

Université de Montréal

Sur les modèles non-linéaires autorégressifs à transition
lisse et le calcul de leurs prévisions

par

Gabrielle Grégoire

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

Mémoire présenté à la Faculté des études supérieures et postdoctorales
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en statistique

2 août 2019

Université de Montréal

Faculté des études supérieures et postdoctorales

Ce mémoire intitulé

Sur les modèles non-linéaires autorégressifs à transition lisse et le calcul de leurs prévisions

présenté par

Gabrielle Grégoire

a été évalué par un jury composé des personnes suivantes :

David Haziza

(président-rapporteur)

Pierre Duchesne

(directeur de recherche)

Maciej Augustyniak

(membre du jury)

Mémoire accepté le :

Sommaire

Ce mémoire porte sur l'étude des données dépendantes. La littérature classique a consacré beaucoup d'énergie dans l'étude de modèles qualifiés de linéaires. Ces modèles sont particulièrement utiles pour des données macroéconomiques mesurées à des périodes finalement assez longues (mois, années, etc). Lorsque les données sont mesurées à une échelle temporelle plus fine, et lorsque les données sont nombreuses, il est alors possible de décrire le processus stochastique sous-jacent par des modèles plus élaborés, permettant de décrire les caractéristiques non-linéaires. C'est dans ce cadre moderne que s'inscrit ce mémoire. Il se propose d'étudier les modèles autorégressifs à transition lisse. Les modèles ont été introduits et popularisés par Teräsvirta (1994), entre autres. Nous concentrons notre étude sur la modélisation et les prévisions pour ces modèles. Ceux-ci étant marqués par la présence de plusieurs régimes ainsi qu'une transition particulière entre ces différents régimes, ils permettent de modéliser plus adéquatement un processus stochastique par des modèles de séries chronologiques qui affichent certains comportements non-linéaires. Notre objectif est de comparer ces modèles aux modèles autorégressifs linéaires classiques, et d'étudier si leur utilisation est marquée par une différence favorable au niveau de la prévision de valeurs futures. Il est à noter qu'une motivation première dans ce mémoire est l'élaboration de prévisions dans ces modèles.

Bien que la formulation des modèles autorégressifs à transition lisse soit particulièrement attrayante, la mise en application entraîne de nombreuses complications. Entre autres, l'estimation des modèles présente des difficultés au niveau de l'obtention de valeurs estimées pour les paramètres, dû à la présence d'une composante non-linéaire dans le modèle, ce qui rend l'estimation plus complexe puisqu'elle se doit alors d'être effectuée par optimisation non-linéaire. Bien sûr, ceci est également vrai dans la classe des modèles autorégressifs à moyennes mobiles (ARMA), reposant sur le choix d'un ordre autorégressif p et moyenne mobile q . Lorsque $q > 0$, il est bien connu que l'optimisation est également non-linéaire. Cependant, l'expérience empirique suggère que les problèmes numériques sont moins difficiles que pour les modèles autorégressifs à transition lisse. L'estimation des erreurs standards des paramètres est également élaborée, mais

possible, puisque l'obtention de la matrice des variances-covariances est souvent marquée par des difficultés calculatoires.

Les prévisions pour les séries temporelles occasionnent également des problèmes dans le cadre non-linéaire. La théorie linéaire classique n'étant pas applicable en raison de la composante non-linéaire du modèle, les prévisions ponctuelles pour les modèles autorégressifs à transition lisse doivent être effectuées à l'aide de différentes méthodes plus ou moins complexes, dont certaines mènent à un biais pour les prévisions ponctuelles aux temps supérieurs à un et d'autres deviennent rapidement difficiles à obtenir lorsque les temps de prévisions sont grands. Pour les intervalles de prévision, ils doivent également être mesurés avec des méthodes de ré-échantillonnage puisque la théorie linéaire n'est pas applicable. En fait, il peut être affirmé qu'une contribution originale du mémoire est une étude détaillée des prévisions et des intervalles de prévision dans nos modèles.

Dans le premier chapitre, nous présentons les séries temporelles ainsi que les modèles autorégressifs linéaires classiques principaux s'y rattachant. Nous élaborons les modèles non-linéaires ainsi que le modèle autorégressif à transition lisse univarié, en plus de ses caractéristiques principales et des conditions de sa stationnarité. Dans le deuxième chapitre, nous développons les techniques d'estimation reliées aux modèles autorégressifs à transition lisse. En particulier, nous élaborons les étapes d'ajustement du modèle, soit la spécification, l'estimation et l'évaluation. La spécification comprend entre autres les tests de linéarité, étape nécessaire afin de justifier l'utilisation des modèles autorégressifs à transition lisse. L'estimation est effectuée par optimisation non-linéaire avec recherche quadrillée pour trouver les valeurs initiales. Dans le troisième chapitre, nous présentons les méthodes de prévision pour les modèles autorégressifs linéaires classiques et pour les modèles autorégressifs à transition lisse. Plus particulièrement, nous élaborons les difficultés d'application des méthodes de prévision habituelles dans le cadre des modèles non-linéaires, et les méthodes permettant de contourner ces difficultés. Nous définissons également les intervalles de prévision et les méthodes pour déterminer ces intervalles. Dans le quatrième chapitre, nous appliquons la théorie définie précédemment lors de simulations empiriques, avec pour but de comparer les modèles linéaires aux modèles autorégressifs à transition lisse, et nous discutons des résultats obtenus. Dans le cinquième chapitre, nous appliquons la théorie à une série temporelle représentant les rendements quotidiens du fonds négocié en bourse SPDR (pour *Standard & Poor's Depositary Receipts*) suivant l'indice boursier S&P 500 (*SPY*), et nous comparons nos résultats avec ceux disponibles dans la littérature, tant au niveau de l'estimation des

modèles autorégressifs à transition lisse qu'à la performance des prévisions ponctuelles et intervalles de prévision. Nous finissons par une conclusion. Tous les codes utiles pour reproduire les résultats de simulations et d'analyse de données sont disponibles sur demande.

Mots-clés : Série temporelle, modèles autorégressifs à transition lisse, stationnarité, prévisions ponctuelles, intervalles de prévision.

Summary

This master's thesis focuses on the study of dependent data. Classical literature has been widely focused on models described as linear. These models find great use when applied to macroeconomics data that are measured with relatively long time span (months, years, etc). When the data are measured using a shorter time span, and when the data are plentiful, it is then possible to describe the underlying stochastic process using more sophisticated models, which allows for proper modelling of the data's nonlinear characteristics. It is in that contemporary setting that this master's thesis takes place. It focuses on smooth transition autoregressive models, introduced and popularized by many authors including Teräsvirta (1994). We focus our study on modelling and forecasting for these models. As they are characterized by multiple regimes and a specific transition between those regimes, the smooth transition autoregressive models may allow for a more adequate modelling of a stochastic process using time series models that show nonlinear components. Our goal is to compare these models with the linear autoregressive models, and to study if they allow for more accurate forecasts. Thus, a main focus of this thesis is the elaboration of a forecasting system for these models.

Although the use of smooth transition autoregressive models may seem enticing, applying the models to data brings its own share of complications, especially regarding parameters estimation. The estimation of the model sometimes has trouble converging due to a nonlinear component in the model, therefore requiring non-linear optimisation which is more complex. Of course, this is also true for the class of autoregressive moving average models (ARMA), which rely on the choice of autoregressive order p and moving average q . When $q > 0$, it is well-known that optimisation is nonlinear as well. However, empirical evidence suggests that numerical problems are less difficult than for smooth transition autoregressive models. The estimation of standard deviations of the parameters may also be difficult to obtain since the calculation of the variance-covariance matrix may be affected by computing issues.

Forecasting for time series also poses issues for nonlinear time series. Classic linear theory is not applicable due to the nonlinear component in the model, and therefore point forecasting for smooth transition autoregressive models requires the use of other methods that vary in complexity, some which cause a bias for forecasts of lead times strictly larger than one which increase drastically in difficulty for longer lead times. For prediction intervals, because the linear theory is not applicable, they must be obtained using resampling methods. In fact, it can be asserted that an original contribution from this thesis is an elaborate and detailed study of point forecasting and forecast intervals for the models.

In the first chapter, we introduce time series and classic linear models that are used to model time series. We elaborate nonlinear models and the univariate smooth transition autoregressive models, as well as the main characteristics and stationarity conditions for these models. In the second chapter, we develop estimation methods for smooth transition autoregressive models. The modelling process includes specification, estimation and evaluation of the model. Specification includes linearity tests for the data, which is required and justifies the use of nonlinear time series models. Estimation is done by nonlinear optimisation, and by finding its starting values using a grid search. In the third chapter, we present forecasting methods for classic linear autoregressive models and smooth transition autoregressive models. We elaborate the problems encountered when applying conventional theory to nonlinear models, and the methods to overcome these issues. We also define prediction intervals and how to obtain them. In the fourth chapter, we apply the theory of prior chapters to empirical simulations, with the goal of comparing linear autoregressive models to smooth transition autoregressive ones, focusing on modelling and both point and interval forecasting. We discuss our results. In the fifth chapter, we apply the theory to a time series representing daily returns of the SPDR S&P 500 stock market index (*SPY*). We compare our results with those available in the literature, for both estimation of the smooth transition autoregressive models and point forecasting and prediction intervals performance. We then conclude the thesis.

Keywords : Time series, smooth transition autoregressive models, stationnarity, forecasting, prediction intervals.

Table des matières

Sommaire	v
Summary	ix
Liste des tableaux	xv
Liste des figures	xvii
Liste des sigles et des abréviations	xix
Remerciements	xxi
Introduction	1
Chapitre 1. Préliminaires	7
1.1. Généralités	7
1.2. Modèles linéaires univariés classiques	9
1.3. Modèles non-linéaires univariés	11
1.3.1. Stationnarité pour le modèle STAR univarié	15
Chapitre 2. Techniques d'estimation	19
2.1. Spécification du modèle	19
2.1.1. Sélection du délai p	20
2.1.2. Tests de linéarité	22
2.1.3. Sélection de la variable de transition s_t	25
2.1.4. Sélection de la fonction de transition g	26
2.2. Estimation des paramètres	27
2.2.1. Problèmes numériques lors de l'estimation	32
2.3. Évaluation du modèle	34
2.3.1. Absence d'autocorrélation des résidus	35
2.3.2. Absence de non-linéarité restante	36

2.3.3. Constance des paramètres.....	37
Chapitre 3. Prévisions ponctuelles et intervalles de prévision.....	39
3.1. Généralités.....	39
3.2. Prévisions ponctuelles pour les modèles linéaires.....	42
3.3. Prévisions ponctuelles pour les modèles STAR(p).....	43
3.3.1. Prévisions directes.....	45
3.3.2. Prévisions exactes.....	45
3.3.3. Prévisions naïves.....	46
3.3.4. Prévisions de Monte Carlo.....	48
3.3.5. Prévisions par la méthode de bootstrap.....	49
3.4. Intervalles de prévision.....	50
Chapitre 4. Étude par simulations : estimation et prévisions dans les modèles STAR.....	55
4.1. Description des simulations.....	56
4.1.1. STAR(1).....	56
4.1.2. STAR(2).....	59
4.2. Estimation dans les modèles STAR.....	60
4.3. Discussion des résultats de simulation sur l'estimation.....	64
4.4. Prévisions ponctuelles et intervalles de prévision.....	67
4.5. Discussion des résultats de simulation sur les prévisions.....	77
4.5.1. Prévisions ponctuelles.....	77
4.5.2. Intervalles de prévision.....	80
Chapitre 5. Analyse de données réelles.....	83
5.1. Description du jeu de données.....	84
5.2. Ajustement d'un modèle STAR(p).....	84
5.2.1. Spécification du modèle.....	84
5.2.1.1. Sélection du délai p	84
5.2.1.2. Test de linéarité, choix de la variable de transition s_t et de la fonction de transition g	85
5.2.2. Estimation des paramètres.....	87

