0,9680
	BIC	0,9019	0,9879	0,2294	0,0074	0,9996	0,9985	0,5220	0,0206	1,0000	1,0000	0,9614	0,1419	1,0000	1,0000	1,0000	0,6704

formules vues au chapitre 2. Puisque les processus générant les données sont connus, nous avons pu calculer le biais des estimateurs ainsi que leur écart-type. Notons que l'écart-type d'un coefficient est considéré comme nul lorsque la matrice des coefficients est posée comme étant nulle. Ainsi, la moyenne des biais et la moyenne des écarts-types ont été calculées et compilées dans les tableaux 3.6 à 3.17.

Certainement, le biais des coefficients ainsi que leur écart-type diminuent alors que la taille échantillonnale augmente. Pour le DGP<sub>1</sub>, le critère BIC produit des estimateurs plus fortement biaisés que les critères AIC et HQ pour la matrice  $\hat{A}_1$  et pour l'élément non nul de la matrice  $\Phi_2$ . Tel que vu dans le tableau 3.3, la matrice  $\hat{A}_2$  est moins biaisée en utilisant le critère BIC car celui-ci l'estime plus souvent correctement comme étant nulle. Il est aussi intéressant de remarquer que le biais de la majorité des coefficients est négatif indiquant ainsi qu'en moyenne, l'estimation sous-estime la vraie valeur des paramètres du DGP<sub>1</sub>. Malgré le fait que, pour ce processus, le critère BIC est celui sélectionnant le bon modèle le moins régulièrement, l'écart-type moyen des coefficients est le plus faible pour ce critère. Pour le DGP<sub>2</sub>, le critère BIC ne choisit pas le bon modèle pour les petites tailles échantillonnales. Pour ces tailles, on retrouve encore que les estimateurs obtenus de ce critère sont plus fortement biaisés mais que leurs écarts-types sont plus faibles. Cependant, à mesure que la taille échantillonnale augmente, le critère BIC choisit plus souvent le vrai modèle jusqu'à le choisir à chaque simulation pour  $n = 400$ . Malgré cela, pour le DGP<sub>2</sub>, le biais des matrices  $\hat{A}_1$  et  $\hat{A}_3$  est le plus grand pour ce critère. La matrice  $\hat{A}_2$  étant toujours mise comme nulle a donc un biais ainsi qu'un écart-type de 0. Pour le DGP<sub>3</sub>, on obtient

**Tableau 3.4.** Fréquence de sélection des modèles pour les VAR

Modèle testé	$n = 50$			$n = 100$			$n = 200$			$n = 400$			
	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC	
DGP <sub>1</sub>	1	756	188	16	707	79	6	820	71	1	961	65	0
	2	4	7	2	0	0	0	0	0	0	0	0	0
	3	3996	2816	1181	5525	3724	1344	7722	6144	2651	8936	9210	6479
	4	911	516	145	663	294	54	310	143	28	44	60	13
	5	5	21	71	0	0	1	0	0	0	0	0	0
	6	4317	6421	8435	3105	5903	8595	1148	3642	7320	59	665	3508
	7	1	5	21	0	0	0	0	0	0	0	0	0
	8	10	26	129	0	0	0	0	0	0	0	0	0
DGP <sub>2</sub>	1	1139	364	31	1122	199	12	1020	114	4	1016	78	0
	2	2	3	3	0	0	0	0	0	0	0	0	0
	3	6325	5720	3693	8393	8566	6798	8977	9844	9699	8984	9922	10000
	4	648	493	226	189	167	83	3	15	16	0	0	0
	5	5	16	46	0	0	0	0	0	0	0	0	0
	6	1875	3376	5868	296	1068	3107	0	27	281	0	0	0
	7	2	6	15	0	0	0	0	0	0	0	0	0
	8	4	22	118	0	0	0	0	0	0	0	0	0
DGP <sub>3</sub>	1	1247	190	6	678	31	0	493	9	0	434	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0
	3	6785	5846	2422	9124	8963	5492	9506	9977	9655	9566	10000	10000
	4	719	495	135	107	86	20	1	3	2	0	0	0
	5	2	14	28	0	0	0	0	0	0	0	0	0
	6	1246	3424	6641	91	920	4488	0	11	343	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0
	8	1	31	768	0	0	0	0	0	0	0	0	0

toujours que le critère BIC produit des estimateurs à faible écart-type. Cependant, la différence entre l'écart-type obtenu pour chaque critère diminue à force que  $n$  augmente. De façon analogue au DGP<sub>2</sub>, les estimateurs les plus biaisés sont ceux de la matrice  $\hat{\mathbf{A}}_1$  ainsi que l'estimateur de l'élément non nul de la matrice  $\hat{\mathbf{A}}_3$ . Nous remarquons que le biais et l'écart-type moyen des coefficients du DGP<sub>1</sub> sont généralement supérieurs à ceux pour du DGP<sub>2</sub>. Encore une fois, nous constatons que les résultats varient fortement selon les éléments des matrices composant les processus générant les données.

Passons à présent à la discussion des résultats pour les processus de type SVARX. Tout comme pour le SVAR correspondant, nous observons que le critère BIC produit des estimateurs plus fortement biaisés que les critères AIC et HQ pour tous les paramètres

non nuls du  $DGP_4$ . Pour le  $DGP_5$ , de mauvaises décisions sont prises pour les matrices  $\hat{\mathbf{A}}_3$  et  $\hat{\mathbf{B}}_0$  de sorte que le biais des coefficients pour le critère BIC est supérieur à celui des autres critères pour ces matrices. Le critère BIC agit de façon similaire au  $DGP_5$  pour le  $DGP_6$ . En effet, les matrices  $\hat{\mathbf{A}}_3$  et  $\hat{\mathbf{B}}_0$  sont celles ayant le plus grand biais lorsque  $n$  augmente. Pour  $n=400$ , le critère BIC parvient à prendre la bonne décision pour toutes les matrices sauf pour la matrice  $\hat{\mathbf{B}}_0$  laquelle est élue comme non nulle pour 67,04% des simulations. Ainsi, le biais de  $\hat{\mathbf{B}}_0$  est le plus grand pour le BIC même pour un large échantillon. Pour  $n = 400$ , le critère AIC peine légèrement à prendre la bonne décision quant à la matrice  $\hat{\mathbf{A}}_2$  alors que le critère HQ peine à prendre la bonne décision par rapport à la matrice  $\hat{\mathbf{B}}_0$ . De façon générale, l'écart-type moyen des coefficients pour le critère BIC se trouve à être inférieur à ceux des deux autres critères évalués pour les DGP de type SVARX également. Il s'agit d'une fausse impression de précision car, comme mentionné précédemment, les estimateurs obtenus avec le critère BIC sont les plus biaisés. Nous relevons également que les biais et écarts-types moyens du  $DGP_5$  sont inférieurs à ceux du  $DGP_4$ .

La section suivante contient les résultats obtenus pour les prévisions.

### 3.6. Prévisions

Dans le cadre de nos simulations, nous avons jugé pertinent d'évaluer les prévisions d'horizon un obtenues selon les critères de sélection et les tailles échantillonales pour chaque processus générant les données. Pour ce faire, lors de la génération des processus,  $n + 1$  observations ont été simulées. Notons qu'afin de générer les  $n + 1$  observations pour les DGP de type VARX,  $n + 1$  observations du vecteur exogène ont du être générées également. La dernière observation du DGP et du vecteur exogène n'ont pas été utilisées pour l'estimation mais ont été réservées pour l'appréciation des prévisions. Dans le cas des VARX, le vecteur exogène est supposé connu pour le calcul des prévisions. Ainsi, pour les prévisions d'horizon un, nous faisons usage de la valeur du vecteur exogène au temps  $n + 1$ . Pour chaque modèle élu par un critère de sélection, les prévisions ont été calculées à l'aide des fonctions *VARpred* et *VARXpred* du *package MTS* selon les formules vues au chapitre 2. Obtenant ainsi les prévisions, nous avons pu les comparer aux vraies valeurs en calculant la matrice de covariance des erreurs de prévision. Comme critère d'évaluation des prévisions, nous utilisons l'erreur quadratique moyenne, qui se trouve à être représentée par les éléments diagonaux de la matrice de covariance des erreurs de prévision. Le résultats se trouvent dans le tableau 3.18.

Pour tous les processus et tous les critères, nous nous attendions à ce que l'erreur quadratique moyenne de la prévision d'horizon un diminue à mesure que la taille échantionnale augmente. Cela est effectivement le cas pour les processus de type SVARX. Cela n'est

cependant pas toujours le cas lorsque les échantillons passent de 200 à 400 observations pour les processus SVAR. Dans l'ensemble, l'erreur quadratique moyenne est plus faible pour les processus SVAR que pour leurs SVARX correspondants. Cependant, pour  $n = 400$ , nous observons que, dans plusieurs cas, l'EQM des processus SVARX est inférieure à celle des processus SVAR. En comparant, l'EQM selon les critères de sélection, nous remarquons qu'il ne semble pas y avoir de régularité. Pour de faible  $n$ , les plus faibles EQM sont partagés entre le critère AIC et le critère BIC. Lorsque  $n$  augmente, on observe quelques situations dans lesquelles le critère HQ produit le plus petit EQM.

**Tableau 3.5.** Fréquence de sélection des modèles pour les VARX

Modèle testé	$n = 50$			$n = 100$			$n = 200$			$n = 400$			
	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC	
DGP <sub>4</sub>	1	232	37	0	471	37	0	804	54	0	953	57	0
	2	4	3	0	0	0	0	0	0	0	0	0	0
	3	1039	443	53	3388	1350	169	7400	5019	1193	8941	9302	6274
	4	315	107	11	449	126	11	297	131	6	50	46	12
	5	1	3	3	0	0	0	0	0	0	0	0	0
	6	1412	1048	444	1953	2059	957	943	2472	2695	53	511	2677
	7	0	0	2	0	0	0	0	0	0	0	0	0
	8	0	1	2	0	0	0	0	0	0	0	0	0
	9	525	151	14	271	55	3	51	19	0	1	0	0
	10	2	3	3	0	0	0	0	0	0	0	0	0
	11	2843	2371	1152	2091	2356	1165	428	1383	1598	2	76	658
	12	615	435	159	235	192	45	16	37	15	0	1	0
	13	4	8	40	0	0	0	0	0	0	0	0	0
	14	3001	5367	8039	1142	3825	7650	61	885	4493	0	7	379
	15	2	6	15	0	0	0	0	0	0	0	0	0
	16	5	17	63	0	0	0	0	0	0	0	0	0
DGP <sub>5</sub>	1	309	59	3	665	75	4	997	85	1	962	68	0
	2	1	0	0	0	0	0	0	0	0	0	0	0
	3	1770	888	164	5202	3094	841	8433	7597	3959	9036	9852	9005
	4	206	80	16	99	42	11	6	7	2	0	0	0
	5	1	2	5	0	0	0	0	0	0	0	0	0
	6	535	486	257	150	286	258	0	7	72	0	0	0
	7	2	4	1	0	0	0	0	0	0	0	0	0
	8	2	1	12	0	0	0	0	0	0	0	0	0
	9	812	292	46	440	103	4	55	36	2	0	0	0
	10	1	0	1	0	0	0	0	0	0	0	0	0
	11	4527	4800	3546	3236	5538	6043	508	2261	5810	2	80	995
	12	459	408	187	82	118	76	1	2	7	0	0	0
	13	3	12	36	0	0	0	0	0	0	0	0	0
	14	1371	2957	5621	126	744	2763	0	5	147	0	0	0
	15	1	4	14	0	0	0	0	0	0	0	0	0
	16	0	7	91	0	0	0	0	0	0	0	0	0
DGP <sub>6</sub>	1	192	13	0	302	8	0	443	3	0	378	2	0
	2	0	0	0	0	0	0	0	0	0	0	0	0
	3	970	340	24	4022	1468	110	8377	5513	1387	9609	9678	6704
	4	143	42	3	36	17	0	0	0	0	0	0	0
	5	0	0	1	0	0	0	0	0	0	0	0	0
	6	295	257	46	58	172	96	0	3	32	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	0	0	0	0
	9	1032	188	6	384	35	0	82	4	0	1	0	0
	10	0	0	0	0	0	0	0	0	0	0	0	0
	11	5720	5264	2224	5077	7409	5110	1098	4473	8227	12	320	3296
	12	579	424	112	61	67	15	0	0	0	0	0	0
	13	0	12	39	0	0	0	0	0	0	0	0	0
	14	1068	3413	6604	60	824	4665	0	4	354	0	0	0
	15	0	1	0	0	0	0	0	0	0	0	0	0
	16	1	46	941	0	0	4	0	0	0	0	0	0