5.2.3.	Évaluation du modèle	91
5.2.3.1.	Absence d'autocorrélation des résidus	91
5.2.3.2.	Absence de non-linéarité restante	91
5.2.3.3.	Constance des paramètres.....	92
5.3.	Prévisions de valeurs futures	93
5.4.	Comparaison avec la littérature	96
5.4.1.	Estimation des modèles STAR(p)	96
5.4.2.	Prévisions ponctuelles pour les modèles STAR(p).....	97
Conclusion	99
Références	101

Liste des tableaux

4. I	Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(1) avec $\gamma = 5$, $c = -0,2$.	61
4. II	Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(1) avec $\gamma = 5$, $c = -0,5$.	61
4. III	Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(1) avec $\gamma = 20$, $c = -0,2$.	61
4. IV	Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(1) avec $\gamma = 20$, $c = -0,5$.	62
4. V	Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(2) avec $\gamma = 2$, $c = 1$.	62
4. VI	Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(2) avec $\gamma = 20$, $c = 1,2$.	62
4. VII	Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 5$ et $c = -0,2$.	68
4. VIII	Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 5$ et $c = -0,5$.	69
4. IX	Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 20$ et $c = -0,2$.	69
4. X	Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 20$ et $c = -0,5$.	69

4. XI	Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(2), AR(p) et STAR(2) avec $\gamma = 2$ et $c = 1$.	70
4. XII	Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(2), AR(p) et STAR(2) avec $\gamma = 20$ et $c = 1,2$.	70
4. XIII	Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 5$ et $c = -0,2$.	71
4. XIV	Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 5$ et $c = -0,5$.	72
4. XV	Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 20$ et $c = -0,2$.	73
4. XVI	Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 20$ et $c = -0,5$.	74
4. XVII	Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(2), AR(p) et STAR(2) avec $\gamma = 2$ et $c = 1$.	75
4. XVIII	Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(2), AR(p) et STAR(2) avec $\gamma = 20$ et $c = 1,2$.	76
5. I	Valeur du critère AIC obtenue selon le délai p dans le modèle AR(p)	85
5. II	Valeur- p du test de linéarité selon le choix de la variable de transition s_t	86
5. III	Estimation des paramètres du modèle STAR(5) ainsi que leurs intervalles de confiance 95%	88
5. IV	Valeur- p obtenue selon le délai m pour le test de Ljung-Box	91
5. V	Biais moyen et mesure d'EQM des prévisions ponctuelles pour dix temps de prévisions du rendement quotidien de l'indice <i>SPY</i>	93
5. VI	Inclusion des vraies valeurs dans les intervalles de prévision pour dix temps de prévisions du rendement quotidien de l'indice <i>SPY</i>	94

Liste des figures

4.1	Forme prise par la fonction de transition selon γ et c	58
4.2	Exemple d'une série STAR(1) selon γ et c , avec $T = 400$	58
4.3	Forme prise par la fonction de transition selon γ et c	59
4.4	Exemple d'une série STAR(2) selon γ et c , avec $T = 400$	60
4.5	Distributions empiriques de $\hat{\gamma}$ pour le modèle STAR(1) avec $\gamma = 5$, $c = -0,2$ selon le nombre de valeurs dans la série. Pour $T = 400$, 69 valeurs sont entre 60 et 59000 et ne sont pas illustrées.	63
4.6	Distributions empiriques de $\hat{\gamma}$ pour le modèle STAR(1) avec $\gamma = 5$, $c = -0,5$ selon le nombre de valeurs dans la série. Pour $T = 400$, 62 valeurs sont entre 60 et 26000 et ne sont pas illustrées.	63
4.7	Distributions empiriques de $\hat{\gamma}$ pour le modèle STAR(1) avec $\gamma = 20$, $c = -0,2$ selon le nombre de valeurs dans la série. Pour $T = 400$, 38 valeurs sont entre 2000 et 16000 et ne sont pas illustrées.	63
4.8	Distributions empiriques de $\hat{\gamma}$ pour le modèle STAR(1) avec $\gamma = 20$, $c = -0,5$ selon le nombre de valeurs dans la série. Pour $T = 400$, 45 valeurs sont entre 2000 et 71000 et ne sont pas illustrées.	64
4.9	Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(1) avec $T = 2000$, $\gamma = 5$, $c = -0,2$	71
4.10	Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(1) avec $T = 2000$, $\gamma = 5$, $c = -0,5$	72
4.11	Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(1) avec $T = 2000$, $\gamma = 20$, $c = -0,2$	73
4.12	Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(1) avec $T = 2000$, $\gamma = 20$, $c = -0,5$	74
4.13	Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(2) avec $T = 2000$, $\gamma = 2$, $c = 1$	75

4.14	Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(2) avec $T = 2000$, $\gamma = 20$, $c = 1,2$	76
5.1	Rendement de l'indice SPDR S&P 500 (SPY) selon la date.	84
5.2	Valeur prise par la fonction g selon le temps pour le jeu de données SPY	90
5.3	Valeur prise par la fonction g selon la valeur s_t de la série pour le jeu de données SPY	90
5.4	Densité des résidus du modèle STAR(5) pour le modèle ajusté	90
5.5	Prévisions ponctuelles selon les différentes méthodes étudiées et vraies valeurs de la série	94
5.6	Prévisions ponctuelles et intervalles de prévision obtenus selon les modèles AR(5) et STAR(5).	95

Liste des sigles et des abréviations

AIC	Critère d'information d'Akaike, de l'anglais <i>Akaike Information Criteria</i>
AR	Autorégressif
ARMA	Autorégressif moyenne mobile, de l'anglais <i>Autoregressive moving average</i>
ARMAX	Autorégressif moyenne mobile avec variables exogènes, de l'anglais <i>Autoregressive moving average with exogenous variables</i>
ARX	Autorégressif avec variables exogènes, de l'anglais <i>Autoregressive with exogenous variables</i>
BS	Bootstrap
EQM	Erreur quadratique moyenne
MC	Monte Carlo
STAR	Autorégressif à transition lisse, de l'anglais <i>Smooth transition autoregressive</i>
TAR	Autorégressif à seuil limite, de l'anglais <i>Threshold autoregressive</i>

Remerciements

Je tiens tout d'abord à remercier mon directeur de mémoire, M. Pierre Duchesne, pour m'avoir accompagné tout au long du processus de planification et de rédaction de ce manuscrit. Ce dernier n'aurait pu voir le jour sans la disponibilité, les encouragements, le soutien financier, les commentaires pertinents, la relecture attentive et la confiance dont il a fait preuve à mon égard.

Je tiens également à témoigner ma gratitude envers tous les professeurs qui ont contribué, de près ou de loin, à mon cheminement académique, et plus particulièrement ceux qui m'ont enseigné au Département de mathématiques et de statistique. Ils ont su transmettre leur savoir avec passion, dévouement et rigueur. Merci également aux membres du jury de l'attention qu'ils portent à mon travail.

À mes parents, qui m'ont toujours supporté dans tout ce que j'entreprends, merci de votre présence, de votre écoute et de vos encouragements constants. Merci d'être de si bons modèles de persévérance et de votre soutien inconditionnel.

À mon frère, Philippe, qui m'a encouragé et guidé tout au long de ma vie, merci pour ton soutien à toute épreuve et ta générosité infinie. À ma belle-soeur, Élise, merci pour tes précieux conseils et pour avoir été un modèle incontestable de persévérance aux cycles supérieurs.

À mes amis et ma famille élargie, qui ont été témoins de ce long parcours scolaire, merci pour vos encouragements, votre appui et votre intérêt à comprendre ce que je faisais.

Ce mémoire n'aurait pas été possible sans vous tous.

Introduction

L'étude des séries chronologiques (ou séries temporelles) porte sur des données qui sont mesurées régulièrement dans le temps. Les mesures sont donc effectuées à des instants égaux, plus ou moins grands. Traditionnellement, les données macroéconomiques étaient mesurées annuellement, trimestriellement ou au mieux mensuellement. Avec la réalité des ordinateurs puissants et les bases de données volumineuses, il est fréquent d'obtenir des données mesurées selon une échelle temporelle plus fine (hebdomadaire, journalière ou même horaire). Dans tous les cas, le phénomène d'intérêt est supposé provenir d'un certain processus stochastique, et par définition, une série chronologique en représente une réalisation finie.

Plusieurs domaines d'études comportent des séries temporelles d'intérêt théorique et pratique, particulièrement dans les applications économiques et financières. En effet, ces domaines comportent plusieurs types de valeurs évoluant dans le temps, par exemple le taux de chômage mensuel en économie ou les rendements quotidiens d'une action en finance. La modélisation de ces séries temporelles permet de représenter le comportement sous-jacent de la valeur évoluant dans le temps, et l'étude de ce comportement peut faciliter l'interprétation de l'élément étudié. La modélisation des séries rend également la prévision de valeurs futures intuitive, puisqu'il est généralement supposé que le comportement sous-jacent identifié se répétera dans le temps. La prévision de valeurs futures pour une série est d'un grand intérêt dans plusieurs domaines, par exemple en finance, où prédire le prix qu'une action prendra dans le futur peut mener à une prise de décision.

La modélisation des séries temporelles est caractérisée par un ensemble de modèles dits linéaires autorégressifs (d'ordre fini ou infini) dont l'utilisation est fréquente et répandue. Cet ensemble comprend les modèles autorégressifs $AR(p)$, autorégressifs avec variables exogènes $ARX(p,s)$, autorégressifs à moyenne mobile $ARMA(p,q)$ et autorégressifs à moyenne mobile avec variables exogènes $ARMAX(p,s,q)$. Leur théorie est grandement élaborée dans la littérature, et leur facilité d'application à tout type de série temporelle fait que ces modèles sont souvent le choix par défaut pour analyser une série. Toutefois, il se peut que ces modèles ne soient pas idéaux pour modéliser une série, particulièrement si la série est marquée par

de la non-linéarité due à la présence de plusieurs modèles sous-jacents ou régimes dans les données, un régime étant par définition la structure dans le temps et l'espace d'un système économique. Des modèles dits non-linéaires ont ainsi été développés dans la littérature, entre autres les modèles autorégressifs à seuil limite $TAR(p)$ et autorégressifs à transition lisse $STAR(p)$, ce dernier étant le principal sujet de ce mémoire. Ils permettent dans certains cas de modéliser plus adéquatement une série qui présenterait un comportement non-linéaire, et ainsi de mener à des prévisions qui pourraient être plus exactes que celles qui seraient obtenues par l'application d'un modèle linéaire classique.

Dans la littérature, les modèles autorégressifs à transition lisse, ou $STAR(p)$, ont été élaborés par Teräsvirta (1994), puis par la suite par Dijk, Teräsvirta et Franses (2002), entre autres. Leurs applications ont principalement lieu dans le domaine de l'économie et se concentrent sur l'ajustement des modèles aux différents jeux de données étudiés ainsi que les difficultés s'y rapportant. Celles-ci comprennent les tests de linéarité et l'estimation des paramètres et de leurs erreurs standards. La littérature s'attarde également dans quelques cas aux prévisions ponctuelles de valeurs futures suite à l'ajustement du modèle, entre autres par Teräsvirta, Dijk et Medeiros (2005) et Lin et Granger (1994). Spécifiquement, les intervalles de prévision ne sont pas élaborés empiriquement par aucun auteur pour les modèles $STAR(p)$. Bien entendu, le calcul des intervalles de prévision est bien développé dans les modèles classiques de séries chronologiques. Il existe des formules explicites qui sont facilement mises en oeuvre, qui permettent d'apprécier l'étendue des valeurs possibles pour une variable aléatoire future. Le calcul de tels intervalles de prévision dans les modèles non-linéaires est souhaitable et le présent mémoire amène des contributions dans cette direction.

Les difficultés de modélisation et d'ajustement des modèles $STAR(p)$ aux données rendent leur application parfois laborieuse. Par conséquent, un thème principal de ce mémoire est la comparaison des modèles $STAR(p)$ aux modèles $AR(p)$, ces derniers étant beaucoup plus simples à déterminer et utiliser. En effet, pour justifier l'utilisation des modèles $STAR(p)$, il faut que ceux-ci présentent un avantage au niveau de la précision des prévisions ponctuelles et intervalles de prévision. Sinon, l'ajustement plus complexe des modèles $STAR(p)$ n'est peut-être pas justifiable si les prévisions ponctuelles ont une performance similaire à celles des modèles $AR(p)$, et que ses intervalles de prévision sont semblables à ceux obtenus par le modèle $AR(p)$. Toutefois, si le modèle $STAR(p)$ présente un avantage pour l'un ou l'autre de ces éléments, alors son utilisation, quoique plus ardue, peut être justifiée.

Ce mémoire est constitué de cinq chapitres. Dans le premier chapitre, nous élaborons le cadre théorique pour l'étude des séries temporelles, en plus des définitions de bruit blanc et de stationnarité. Nous présentons les modèles autorégressifs linéaires classiques, soit les modèles $AR(p)$, $ARX(p,s)$, $ARMA(p,q)$ et $ARMAX(p,s,q)$, suivis des modèles non-linéaires généraux, des modèles $TAR(p)$ et des modèles $STAR(p)$ à deux et m régimes. Pour ces derniers, nous présentons la fonction de transition g et ses formes habituelles ainsi que la variable de transition s_t , en plus d'élaborer la stationnarité et les conditions pour que celle-ci soit satisfaite pour les modèles $STAR(p)$.

Dans le deuxième chapitre, nous développons le processus d'ajustement des modèles $STAR(p)$, qui comprend les étapes de spécification, d'estimation et d'évaluation du modèle. Pour la spécification, nous présentons le choix du délai p , de la fonction g et de la variable de transition s_t , en plus des tests de linéarité. Le rejet de la linéarité est requis pour justifier l'utilisation d'un modèle $STAR(p)$. Pour l'estimation des paramètres, nous posons le théorème de convergence des paramètres obtenus par optimisation non-linéaire, en plus de définir leur matrice de variances-covariances et l'estimation de celle-ci. Nous nous attardons aux problèmes calculatoires et de convergence des algorithmes fréquemment rencontrés lors de l'estimation d'un modèle $STAR(p)$ et de sa matrice de variances-covariances, en plus d'élaborer la recherche quadrillée permettant d'obtenir des valeurs initiales pour l'algorithme. Pour l'évaluation du modèle, nous élaborons les tests d'absence d'autocorrélation des résidus, d'absence de non-linéarité restante et de constance des paramètres.

Dans le troisième chapitre, nous nous attardons à la composante prévisionnelle de l'analyse complète des séries temporelles. Nous définissons tout d'abord le concept de prévisions ainsi que les méthodes d'évaluation de la performance de celles-ci. Les mesures empiriques naturelles sont les biais et erreurs quadratiques moyennes calculés avec la méthode de Monte Carlo. Nous élaborons ensuite les prévisions ponctuelles pour les modèles autorégressifs linéaires classiques, et plus particulièrement les modèles $AR(p)$, suivies de celles pour les modèles $STAR(p)$. Les méthodes de prévisions ponctuelles pour les modèles $STAR(p)$ aux temps de prévisions supérieurs à un comportent la méthode naïve, qui mène à un biais systématique, ainsi que les méthodes directes et exactes, qui sont passablement plus complexes à mettre en oeuvre comparativement aux méthodes dans les modèles linéaires, et les méthodes de Monte Carlo et bootstrap qui sont plus simples à appliquer ainsi que significativement moins biaisées, du moins ce sont les conclusions suggérées par nos expériences empiriques. Nous élaborons ensuite les intervalles de prévision pour les

modèles $AR(p)$ ainsi que ceux pour les modèles $STAR(p)$, qui doivent être effectués par les méthodes de Monte Carlo et bootstrap.

Le quatrième chapitre est consacré à des simulations Monte Carlo pour les modèles $STAR(p)$. Nous concentrons notre étude empirique sur plusieurs modèles $STAR(1)$ et $STAR(2)$ ayant des valeurs de paramètres non-linéaires variées. Nous simulons ainsi des séries suivant un modèle $STAR$, puis nous estimons les paramètres de ces modèles et discutons des difficultés d'estimation présentes selon le nombre de valeurs dans les séries, soit le fait que l'estimation des paramètres du modèle peuvent être loins des vraies valeurs si la série comporte un trop petit nombre de valeurs. Par la suite, nous calculons les prévisions et intervalles de prévision avec les différentes méthodes élaborées au chapitre 3. Nous mesurons le biais et l'erreur quadratique moyenne pour les prévisions ponctuelles, en plus des longueurs des intervalles de prévision selon les différents taux de couverture considérés. Finalement, nous discutons des résultats et de la façon dont les prévisions ponctuelles et intervalles de prévision se comparent à ceux qui seraient obtenus si le modèle ajusté était un $AR(p)$. Nous observons ainsi que les prévisions ponctuelles des modèles $STAR(p)$ ne sont pas toujours meilleures que celles obtenues par le modèle $AR(p)$, mais que les intervalles de prévision sont constamment plus petits en moyenne pour le modèle $STAR$ que pour le modèle AR , pour des taux de couverture comparables.

Le cinquième chapitre comporte une application de la théorie des modèles $STAR(p)$ élaborée dans les chapitres 1, 2 et 3 à un vrai jeu de données comportant les rendements quotidiens du fonds négocié en bourse SPDR (pour *Standard & Poor's Depository Receipts*) suivant l'indice boursier S&P 500 (*SPY*) pour les jours ouvrables du 1er janvier 2010 au 31 décembre 2018, dont les valeurs sont obtenues sur Yahoo Finance. Nous ajustons donc à ces données un modèle $STAR(p)$ en procédant par les étapes mentionnées au chapitre 2, ce qui inclut un test de linéarité ainsi que l'estimation des paramètres et de leurs erreurs standards. Par la suite, nous calculons les prévisions ponctuelles et les intervalles de prévision selon les différentes méthodes applicables pour le modèle $STAR(p)$, en plus de calculer celles qui sont obtenues par l'ajustement d'un modèle $AR(p)$. Nous discutons des résultats obtenus, à savoir si le modèle $STAR(p)$ mène à une réelle différence au niveau des prévisions et intervalles dans le cadre de données réelles, et nous concluons que ce n'est pas le cas pour notre jeu de données. Nous comparons également nos résultats à ceux élaborés dans la littérature, et trouvons que les processus d'ajustement pour les modèles $STAR(p)$ varient grandement selon les auteurs, et que les prévisions ponctuelles sont rarement significativement meilleures pour

le modèle STAR(p) que pour le modèle AR(p) dans la plupart des jeux de données étudiés dans la littérature.

Chapitre 1

Préliminaires

Dans ce chapitre, nous définissons tout d'abord certains concepts et propriétés fondamentales dans un contexte d'analyse des séries chronologiques. Ensuite, nous présentons les modèles linéaires classiques univariés ainsi que leurs limites. Ces modèles comprennent les modèles autorégressifs $AR(p)$, autorégressifs avec variables exogènes $ARX(p,s)$, autorégressifs à moyenne mobile $ARMA(p,q)$ et autorégressifs à moyenne mobile avec variables exogènes $ARMAX(p,s,q)$. Ensuite, nous présentons les modèles autorégressifs non-linéaires univariés, dont l'utilisation permet de combler les lacunes présentes chez les modèles linéaires univariés classiques. Les modèles univariés non-linéaires comprennent entre autres le modèle autorégressif à seuil limite ainsi que le modèle autorégressif à transition lisse, principal sujet de ce mémoire. Ce modèle comprend deux ou plusieurs régimes utilisant une fonction de transition entre les différents régimes. Nous élaborons ainsi le modèle autorégressif à transition lisse à deux et m régimes, ainsi que les formes les plus communes de la fonction de transition, et ses paramètres influençant le changement entre les différents régimes. Nous étudions finalement les conditions de stationnarité pour le modèle autorégressif à transition lisse à deux régimes.

1.1. Généralités

Tout d'abord, nous présentons le concept de série chronologique. Une série chronologique, ou série temporelle, est une réalisation finie d'un processus stochastique $\{\mathbf{y}_t, t \in \mathbb{Z}\}$ de taille T , avec chaque élément de la série observé typiquement de façon équidistante aux instants $t = 1, \dots, T$. \mathbf{y}_t est un vecteur aléatoire de dimension $k \times 1$ et est tel que $\mathbf{y}_t = (y_{t1}, y_{t2}, \dots, y_{tk})'$. Toutes les définitions ci-dessous proviennent de Brockwell et Davis (2016) et Shumway et Stoffer (2010).

Une propriété importante et fréquemment utilisée dans le cadre de l'étude des processus stochastiques (et donc des séries chronologiques) est le concept de stationnarité. Ce concept

est fondamental car lorsque cette propriété est remplie l'inférence statistique est facilitée. On distingue deux concepts : la stationnarité au sens strict et celle au sens large. La stationnarité au sens strict fait l'objet de la définition suivante.

Définition 1.1.1. *Un processus $\{\mathbf{y}_t\}$ est dit stationnaire au sens strict si*

$$(\mathbf{y}'_{t_1}, \dots, \mathbf{y}'_{t_k})' \sim (\mathbf{y}'_{t_1+h}, \dots, \mathbf{y}'_{t_k+h})',$$

où \sim représente l'égalité en loi, c'est-à-dire que la loi de probabilité d'un ensemble de valeurs $(\mathbf{y}'_{t_1}, \dots, \mathbf{y}'_{t_k})'$ est identique à celui d'un ensemble $(\mathbf{y}'_{t_1+h}, \dots, \mathbf{y}'_{t_k+h})'$, pour tout ensemble d'indices (t_1, \dots, t_k) , et pour tout $h \in \mathbb{Z}$.

La stationnarité au sens strict est peu utilisée dans les applications car elle est difficile à vérifier dans une réalisation pratique. De plus, elle implique d'imposer des conditions sur toutes les distributions conjointes du processus stochastique. Par conséquent, on utilise plutôt la stationnarité au sens large lors de l'application des séries chronologiques, puisque celle-ci comporte moins de conditions que la stationnarité au sens strict pour être satisfaite. Ceci étant dit, même si la stationnarité au sens large peut sembler moins restrictive car se concentrant sur les deux premiers moments du processus stochastique, elle exige l'existence de ces moments. On note que la stationnarité au sens strict pourrait être satisfaite sans que le processus stochastique soit composé de variables aléatoires avec des moments finis.