**Tableau 3.6.** Biais moyen des coefficients obtenus pour le DGP<sub>1</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$
$n = 50$	AIC	$\begin{bmatrix} -0,0658 & -0,0148 \\ 0,0100 & -0,0255 \end{bmatrix}$	$\begin{bmatrix} -0,0115 & -0,0036 \\ 0,0164 & -0,0026 \end{bmatrix}$	$\begin{bmatrix} -0,0138 & -0,0191 \\ -0,0785 & -0,0090 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0686 & -0,0202 \\ 0,0170 & -0,0089 \end{bmatrix}$	$\begin{bmatrix} -0,0059 & -0,0027 \\ 0,0114 & -0,0004 \end{bmatrix}$	$\begin{bmatrix} -0,0109 & -0,0128 \\ -0,1318 & -0,0071 \end{bmatrix}$
	BIC	$\begin{bmatrix} -0,0725 & -0,0276 \\ 0,0227 & 0,0073 \end{bmatrix}$	$\begin{bmatrix} -0,0009 & -0,0006 \\ 0,0047 & 0,0012 \end{bmatrix}$	$\begin{bmatrix} -0,0056 & -0,0061 \\ -0,1940 & -0,0026 \end{bmatrix}$
$n = 100$	AIC	$\begin{bmatrix} -0,0305 & -0,0050 \\ 0,0087 & -0,0067 \end{bmatrix}$	$\begin{bmatrix} -0,0036 & -0,0004 \\ 0,0094 & 0,0007 \end{bmatrix}$	$\begin{bmatrix} -0,0099 & -0,0099 \\ -0,0552 & -0,0057 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0316 & -0,0078 \\ 0,0199 & 0,0157 \end{bmatrix}$	$\begin{bmatrix} -0,0013 & -0,0010 \\ 0,0053 & 0,0017 \end{bmatrix}$	$\begin{bmatrix} -0,0078 & -0,0060 \\ -0,1202 & -0,0049 \end{bmatrix}$
	BIC	$\begin{bmatrix} -0,0311 & -0,0120 \\ 0,0325 & 0,0438 \end{bmatrix}$	$\begin{bmatrix} -0,0004 & -0,0004 \\ 0,0012 & 0,0004 \end{bmatrix}$	$\begin{bmatrix} -0,0031 & -0,0025 \\ -0,1973 & -0,0021 \end{bmatrix}$
$n = 200$	AIC	$\begin{bmatrix} -0,0144 & -0,0008 \\ 0,0044 & -0,0064 \end{bmatrix}$	$\begin{bmatrix} -0,0026 & -0,0006 \\ 0,0036 & 0,0008 \end{bmatrix}$	$\begin{bmatrix} -0,007 & -0,0056 \\ -0,0174 & -0,0036 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0157 & -0,0024 \\ 0,0131 & 0,0105 \end{bmatrix}$	$\begin{bmatrix} -0,0006 & -0,0003 \\ 0,0025 & 0,0014 \end{bmatrix}$	$\begin{bmatrix} -0,0059 & -0,0046 \\ -0,0680 & -0,0030 \end{bmatrix}$
	BIC	$\begin{bmatrix} -0,0156 & -0,0049 \\ 0,0288 & 0,0442 \end{bmatrix}$	$\begin{bmatrix} 0,0000 & 0,0000 \\ 0,0005 & 0,0004 \end{bmatrix}$	$\begin{bmatrix} -0,0032 & -0,0026 \\ -0,1624 & -0,0015 \end{bmatrix}$
$n = 400$	AIC	$\begin{bmatrix} -0,0003 & -0,0012 \\ -0,0059 & 0,0009 \end{bmatrix}$	$\begin{bmatrix} -0,0003 & -0,0031 \\ -0,0003 & 0,0016 \end{bmatrix}$	$\begin{bmatrix} -0,0029 & 0,0000 \\ -0,0022 & 0,0000 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0005 & -0,0003 \\ -0,0028 & 0,0008 \end{bmatrix}$	$\begin{bmatrix} -0,0001 & -0,0034 \\ 0,0005 & -0,0091 \end{bmatrix}$	$\begin{bmatrix} -0,0029 & 0,0000 \\ -0,0023 & 0,0000 \end{bmatrix}$
	BIC	$\begin{bmatrix} -0,0014 & -0,0001 \\ 0,0185 & 0,0003 \end{bmatrix}$	$\begin{bmatrix} 0,0000 & -0,0025 \\ 0,0001 & -0,0694 \end{bmatrix}$	$\begin{bmatrix} -0,0022 & 0,0000 \\ -0,0018 & 0,0000 \end{bmatrix}$

**Tableau 3.7.** Écart-type moyen des coefficients obtenus pour le DGP<sub>1</sub>

		$\hat{A}_1$		$\hat{A}_2$		$\hat{A}_3$	
$n = 50$	AIC	0,1384	0,1208	0,0280	0,0244	0,0736	0,0569
		0,1412	0,1208	0,0285	0,0244	0,0729	0,0556
	HQ	0,1353	0,1130	0,0116	0,0098	0,0462	0,0355
		0,1388	0,1138	0,0119	0,0099	0,0453	0,0343
	BIC	0,1312	0,1048	0,0029	0,0023	0,0190	0,0146
		0,1357	0,1067	0,0030	0,0024	0,0184	0,0140
$n = 100$	AIC	0,0948	0,0836	0,0160	0,0140	0,0656	0,0497
		0,0957	0,0836	0,0162	0,0140	0,0652	0,0491
	HQ	0,0930	0,0774	0,0042	0,0035	0,0396	0,0297
		0,0946	0,0779	0,0043	0,0035	0,0390	0,0292
	BIC	0,0920	0,0723	0,0007	0,0006	0,0140	0,0105
		0,0945	0,0737	0,0007	0,0005	0,0136	0,0101
$n = 200$	AIC	0,0661	0,0599	0,0094	0,0084	0,0624	0,0468
		0,0664	0,0599	0,0095	0,0084	0,0623	0,0467
	HQ	0,0650	0,0561	0,0017	0,0014	0,0450	0,0336
		0,0656	0,0563	0,0018	0,0015	0,0448	0,0333
	BIC	0,0642	0,0514	0,0002	0,0002	0,0192	0,0143
		0,0656	0,0522	0,0002	0,0002	0,0189	0,0140
$n = 400$	AIC	0,0463	0,0427	0,0060	0,0055	0,0507	0,0378
		0,0464	0,0427	0,0060	0,0055	0,0507	0,0378
	HQ	0,0458	0,0416	0,0007	0,0006	0,0470	0,0349
		0,0459	0,0416	0,0007	0,0006	0,0470	0,0349
	BIC	0,0454	0,0391	0,0001	0,0001	0,0328	0,0244
		0,0458	0,0393	0,0001	0,0001	0,0327	0,0242

**Tableau 3.8.** Biais moyen des coefficients obtenus pour le DGP<sub>2</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$
$n = 50$	AIC	$\begin{bmatrix} -0,0465 & 0,0116 \\ 0,0026 & -0,0657 \end{bmatrix}$	$\begin{bmatrix} -0,0136 & 0,0112 \\ 0,0007 & -0,0098 \end{bmatrix}$	$\begin{bmatrix} 0,0261 & 0,0060 \\ -0,0623 & -0,0696 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0459 & 0,0221 \\ 0,0103 & -0,0676 \end{bmatrix}$	$\begin{bmatrix} -0,0098 & 0,0095 \\ 0,0041 & -0,0035 \end{bmatrix}$	$\begin{bmatrix} 0,0502 & -0,0096 \\ -0,0813 & -0,0971 \end{bmatrix}$
	BIC	$\begin{bmatrix} -0,0439 & 0,0401 \\ 0,0331 & -0,0694 \end{bmatrix}$	$\begin{bmatrix} -0,0050 & 0,0055 \\ 0,0037 & -0,0004 \end{bmatrix}$	$\begin{bmatrix} 0,0989 & -0,0400 \\ -0,1200 & -0,1487 \end{bmatrix}$
$n = 100$	AIC	$\begin{bmatrix} -0,0238 & 0,0046 \\ -0,0021 & -0,0314 \end{bmatrix}$	$\begin{bmatrix} -0,0033 & 0,0033 \\ 0,0005 & -0,0047 \end{bmatrix}$	$\begin{bmatrix} 0,0012 & 0,0085 \\ -0,0230 & -0,0233 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0241 & 0,0085 \\ 0,0014 & -0,0323 \end{bmatrix}$	$\begin{bmatrix} -0,0028 & 0,0029 \\ 0,0026 & -0,0007 \end{bmatrix}$	$\begin{bmatrix} 0,0114 & 0,0026 \\ -0,0323 & -0,0375 \end{bmatrix}$
	BIC	$\begin{bmatrix} -0,0218 & 0,0204 \\ 0,0168 & -0,0302 \end{bmatrix}$	$\begin{bmatrix} -0,0015 & 0,0017 \\ 0,0021 & 0,0002 \end{bmatrix}$	$\begin{bmatrix} 0,0441 & -0,0160 \\ -0,0626 & -0,0762 \end{bmatrix}$
$n = 200$	AIC	$\begin{bmatrix} -0,0112 & 0,0028 \\ -0,0009 & -0,0165 \end{bmatrix}$	$\begin{bmatrix} -0,0016 & 0,0006 \\ -0,0016 & -0,0018 \end{bmatrix}$	$\begin{bmatrix} -0,0021 & 0,0060 \\ -0,0090 & -0,0086 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0123 & 0,0032 \\ -0,0015 & -0,0172 \end{bmatrix}$	$\begin{bmatrix} -0,0004 & 0,0003 \\ -0,0001 & 0,0001 \end{bmatrix}$	$\begin{bmatrix} -0,0025 & 0,0058 \\ -0,0095 & -0,0101 \end{bmatrix}$
	BIC	$\begin{bmatrix} -0,0122 & 0,0045 \\ 0,0000 & -0,0170 \end{bmatrix}$	$\begin{bmatrix} -0,0003 & 0,0002 \\ 0,0003 & 0,0002 \end{bmatrix}$	$\begin{bmatrix} 0,0010 & 0,0041 \\ -0,0129 & -0,0143 \end{bmatrix}$
$n = 400$	AIC	$\begin{bmatrix} -0,0059 & 0,0013 \\ 0,0000 & -0,0082 \end{bmatrix}$	$\begin{bmatrix} -0,0003 & 0,0001 \\ -0,0008 & -0,0009 \end{bmatrix}$	$\begin{bmatrix} -0,0011 & 0,0030 \\ -0,0046 & -0,0046 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0063 & 0,0014 \\ -0,0004 & -0,0086 \end{bmatrix}$	$\begin{bmatrix} -0,0002 & -0,0001 \\ -0,0001 & -0,0001 \end{bmatrix}$	$\begin{bmatrix} -0,0013 & 0,0031 \\ -0,0046 & -0,0051 \end{bmatrix}$
	BIC	$\begin{bmatrix} -0,0064 & 0,0015 \\ -0,0004 & -0,0086 \end{bmatrix}$	$\begin{bmatrix} 0,0000 & 0,0000 \\ 0,0000 & 0,0000 \end{bmatrix}$	$\begin{bmatrix} -0,0013 & 0,0031 \\ -0,0046 & -0,0051 \end{bmatrix}$

**Tableau 3.9.** Écart-type moyen des coefficients obtenus pour le DGP<sub>2</sub>

		$\hat{A}_1$		$\hat{A}_2$		$\hat{A}_3$	
$n = 50$	AIC	0,1328	0,1287	0,0293	0,0277	0,0978	0,0980
		0,1342	0,1275	0,0294	0,0273	0,0984	0,0973
	HQ	0,1295	0,1249	0,0136	0,0128	0,0776	0,0787
		0,1307	0,1237	0,0137	0,0127	0,0782	0,0782
	BIC	0,1246	0,1200	0,0040	0,0038	0,0465	0,0477
		0,1259	0,1190	0,0041	0,0038	0,0468	0,0475
$n = 100$	AIC	0,091	0,0861	0,0152	0,0143	0,0842	0,0837
		0,0914	0,0857	0,0153	0,0142	0,0845	0,0833
	HQ	0,0895	0,0843	0,0040	0,0038	0,0758	0,0760
		0,0899	0,0840	0,0040	0,0038	0,0760	0,0757
	BIC	0,0882	0,0830	0,0010	0,0009	0,0581	0,0586
		0,0888	0,0829	0,0010	0,0009	0,0582	0,0583
$n = 200$	AIC	0,0629	0,0587	0,0085	0,0079	0,0609	0,0603
		0,0631	0,0586	0,0085	0,0079	0,0610	0,0603
	HQ	0,0621	0,0578	0,0010	0,0010	0,0596	0,0594
		0,0623	0,0577	0,0010	0,0010	0,0597	0,0594
	BIC	0,0620	0,0576	0,0001	0,0001	0,0578	0,0578
		0,0622	0,0575	0,0001	0,0001	0,0580	0,0577
$n = 400$	AIC	0,0440	0,0409	0,0059	0,0055	0,0425	0,0421
		0,0441	0,0408	0,0059	0,0055	0,0426	0,0420
	HQ	0,0435	0,0401	0,0005	0,0004	0,0418	0,0416
		0,0435	0,0401	0,0005	0,0004	0,0418	0,0415
	BIC	0,0434	0,0401	0,0000	0,0000	0,0417	0,0415
		0,0435	0,0400	0,0000	0,0000	0,0417	0,0415