Définition 1.1.2. *Un processus $\{\mathbf{y}_t\}$ est dit stationnaire au sens large s'il répond aux deux conditions suivantes :*

- (1) *La fonction moyenne $E(\mathbf{y}_t) = \boldsymbol{\mu}$ est constante et ne dépend pas du temps t .*
- (2) *La fonction d'autocovariance $E\{(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t+h} - \boldsymbol{\mu})'\} = \boldsymbol{\Gamma}_h$ est une matrice comprenant les covariances $\gamma_{ij}(h) = Cov(y_i, y_{j+h})$ où $h \in \mathbb{Z}$ et $i, j = 1, \dots, k$.*

La stationnarité au sens large équivaut donc à poser des conditions sur les deux premiers moments du processus, contrairement à la stationnarité au sens strict qui impose des restrictions sur les distributions conjointes. Elle est donc plus facile à démontrer et suffisante pour la plupart des jeux de données. Dans ce mémoire, le terme stationnaire sous-entend la stationnarité au sens large. La stationnarité d'une série permet de développer l'inférence statistique qui fait l'objet de la définition 1.1.3 et de développer des prévisions pour les valeurs futures de la série.

Un premier exemple de processus stationnaire est le bruit blanc. Il s'agit d'un exemple de base et représente essentiellement un terme d'erreur en statistique classique. Le nom provient des applications touchant l'ingénierie : si on utilise le concept de densité spectrale, la densité spectrale du bruit blanc est constante, avec des fréquences égales, d'où le nom de bruit blanc. Il est fréquemment utilisé comme bruit dans les différents modèles. La définition 1.1.3 donne une définition formelle.

Définition 1.1.3. *Un bruit blanc $\{\boldsymbol{\varepsilon}_t\}$ est un processus stochastique satisfaisant les conditions suivantes :*

$$(1) E(\boldsymbol{\varepsilon}_t) = \mathbf{0},$$

$$(2) Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Omega}_\varepsilon,$$

$$(3) E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t-i}) = \mathbf{0}, \text{ lorsque } i \neq 0,$$

où $\boldsymbol{\Omega}_\varepsilon$ est la matrice de variances-covariances ne dépendant pas du temps.

Un bruit blanc est ainsi un processus univarié ou multivarié de dimension $k \times 1$ comprenant des variables non-corrélées, d'espérance $\mathbf{0}$ et de variance finie constante. Lorsque les variables sont également indépendantes et identiquement distribuées, le bruit blanc est dit fort, autrement il est dit faible. De par sa définition, le bruit blanc est un processus stationnaire, et lorsque $\{\boldsymbol{\varepsilon}_t\}$ est constitué de variables aléatoires normales et indépendantes, le bruit blanc est appelé un bruit blanc Gaussien.

1.2. Modèles linéaires univariés classiques

Lors de l'étude de séries chronologiques, il est pertinent d'ajuster un modèle aux données observées afin de permettre l'analyse de la série et la prévision de valeurs futures à l'aide du modèle. Dans le cas univarié, les modèles les plus fréquemment utilisés sont les modèles linéaires classiques, soit les modèles $AR(p)$ et $ARMA(p,q)$, ainsi que leurs extensions à l'inclusion de variables exogènes, comme par exemple les modèles $ARX(p,s)$ et $ARMAX(p,s,q)$, élaborés entre autres par Brockwell et Davis (2016), Franses, Dijk et Opschoor (2014), Hannan et Deistler (2012) et Shumway et Stoffer (2010).

Nous définissons tout d'abord le modèle autorégressif $AR(p)$.

Définition 1.2.1. *Un modèle autorégressif d'ordre p , noté $AR(p)$, est un processus qui satisfait l'équation aux différences stochastiques :*

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad (1.2.1)$$

où $\{y_t\}$ est stationnaire ; $\phi_1, \phi_2, \dots, \phi_p$ sont des paramètres, $\phi_p \neq 0$; et $\{\varepsilon_t\}$ est un bruit blanc.

Le modèle $AR(p)$ peut être étendu pour inclure des variables exogènes x_1, \dots, x_s permettant d'expliquer la variable réponse y_t , ces variables étant également respectivement des processus. On obtient alors le modèle $ARX(p,s)$.

Définition 1.2.2. *Un modèle autorégressif avec variables exogènes d'ordre (p,s) , noté $ARX(p,s)$, satisfait l'équation aux différences stochastiques :*

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \gamma_1 x_{1t} + \dots + \gamma_s x_{st} + \varepsilon_t, \quad (1.2.2)$$

où $\{y_t\}$ est stationnaire ; $\{x_{1t}\}, \dots, \{x_{st}\}$ sont des processus stochastiques exogènes ; $\phi_1, \phi_2, \dots, \phi_p, \gamma_1, \dots, \gamma_s$ sont des paramètres tels que $\phi_p \neq 0$ et $\gamma_s \neq 0$; et $\{\varepsilon_t\}$ est un bruit blanc.

Ces deux modèles autorégressifs univariés sont les formes linéaires classiques utilisées lorsque la composante autorégressive du modèle univarié se base uniquement sur les valeurs précédentes de la variable modélisée. Toutefois, ces modèles peuvent également être élargis afin d'inclure une composante moyenne mobile. Cette inclusion mène aux modèles linéaires $ARMA(p,q)$ et $ARMAX(p,s,q)$.

Définition 1.2.3. *Un modèle autorégressif à moyenne mobile d'ordre (p,q) , noté $ARMA(p,q)$, est un processus satisfaisant l'équation :*

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (1.2.3)$$

où $\{y_t\}$ est stationnaire ; $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ sont des paramètres, avec $\phi_p \neq 0$ et $\theta_q \neq 0$; et $\{\varepsilon_t\}$ est un bruit blanc.

Comme pour le modèle $AR(p)$, le modèle $ARMA(p,q)$ peut être étendu au cas comportant des variables exogènes.

Définition 1.2.4. *Un modèle autorégressif à moyenne mobile avec variables exogènes d'ordre (p,s,q) , noté $ARMAX(p,s,q)$, satisfait l'équation :*

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \gamma_1 x_{1t} + \dots + \gamma_s x_{st} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (1.2.4)$$

où $\{y_t\}$ est stationnaire; $\{x_{1t}\}, \dots, \{x_{st}\}$ sont des processus stochastiques exogènes; $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \gamma_1, \dots, \gamma_s$ sont des paramètres, avec $\phi_p \neq 0$, $\theta_q \neq 0$ et $\gamma_s \neq 0$; et $\{\varepsilon_t\}$ est un bruit blanc.

Le modèle $ARMAX(p,s,q)$ est ainsi une extension du modèle $ARMA(p,q)$ acceptant aussi des variables exogènes x_1, \dots, x_s pouvant influencer le modèle. Les méthodes d'estimation pour ces modèles linéaires sont élaborées par plusieurs auteurs, notamment Shumway et Stoffer (2010) et Hannan et Deistler (2012).

Les modèles linéaires classiques sont appropriés et suffisants dans plusieurs applications puisqu'ils parviennent généralement à adéquatement représenter les données macroéconomiques ou encore observées à des espaces temporels relativement grands (mois, années, etc). Toutefois, il existe certains cas où les modèles linéaires ne peuvent pas bien représenter le vrai modèle sous-jacent aux données, particulièrement dans le cas où il y a plusieurs régimes. Un régime est par définition la structure dans le temps et l'espace d'un système économique. Ainsi, la présence de plusieurs régimes sous-entend que les données comportent plusieurs structures sous-jacentes dans le temps et l'espace. Cela se voit notamment en économie, où plusieurs types de jeux de données comportent deux régimes ou plus, avec une transition entre les deux. Il convient alors d'utiliser un modèle plus approprié que les modèles linéaires classiques pour ces données, ce qui nous mène à l'étude des modèles non-linéaires pouvant modéliser les séries temporelles.

1.3. Modèles non-linéaires univariés

Comme mentionné à la section précédente, les modèles linéaires comportent des limites pouvant restreindre leur pertinence d'application. Il est ainsi justifié d'étendre l'étude aux modèles non-linéaires. Le modèle autorégressif univarié général est de la forme :

$$y_t = f(\mathbf{z}_t) + \varepsilon_t,$$

où \mathbf{z}_t est un vecteur pouvant inclure les variables retardées y_{t-1}, \dots, y_{t-p} et les variables exogènes x_{1t}, \dots, x_{st} . Dans le cas où la fonction f n'est pas linéaire, le modèle définit un

modèle autorégressif non-linéaire, dans la mesure que y_t s'exprime en fonction de valeurs retardées y_{t-1}, \dots, y_{t-p} , mais que les coefficients n'expliquent pas de manière linéaire la variable réponse. Plusieurs formes de modèles non-linéaires sont explorées dans la littérature, voir par exemple Teräsvirta, Tjostheim et Granger (2010) et De Gooijer (2017). Parmi ceux-ci, le modèle non-linéaire le plus élémentaire est le modèle autorégressif à seuil limite, noté TAR pour *Threshold Autoregressive*, élaboré par Tsay (1989). Il est défini dans la définition 1.3.1.

Définition 1.3.1. *Le modèle autorégressif à seuil limite, noté $TAR(p)$, satisfait l'équation :*

$$y_t = (\phi_{1,0} + \phi_{1,1}y_{t-1} + \dots + \phi_{1,p}y_{t-p})\mathbb{I}(s_t \leq c) + (\phi_{2,0} + \phi_{2,1}y_{t-1} + \dots + \phi_{2,p}y_{t-p})\mathbb{I}(s_t > c) + \varepsilon_t, \quad (1.3.1)$$

où $\phi_{1,i}, \phi_{2,i}, i \in \{0, \dots, p\}$ sont des constantes; \mathbb{I} dénote la fonction indicatrice; $\{\varepsilon_t\}$ est un bruit blanc; et c est le paramètre de seuil. La variable s_t est la variable qui permet de déterminer dans quel régime le modèle se trouve, et prend généralement une valeur autorégressive de forme y_{t-d} . La fonction indicatrice \mathbb{I} est, par définition, une fonction prenant la valeur de un lorsque sa condition est satisfaite, et de zéro sinon.

La forme la plus simple du modèle autorégressif à seuil limite, soit avec $p = 1$, $\phi_{1,0} = 0$, $\phi_{2,0} = 0$ et $d = 1$ soit $s_t = y_{t-1}$, prend ainsi la forme :

$$y_t = \phi_1 y_{t-1} \mathbb{I}(y_{t-1} \leq c) + \phi_2 y_{t-1} \mathbb{I}(y_{t-1} > c) + \varepsilon_t.$$

Ici, la variable de transition s_t prend la valeur de y_{t-1} . Dans ce cas-ci, le modèle TAR génère une valeur y_t à partir d'un modèle autorégressif linéaire différent selon la valeur de y_{t-1} . Chaque modèle autorégressif représente un régime, et le changement de régime se fait en fonction de la valeur du seuil c . Le modèle TAR est également généralisable à plusieurs cas, dont le cas avec présence de plus de deux régimes. L'estimation des modèles TAR est élaborée par Tsay (1989). Celui-ci suggère de sélectionner les seuils à l'aide de diagrammes de dispersion des valeurs, et de par la suite compléter l'estimation des paramètres par la méthode des moindres carrés ordinaires.

La principale limite du modèle TAR est que la transition entre les deux états ou régimes se produit instantanément à la valeur de seuil c . Par conséquent, si le changement entre deux régimes se fait graduellement dans le temps plutôt qu'instantanément lorsque la valeur

c est atteinte, le modèle TAR ne parvient pas à bien représenter cette transition graduelle entre les deux régimes.

L'un des modèles non-linéaires pouvant répondre à cette faiblesse du modèle TAR est le modèle autorégressif à transition lisse, noté STAR pour *Smooth Transition Autoregressive*, élaboré par Teräsvirta (1994) et Dijk, Teräsvirta et Franses (2002).

Ce modèle comprend une fonction de transition $g(s_t, \gamma, c)$, représentant la fonction de transition entre les régimes, qui permet de modéliser la transition graduelle entre ceux-ci. Elle est fonction des paramètres s_t , γ et c et prend des valeurs de 0 à 1. La variable s_t est celle dite de transition, pouvant soit prendre la forme d'une variable endogène (par exemple y_{t-d} , pour une certaine valeur de d), exogène (comme x_{it} , pour une certaine valeur de i , voir la définition 1.2.4), ou encore d'une tendance linéaire (s'exprimant comme une fonction de l'indice temporel t). Les paramètres γ et c caractérisent le comportement de la fonction g selon leurs valeurs et la forme de la fonction g en vigueur dans le modèle. En pratique, la fonction g est généralement de forme exponentielle ou logistique. Les deux définitions qui suivent donnent explicitement les expressions pour ces classes de fonctions.

Définition 1.3.2. Une fonction $g(x, \gamma, c)$ logistique est de la forme :

$$g(x, \gamma, c) = [1 + \exp(-\gamma(x - c))]^{-1},$$

avec $\gamma > 0$ et c une constante réelle.

Définition 1.3.3. Une fonction $g(x, \gamma, c)$ exponentielle est de la forme :

$$g(x, \gamma, c) = 1 - \exp(-\gamma(x - c)^2),$$

avec $\gamma > 0$ et c une constante réelle.

Le modèle STAR(p) est défini dans la définition 1.3.4.

Définition 1.3.4. Dans le cas univarié à deux régimes, le modèle autorégressif à transition lisse, noté STAR(p), satisfait l'équation :

$$\begin{aligned}
y_t &= (\phi_{1,0} + \phi_{1,1}y_{t-1} + \dots + \phi_{1,p}y_{t-p})[1 - g(s_t, \gamma, c)] \\
&\quad + (\phi_{2,0} + \phi_{2,1}y_{t-1} + \dots + \phi_{2,p}y_{t-p})g(s_t, \gamma, c) + \varepsilon_t \\
&= \boldsymbol{\phi}'_1 \mathbf{x}_t [1 - g(s_t, \gamma, c)] + \boldsymbol{\phi}'_2 \mathbf{x}_t [g(s_t, \gamma, c)] + \varepsilon_t,
\end{aligned} \tag{1.3.2}$$

où $\boldsymbol{\phi}_i = (\phi_{i,0}, \phi_{i,1}, \dots, \phi_{i,p})'$, $i=1,2$, sont les vecteurs des paramètres ; $\mathbf{x}_t = (1, y_{t-1}, \dots, y_{t-p})'$ est le vecteur des valeurs autorégressives ; $\{\varepsilon_t\}$ est un bruit blanc ; et la fonction $g(s_t, \gamma, c)$ est une fonction définie dans les définitions 1.3.2 ou 1.3.3.

Lorsque la fonction g est de forme logistique, le paramètre γ détermine le lissage de la transition entre les deux régimes, et le paramètre c le point milieu du changement entre les deux régimes. Lorsque γ est grand, la fonction g se rapproche d'une fonction indicatrice, avec changement de régime ainsi effectué à $s_t = c$. Le modèle STAR(p) tend alors vers le modèle TAR(p) élaboré précédemment. À l'opposé, plus γ est petit, plus la transition entre les deux régimes est lente. Lorsque $\gamma \rightarrow 0$, alors la fonction se rapproche d'une constante, et lorsque $\gamma = 0$ le modèle devient linéaire autorégressif AR(p).

Lorsque la fonction g est de forme exponentielle, la valeur de la fonction g se rapproche d'une constante de valeur 0 ou 1 lorsque $\gamma \rightarrow 0$ ou $\gamma \rightarrow \infty$. Dans ces cas, le modèle devient donc linéaire autorégressif AR(p) puisque c'est alors un seul des deux régimes qui détermine les valeurs de la série. Lorsque $s_t = c$, la fonction g est nulle, et elle tend vers 1 lorsque $s_t \rightarrow \infty$ ou $s_t \rightarrow -\infty$. Ainsi, pour la fonction de forme exponentielle, le changement de régime est associé aux valeurs absolues de s_t relativement à c .

Le modèle STAR(p) avec deux régimes présenté dans la définition 1.3.4 peut être reparamétrisé sous la forme suivante :

$$\begin{aligned}
y_t &= (\phi_{1,0} + \phi_{1,1}y_{t-1} + \dots + \phi_{1,p}y_{t-p}) \\
&\quad + (\psi_{1,0} + \psi_{1,1}y_{t-1} + \dots + \psi_{1,p}y_{t-p})g(s_t, \gamma, c) + \varepsilon_t \\
&= \boldsymbol{\phi}'_1 \mathbf{x}_t + \boldsymbol{\psi}'_1 \mathbf{x}_t g(s_t, \gamma, c) + \varepsilon_t,
\end{aligned} \tag{1.3.3}$$

où $\psi_{1,j} = \phi_{2,j} - \phi_{1,j}$, et donc $\boldsymbol{\psi}_1 = \boldsymbol{\phi}_2 - \boldsymbol{\phi}_1$ avec $\boldsymbol{\phi}_i = (\phi_{i,0}, \phi_{i,1}, \dots, \phi_{i,p})'$ et $\boldsymbol{\psi}_1 = (\psi_{1,0}, \psi_{1,1}, \dots, \psi_{1,p})'$.

Cette reparamétrisation permet de faciliter l'estimation des paramètres du modèle. Cela sera discuté dans le chapitre 2.

Le modèle $\text{STAR}(p)$ peut être étendu afin que le modèle comprenne plus de deux régimes. Dans ce cas, plusieurs fonctions g_i composent le modèle pour $i = 1, \dots, m - 1$, avec m dénotant le nombre total de régimes. Chaque fonction g_i marque la transition entre deux régimes distincts. Autrement dit, la fonction g_i modélise la transition entre les régimes $i - 1$ et i . Le modèle $\text{STAR}(p)$ à m régimes prend ainsi la forme suivante :

$$y_t = \phi'_1 \mathbf{x}_t + \psi'_1 \mathbf{x}_t g_1(s_{1,t}, \gamma_1, c_1) + \dots + \psi'_{m-1} \mathbf{x}_t g_{m-1}(s_{(m-1),t}, \gamma_{m-1}, c_{m-1}) + \varepsilon_t, \quad (1.3.4)$$

où les fonctions g_i , $i = 1, \dots, m - 1$ sont généralement de forme logistique ou exponentielle, comme dans le cas du modèle $\text{STAR}(p)$ avec deux régimes tel que défini plus haut, et représentent la transition lisse entre les régimes $i - 1$ et i . Ce modèle général à m régimes est abordé par Dijk et Franses (1999), entre autres.

1.3.1. Stationnarité pour le modèle STAR univarié

Dans le cadre des modèles non-linéaires, l'hypothèse de stationnarité est, comme dans le cas linéaire, souvent nécessaire à l'analyse des séries. Toutefois, il est beaucoup plus complexe d'étudier la stationnarité dans le cas non-linéaire que dans le cas linéaire. Dans le cas linéaire, il existe des résultats permettant de vérifier aisément les conditions de stationnarité au sens large énoncées dans la définition 1.1.2. Ces conditions sont bien connues et se trouvent dans Brockwell et Davis (2016). Par exemple, dans les modèles $\text{AR}(p)$, il suffit d'étudier les racines du polynôme autorégressif afin de conclure si une solution stationnaire existe. En absence de racines sur le disque unitaire (disque défini dans le plan complexe), il est possible de définir un opérateur autorégressif inverse et formuler un processus $\text{AR}(p)$ comme un processus linéaire. Or, dans le cas non-linéaire et plus particulièrement pour le modèle $\text{STAR}(p)$, démontrer ces conditions se révèle nettement plus compliqué en raison de la présence d'une composante autorégressive dans la fonction de transition g entre les régimes. Il n'y a donc pas un équivalent simple et comparable aux résultats disponibles dans les modèles AR (ou ARMA).

L'élaboration et l'étude des modèles $\text{STAR}(p)$ dans la littérature suppose généralement, explicitement ou implicitement, que la stationnarité est satisfaite, voir par exemple Teräsvirta (1994) ainsi que Dijk, Teräsvirta et Franses (2002). La stationnarité des modèles $\text{STAR}(p)$ est encore un sujet de recherche actif, et a récemment été explorée par Kheifets et

Saikkonen (2018). Ceux-ci ont défini des conditions suffisantes pour démontrer la stationnarité des modèles $\text{STAR}(p)$ multivariés ou vectoriels en utilisant le fait que le modèle est une chaîne de Markov. Nous élaborons ci-bas l'étude de la stationnarité pour le cas où le modèle vectoriel $\text{STAR}(p)$ défini par Kheifets et Saikkonen (2018) comporte seulement une variable et devient ainsi le modèle $\text{STAR}(p)$ à deux régimes présenté à la définition 1.3.4. Pour le cas univarié général à m régimes, nous référons à Kheifets et Saikkonen (2018). Bien que l'étude de la stationnarité d'un $\text{STAR}(p)$ univarié découle des résultats généraux de Kheifets et Saikkonen (2018), au mieux de notre connaissance c'est la première fois que cette question est explicitement abordée dans la littérature. Ce mémoire amène donc ceci comme contribution originale à la théorie des modèles STAR.

Si l'on considère un modèle $\text{STAR}(p)$ avec deux régimes prenant la forme reparamétrisée élaborée en (1.3.3), les matrices \mathbf{C}_1 et \mathbf{C}_2 définies par Kheifets et Saikkonen (2018) sont alors toutes deux de taille $p \times p$ et sont les suivantes :

$$\mathbf{C}_1 = \begin{pmatrix} \phi_{1,1} + \psi_{1,1} & \phi_{1,2} + \psi_{1,2} & \dots & \phi_{1,p-1} + \psi_{1,p-1} & \phi_{1,p} + \psi_{1,p} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

et

$$\mathbf{C}_2 = \begin{pmatrix} \phi_{1,1} & \phi_{1,2} & \dots & \phi_{1,p-1} & \phi_{1,p} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix},$$

où les $\phi_{1,i}$ et $\psi_{1,j}$ sont les paramètres du modèle $\text{STAR}(p)$ reparamétrisé élaboré en (1.3.3).

Kheifets et Saikkonen (2018) montrent que la stationnarité du modèle $\text{STAR}(p)$ peut être étudiée en considérant le rayon spectral conjoint des matrices \mathbf{C}_1 et \mathbf{C}_2 . Le rayon spectral conjoint est défini comme suit :

$$\rho(\{\mathbf{C}_1, \mathbf{C}_2\}) = \limsup_{j \rightarrow \infty} \left\{ \sup_{A \in A^j} \rho(A) \right\}^{1/j},$$

où $A^j = \{A_1 A_2 \dots A_j : A_i \in \{\mathbf{C}_1, \mathbf{C}_2\}, i = 1, \dots, j\}$ et $\rho(A)$ est le rayon spectral de la matrice A , c'est-à-dire sa plus grande valeur propre.