**Tableau 3.10.** Biais moyen des coefficients obtenus pour le DGP<sub>3</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$		
$n = 50$	AIC	$\begin{bmatrix} -0,0166 & -0,0010 & 0,0007 \\ -0,0107 & -0,0082 & -0,0033 \\ -0,0078 & -0,0073 & -0,0194 \end{bmatrix}$	$\begin{bmatrix} -0,0263 & -0,0082 & 0,0162 \\ 0,0018 & -0,0161 & 0,0098 \\ 0,0033 & 0,0029 & -0,0165 \end{bmatrix}$	$\begin{bmatrix} -0,0281 & -0,0492 & 0,0640 \\ -0,0042 & -0,0350 & -0,0182 \\ -0,0051 & -0,0023 & -0,0517 \end{bmatrix}$		
		HQ	$\begin{bmatrix} -0,0106 & -0,0014 & -0,0038 \\ -0,0093 & -0,0024 & -0,0038 \\ -0,0115 & -0,0093 & -0,0206 \end{bmatrix}$	$\begin{bmatrix} -0,0156 & -0,0083 & 0,0117 \\ 0,0019 & -0,0054 & 0,0057 \\ 0,0019 & 0,0023 & -0,0080 \end{bmatrix}$	$\begin{bmatrix} -0,0450 & -0,1151 & 0,1310 \\ -0,0014 & -0,0385 & -0,0427 \\ -0,0036 & 0,0150 & -0,0612 \end{bmatrix}$	
			BIC	$\begin{bmatrix} 0,0139 & -0,0092 & -0,0678 \\ -0,0094 & 0,0115 & -0,0164 \\ -0,0122 & -0,0242 & -0,0372 \end{bmatrix}$	$\begin{bmatrix} -0,0048 & -0,0031 & 0,0046 \\ 0,0010 & -0,0014 & 0,0016 \\ 0,0006 & 0,0007 & -0,0026 \end{bmatrix}$	$\begin{bmatrix} -0,0861 & -0,2488 & 0,2869 \\ 0,0040 & -0,0541 & -0,0853 \\ -0,0008 & 0,0590 & -0,0771 \end{bmatrix}$
	$n = 100$	AIC	$\begin{bmatrix} -0,0037 & -0,0003 & 0,0008 \\ -0,0043 & -0,0042 & -0,0020 \\ -0,0062 & -0,0034 & -0,0092 \end{bmatrix}$	$\begin{bmatrix} -0,0054 & -0,0004 & 0,0028 \\ -0,0002 & -0,0035 & 0,0017 \\ 0,0006 & 0,0004 & -0,0040 \end{bmatrix}$	$\begin{bmatrix} -0,0068 & -0,0046 & 0,0078 \\ -0,0025 & -0,0126 & -0,0013 \\ -0,0035 & -0,0066 & -0,0240 \end{bmatrix}$	
			HQ	$\begin{bmatrix} -0,0022 & -0,0007 & 0,0000 \\ -0,0038 & -0,0024 & -0,0018 \\ -0,0074 & -0,0038 & -0,0090 \end{bmatrix}$	$\begin{bmatrix} -0,0024 & -0,0017 & 0,0017 \\ 0,0004 & -0,0005 & 0,0010 \\ 0,0005 & 0,0006 & -0,0008 \end{bmatrix}$	$\begin{bmatrix} -0,0130 & -0,0274 & 0,0315 \\ -0,0018 & -0,0146 & -0,0101 \\ -0,0032 & -0,0018 & -0,0288 \end{bmatrix}$
				BIC	$\begin{bmatrix} -0,0016 & 0,0002 & -0,0052 \\ -0,0040 & 0,0029 & 0,0032 \\ -0,0080 & -0,0051 & -0,0089 \end{bmatrix}$	$\begin{bmatrix} -0,0006 & -0,0005 & 0,0005 \\ 0,0001 & -0,0001 & 0,0004 \\ 0,0001 & 0,0001 & -0,0001 \end{bmatrix}$
$n = 200$	AIC	$\begin{bmatrix} -0,0009 & 0,0007 & 0,0000 \\ -0,0019 & -0,0015 & -0,0019 \\ -0,0039 & -0,0014 & -0,0044 \end{bmatrix}$	$\begin{bmatrix} -0,0011 & 0,0002 & 0,0003 \\ -0,0002 & -0,0015 & 0,0005 \\ 0,0000 & -0,0003 & -0,0013 \end{bmatrix}$	$\begin{bmatrix} -0,0021 & -0,0005 & -0,0007 \\ -0,0012 & -0,0066 & -0,0003 \\ -0,0014 & -0,0044 & -0,0107 \end{bmatrix}$		
		HQ	$\begin{bmatrix} -0,0005 & 0,0007 & 0,0001 \\ -0,0017 & -0,0010 & -0,0022 \\ -0,0045 & -0,0015 & -0,0044 \end{bmatrix}$	$\begin{bmatrix} -0,0001 & 0,0000 & 0,0001 \\ 0,0000 & 0,0000 & 0,0001 \\ 0,0000 & 0,0000 & 0,0000 \end{bmatrix}$	$\begin{bmatrix} -0,0017 & -0,0010 & -0,0014 \\ -0,0012 & -0,0062 & -0,0010 \\ -0,0015 & -0,0047 & -0,0112 \end{bmatrix}$	
			BIC	$\begin{bmatrix} -0,0003 & 0,0007 & -0,0007 \\ -0,0017 & -0,0006 & -0,0018 \\ -0,0045 & -0,0015 & -0,0043 \end{bmatrix}$	$\begin{bmatrix} -0,0001 & 0,0000 & 0,0001 \\ 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \end{bmatrix}$	$\begin{bmatrix} -0,0046 & -0,0100 & 0,0092 \\ -0,0009 & -0,0076 & -0,0042 \\ -0,0014 & -0,0023 & -0,0131 \end{bmatrix}$
	$n = 400$	AIC	$\begin{bmatrix} -0,0011 & 0,0008 & -0,0005 \\ -0,0008 & -0,0010 & -0,0008 \\ -0,0017 & -0,0011 & -0,0019 \end{bmatrix}$	$\begin{bmatrix} -0,0007 & -0,0001 & 0,0004 \\ -0,0003 & -0,0008 & 0,0004 \\ 0,0001 & 0,0000 & -0,0005 \end{bmatrix}$	$\begin{bmatrix} -0,0009 & -0,0003 & -0,0011 \\ -0,0005 & -0,0035 & -0,0004 \\ -0,0011 & -0,0022 & -0,0052 \end{bmatrix}$	
			HQ	$\begin{bmatrix} -0,0007 & 0,0008 & -0,0005 \\ -0,0006 & -0,0008 & -0,0010 \\ -0,0019 & -0,0011 & -0,0019 \end{bmatrix}$	$\begin{bmatrix} 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \end{bmatrix}$	$\begin{bmatrix} -0,0006 & -0,0003 & -0,0019 \\ -0,0004 & -0,0033 & -0,0009 \\ -0,0011 & -0,0023 & -0,0053 \end{bmatrix}$
				BIC	$\begin{bmatrix} -0,0007 & 0,0008 & -0,0005 \\ -0,0006 & -0,0008 & -0,0010 \\ -0,0019 & -0,0011 & -0,0019 \end{bmatrix}$	$\begin{bmatrix} 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \end{bmatrix}$

**Tableau 3.11.** Écart-type moyen des coefficients obtenus pour le DGP<sub>3</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$
$n = 50$	AIC	[0,1085 0,1522 0,1518]	[0,0266 0,0320 0,0430]	[0,0814 0,1199 0,1272]
		[0,1080 0,1491 0,1502]	[0,0266 0,0313 0,0426]	[0,0825 0,1200 0,1285]
		[0,1079 0,1508 0,1487]	[0,0267 0,0319 0,0427]	[0,0823 0,1211 0,1274]
	HQ	[0,1033 0,1514 0,1510]	[0,0081 0,0108 0,0144]	[0,0586 0,0878 0,0891]
		[0,1006 0,1453 0,1463]	[0,0081 0,0105 0,0142]	[0,0603 0,0892 0,0913]
		[0,1004 0,1469 0,1449]	[0,0081 0,0107 0,0144]	[0,0602 0,0900 0,0904]
	BIC	[0,0959 0,1435 0,1425]	[0,0014 0,0021 0,0028]	[0,0224 0,0340 0,0340]
		[0,0887 0,1308 0,1312]	[0,0015 0,0021 0,0028]	[0,0238 0,0357 0,0360]
		[0,0890 0,1329 0,1308]	[0,0015 0,0022 0,0029]	[0,0238 0,0360 0,0357]
$n = 100$	AIC	[0,0696 0,1011 0,1006]	[0,0076 0,0087 0,0116]	[0,0668 0,0983 0,1009]
		[0,0696 0,1004 0,1004]	[0,0075 0,0085 0,0115]	[0,0669 0,0978 0,1009]
		[0,0696 0,1009 0,1000]	[0,0075 0,0086 0,0115]	[0,0668 0,0983 0,1005]
	HQ	[0,0681 0,1012 0,1006]	[0,0010 0,0013 0,0017]	[0,0603 0,0894 0,0897]
		[0,0676 0,0998 0,0997]	[0,0009 0,0012 0,0016]	[0,0606 0,0893 0,0899]
		[0,0676 0,1003 0,0993]	[0,0009 0,0012 0,0017]	[0,0606 0,0898 0,0896]
	BIC	[0,0702 0,1042 0,1037]	[0,0001 0,0002 0,0003]	[0,0358 0,0534 0,0534]
		[0,0668 0,0986 0,0986]	[0,0001 0,0002 0,0003]	[0,0366 0,0543 0,0545]
		[0,0667 0,0991 0,0982]	[0,0001 0,0002 0,0003]	[0,0367 0,0546 0,0544]
$n = 200$	AIC	[0,0474 0,0695 0,0690]	[0,0034 0,0037 0,0050]	[0,0469 0,0690 0,0702]
		[0,0475 0,0693 0,0690]	[0,0034 0,0037 0,0050]	[0,0469 0,0688 0,0702]
		[0,0474 0,0695 0,0688]	[0,0034 0,0037 0,0050]	[0,0469 0,0690 0,0700]
	HQ	[0,0465 0,0693 0,0688]	[0,0001 0,0001 0,0001]	[0,0465 0,0687 0,0688]
		[0,0466 0,0691 0,0688]	[0,0001 0,0001 0,0001]	[0,0466 0,0686 0,0687]
		[0,0465 0,0693 0,0686]	[0,0001 0,0001 0,0001]	[0,0466 0,0687 0,0686]
	BIC	[0,0466 0,0695 0,0690]	[0,0000 0,0000 0,0000]	[0,0449 0,0663 0,0663]
		[0,0465 0,0691 0,0687]	[0,0000 0,0000 0,0000]	[0,0450 0,0663 0,0664]
		[0,0465 0,0693 0,0686]	[0,0000 0,0000 0,0000]	[0,0450 0,0664 0,0663]
$n = 400$	AIC	[0,0330 0,0485 0,0481]	[0,0021 0,0023 0,0031]	[0,0327 0,0482 0,0489]
		[0,0330 0,0484 0,0481]	[0,0021 0,0023 0,0031]	[0,0327 0,0481 0,0488]
		[0,0330 0,0485 0,0480]	[0,0021 0,0023 0,0031]	[0,0327 0,0481 0,0488]
	HQ	[0,0325 0,0484 0,0480]	[0,0000 0,0000 0,0000]	[0,0325 0,0481 0,0480]
		[0,0325 0,0483 0,0480]	[0,0000 0,0000 0,0000]	[0,0325 0,0480 0,0480]
		[0,0325 0,0484 0,0479]	[0,0000 0,0000 0,0000]	[0,0325 0,0480 0,0480]
	BIC	[0,0325 0,0484 0,0480]	[0,0000 0,0000 0,0000]	[0,0325 0,0481 0,0480]
		[0,0325 0,0483 0,0480]	[0,0000 0,0000 0,0000]	[0,0325 0,0480 0,0480]
		[0,0325 0,0484 0,0479]	[0,0000 0,0000 0,0000]	[0,0325 0,0480 0,0480]

**Tableau 3.12.** Biais moyen des coefficients obtenus pour le DGP<sub>4</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{B}_0$
$n = 50$	AIC	-0,0553 -0,0095	-0,0144 -0,0012	-0,0156 -0,0161	-0,6396
		0,0159 -0,0225	0,0167 -0,0012	-0,0829 -0,0085	-0,4404
	HQ	-0,0525 -0,0134	-0,0083 -0,0024	-0,0120 -0,0161	-0,8644
		0,0253 -0,0036	0,0108 0,0012	-0,1349 -0,0071	-0,5926
	BIC	-0,0497 -0,0209	-0,0023 -0,0012	-0,0069 -0,0058	-1,0808
		0,0342 0,0185	0,0050 0,0011	-0,1964 -0,0030	-0,7291
$n = 100$	AIC	-0,0270 -0,0022	-0,0052 0,0005	-0,0095 -0,0094	-0,2906
		0,0113 -0,0046	0,0106 0,0014	-0,0564 -0,0050	-0,2057
	HQ	-0,0229 -0,0035	-0,0026 -0,0007	-0,0079 -0,0094	-0,6235
		0,0246 0,0208	0,0062 0,0018	-0,1216 -0,0049	-0,4333
	BIC	-0,0160 -0,0075	-0,0007 -0,0002	-0,0039 -0,0038	-0,9889
		0,0402 0,0522	0,0014 0,0002	-0,1994 -0,0020	-0,6759
$n = 200$	AIC	-0,0157 0,0000	-0,0017 -0,0003	-0,0064 -0,0036	-0,0278
		0,0040 -0,0076	0,0051 0,0006	-0,0138 -0,0031	-0,0232
	HQ	-0,0143 -0,0001	-0,0006 -0,0001	-0,0071 -0,0036	-0,1903
		0,0132 0,0104	0,0028 0,0012	-0,0621 -0,0032	-0,1406
	BIC	-0,0073 -0,0011	0,0000 -0,0001	-0,0050 -0,0024	-0,6285
		0,0325 0,0486	0,0005 0,0002	-0,1595 -0,0020	-0,4380
$n = 400$	AIC	-0,0075 -0,0001	-0,0007 0,0001	-0,0028 -0,0023	0,0045
		0,0016 -0,0057	0,0010 -0,0005	0,0020 -0,0019	0,0037
	HQ	-0,0080 -0,0001	0,0000 -0,0001	-0,0032 -0,0023	-0,0008
		0,0030 -0,0033	0,0007 0,0003	-0,0059 -0,0022	-0,0018
	BIC	-0,0068 -0,0002	0,0000 0,0000	-0,0027 -0,0020	-0,0862
		0,0121 0,0167	0,0002 0,0001	-0,0590 -0,0019	-0,0693