Les résultats disponibles font l'hypothèse que le rayon spectral conjoint est strictement inférieur à un, soit que $\rho(\{\mathbf{C}_1, \mathbf{C}_2\}) < 1$. Toutefois, il n'est pas possible de poser des conditions sur les matrices \mathbf{C}_1 et \mathbf{C}_2 qui garantissent que leur rayon spectral conjoint sera inférieur à un. Selon le Lemme 1 de Kheifets et Saikkonen (2018),

$$\rho(\lambda \mathbf{C}_1 + (1 - \lambda) \mathbf{C}_2) \leq \rho(\{\mathbf{C}_1, \mathbf{C}_2\}),$$

pour $\lambda \in [0, 1]$. Ainsi, en posant $\lambda = 0$ et $\lambda = 1$, on trouve les relations :

$$\rho(\mathbf{C}_1) \leq \rho(\{\mathbf{C}_1, \mathbf{C}_2\}),$$

$$\rho(\mathbf{C}_2) \leq \rho(\{\mathbf{C}_1, \mathbf{C}_2\}),$$

et donc

$$\max\{\rho(\mathbf{C}_1), \rho(\mathbf{C}_2)\} < 1,$$

si le rayon spectral conjoint est inférieur à un. Ainsi, sous l'hypothèse de départ que $\rho(\{\mathbf{C}_1, \mathbf{C}_2\}) < 1$, on trouve que les rayons spectraux de \mathbf{C}_1 et \mathbf{C}_2 sont aussi inférieurs à un. Leur rayon spectral conjoint doit ensuite être approximé afin de vérifier qu'il est bien inférieur à un afin de garantir la stationnarité du modèle $\text{STAR}(p)$. En fait, c'est l'intérêt de s'intéresser au rayon spectral conjoint, car il peut être montré qu'une condition suffisante pour qu'une solution stationnaire au sens large (et même au sens strict) existe est que le rayon spectral conjoint soit inférieur à un.

Nous pouvons élaborer et étudier le cas spécifique où $p = 1$, ce qui rend les matrices \mathbf{C}_1 et \mathbf{C}_2 de dimension 1×1 , donc scalaires, et donc où $\mathbf{C}_1 = \phi_{1,1} + \psi_{1,1} = \phi_{2,1}$ et $\mathbf{C}_2 = \phi_{1,1}$. On a alors $A^j = \{A_1 A_2 \dots A_j : A_i \in \{\phi_{1,1} + \psi_{1,1}, \phi_{1,1}\}, i = 1, \dots, j\}$, ce qui fait que l'ensemble des matrices A^j possible est de la forme :

$$(\phi_{1,1} + \psi_{1,1})^a (\phi_{1,1})^b,$$

avec $a + b = j$, où $a, b \geq 0$. Comme cette matrice est de dimension 1×1 , son rayon spectral correspond à sa valeur propre en valeur absolue, qui est la valeur absolue de la matrice en soi. Par conséquent, le rayon spectral conjoint est la combinaison de $a + b = j \rightarrow \infty$ qui maximise $(\phi_{1,1} + \psi_{1,1})^a (\phi_{1,1})^b$ et se doit d'être inférieur à 1 pour que le processus soit

stationnaire. Ainsi, une condition suffisante pour la stationnarité dans le cas où $p = 1$ est que $|\phi_{1,1}| < 1$ et que $|\phi_{1,1} + \psi_{1,1}| = |\phi_{2,1}| < 1$.

Le cadre théorique des séries chronologiques et plus particulièrement des modèles STAR(p) ainsi que leurs principales caractéristiques étant maintenant définis, nous procédons à l'élaboration des techniques d'estimation pour ceux-ci. En raison de la non-linéarité des modèles STAR(p) et de la présence de la fonction de transition g , l'ajustement et l'estimation de ces modèles comportent des éléments auxquels il est important de porter une attention particulière, notamment les tests de linéarité et les difficultés calculatoires présentes lors de l'estimation des paramètres et lors du calcul des erreurs standards. Le prochain chapitre présente donc les techniques d'estimation pour les modèles STAR(p).

Chapitre 2

Techniques d'estimation

Dans ce chapitre, nous présentons le processus d'ajustement et les techniques d'estimation pour les modèles STAR. L'ajustement d'un modèle STAR(p) à un jeu de données comporte trois étapes : la spécification du modèle, l'estimation de ses paramètres, et l'évaluation du modèle. La spécification du modèle implique de choisir le délai p qui représente le mieux les données et qui sera ainsi utilisé pour la modélisation. Par la suite, il convient de tester le jeu de données pour la linéarité. Une fois la linéarité rejetée, il est alors justifié de procéder à l'ajustement d'un modèle STAR(p) au jeu de données, en commençant par déterminer la variable de transition s_t et la fonction g appropriées pour le modèle, ce qui est relativement simple. Par la suite, on procède à l'estimation des paramètres du modèle. L'estimation des paramètres du modèle STAR(p) est faite à partir d'un algorithme d'optimisation non-linéaire. En pratique, cet algorithme nécessite des valeurs initiales qui sont trouvées par recherche quadrillée sur des ensembles de valeurs possibles pour les paramètres non-linéaires. L'optimisation non-linéaire subséquente permet alors d'obtenir des estimés pour les paramètres ainsi que leur matrice de variances-covariances. Finalement, une fois le modèle STAR(p) déterminé, on procède à l'évaluation de celui-ci. L'évaluation consiste à valider les hypothèses faites lors de l'ajustement du modèle, soit l'absence d'autocorrélation des résidus, l'absence de non-linéarité restante et la constance des paramètres.

2.1. Spécification du modèle

La première étape caractérisant l'ajustement d'un modèle autorégressif à transition lisse est la spécification du modèle. Dans ce mémoire, nous concentrons notre étude sur les modèles STAR(p) comprenant deux régimes seulement comme dans la définition 1.3.4, et nous élaborons l'ajustement des modèles pour ce cas spécifique. Les modèles STAR(p) avec $m > 2$ régimes peuvent également être ajustés, mais ceux-ci semblent peu utilisés dans la littérature en raison de la difficulté supplémentaire entraînée par la présence de régimes additionnels.

Nous élaborons ainsi les étapes de la spécification du modèle pour les STAR(p) comprenant deux régimes. En observant les données, il convient tout d'abord de déterminer le délai p qui représente le mieux le processus. Ceci permet de déterminer la composante autorégressive $\mathbf{y}_1, \dots, \mathbf{y}_p$ devant être considérée lors de l'ajustement du modèle. Ensuite, il convient d'effectuer un test de linéarité pour les données, l'ajustement d'un modèle STAR à celles-ci étant non-justifié si la linéarité des données n'est pas rejetée, puisqu'alors il est plus pertinent d'ajuster simplement un modèle linéaire classique, comme un modèle de type ARMA(p, q) ou même un modèle AR(p) avec un p plus ou moins élevé. Finalement, la spécification du modèle implique de déterminer la variable de transition s_t et la forme de la fonction g les plus appropriées pour les données, ce qui permettra ensuite de procéder à l'estimation des paramètres.

2.1.1. Sélection du délai p

Pour déterminer le délai p le plus approprié pour les données, plusieurs méthodes sont possibles. Teräsvirta (1994) propose ainsi de déterminer le délai p à l'aide du critère d'information d'Akaike, ou AIC, développé par Akaike (1974), et en se basant sur des modèles autorégressifs AR(p) ajustés aux données. Le critère AIC pour un modèle prédéterminé est calculé à partir de la valeur de vraisemblance du modèle ajusté.

Formuler la fonction de vraisemblance exacte dans les modèles ARMA présumant que le terme d'innovation est normal demeure compliqué. La fonction de vraisemblance est fournie dans Brockwell et Davis (2016) et correspond à :

$$L_C(\theta) = \frac{1}{\sqrt{(2\pi)^n v_0 \dots v_{n-1}}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\varepsilon_t)^2 / v_{t-1} \right\},$$

où v_{j-1} , $j = 1, \dots, n$, sont les variances des erreurs déterminées selon l'algorithme des innovations, voir la section 5.2 de Brockwell et Davis (2016) pour plus de détails.

Le critère AIC pour un modèle AR(p) est ainsi calculé en utilisant la vraisemblance de ce modèle AR(p) ajusté aux données et le nombre de délais p inclus dans le modèle :

$$AIC = -2 \log(L) + 2(p + 1), \tag{2.1.1}$$

où L est la valeur de vraisemblance du modèle AR(p) ajusté aux données, et p est le nombre de délais inclus dans le modèle. Ainsi, plusieurs valeurs de p qui semblent intuitives selon les

données sont testées, et la valeur de p du modèle $\text{AR}(p)$ qui minimise la valeur du critère AIC est choisie comme délai p pour le modèle $\text{STAR}(p)$ ajusté.

Il est à noter que si le modèle était $\text{STAR}(p)$ plutôt qu'un modèle $\text{AR}(p)$, la fonction de vraisemblance différerait légèrement. Ainsi, un critère de vraisemblance conditionnelle est considéré par Teräsvirta (1994) pour les modèles STAR, présumant que des valeurs initiales sont disponibles et fixées. Utilisant des résultats généraux sur l'estimation à vraisemblance maximale, si des arguments asymptotiques sont utilisés, alors considérer une vraisemblance exacte ou conditionnelle devrait mener à des résultats similaires si le nombre d'observations est grand. Puisque l'utilisation de modèles non-linéaires est justement associée à de grands échantillons, une vraisemblance conditionnelle devrait être suffisante dans les applications. Pour le modèle donné dans la définition 1.3.4, présumant que $\{\varepsilon_t\}$ est un bruit blanc Gaussien, la fonction de vraisemblance conditionnelle est simplement :

$$L_C(\theta) \equiv L_C(\theta|y_{-p+1}, y_{-p+2}, \dots, y_0) = \prod_{t=1}^T \phi_\varepsilon(\varepsilon_t),$$

où y_{-p+1}, \dots, y_0 sont considérées des valeurs initiales fixées, et

$$\varepsilon_t = y_t - \phi_1' \mathbf{x}_t - \psi_1' \mathbf{x}_t g(s_t, \gamma, c)$$

est défini dans l'expression (1.3.3), et finalement, $\phi_\varepsilon(x)$ correspond à la densité d'une normale, c'est-à-dire $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

En utilisant des arguments élémentaires, la fonction de log-vraisemblance conditionnelle peut s'exprimer comme suit :

$$\log L_c(\theta) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{t=1}^T \{y_t - F(\mathbf{x}_t, \theta)\}^2,$$

où

$$\begin{aligned} F(\mathbf{x}_t, \theta) &= \phi_1' \mathbf{x}_t [1 - g(s_t, \gamma, c)] + \phi_2' \mathbf{x}_t [g(s_t, \gamma, c)], \\ &= \phi_1' \mathbf{x}_t + \psi_1' \mathbf{x}_t [g(s_t, \gamma, c)]. \end{aligned} \tag{2.1.2}$$

Le critère AIC pour un modèle $\text{STAR}(p)$ est par la suite calculé de la même façon que pour un modèle $\text{AR}(p)$, soit avec l'équation (2.1.1). Comme dans la littérature, lorsque nous parlerons de méthode du maximum de vraisemblance, il sera sous-entendu qu'il est question de vraisemblance conditionnelle. Voir aussi Granger et Teräsvirta (1993) et Luukkonen, Saikkonen et Teräsvirta (1988). Toutefois, à cette étape de l'ajustement du modèle, soit pour

la détermination du délai p , seul le critère AIC pour le modèle $AR(p)$ est considéré puisque le modèle $STAR(p)$ n'est pas encore ajusté, et qu'il est donc impossible de calculer sa vraisemblance à cette étape du processus d'ajustement. Ainsi, le choix du délai p s'effectue selon le critère AIC des modèles $AR(p)$.

Suite à la sélection d'un délai p pour les données, il est souhaitable de tester les données pour la linéarité. Cela revient à tester le modèle $STAR(p)$ contre son alternative linéaire, le modèle $AR(p)$ classique.

2.1.2. Tests de linéarité

Avant d'appliquer un modèle non-linéaire $STAR(p)$ aux données, il convient de tester si le modèle sous-jacent dont les données ont été générées est compatible ou non avec l'hypothèse de linéarité. En effet, si la linéarité n'est pas rejetée, il est préférable d'appliquer un modèle linéaire classique aux données plutôt qu'un modèle non-linéaire, en raison de la plus grande simplicité d'estimation de ceux-ci ainsi que de la théorie approfondie développée pour ces modèles. Dans le cadre de ce mémoire, les modèles linéaires considérés sont les modèles $AR(p)$ tel que présentés dans la définition en 1.2.1, équivalents linéaires aux modèles non-linéaires $STAR(p)$ de la définition 1.3.4.

Nous élaborons ici les tests de linéarité pour un modèle $STAR(p)$ comportant deux régimes. Pour ces modèles, un test de linéarité revient à tester l'hypothèse d'égalité des paramètres pour les deux régimes. Autrement dit, si le régime $STAR(p)$ prend la forme suivante :

$$\begin{aligned} y_t &= (\phi_{1,0} + \phi_{1,1}y_{t-1} + \dots + \phi_{1,p}y_{t-p})[1 - g(s_t, \gamma, c)] \\ &\quad + (\phi_{2,0} + \phi_{2,1}y_{t-1} + \dots + \phi_{2,p}y_{t-p})g(s_t, \gamma, c) + \varepsilon_t \\ &= \phi_1' \mathbf{x}_t [1 - g(s_t, \gamma, c)] + \phi_2' \mathbf{x}_t [g(s_t, \gamma, c)] + \varepsilon_t. \end{aligned} \tag{2.1.3}$$

Ainsi, le test de linéarité revient à confronter les hypothèses suivantes :

$$H_0 : \phi_1 = \phi_2,$$

$$H_1 : \phi_1 \neq \phi_2.$$

On remarque que sous H_0 , $\phi_{1,j} = \phi_{2,j}$, pour tout $j \in \{0, \dots, p\}$. D'un autre côté, sous l'hypothèse alternative H_1 , il existe au moins un j pour lequel ce n'est pas le cas, c'est-à-dire que $\phi_{1,j} \neq \phi_{2,j}$ pour au moins un $j \in \{0, \dots, p\}$. Toutefois, on note que l'on fait face à un

problème non-standard, car sous l’hypothèse nulle les paramètres γ et c intervenant dans la fonction g ne sont pas identifiables.

Une autre façon possible de présenter l’hypothèse nulle de linéarité des données est $H'_0 : \gamma = 0$. Cette hypothèse mène également à un modèle linéaire, tel qu’expliqué à la section 1.3, toutefois en utilisant cette hypothèse ce sont alors les paramètres ϕ_1 , ϕ_2 et c qui sont non-identifiés, plutôt que γ et c pour H_0 . Ainsi, en raison de la présence de paramètres de nuisance peu importe l’hypothèse utilisée pour tester la linéarité, il n’est pas possible d’utiliser la théorie statistique conventionnelle pour tester ces hypothèses de linéarité du modèle.

Pour contourner ce problème, Luukkonen, Saikkonen et Teräsvirta (1988) proposent de remplacer la fonction $g(s_t, \gamma, c)$ dans la forme reparamétrisée du modèle STAR(p) (1.3.3) par une approximation par série de Taylor d’ordre un de la fonction autour de $\gamma = 0$. Ainsi, si une fonction $G(z)$ est monotone croissante possédant une dérivée d’ordre un non-nulle dans l’intervalle $(-a, a)$, $a > 0$, alors une expansion de Taylor au premier ordre autour du point $z = 0$ fournit :

$$G(z) \approx G(0) + G'(0)z,$$

où $G'(0)$ dénote la dérivée d’ordre un de la fonction G évaluée en 0.

En appliquant l’argument mais avec $z = \gamma(s_t - c)$, la fonction $g(s_t, \gamma, c)$ est remplacée par son approximation ce qui modifie le modèle étudié initialement. Par la suite, on reparamétrise le modèle tel qu’effectué par Luukkonen, Saikkonen et Teräsvirta (1988) afin d’obtenir une équation dans laquelle les paramètres non-linéaires γ et c sont absents, ce qui enlève le problème des paramètres non-restreints par l’hypothèse nulle.

Si la fonction g est de forme logistique, un développement de Taylor à l’ordre un tel que dans l’argumentation précédente fournit une approximation suffisante afin d’obtenir des procédures puissantes. Ainsi, il ne semble pas nécessaire d’effectuer des développements de Taylor à des ordres supérieurs car les gains ne seraient pas substantiels. Voir Dijk, Teräsvirta et Franses (2002). Le modèle linéaire ajusté devient alors :

$$y_t = \beta'_0 \mathbf{x}_t + \beta'_1 \mathbf{x}_t s_t + e_t, \tag{2.1.4}$$

où $\beta_i = (\beta_{i,0}, \dots, \beta_{i,p})'$ sont les paramètres linéaires ajustés ; \mathbf{x}_t est tel que défini dans le modèle STAR(p) en 1.3.4 ; s_t est la variable de transition ; et e_t joue le rôle du terme d’erreur

de la régression lorsque l'ordre de développement de Taylor est de un. Lorsque la variable s_t correspond à une valeur autorégressive, soit y_{t-d} avec $d \in \{1, \dots, p\}$, le terme $\beta_{1,0}s_t$ doit être retiré de la régression afin d'éviter la multicollinéarité, puisque le terme $\beta_{0,d}y_{t-d}$ est déjà présent dans l'équation. Le choix de la forme de la variable de transition s_t et de la valeur de d correspondante sont intrinsèquement liés au test de linéarité. Nous discutons ces considérations dans la section 2.1.3.

Luukkonen, Saikkonen et Teräsvirta (1988) trouvent qu'un développement de Taylor d'ordre un pour la fonction g n'est pas suffisant en puissance pour le cas où seulement la constante diffère entre les deux régimes, soit lorsque $\phi_{1,0} \neq \phi_{2,0}$ mais que $\phi_{1,j} = \phi_{2,j}$ pour tout $j \in \{1, \dots, p\}$. Pour obtenir une puissance suffisante afin de détecter la non-linéarité dans ces modèles, il est nécessaire d'augmenter l'ordre de développement de Taylor utilisé pour la fonction g . Ainsi, Luukkonen, Saikkonen et Teräsvirta (1988) déterminent qu'utiliser un développement de Taylor d'ordre trois pour la fonction g lorsque seule la constante diffère entre les modèles s'avère suffisant pour détecter la non-linéarité. Dans le cadre de ce mémoire, un ordre de un sera suffisant puisque tous les paramètres différeront entre les régimes pour les données considérées.

Ainsi, dans le cas où la fonction g est de forme logistique, le test de linéarité revient tout d'abord à ajuster un modèle sous l'hypothèse de linéarité, soit une régression de y_t sur \mathbf{x}_t , puis de calculer la somme des résidus au carré pour ce modèle ajusté, ce qui correspond à SSR_0 . Ensuite, on ajuste un modèle avec la fonction g développée à l'ordre un telle qu'élaboré en (2.1.4), puis l'on calcule la somme des résidus au carré pour ce deuxième modèle, ce qui correspond à SSR_1 .

La statistique calculée dépend de la taille échantillonnale de la série temporelle, tel qu'élaboré par Dijk, Teräsvirta et Franses (2002). En effet, sous l'hypothèse nulle de linéarité, le test du multiplicateur de Lagrange est effectué à partir d'une distribution χ^2 asymptotique. Lorsque la taille est suffisamment grande, la statistique de test correspond alors à :

$$LM = T(SSR_0 - SSR_1)/SSR_0, \quad (2.1.5)$$

et cette statistique LM suit approximativement une loi χ^2 à $p + 1$ degrés de liberté. L'hypothèse nulle de linéarité est alors rejetée si $LM > \chi_{0,05;p+1}^2$ lorsque le niveau du test

adopté est de 5%. Cette statistique est celle élaborée et utilisée par la plupart des auteurs, dont Luukkonen, Saikkonen et Teräsvirta (1988).

Toutefois, lorsque la taille échantillonnale de la série est petite, Dijk, Teräsvirta et Franses (2002) recommandent d'effectuer la version F du test du multiplicateur de Lagrange, puisque la variante χ^2 du test peut grandement surestimer l'erreur de type un. La statistique de test correspond dans ce cas à :

$$LM = \frac{(SSR_0 - SSR_1)/3(p+1)}{SSR_1/(T - 4(p+1))}, \quad (2.1.6)$$

et cette statistique suit approximativement une loi F avec $3(p+1)$ et $T - 4(p+1)$ degrés de liberté.

Dans le cadre de ce mémoire, la taille échantillonnale des séries temporelles en cause est suffisamment grande pour que l'utilisation de la version F du test ne soit pas justifiée, ainsi nous utilisons la variante χ^2 du test basée sur la statistique en (2.1.5) et une distribution χ^2 asymptotique. C'est également le cas dans la majeure partie de la littérature, au mieux de nos recherches.

Si la fonction g est plutôt de forme exponentielle, un développement de Taylor d'ordre un pour la fonction est généralement suffisant pour détecter la non-linéarité. Toutefois, dans certains cas un développement d'ordre deux permet de mieux représenter la forme de la fonction exponentielle, voir Dijk, Teräsvirta et Franses (2002) pour plus de détails. Toutefois, peu importe l'ordre du développement utilisé, la puissance du test semble être comparable pour un développement d'ordre un et un d'ordre deux. Le test de linéarité se fait ainsi de la même façon que pour la fonction logistique ci-haut si la fonction g est de forme exponentielle.

Un exemple d'application de ce test de linéarité est élaboré au chapitre 5 pour le modèle STAR(5), lorsque nous appliquons la méthodologie aux données comprenant les rendements quotidiens de l'indice SPDR S&P 500 (SPY).

2.1.3. Sélection de la variable de transition s_t

Comme mentionné à la section 1.3, la variable de transition s_t peut être endogène, exogène ou une fonction de l'indice temporel t . Le choix de la forme de cette variable peut être basé sur la théorie économique de même que sur l'observation des données. Dans le cadre

des modèles $\text{STAR}(p)$ dits purs, nous considérons que la variable s_t sera nécessairement endogène, soit de la forme y_{t-d} avec une valeur de d se devant d'être déterminée. Cette valeur fera toutefois intuitivement partie de l'ensemble $\{1, \dots, p\}$. Si le choix de la valeur de d ne semble pas évident, ce qui est généralement le cas, il est possible de la trouver en procédant par le test de linéarité décrit à la section 2.1.2, tel qu'effectué par Dijk, Teräsvirta et Franses (2002). Ce test est alors effectué pour toutes les valeurs de d vraisemblables, soit souvent avec $d \in \{1, \dots, p\}$. Ensuite, on sélectionne la valeur de d qui mène à la plus petite valeur- p pour le test de linéarité, soit la variable s_t de forme autorégressive qui rejette le plus fortement la linéarité des données.