**Tableau 3.13.** Écart-type moyen des coefficients obtenus pour le DGP<sub>4</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{B}_0$
$n = 50$	AIC	0,1337 0,1174	0,0273 0,0240	0,0697 0,0536	0,1926
		0,1353 0,1166	0,0275 0,0237	0,0685 0,0520	0,1975
	HQ	0,1328 0,1115	0,0118 0,0100	0,0451 0,0344	0,1026
		0,1347 0,1111	0,0119 0,0099	0,0437 0,0330	0,1070
	BIC	0,1315 0,1053	0,0032 0,0027	0,0189 0,0144	0,0309
		0,1338 0,1054	0,0032 0,0026	0,0180 0,0136	0,0334
$n = 100$	AIC	0,0912 0,0809	0,0160 0,0141	0,0630 0,0475	0,2914
		0,0920 0,0809	0,0161 0,0141	0,0625 0,0469	0,2940
	HQ	0,0908 0,0757	0,0046 0,0038	0,0386 0,0289	0,1637
		0,0919 0,0759	0,0046 0,0038	0,0378 0,0281	0,1671
	BIC	0,0912 0,0713	0,0006 0,0005	0,0137 0,0103	0,0508
		0,0926 0,0719	0,0006 0,0005	0,0132 0,0098	0,0531
$n = 200$	AIC	0,0635 0,0584	0,0093 0,0085	0,0611 0,0456	0,3119
		0,0637 0,0584	0,0094 0,0085	0,0611 0,0455	0,3125
	HQ	0,0629 0,0549	0,0019 0,0016	0,0454 0,0337	0,2517
		0,0633 0,0550	0,0019 0,0016	0,0450 0,0334	0,2537
	BIC	0,0631 0,0506	0,0002 0,0001	0,0198 0,0147	0,1259
		0,0639 0,0509	0,0002 0,0001	0,0193 0,0143	0,1289
$n = 400$	AIC	0,0447 0,0418	0,0058 0,0054	0,0491 0,0365	0,2326
		0,0447 0,0418	0,0058 0,0054	0,0491 0,0365	0,2325
	HQ	0,0442 0,0408	0,0006 0,0005	0,0464 0,0342	0,2307
		0,0443 0,0408	0,0006 0,0005	0,0463 0,0341	0,2307
	BIC	0,0441 0,0386	0,0001 0,0000	0,0341 0,0252	0,2078
		0,0443 0,0387	0,0001 0,0001	0,0339 0,0250	0,2090

**Tableau 3.14.** Biais moyen des coefficients obtenus pour le DGP<sub>5</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{B}_0$
$n = 50$	AIC	-0,0390 0,0164	-0,0143 0,0114	0,0253 0,0020	-0,6714
		0,0084 -0,0680	0,0007 -0,0105	-0,0638 -0,0705	-0,4451
	HQ	-0,0337 0,0293	-0,0113 0,0074	0,0495 0,0020	-0,8895
		0,0194 -0,0670	0,0039 -0,0046	-0,0829 -0,0995	-0,5897
	BIC	-0,0283 0,0491	-0,0052 0,0041	0,0965 -0,0416	-1,0965
		0,0442 -0,0676	0,0036 -0,0012	-0,1212 -0,1493	-0,7312
$n = 100$	AIC	-0,0206 0,0053	-0,0055 0,0044	0,0020 0,0070	-0,3113
		0,0011 -0,0317	0,0000 -0,0055	-0,0211 -0,0221	-0,2025
	HQ	-0,0158 0,0120	-0,0036 0,0027	0,0108 0,0070	-0,6361
		0,0074 -0,0305	0,0028 -0,0005	-0,0305 -0,0377	-0,4248
	BIC	-0,0071 0,0268	-0,0022 0,0017	0,0438 -0,0199	-0,9955
		0,0272 -0,0263	0,0022 0,0003	-0,0599 -0,0762	-0,6661
$n = 200$	AIC	-0,0131 0,0023	-0,0011 0,0002	-0,0034 0,0072	-0,0314
		-0,0005 -0,0168	-0,0014 -0,0019	-0,0073 -0,0086	-0,0177
	HQ	-0,0118 0,0037	-0,0001 0,0000	-0,0038 0,0072	-0,1903
		0,0004 -0,0165	-0,0001 -0,0002	-0,0075 -0,0102	-0,1257
	BIC	-0,0043 0,0079	-0,0002 0,0001	-0,0009 0,0020	-0,6144
		0,0063 -0,0137	0,0002 0,0000	-0,0099 -0,0156	-0,4068
$n = 400$	AIC	-0,0062 0,0015	-0,0004 0,0000	-0,0011 0,0032	0,0027
		0,0004 -0,0080	-0,0006 -0,0012	-0,0044 -0,0045	0,0050
	HQ	-0,0066 0,0016	-0,0001 0,0000	-0,0013 0,0032	-0,0023
		0,0001 -0,0084	-0,0001 -0,0001	-0,0043 -0,0051	0,0013
	BIC	-0,0055 0,0022	0,0000 0,0000	-0,0012 0,0028	-0,0841
		0,0008 -0,0080	0,0000 0,0000	-0,0042 -0,0054	-0,0536

**Tableau 3.15.** Écart-type moyen des coefficients obtenus pour le DGP<sub>5</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{B}_0$
$n = 50$	AIC	0,1272 0,1268	0,0283 0,0270	0,0930 0,0956	0,1841
		0,1274 0,1245	0,0282 0,0265	0,0927 0,0941	0,1861
	HQ	0,1256 0,1250	0,0130 0,0124	0,0745 0,0779	0,0973
		0,1254 0,1224	0,0129 0,0121	0,0740 0,0763	0,0993
	BIC	0,1225 0,1219	0,0039 0,0038	0,0460 0,0486	0,0285
		0,1218 0,1191	0,0038 0,0037	0,0457 0,0476	0,0296
$n = 100$	AIC	0,0866 0,0837	0,0144 0,0136	0,0801 0,0815	0,2865
		0,0868 0,0832	0,0144 0,0135	0,0801 0,0810	0,2872
	HQ	0,0862 0,0832	0,0036 0,0034	0,0732 0,0754	0,1620
		0,0861 0,0824	0,0036 0,0034	0,0729 0,0746	0,1632
	BIC	0,0863 0,0834	0,0010 0,0009	0,0575 0,0595	0,0506
		0,0859 0,0824	0,0010 0,0009	0,0568 0,0585	0,0514
$n = 200$	AIC	0,0598 0,0572	0,0084 0,0079	0,0577 0,0587	0,3120
		0,0599 0,0571	0,0085 0,0079	0,0578 0,0586	0,3120
	HQ	0,0593 0,0566	0,0010 0,0010	0,0568 0,0583	0,2533
		0,0593 0,0563	0,0010 0,0010	0,0568 0,0581	0,2537
	BIC	0,0599 0,0572	0,0001 0,0001	0,0562 0,0578	0,1313
		0,0596 0,0567	0,0001 0,0001	0,0559 0,0573	0,1320
$n = 400$	AIC	0,0420 0,0400	0,0054 0,0051	0,0405 0,0412	0,2326
		0,0420 0,0399	0,0054 0,0051	0,0405 0,0411	0,2325
	HQ	0,0415 0,0393	0,0004 0,0004	0,0397 0,0407	0,2309
		0,0415 0,0393	0,0004 0,0004	0,0397 0,0407	0,2307
	BIC	0,0415 0,0394	0,0000 0,0000	0,0397 0,0408	0,2091
		0,0415 0,0393	0,0000 0,0000	0,0397 0,0407	0,2091

**Tableau 3.16.** Biais moyen des coefficients obtenus pour le DGP<sub>6</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{B}_0$	
$n = 50$	AIC	$\begin{bmatrix} -0,0100 & 0,0086 & 0,0083 \\ -0,0037 & -0,0047 & 0,0008 \\ -0,0048 & -0,0004 & -0,0224 \end{bmatrix}$	$\begin{bmatrix} -0,0243 & -0,0075 & 0,0133 \\ 0,0020 & -0,0133 & 0,0063 \\ 0,0056 & 0,0035 & -0,0208 \end{bmatrix}$	$\begin{bmatrix} -0,0312 & -0,0566 & 0,0648 \\ -0,0045 & -0,0335 & -0,0196 \\ -0,0039 & -0,0003 & -0,0538 \end{bmatrix}$	$\begin{bmatrix} -0,8911 \\ -0,5857 \\ -0,2957 \end{bmatrix}$	
		HQ	$\begin{bmatrix} -0,0018 & 0,0091 & 0,0025 \\ -0,0013 & 0,0038 & 0,0005 \\ -0,0089 & -0,0024 & -0,0225 \end{bmatrix}$	$\begin{bmatrix} -0,0144 & -0,0075 & 0,0096 \\ 0,0025 & -0,0046 & 0,0030 \\ 0,0029 & 0,0021 & -0,0099 \end{bmatrix}$	$\begin{bmatrix} -0,0491 & -0,1261 & 0,1392 \\ -0,0006 & -0,0397 & -0,0455 \\ -0,0023 & 0,0184 & -0,0613 \end{bmatrix}$	$\begin{bmatrix} -1,0627 \\ -0,7036 \\ -0,3519 \end{bmatrix}$
			BIC	$\begin{bmatrix} 0,0296 & -0,0009 & -0,0740 \\ -0,0028 & 0,0210 & -0,0165 \\ -0,0101 & -0,0200 & -0,0408 \end{bmatrix}$	$\begin{bmatrix} -0,0037 & -0,0026 & 0,0031 \\ 0,0009 & -0,0010 & 0,0007 \\ 0,0007 & 0,0008 & -0,0027 \end{bmatrix}$	$\begin{bmatrix} -0,0888 & -0,2564 & 0,2924 \\ 0,0045 & -0,0560 & -0,0891 \\ 0,0000 & 0,0616 & -0,0780 \end{bmatrix}$
	AIC			$\begin{bmatrix} -0,0007 & 0,0061 & 0,0032 \\ -0,0007 & -0,0020 & 0,0005 \\ -0,0045 & 0,0001 & -0,0106 \end{bmatrix}$	$\begin{bmatrix} -0,0044 & -0,0001 & 0,0027 \\ 0,0000 & -0,0039 & 0,0008 \\ 0,0011 & -0,0001 & -0,0051 \end{bmatrix}$	$\begin{bmatrix} -0,0061 & -0,0036 & 0,0046 \\ -0,0030 & -0,0126 & -0,0026 \\ -0,0027 & -0,0062 & -0,0239 \end{bmatrix}$
		HQ		$\begin{bmatrix} 0,0048 & 0,0101 & 0,0048 \\ 0,0017 & 0,0041 & 0,0024 \\ -0,0053 & 0,0012 & -0,0086 \end{bmatrix}$	$\begin{bmatrix} -0,0020 & -0,0010 & 0,0012 \\ 0,0006 & -0,0004 & 0,0002 \\ 0,0004 & 0,0003 & -0,0010 \end{bmatrix}$	$\begin{bmatrix} -0,0128 & -0,0281 & 0,0299 \\ -0,0021 & -0,0152 & -0,0130 \\ -0,0025 & -0,0007 & -0,0296 \end{bmatrix}$
			BIC	$\begin{bmatrix} 0,0082 & 0,0149 & 0,0014 \\ 0,0029 & 0,0124 & 0,0088 \\ -0,0053 & 0,0009 & -0,0077 \end{bmatrix}$	$\begin{bmatrix} -0,0004 & -0,0003 & 0,0003 \\ 0,0002 & -0,0001 & 0,0001 \\ 0,0001 & 0,0002 & -0,0002 \end{bmatrix}$	$\begin{bmatrix} -0,0524 & -0,1507 & 0,1718 \\ 0,0020 & -0,0334 & -0,0532 \\ -0,0008 & 0,0341 & -0,0525 \end{bmatrix}$
	AIC			$\begin{bmatrix} -0,0015 & 0,0012 & 0,0016 \\ -0,0019 & -0,0022 & -0,0009 \\ -0,0024 & -0,0014 & -0,0063 \end{bmatrix}$	$\begin{bmatrix} -0,0013 & -0,0001 & 0,0002 \\ 0,0001 & -0,0016 & 0,0003 \\ 0,0003 & 0,0000 & -0,0017 \end{bmatrix}$	$\begin{bmatrix} -0,0032 & 0,0007 & -0,0013 \\ -0,0005 & -0,0049 & -0,0017 \\ -0,0020 & -0,0019 & -0,0116 \end{bmatrix}$
		HQ		$\begin{bmatrix} 0,0012 & 0,0041 & 0,0031 \\ -0,0005 & 0,0008 & 0,0000 \\ -0,0024 & -0,0005 & -0,0052 \end{bmatrix}$	$\begin{bmatrix} 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0001 \\ 0,0000 & 0,0001 & -0,0001 \end{bmatrix}$	$\begin{bmatrix} -0,0024 & 0,0003 & -0,0032 \\ -0,0006 & -0,0047 & -0,0026 \\ -0,0021 & -0,0023 & -0,0120 \end{bmatrix}$
			BIC	$\begin{bmatrix} 0,0058 & 0,0102 & 0,0054 \\ 0,0022 & 0,0060 & 0,0027 \\ -0,0011 & 0,0015 & -0,0032 \end{bmatrix}$	$\begin{bmatrix} 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \end{bmatrix}$	$\begin{bmatrix} -0,0056 & -0,0105 & 0,0082 \\ 0,0001 & -0,0064 & -0,0068 \\ -0,0019 & 0,0001 & -0,0145 \end{bmatrix}$
AIC	$\begin{bmatrix} -0,0011 & 0,0011 & 0,0006 \\ -0,0006 & -0,0018 & -0,0016 \\ -0,0011 & -0,0019 & -0,0030 \end{bmatrix}$			$\begin{bmatrix} -0,0004 & -0,0001 & 0,0003 \\ -0,0001 & -0,0005 & 0,0002 \\ 0,0002 & 0,0001 & -0,0004 \end{bmatrix}$	$\begin{bmatrix} -0,0010 & 0,0001 & -0,0011 \\ -0,0004 & -0,0030 & -0,0002 \\ -0,0006 & -0,0014 & -0,0057 \end{bmatrix}$	$\begin{bmatrix} -0,0005 \\ 0,0024 \\ 0,0013 \end{bmatrix}$
	HQ	$\begin{bmatrix} -0,0008 & 0,0013 & 0,0006 \\ -0,0004 & -0,0015 & -0,0016 \\ -0,0013 & -0,0019 & -0,0029 \end{bmatrix}$		$\begin{bmatrix} 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \end{bmatrix}$	$\begin{bmatrix} -0,0008 & 0,0001 & -0,0016 \\ -0,0004 & -0,0029 & -0,0004 \\ -0,0007 & -0,0015 & -0,0057 \end{bmatrix}$	$\begin{bmatrix} -0,0257 \\ -0,0139 \\ -0,0065 \end{bmatrix}$
		BIC	$\begin{bmatrix} 0,0012 & 0,0040 & 0,0021 \\ 0,0008 & 0,0005 & -0,0006 \\ -0,0007 & -0,0009 & -0,0021 \end{bmatrix}$	$\begin{bmatrix} 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 \end{bmatrix}$	$\begin{bmatrix} -0,0006 & -0,0001 & -0,0021 \\ -0,0003 & -0,0030 & -0,0007 \\ -0,0006 & -0,0016 & -0,0058 \end{bmatrix}$	$\begin{bmatrix} -0,3328 \\ -0,2186 \\ -0,1096 \end{bmatrix}$

**Tableau 3.17.** Écart-type moyen des coefficients obtenus pour le DGP<sub>6</sub>

		$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{B}_0$
$n = 50$	AIC	[0,1076 0,1514 0,1524]	[0,0256 0,0314 0,0421]	[0,0797 0,1176 0,1250]	[0,1023]
		[0,1059 0,1467 0,1491]	[0,0253 0,0304 0,0412]	[0,0800 0,1165 0,1251]	[0,1019]
		[0,1049 0,1471 0,1464]	[0,0251 0,0305 0,0409]	[0,0792 0,1166 0,1230]	[0,1026]
	HQ	[0,1042 0,1527 0,1535]	[0,0079 0,0106 0,0142]	[0,0570 0,0852 0,0868]	[0,0421]
		[0,0999 0,1444 0,1464]	[0,0077 0,0102 0,0138]	[0,0579 0,0854 0,0878]	[0,0410]
		[0,0987 0,1444 0,1436]	[0,0077 0,0102 0,0137]	[0,0572 0,0854 0,0863]	[0,0415]
	BIC	[0,0955 0,1426 0,1426]	[0,0012 0,0019 0,0025]	[0,0215 0,0324 0,0327]	[0,0048]
		[0,0872 0,1285 0,1297]	[0,0012 0,0018 0,0025]	[0,0224 0,0334 0,0339]	[0,0046]
		[0,0864 0,1289 0,1277]	[0,0012 0,0018 0,0024]	[0,0221 0,0333 0,0332]	[0,0046]
$n = 100$	AIC	[0,0681 0,0994 0,0994]	[0,0072 0,0084 0,0112]	[0,0653 0,0963 0,0994]	[0,2028]
		[0,0678 0,0983 0,0988]	[0,0071 0,0082 0,0111]	[0,0651 0,0954 0,0989]	[0,2032]
		[0,0675 0,0984 0,0980]	[0,0071 0,0083 0,0110]	[0,0648 0,0956 0,0982]	[0,2044]
	HQ	[0,0683 0,1017 0,1017]	[0,0010 0,0014 0,0018]	[0,0600 0,0891 0,0898]	[0,0756]
		[0,0670 0,0993 0,0997]	[0,0010 0,0013 0,0018]	[0,0596 0,0880 0,0892]	[0,0755]
		[0,0665 0,0990 0,0985]	[0,0010 0,0013 0,0018]	[0,0592 0,0879 0,0882]	[0,0762]
	BIC	[0,0713 0,1062 0,1063]	[0,0001 0,0002 0,0002]	[0,0346 0,0517 0,0520]	[0,0095]
		[0,0668 0,0989 0,0994]	[0,0001 0,0002 0,0002]	[0,0348 0,0517 0,0522]	[0,0092]
		[0,0660 0,0983 0,0979]	[0,0001 0,0002 0,0002]	[0,0345 0,0516 0,0516]	[0,0094]
$n = 200$	AIC	[0,0462 0,0679 0,0678]	[0,0034 0,0038 0,0051]	[0,0457 0,0674 0,0690]	[0,2894]
		[0,0462 0,0676 0,0677]	[0,0034 0,0038 0,0051]	[0,0456 0,0671 0,0689]	[0,2892]
		[0,0461 0,0677 0,0675]	[0,0034 0,0038 0,0051]	[0,0456 0,0672 0,0687]	[0,2892]
	HQ	[0,0460 0,0687 0,0686]	[0,0000 0,0001 0,0001]	[0,0460 0,0681 0,0686]	[0,1794]
		[0,0457 0,0681 0,0682]	[0,0000 0,0001 0,0001]	[0,0458 0,0675 0,0681]	[0,1796]
		[0,0456 0,0681 0,0678]	[0,0000 0,0001 0,0001]	[0,0456 0,0675 0,0678]	[0,1799]
	BIC	[0,0471 0,0704 0,0703]	[0,0000 0,0000 0,0000]	[0,0452 0,0669 0,0673]	[0,0452]
		[0,0463 0,0690 0,0691]	[0,0000 0,0000 0,0000]	[0,0446 0,0659 0,0664]	[0,0455]
		[0,0459 0,0686 0,0684]	[0,0000 0,0000 0,0000]	[0,0442 0,0655 0,0657]	[0,0457]
$n = 400$	AIC	[0,0323 0,0476 0,0475]	[0,0017 0,0019 0,0026]	[0,0320 0,0472 0,0481]	[0,2312]
		[0,0323 0,0475 0,0475]	[0,0017 0,0019 0,0026]	[0,0320 0,0471 0,0481]	[0,2310]
		[0,0323 0,0476 0,0475]	[0,0017 0,0019 0,0026]	[0,0320 0,0472 0,0480]	[0,2311]
	HQ	[0,0319 0,0476 0,0475]	[0,0000 0,0000 0,0000]	[0,0319 0,0472 0,0475]	[0,2239]
		[0,0318 0,0475 0,0474]	[0,0000 0,0000 0,0000]	[0,0319 0,0471 0,0474]	[0,2237]
		[0,0319 0,0476 0,0474]	[0,0000 0,0000 0,0000]	[0,0319 0,0472 0,0474]	[0,2239]
	BIC	[0,0322 0,0481 0,0479]	[0,0000 0,0000 0,0000]	[0,0322 0,0477 0,0479]	[0,1541]
		[0,0320 0,0478 0,0477]	[0,0000 0,0000 0,0000]	[0,0321 0,0474 0,0477]	[0,1541]
		[0,0320 0,0477 0,0475]	[0,0000 0,0000 0,0000]	[0,0320 0,0473 0,0475]	[0,1544]

**Tableau 3.18.** Erreur quadratique moyenne de la prévision d'horizon un

		$n = 50$			$n = 100$			$n = 200$			$n = 400$		
		AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC	AIC	HQ	BIC
DGP <sub>1</sub>	Y(1)	1,1368	1,1170	1,1114	1,0535	1,0494	1,0395	1,0192	1,0171	1,0092	1,0167	1,0172	1,0133
	Y(2)	1,1603	1,1788	1,1665	1,0990	1,1175	1,1123	1,0283	1,0519	1,0548	1,0384	1,0486	1,0602
DGP <sub>2</sub>	Y(1)	1,1586	1,1617	1,1514	1,0624	1,0699	1,0749	1,0148	1,0195	1,0164	1,0184	1,0191	1,0163
	Y(2)	1,1760	1,1916	1,1999	1,0897	1,1029	1,1188	1,0217	1,0290	1,0209	1,0371	1,0449	1,0370
DGP <sub>3</sub>	Y(1)	1,2475	1,2932	1,4823	1,1057	1,1241	1,2366	1,0368	1,0348	1,0446	1,0107	1,0094	1,0094
	Y(2)	1,2131	1,1781	1,1564	1,1145	1,1086	1,1041	1,0515	1,0488	1,0492	1,0348	1,0340	1,0340
	Y(3)	1,1996	1,1685	1,1395	1,0774	1,0712	1,0677	1,0230	1,0216	1,0220	1,0272	1,0269	1,0269
DGP <sub>4</sub>	Y(1)	1,2001	1,1850	1,1684	1,0876	1,0957	1,1023	1,0412	1,0448	1,0513	1,0140	1,0148	1,0200
	Y(2)	1,2065	1,2046	1,1936	1,0808	1,0938	1,1056	1,0575	1,0733	1,0931	1,0340	1,0357	1,0558
DGP <sub>5</sub>	Y(1)	1,2344	1,2398	1,2479	1,0887	1,1061	1,1325	1,0351	1,0456	1,0712	1,0127	1,0124	1,0172
	Y(2)	1,2250	1,2298	1,2380	1,0696	1,0804	1,0978	1,0441	1,0442	1,0554	1,0339	1,0338	1,0356
DGP <sub>6</sub>	Y(1)	1,3643	1,4253	1,6302	1,1346	1,1785	1,3372	1,0565	1,0763	1,1198	1,0231	1,0244	1,0441
	Y(2)	1,2249	1,2009	1,1679	1,1380	1,1376	1,1285	1,0323	1,0362	1,0439	1,0037	1,0034	1,0102
	Y(3)	1,2494	1,2147	1,1914	1,1345	1,1266	1,1217	1,0301	1,0317	1,0331	1,0224	1,0218	1,0252

# Chapitre 4

---

## Analyse de données réelles

Dans ce chapitre, nous appliquons la théorie des modèles vectoriels autorégressifs à variables exogènes présentée aux chapitres 1 et 2 sur des données macroéconomiques réelles. Les données proviennent de la base de données du système canadien d'information socio-économique CANSIM. Elles représentent les productions mensuelles de fromage mozzarella, cheddar et ricotta au Canada, ainsi que les prix mensuels du lait de bovin non transformé pour trois provinces, à savoir le Québec, l'Ontario et la Colombie-Britannique. La période couverte est récente, allant de janvier 2003 à juillet 2021.

Nous illustrons nos méthodes en spécifiant dans un premier temps les ordres maximaux du modèle  $p$  et  $s$ . Par la suite, nous procédons à l'estimation dans tous les modèles à sous-ensemble d'indices dont les ordres maximaux sont  $(p,s)$ . Le modèle choisi est celui minimisant le critère AIC. Nous comparons les résultats obtenus pour le modèle complet à ceux obtenus pour le modèle restreint déterminé par le critère de sélection de modèle. Nous effectuons également l'évaluation du modèle choisi. Finalement, nous étudions les prévisions d'horizon un pour les séries endogènes, en appliquant le modèle complet VARX( $p,s$ ) ainsi que le modèle SVARX( $p,s$ ) aux données. Pour ce faire, nous comparons les prévisions ponctuelles à l'aide des mesures du biais et de l'erreur quadratique moyenne.

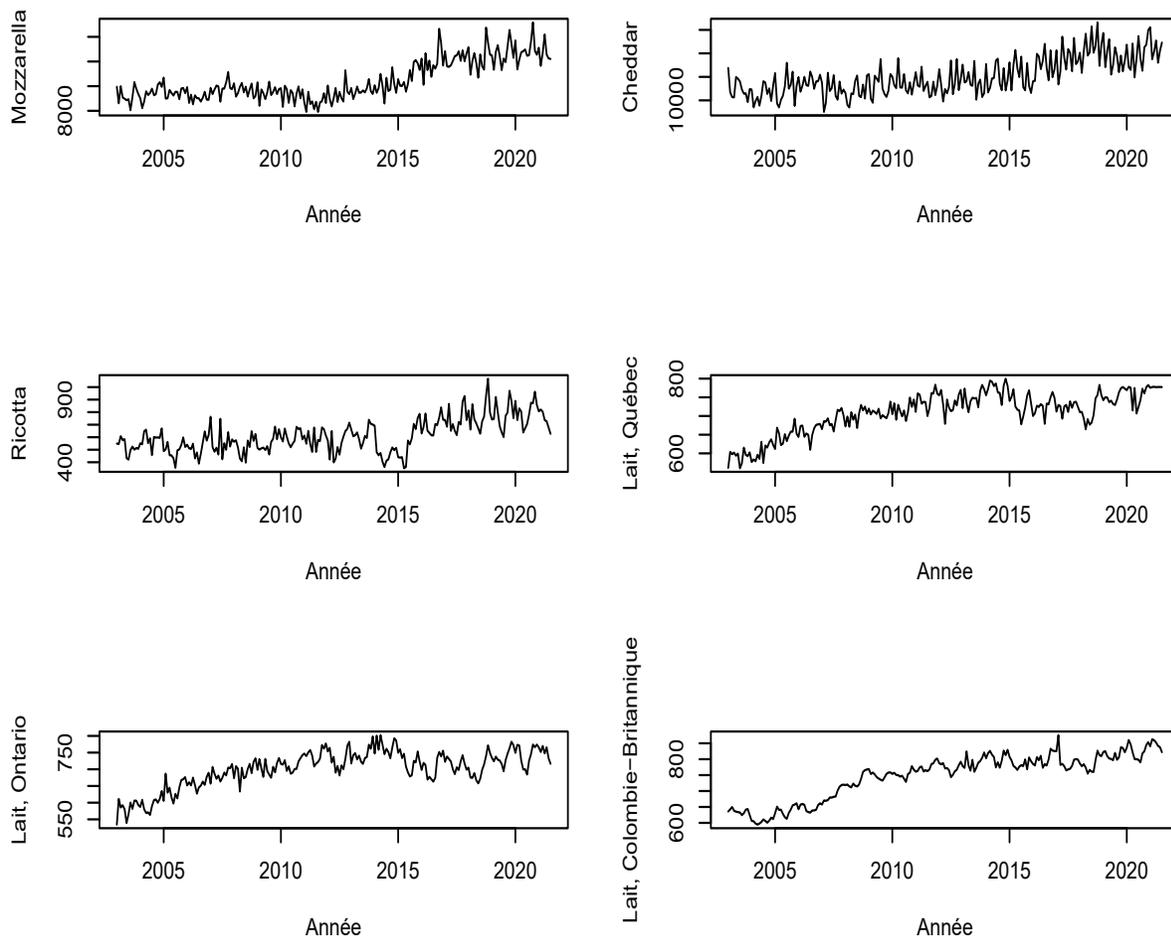
La prochaine section fait une description complète des données utilisées aux fins de notre analyse de données.

### 4.1. Description du jeu de données

Comme mentionné précédemment, les données utilisées proviennent de CANSIM, une base de données statistique informatisée créée et mise à jour par Statistique Canada. Plus précisément, les variables endogènes sont constituées de données sur la production mensuelle de fromage mozzarella, cheddar et ricotta au Canada pour la période de janvier 2003 à juillet 2021 (les numéros des séries CANSIM sont V382880, V382649 et V29850281, respectivement). Pour les fins de notre illustration, aucune donnée manquante n'était présente pour

la période choisie. Certaines séries étaient cependant plus longues sur CANSIM. L'unité de mesure est la tonne avec une taille d'échantillon modérée de  $n = 223$ . Les variables exogènes incluses dans l'étude sont les prix mensuels du lait de bovin non transformé (dollars par kilolitre) dans trois provinces, soient le Québec et l'Ontario et la Colombie-Britannique, pour la même période (séries avec numéros CANSIM V31212138, V31212164 et V31212300). Nous aurions pu inclure davantage de provinces, ou moins, mais pour les fins de l'étude nous avons tenté de choisir des provinces et des séries représentatives. Ces séries chronologiques produisent des données provinciales sur les prix moyens reçus par les agriculteurs pour leurs produits agricoles fabriqués et vendus.

**Figure 4.1.** Production mensuelle de fromage mozzarella, cheddar et ricotta au Canada et prix du lait de bovin non transformé au Québec, en Ontario et en Colombie-Britannique (janvier 2003 à juillet 2021).



Les séries chronologiques sont présentées à la figure 4.1 avec sur l'axe vertical la production et l'axe horizontal le temps (mois). Afin de réduire la variance et plus spécifiquement l'étendue des variables, le logarithme a été appliqué aux variables. Cette méthodologie peut être justifiée avec l'étude de transformations, comme les transformations puissances ou encore les transformations de type Box-Cox. Cependant, avec des données macroéconomiques, et surtout compte tenu de la nature des données en cause, la transformation  $\log(\cdot)$  est souvent recommandée dans les ouvrages d'économétrie. Suivant la méthodologie de Box et Jenkins, puisqu'une tendance linéaire générale est présente pour chaque série temporelle, une première différence semble également nécessaire afin d'atteindre la stationnarité. Par conséquent, le filtre  $1 - B$  est appliqué à chaque série chronologique d'origine.

Soient  $MOZ_t$ ,  $CHED_t$  et  $RICO_t$ ,  $t = 1, \dots, n$  les variables de sorties associées aux fromages mozzarella, cheddar et ricotta. Nous avons considéré les variables suivantes :

$$\mathbf{Y}_t = \begin{pmatrix} Y_t(1) \\ Y_t(2) \\ Y_t(3) \end{pmatrix} = \begin{pmatrix} (1 - B) \log(MOZ_t) \\ (1 - B) \log(CHED_t) \\ (1 - B) \log(RICO_t) \end{pmatrix},$$

$t = 2, \dots, n$ . De même, soit  $LAIT.QC_t$ ,  $LAIT.ON_t$  et  $LAIT.CB_t$  les prix mensuels du lait au Québec, en Ontario et en Colombie-Britannique, respectivement. Nous définissons les variables exogènes :

$$\mathbf{X}_t = \begin{pmatrix} X_t(1) \\ X_t(2) \\ X_t(3) \end{pmatrix} = \begin{pmatrix} (1 - B) \log(LAIT.QC_t) \\ (1 - B) \log(LAIT.ON_t) \\ (1 - B) \log(LAIT.CB_t) \end{pmatrix},$$

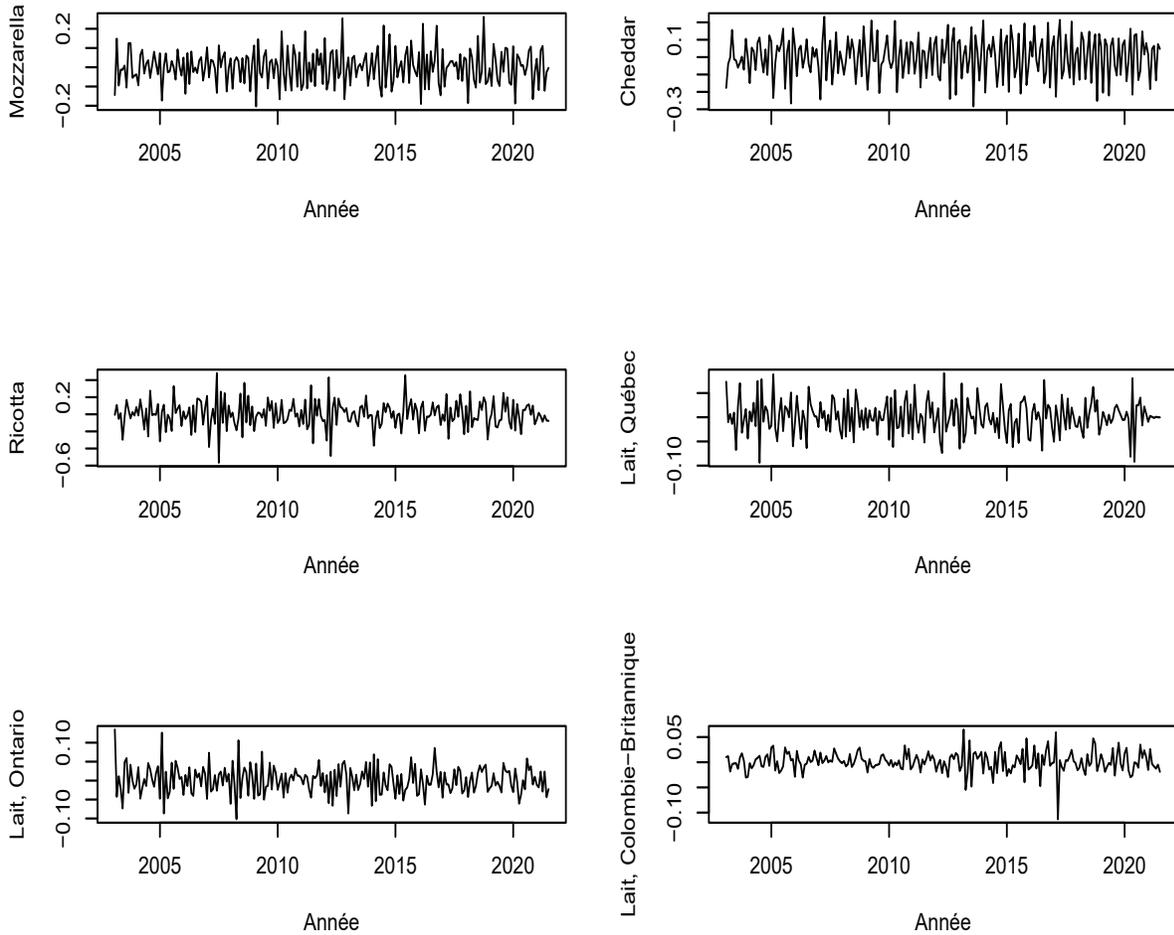
$t = 2, \dots, n$ . Les séries transformées sont présentées à la figure 4.2. On constate que les hypothèses de base, à savoir la stationnarité en moyenne et en variance, semblent raisonnablement satisfaites.

Une fois les séries transformées obtenues, nous pouvons passer à l'ajustement du modèle. C'est ce qui est effectué à la prochaine section.

## 4.2. Ajustement du modèle

L'ajustement d'un modèle  $SVARX(p, s)$  s'effectue en respectant les étapes décrites dans le chapitre 2. Dans une première étape, nous commençons par la spécification des ordres maximaux, et par la suite nous procédons à l'estimation des paramètres et de l'étape de sélection des indices qui expliquent le mieux les données. Comme dernière étape, nous calculons des statistiques de test pour valider le modèle et nous examinons la dépendance dans les résidus.

**Figure 4.2.** Première différence des logarithmes, pour la production mensuelle de fromage mozzarella, cheddar et ricotta au Canada, et pour les prix du lait de bovin non transformé au Québec, en Ontario et en Colombie-Britannique (janvier 2003 à juillet 2021).



#### 4.2.1. Sélection des ordres maximaux $p$ et $s$

Le modèle VARX( $p,s$ ) suivant a été postulé :

$$\mathbf{Y}_t = \phi_0 + \sum_{i=1}^p \Phi_i \mathbf{Y}_{t-i} + \sum_{i=0}^s \mathbf{B}_i \mathbf{X}_{t-i} + \epsilon_t,$$

où

$$\phi_0 = \left( \mathbf{I}_k - \sum_{i=1}^p \Phi_i \right) \boldsymbol{\mu},$$

$\mu$  étant la moyenne du processus  $\{\mathbf{Y}_t\}$ . Afin de choisir les ordres  $(p,s)$ , c'est-à-dire les ordres autorégressif et exogène maximums, nous calculons le critère AIC :

$$\text{AIC} = \log \det \hat{\Sigma}_{\epsilon} + \frac{2}{n - \max(p,s) - 1} (k^2 p + k k_x s),$$

où  $\hat{\Sigma}_{\epsilon} = \sum \hat{\epsilon}_t \hat{\epsilon}_t^{\top} / (n - \max(p,s) - 1)$ .

Concernant le choix du critère de sélection de modèle, on note que dans les simulations du chapitre précédent nous avons inclus également les critères BIC et HQ. Cependant, le critère AIC semble largement accepté dans la littérature sur les séries chronologiques, voir, par exemple, Brockwell et Davis (1991). Tel que discuté par Tsay (2014), les propriétés de convergence des critères BIC ou HQ s'appuient sur des hypothèses précises, telles que l'existence d'un vrai modèle. Or, dans les applications de données réelles, il n'y a pas de vrais modèles. Voir la discussion de Brockwell et Davis (2016, pp. 152-153) ou Tsay (2014, p. 63). Un des avantages du critère AIC est l'efficacité asymptotique dans les processus autorégressifs, voir Shibata (1980), Hurvich et Tsai (1989) ou Brockwell et Davis (1991, Section 10.3). Pour ces raisons, nous avons préconisé le critère AIC et avons trouvé  $p = 12$  et  $s = 3$ . Conséquemment, nous avons postulé  $(p,s) = (12,3)$  comme étant les ordres maximums. On note que compte tenu de la nature des données de CANSIM, ayant à notre disposition des données mensuelles, le choix  $p = 12$  semble raisonnable pour modéliser des effets saisonniers potentiels. Il est attendu que les prix du lait expliquent la productivité du fromage et  $s = 3$  permet des délais contemporains ainsi que des délais allant jusqu'à trois mois. Il y a donc un total de douze matrices  $3 \times 3$  pour l'ordre VAR et quatre matrices  $3 \times 3$  pour les variables exogènes (trois retards et un effet contemporain). Incluant le terme constant, cela représente 147 paramètres libres.

Les ordres maximums  $(p,s)$  étant déterminés, nous devons choisir les matrices qui seront nulles dans notre modèle. La prochaine section porte sur l'ajustement du modèle SVARX le plus parcimonieux pour les données à notre disposition.

#### 4.2.2. Ajustement d'un SVARX( $p,s$ )

Afin d'avoir un modèle plus parcimonieux, nous avons ajusté des modèles de type SVARX. Nous avons considéré l'ajustement des  $2^{11} \times 2^3 = 16384$  modèles et avons sélectionné le modèle final selon le critère AIC :

$$\text{AIC}(\mathcal{I}, \mathcal{J}) = \log \det \hat{\Sigma}_{\epsilon} + \frac{2}{n - p - 1} (k^2 \text{card}(\mathcal{I}) + k k_x \text{card}(\mathcal{J})).$$

Un modèle VARX à sous-ensemble d'indices à été sélectionné, de la forme SVARX( $\{1,2,3,4,5,7,9,12\}, \{0,3\}$ ). Notons que 16384 est un nombre élevé de modèles à ajuster mais qui demeure nettement inférieur aux  $2^{11k^2 + 3kk_x} = 2^{42}$  modèles possibles. Ici, en se concentrant sur les indices, on constate que la sélection de variables suggère une certaine

**Tableau 4.1.** Résultats de l'estimation du modèle SVARX( $\{1,2,3,4,5,7,9,12\},\{0,3\}$ ) pour la première différence des logarithmes, pour la production mensuelle de fromage expliquée par les prix du lait de bovin non transformé au Québec, en Ontario et en Colombie-Britannique. Les écarts-types sont indiquées entre parenthèses.

$$\begin{aligned}
 \hat{\phi}_0 &= \begin{pmatrix} 0,002 \\ (0,004) \\ 0,003 \\ (0,004) \\ 0,001 \\ (0,008) \end{pmatrix}, & \hat{\Phi}_1 &= \begin{pmatrix} -0,681 & 0,018 & 0,083 \\ (0,062) & (0,053) & (0,027) \\ 0,002 & -0,435 & -0,019 \\ (0,074) & (0,063) & (0,033) \\ 0,168 & -0,111 & -0,401 \\ (0,135) & (0,114) & (0,059) \end{pmatrix}, & \hat{\Phi}_2 &= \begin{pmatrix} -0,500 & -0,108 & 0,064 \\ (0,074) & (0,057) & (0,032) \\ 0,130 & -0,523 & -0,083 \\ (0,088) & (0,068) & (0,038) \\ -0,021 & 0,041 & -0,014 \\ (0,161) & (0,124) & (0,070) \end{pmatrix}, \\
 \hat{\Phi}_3 &= \begin{pmatrix} -0,332 & 0,030 & 0,034 \\ (0,079) & (0,064) & (0,032) \\ 0,163 & -0,235 & -0,077 \\ (0,094) & (0,076) & (0,038) \\ 0,366 & -0,062 & -0,046 \\ (0,170) & (0,139) & (0,069) \end{pmatrix}, & \hat{\Phi}_4 &= \begin{pmatrix} -0,367 & 0,007 & 0,029 \\ (0,076) & (0,058) & (0,031) \\ -0,096 & -0,179 & 0,001 \\ (0,090) & (0,069) & (0,037) \\ 0,300 & -0,071 & -0,139 \\ (0,163) & (0,125) & (0,067) \end{pmatrix}, & \hat{\Phi}_5 &= \begin{pmatrix} -0,296 & 0,018 & 0,012 \\ (0,064) & (0,057) & (0,029) \\ -0,029 & -0,155 & -0,034 \\ (0,076) & (0,067) & (0,034) \\ 0,372 & -0,013 & 0,029 \\ (0,138) & (0,122) & (0,062) \end{pmatrix}, \\
 \hat{\Phi}_7 &= \begin{pmatrix} 0,006 & 0,026 & 0,066 \\ (0,053) & (0,049) & (0,026) \\ -0,043 & 0,023 & 0,045 \\ (0,063) & (0,058) & (0,031) \\ -0,040 & -0,012 & -0,170 \\ (0,114) & (0,105) & (0,056) \end{pmatrix}, & \hat{\Phi}_9 &= \begin{pmatrix} -0,049 & 0,037 & 0,032 \\ (0,054) & (0,047) & (0,026) \\ 0,127 & 0,039 & 0,031 \\ (0,064) & (0,056) & (0,031) \\ -0,398 & 0,051 & 0,093 \\ (0,116) & (0,102) & (0,056) \end{pmatrix}, & \hat{\Phi}_{12} &= \begin{pmatrix} 0,115 & 0,140 & 0,059 \\ (0,052) & (0,046) & (0,026) \\ -0,040 & 0,311 & -0,035 \\ (0,062) & (0,055) & (0,030) \\ 0,450 & 0,103 & 0,268 \\ (0,112) & (0,099) & (0,055) \end{pmatrix}, \\
 \hat{B}_0 &= \begin{pmatrix} 0,343 & 0,075 & 0,657 \\ (0,124) & (0,113) & (0,202) \\ 0,078 & -0,010 & -0,205 \\ (0,147) & (0,134) & (0,240) \\ 0,632 & -0,003 & 0,529 \\ (0,267) & (0,244) & (0,436) \end{pmatrix}, & \hat{B}_3 &= \begin{pmatrix} 0,124 & 0,052 & -0,288 \\ (0,124) & (0,116) & (0,200) \\ -0,105 & 0,068 & 0,389 \\ (0,147) & (0,137) & (0,237) \\ -0,262 & -0,183 & -1,538 \\ (0,268) & (0,250) & (0,431) \end{pmatrix}.
 \end{aligned}$$

dépendance locale, étant donné que des retards d'ordre faible ont été sélectionnés, et  $p = 12$  révèle un effet saisonnier. Les variables exogènes expliquent la sortie, et le modèle final garde la dépendance instantanée ainsi qu'un délai de trois mois. Les résultats sont présentés dans la table 4.1.

La prochaine section montre l'estimateur des moindres carrés obtenu.

### 4.2.3. Estimation des paramètres

Le modèle VARX à sous-ensemble d'indices proposé repose sur huit matrices de dimension  $3 \times 3$  pour la partie VAR(12) ainsi que deux matrices  $3 \times 3$  pour les variables exogènes. Incluant le terme constant, cela donne 93 paramètres, ce qui représente une réduction de 36,7% par

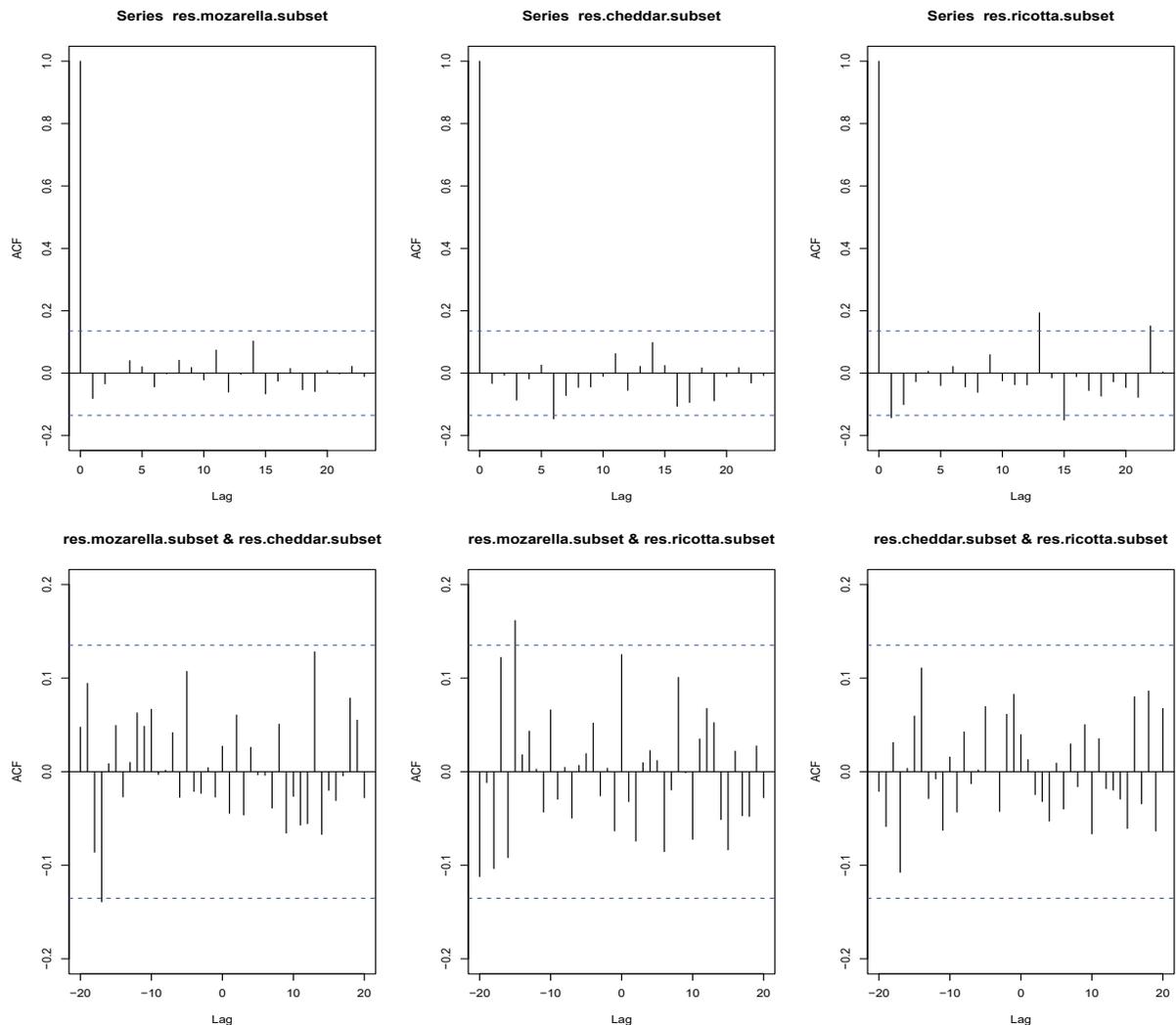
rapport au modèle complet. L'AIC du modèle complet était de -14,93 alors que celui du modèle SVARX était de -15,11.

Afin de déterminer si le modèle SVARX ajusté est approprié pour les données, nous effectuons, à la section suivante, l'évaluation du modèle.

#### 4.2.4. Évaluation du modèle

Dans la figure 4.3, nous rapportons les autocorrélations et les corrélations croisées résiduelles à des délais individuels. Globalement, particulièrement pour les retards d'ordre faible, l'ajustement semble raisonnable. Notons qu'un grand nombre de tests individuels sont effectués sur ce genre de graphique. Bien qu'en regardant ces tableaux nous observons certaines autocorrélations et corrélations croisées résiduelles qui semblent proches d'être significatives, elles n'indiquent pas un ajustement fortement insatisfaisant.

**Figure 4.3.** Autocorrélations et corrélations croisées des résidus, modèle restreint



Enfin, nous avons également calculé la statistique de test  $T_n$  présentée à la sous-section 2.2.3 en utilisant les noyaux uniforme tronqué, de Bartlett et de Daniell. Nous avons utilisé les mêmes taux (les valeurs de  $P_n$ ) que ceux étudiés et suggérés dans Duchesne et Roy (2004), c'est-à-dire que nous avons inclus la dépendance d'ordre faible  $P_n = 2, 3$ , et nous avons également considéré  $[\log(n)]$ ,  $[3,5n^{0.2}]$  et  $[3n^{0.3}]$ , où  $[\cdot]$  désigne la partie entière. Avec  $n' = n - 1 = 222$ , cela donne  $P_n = 2, 3, 5, 10, 15$ . Les valeurs de la statistique de test sont données dans le tableau 4.2. Les statistiques de test  $T_n$  convergent en loi vers une distribution normale centrée réduite  $\mathcal{N}(0,1)$  sous l'hypothèse nulle d'adéquation. Étant une forme quadratique standardisée, ces statistiques rejettent pour de grandes valeurs. Étant donné que le 90<sup>e</sup> quantile d'une loi normale  $\mathcal{N}(0,1)$  est 1,282, les valeurs obtenues n'indiquent pas un ajustement inapproprié au niveau nominal de 10%.

**Tableau 4.2.** Valeurs de la statistique de test portemanteau basée sur le noyau  $T_n$  lorsque  $\mathcal{K}$  est le noyau uniforme tronqué, le noyau Bartlett et le noyau Daniell pour  $P_n \in \{2,3,5,10,15\}$ .

Kernel $\mathcal{K}$	$P_n = 2$	$P_n = 3$	$P_n = 5$	$P_n = 10$	$P_n = 15$
Truncated Uniform	-0,75	-1,46	-2,40	-3,34	-2,41
Bartlett	-0,08	-0,31	-0,93	-2,06	-2,72
Daniell	-0,32	-0,51	-1,12	-2,18	-2,93

Enfin, nous présentons les résultats concernant les prévisions d'horizon un à la prochaine section.

### 4.3. Prévisions de valeurs futures

Nous débutons par obtenir les vraies valeurs futures des séries endogènes et exogènes. Les valeurs futures des séries exogènes sont essentielles afin d'effectuer le calcul des prévisions alors que les valeurs futures des séries endogènes sont nécessaires au calcul des biais ainsi que des erreurs quadratiques moyennes des prévisions effectuées. Les valeurs futures originales utilisées sont celles des prix mensuels du lait de bovin non transformé au Québec, en Ontario et en Colombie-Britannique ainsi que la production mensuelle de fromage mozzarella, cheddar et ricotta au Canada pour le mois d'août 2021. Les valeurs futures originales ont ensuite été remaniées en prenant la première différence du logarithme afin qu'elles soient cohérentes avec les données précédemment utilisées pour l'analyse. Certainement, l'intérêt premier est de prévoir les variables dans l'échelle originale. En transformant les variables, nous trouvons un modèle exprimé pour des variables correspondant à  $(1 - B) * \log$  des variables originales. En pratique, il faudra retransformer les variables afin de revenir à l'échelle originale.

Nous comparons les prévisions obtenues avec le modèle VARX(12,3) complet à celles obtenues avec le modèle à sous-ensemble d'indices SVARX( $\{1,2,3,4,5,7,9,12\}, \{0,3\}$ ). Nous observons des mesures de biais et d'erreur quadratique moyenne.

Les résultats sont présentés dans le tableau 4.3.

**Tableau 4.3.** Biais moyen et mesure d'EQM des prévisions ponctuelles pour les prévisions d'horizon un de la production mensuelle de fromage mozzarella, cheddar et ricotta au Canada transformée

Modèle ajusté	Biais	EQM	
VARX(12,3)	$\begin{bmatrix} -0,0070 \\ -0,2614 \\ 2,8949 \end{bmatrix}$	$\begin{bmatrix} 0,0024 \\ 0,0716 \\ 8,3913 \end{bmatrix}$	
	SVARX( $\{1,2,3,4,5,7,9,12\},\{0,3\}$ )	$\begin{bmatrix} -0,0310 \\ -0,2494 \\ 2,8871 \end{bmatrix}$	$\begin{bmatrix} 0,0035 \\ 0,0659 \\ 8,3477 \end{bmatrix}$

On remarque au tableau 4.3 que le modèle SVARX( $\{1,2,3,4,5,7,9,12\},\{0,3\}$ ) est celui donnant le biais le plus grand pour la première variable. En effet, pour la production mensuelle de fromage mozzarella, le biais obtenu avec le modèle restreint est quatre fois plus grand que celui du modèle complet. En revanche, le modèle VARX(12,3) est celui ayant le plus grand biais pour les deuxième et troisième variables. On observe cependant que les biais obtenus pour les deux modèles sont assez proches. Cela étant dit, en moyenne, le modèle à sous-ensemble d'indices donne un biais inférieur à celui du modèle complet. Il en est de même en ce qui concerne la mesure de l'écart quadratique moyen, le modèle à sous-ensemble d'indices donne de meilleurs résultats pour les variables deuxième et troisième variables. Effectivement, bien que les EQM obtenus à l'aide des deux modèles soient près, la moyenne des EQM du modèle SVARX( $\{1,2,3,4,5,7,9,12\},\{0,3\}$ ) est inférieure à celle du modèle complet. On peut donc dire que, pour les données utilisées, le modèle restreint, tout en ayant moins de paramètres, a mené à des résultats similaires à ceux du modèle complet.



## Conclusion

---

Dans ce mémoire, nous avons étudié les modèles vectoriels autorégressifs avec variables exogènes à sélection d'indices. Nous avons étudié des aspects théoriques entourant l'estimation par moindres carrés dans ces modèles. La vérification de la qualité de l'ajustement avec des statistiques de test reposant sur la méthode du noyau et le calcul des prévisions a été considéré. Des simulations de Monte Carlo ont été mises en oeuvre et une analyse de données réelles a été effectuée.

Premièrement, nous avons présenté la théorie portant sur les modèles vectoriels autorégressifs incluant les conditions de stationnarité et d'inversibilité de ceux-ci. Nous avons également exposé les propriétés asymptotiques des estimateurs ainsi qu'une statistique de test servant à l'évaluation des modèles  $\text{SVARX}(p,s)$ . Nous avons ensuite élaboré des prévisions ponctuelles dans les modèles  $\text{SVARX}(p,s)$ , distinguant les particularités dans ces modèles par rapport aux modèles autorégressifs purs.

Deuxièmement, nous avons effectué des simulations de Monte Carlo de plusieurs modèles de type  $\text{SVAR}(p)$  et  $\text{SVARX}(p,s)$ . Nous avons effectué l'estimation de divers modèles et avons calculé la valeur de différents critères de sélection afin d'étudier empiriquement la proportion de fois où le bon modèle était choisi, le nombre de fois que la bonne décision était prise concernant chaque matrice du modèle ainsi que la performance des prévisions ponctuelles. Les simulations ont été effectuées pour quatre tailles échantillonales dans les séries temporelles, soit  $n = 50$ ,  $n = 100$ ,  $n = 200$  et  $n = 400$ , ainsi que pour les critères AIC, HQ et BIC, permettant de comparer les estimations des paramètres, particulièrement le biais et la variance des estimateurs. Nous trouvons que les résultats varient grandement selon les matrices composant les processus générés. Nous trouvons également que les modèles  $\text{SVARX}$  performant moins bien que les  $\text{SVAR}$  correspondants en ce qui concerne les bonnes décisions alors que ce n'est pas systématiquement le cas lorsqu'on se concentre sur l'EQM des prévisions.

Troisièmement, nous avons appliqué la théorie étudiée à des données macroéconomiques en utilisant la production mensuelle des fromages cheddar, mozzarella et ricotta au Canada en tant que variables endogènes et le prix mensuel du lait au Québec, en Ontario et en Colombie-Britannique en tant que variables exogènes. Après avoir transformé les données

pour fins de stationnarité, nous avons effectué le processus d'ajustement en commençant par la spécification du modèle, qui a menée au choix d'un modèle VARX(12,3). Ensuite, nous avons fait l'estimation des paramètres pour tous les modèles SVARX dont les ordres maximaux sont  $p = 12$  et  $s = 3$ . En choisissant celui qui minimisait le critère AIC, nous sommes parvenus à un modèle SVARX( $\{1,2,3,4,5,7,9,12\},\{0,3\}$ ). Nous avons terminé par l'évaluation du modèle et avons conclu, avec la vérification diagnostic, que le modèle était approprié aux données. Par la suite, nous avons calculé les prévisions ponctuelles et avons conclu que l'ajustement d'un SVARX( $p,s$ ) semble mener à de meilleures prévisions qu'un modèle VARX( $p,s$ ) dans le cadre de ce jeu de données réelles.

Parmi les avenues de recherche possibles, il serait intéressant de développer plus en détail le calcul des prévisions pour des horizons supérieurs à un. Le calcul des prévisions ponctuelles peut se faire comme décrit dans ce mémoire mais plus de travail est nécessaire pour élaborer les bandes de prévision.

Des méthodes de sélection de modèles comme le LASSO mériteraient d'être adaptées aux modèles VARX( $p,s$ ). Des comparaisons avec les méthodes de sélection de modèles traditionnelles, comme les critères de type AIC, BIC et HQ seraient bienvenues. Nous laissons ces voies de recherche pour des travaux à venir.

## Références

---

- Akaike, H. (1974). "A new look at the statistical model identification", *IEEE Transactions on Automatic Control* **19**, 716–723.
- Boutahar, M. et Deniau, C. (1995). "A proof of asymptotic normality for some VARX models", *Metrika* **5**, 331–339.
- Brockwell, P. J. et Davis, R. A. (1991). *Time Series : Theory and Methods*. 2e édition. Springer-Verlag, New York.
- Brockwell, P. J. et Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. 3e édition. Springer Texts in Statistics. Springer-Verlag, New York.
- Brockwell, P. J., Davis, R. A. et Alexandre Trindade, A. (2004). "Asymptotic properties of some subset vector autoregressive process estimators", *Journal of Multivariate Analysis* **90**, 327–347.
- Brown, B. M. (1971). "Martingale central limit theorems", *Annals of Mathematical Statistics* **42**, 59–66.
- Brüggemann, R. Krolzig, H.-M. et Lütkepohl, H. (2002), "Comparison of Model Reduction Methods for VAR Processes", SFB 373 Discussion Paper No. 2002,80, Universität Humboldt de Berlin.
- Brüggemann, R. (2004). *Model Reduction Methods for Vector Autoregressive Processes*. Lecture Notes in Economics and Mathematical Systems, no. 536. Springer-Verlag, New York.
- Brüggemann, R. et Lütkepohl, H. (2001). "Lag selection in subset VAR models with an application to a U.S. monetary system", in R. Friedmann, L. Knüppel and H. Lütkepohl (eds), *Econometric Studies - A Festschrift in Honour of Joachim Frohn*, LIT : Münster, 107–128.
- Chen, C. W. S. (1999). "Subset selection of autoregressive time series models", *Journal of Forecasting* **18**, 505–516.

- da Silva, A. J. , da Fonseca Neto, J. V. et Nagem N. F. (2009). "Parametric ARX modeling of the electrolytic smelter pot", *2009 11th International Conference on Computer Modelling and Simulation*, 217–222.
- Diversi, R., Guidorzi, R. et Soverini, U. (2007). "Identification of ARX models with noisy input and output", *2007 European Control Conference (ECC)*, 4073–4078.
- Duchesne, P. (2005). "On the asymptotic distribution of residual autocovariances in VARX models with applications", *Test* **14**, 449–473.
- Duchesne, P. et Lafaye de Micheaux, P. (2013), "Distributions for residual autocovariances in parsimonious periodic vector autoregressive models with applications", *Journal of Time Series Analysis* **34**, 496–507.
- Duchesne, P. et Roy, R. (2004). "On consistent testing for serial correlation of unknown form in vector time series models", *Journal of Multivariate Analysis* **89**, 148–180.
- Francq, C. et Zakoïan, J. (2010). *GARCH Models : Structure, Statistical Inference and Financial Applications*. John Wiley & Sons Ltd., Hoboken, New Jersey.
- Fuller, W. A. (1996). *Introduction to Statistical Time Series*. 2e édition. Wiley Series in Probability and Statistics. John Wiley & Sons, New York.
- Guo, L., Huang, D. W. et Hannan, E. J. (1990). "On ARX( $\infty$ ) approximation", *Journal of Multivariate Analysis* **32**, 17–47.
- Haggan, V. et Oyetunji, O. B. (1984). "On the selection of subset autoregressive time series models", *Journal of Time Series Analysis* **5**, 103–113.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*, Academic Press : New York.
- Hannan, E. J. et Deistler, M. (1988). *The Statistical Theory of Linear Systems*, Wiley :New-York.
- Hannan, E. J., Dunsmuir, W. T. M. et Deistler, M. (1980). "Estimation of vector ARMAX models", *Journal of Multivariate Analysis* **10**, 275–295.
- Hannan, E. J. et Quinn, B. G. (1979). "The determination of the order of an autoregression", *Journal of the Royal Statistical Society, Series B* **41**, 190–195.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag :Berlin.

- Hurvich, C. M. et Tsai, C. L. (1989). "Regression and time series model selection in small samples", *Biometrika* **76**, 297–307.
- Krolzig, H.-M. et Hendry, D. F. (2001). "Computer automation of general-to-specific model selection procedures", *Journal of Economic Dynamics and Control* **25**, 831–866.
- Lütkepohl, H. (2005). *New introduction to Multiple Time Series Analysis*, Springer-Verlag :Berlin.
- Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*, 2e édition, Springer-Verlag :Berlin.
- McClave, J. T. (1975). "Subset autoregression", *Technometrics* **17**, 213–220.
- McLeod, A. I. et Zhang, Y. (2006). "Partial autoregression parameterization for subset autoregression", *Journal of Time Series Analysis* **27**, 599–612.
- McLeod, I. et Zhang, Y. (2008). "Improved subset autoregression : With R package". *Journal of Statistical Software* **28**, 1–28.
- Nardi, Y. et Rinaldo, A. (2011). "Autoregressive process modeling via the Lasso procedure", *Journal of Multivariate Analysis* **102**, 528–549.
- Nicholson, W. B., Matteson, D. S. et Bien, J. (2017). "VARX-L : Structured regularization for large vector autoregressions with exogenous variables", *International Journal of Forecasting* **33**, 627–651.
- Ocampo, S. et Rodríguez, N (2012). "An introductory review of a structural VAR-X estimation and applications", *Colombian Journal of Statistics* **35**, 479–508.
- Penm, J. H. W., Penm, J. H. et Terrell, R. D. (1993). "The recursive fitting of subset VARX models", *Journal of Time Series Analysis* **6**, 603–619.
- Quinn, B. G. (1980). "Order determination for a multivariate autoregression", *Journal of the Royal Statistical Society, Series B* **42**, 182–185.
- Reinsel, G. R. (1997). *Elements of Multivariate Time Series Analysis*. 2e édition, Springer :New York.
- Sanandaji, B. M., Vincent, T. L., Wakin, M. B., Tóth, R. et Poolla, K. (2011). "Compressive System Identification of LTI and LTV ARX models", *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 791–798.
- Schwarz, G. (1978). "Estimating the dimension of a model", *Annals of Statistics* **6**, 461–464.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley :New York.

- Shibata, R. (1980). "Asymptotically efficient selection of the order of the model for estimating parameters of a linear process", *Annals of Statistics* **8**, 147–164.
- Shumway, R. H. et Stoffer, D. S. (2017). *Time Series Analysis and Its Applications. With R Examples*. 4e édition. Springer Texts in Statistics. Springer-Verlag, New York.
- Tsay, R. S. (2014), *Multivariate Time Series Analysis. With R and Financial Applications*, Wiley :New York.
- Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267–288.
- Unnikrishnan, N. K. (2004). "Bayesian subset model selection for time series", *Journal of Time Series Analysis* **25**, 671–690.
- Ursu, E. et Duchesne, P. (2009a). "On multiplicative seasonal modelling for vector time series", *Statistics and Probability Letters* **79**, 2045-2052.
- Ursu, E. et Duchesne, P. (2009b). "Estimation and model adequacy checking for multivariate seasonal autoregressive time series models with periodically varying parameters", *Statistica Neerlandica* **63**, 183-212.
- Ursu, E. et Duchesne, P. (2009c). "On modeling and diagnostic checking of vector periodic autoregressive time series models", *Journal of Time Series Analysis* **30**, 70-96.
- Zhang, X. et Terrell, R. D. (1997). "Prediction modulus : a new direction for selecting subset autoregressive models", *Journal of Time Series Analysis* **18**, 195–212.