2.1.4. Sélection de la fonction de transition g

En théorie, le choix de la fonction de transition g telle qu'apparaissant dans l'équation 2.1.3 est complètement laissé à l'utilisateur. Cependant, dans le cadre des modèles $\text{STAR}(p)$ utilisés dans la littérature, il semblerait que les deux options présentées précédemment, à savoir une fonction g dans la classe logistique ou exponentielle, telles que définies en 1.3.2 et 1.3.3 respectivement, représentent les options les plus populaires. Teräsvirta (1994) présente une série de tests d'hypothèses permettant d'effectuer le choix de cette fonction avant de procéder à l'estimation des paramètres.

Toutefois, il est également possible d'ajuster deux modèles $\text{STAR}(p)$ aux données, soit un ajustement avec la fonction de forme exponentielle et l'autre avec la fonction de forme logistique, et de repousser le choix de la forme de la fonction g à la fin de l'ajustement du modèle $\text{STAR}(p)$, soit à l'étape de l'évaluation du modèle. À ce stade, si l'une des fonctions permet d'adéquatement modéliser les données et que les hypothèses sont toutes validées, alors cette fonction peut être considérée comme convenable. Il n'est pas nécessairement obligatoire de considérer l'autre option si le modèle semble approprié avec la première fonction testée. C'est cette stratégie que nous privilégions ici.

2.2. Estimation des paramètres

Suite au test de linéarité et au choix du délai p , de la variable de transition s_t et de la fonction de transition $g(s_t, \gamma, c)$ élaborés à la section 2.1, l'ajustement d'un modèle STAR(p) à un jeu de données nécessite l'estimation des paramètres du modèle. Le modèle STAR(p) à deux régimes présenté à la définition 1.3.4 comporte en effet plusieurs paramètres devant être estimés, soit l'ensemble $\theta = (\phi'_1, \phi'_2, \gamma, c)'$. Cet ensemble contient des paramètres associés à la partie linéaire du modèle, soit ϕ'_1 et ϕ'_2 , et les paramètres γ et c , qui interviennent de façon non-linéaire sur la variable d'intérêt y_t . Ces paramètres sont contenus dans la fonction g , qui sera à toute fin pratique exclusivement non-linéaire.

Comme le modèle comporte une composante non-linéaire, l'estimation des paramètres se fait par la méthode des moindres carrés non-linéaire. Ceci revient à minimiser :

$$Q_T(\theta) = \sum_{t=1}^T \{y_t - F(\mathbf{x}_t, \theta)\}^2, \quad (2.2.1)$$

où $F(\mathbf{x}_t, \theta)$ correspond au squelette du modèle STAR(p) présenté dans la formule (2.1.2). On remarque que la méthode des moindres carrés menant à minimiser $Q_T(\theta)$ dans la formule (2.2.1) est intimement liée au critère de vraisemblance conditionnelle dans la section 2.1.

En théorie, la fonction (2.2.1) définit une fonction à optimiser et l'on recherche le point de l'espace paramétrique qui permet d'atteindre le minimum. Malheureusement, même si la formulation du critère est plutôt simple (un critère de type moindres carrés), la mise en oeuvre pratique avec les routines d'optimisation disponibles est particulièrement ardue. Il faut donc des stratégies particulières afin d'avoir une bonne approximation de l'optimum. En particulier, la deuxième forme du modèle, soit avec ϕ'_1 et ψ'_1 telle que définie par (2.1.2), est utilisée lors de l'optimisation car elle rend l'estimation des paramètres plus aisée. En effet, cette forme implique que la fonction $g(s_t, \gamma, c)$ et donc les paramètres non-linéaires γ et c n'apparaissent qu'une seule fois dans le modèle, ce qui facilite les calculs pour l'obtention de la matrice de variances-covariances élaborée au théorème 2.2.1. Ainsi, l'ensemble des paramètres à minimiser correspond maintenant à la formulation alternative $\theta = (\phi'_1, \psi'_1, \gamma, c)'$.

La minimisation de $Q_T(\theta)$ doit être effectuée avec un algorithme d'optimisation non-linéaire en raison de la présence d'une composante non-linéaire dans le modèle. Plusieurs types d'algorithmes non-linéaires existent. Wooldridge (1994) élabore le théorème 2.2.1 de

convergence asymptotique des estimations $\hat{\theta}$ obtenues par optimisation non-linéaire pour les paramètres. Trois conditions sont nécessaires pour que ce théorème s'applique. Premièrement, l'espérance du modèle doit correspondre à son squelette présenté à l'équation (2.1.2), soit que :

$$E(y_t|\Omega_t) = F(\mathbf{x}_t, \theta_0), \quad (2.2.2)$$

pour un certain paramètre θ_0 dans un espace paramétrique Θ et pour $t \in \{1, 2, \dots\}$, avec Ω_t étant l'historique de la série, soit $\Omega_t = \{y_t, \dots, y_1\}$. Deuxièmement, les conditions de régularité des M -estimateurs, telles qu'élaborées par Wooldridge (1994) et Pötscher et Prucha (1997), doivent être satisfaites. Troisièmement, il ne doit pas y avoir d'autocorrélation entre les erreurs, c'est-à-dire que la condition suivante doit être remplie :

$$E(\varepsilon_t \varepsilon'_{t+j} | \mathbf{x}_t \mathbf{x}_{t+j}) = \mathbf{0}, \quad (2.2.3)$$

pour $j \geq 1$. On présente le résultat général dans le théorème suivant.

Théorème 2.2.1. *Si les conditions (2.2.2), (2.2.3) et les conditions de régularité des M -estimateurs sont satisfaites, les estimations $\hat{\theta}$ obtenues par la minimisation de $Q_T(\theta)$ satisfont :*

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, C),$$

où θ_0 sont les vraies valeurs des paramètres appartenants à l'intérieur de Θ . La matrice C des variances-covariances correspond à $A^{-1}BA^{-1}$, avec :

$$A = \text{plim} \left[-T^{-1} \sum_{t=1}^T E \{ \nabla^2 q_t(\theta_0) \} \right],$$

où $q_T(\theta_0) = \{y_t - F(\mathbf{x}_t; \theta_0)\}^2$. Comme lorsque $\hat{\theta} = \theta_0$ le vecteur des erreurs ε_t a une espérance conditionnelle nulle, la matrice A se réduit à :

$$A = \text{plim} \left[-T^{-1} \sum_{t=1}^T E \{ \nabla F_t(\mathbf{x}_t, \theta_0) \nabla F_t(\mathbf{x}_t, \theta_0)' \} \right].$$

La matrice B correspond à :

$$B = \text{plim} \left[T^{-1} \sum_{t=1}^T E \{ \nabla q_t(\theta_0) \nabla q_t(\theta_0)' \} \right],$$

où $\text{plim}\{X_n\}$ dénote la limite probabiliste de $X_n, n \geq 1$, soit si $\text{plim}\{X_n\} = X$, alors $X_n \rightarrow X$ en probabilité.

La matrice asymptotique des variances-covariances C présentée au théorème 2.2.1 peut être estimée de façon convergente pour le cas spécifique des modèles STAR(p). Dijk, Teräsvirta et Franses (2002) l'estiment ainsi par $\widehat{A}_T^{-1} \widehat{B}_T \widehat{A}_T^{-1}$, avec :

$$\begin{aligned}\widehat{A}_T &= -T^{-1} \sum_{t=1}^T \left\{ \nabla^2 q_t(\hat{\theta}) \right\} = T^{-1} \sum_{t=1}^T \left\{ \nabla F(\mathbf{x}_t; \hat{\theta}) \nabla F(\mathbf{x}_t; \hat{\theta})' - \nabla^2 F(\mathbf{x}_t; \hat{\theta}) \widehat{\varepsilon}_t \right\}, \\ \widehat{B}_T &= T^{-1} \sum_{t=1}^T \left\{ \nabla q_t(\hat{\theta}) \nabla q_t(\hat{\theta})' \right\} = T^{-1} \sum_{t=1}^T \left\{ \widehat{\varepsilon}_t^2 \nabla F(\mathbf{x}_t; \hat{\theta}) \nabla F(\mathbf{x}_t; \hat{\theta})' \right\},\end{aligned}\tag{2.2.4}$$

où $q_t(\hat{\theta}) = \left\{ y_t - F(\mathbf{x}_t; \hat{\theta}) \right\}^2$, et T est la taille de la série temporelle.

Nous développons le calcul de ces matrices pour les modèles STAR(p) à deux régimes présenté à l'équation 1.3.3 afin d'obtenir un estimé des erreurs standards pour le vecteur de paramètres θ . Le calcul de ces matrices passe par l'obtention du gradient $\nabla F(\mathbf{x}_t; \hat{\theta})$ et de la matrice hessienne $\nabla^2 F(\mathbf{x}_t; \hat{\theta})$ du squelette du modèle STAR(p) tel qu'élaboré en 2.1.2. Le gradient du squelette d'un modèle STAR(p) pour un temps $t \in \{1, \dots, T\}$ correspond à :

$$\begin{aligned}\nabla F(\mathbf{x}_t; \hat{\theta}) &= \frac{\partial(F(\mathbf{x}_t; \hat{\theta}))}{\partial \hat{\theta}} \\ &= \left(\frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,0}}, \dots, \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,p}}, \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,0}}, \dots, \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,p}}, \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\gamma}}, \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{c}} \right)'\end{aligned}$$

où

$$\begin{aligned}\frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,0}} &= 1, \quad \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,1}} = y_{t-1}, \dots, \quad \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,p}} = y_{t-p}, \\ \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,0}} &= g(s_t, \hat{\gamma}, \hat{c}), \quad \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,1}} = y_{t-1} g(s_t, \hat{\gamma}, \hat{c}), \dots, \quad \frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,p}} = y_{t-p} g(s_t, \hat{\gamma}, \hat{c}),\end{aligned}$$

et, pour les paramètres nonlinéaires lorsque la fonction g est de forme logistique,

$$\frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\gamma}} = \hat{\psi}'_1 \mathbf{x}_t \frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{\gamma}} = \hat{\psi}'_1 \mathbf{x}_t \frac{(s_t - \hat{c}) \exp(-\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^2},$$

et

$$\frac{\partial F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{c}} = \hat{\psi}'_1 \mathbf{x}_t \frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{c}} = \hat{\psi}'_1 \mathbf{x}_t \frac{-\hat{\gamma} \exp(-\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^2}.$$

De façon similaire, la matrice hessienne du squelette du modèle STAR(p) correspond à :

$$\nabla^2 F(\mathbf{x}_t; \hat{\theta}) = \begin{pmatrix} \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,0}^2} & \cdots & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,0} \partial \hat{\psi}_{1,p}} & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,0} \partial \hat{\gamma}} & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\phi}_{1,0} \partial \hat{c}} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,p} \partial \hat{\phi}_{1,0}} & \cdots & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,p}^2} & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,p} \partial \hat{\gamma}} & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,p} \partial \hat{c}} \\ \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\gamma} \partial \hat{\phi}_{1,0}} & \cdots & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\gamma} \partial \hat{\psi}_{1,p}} & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\gamma}^2} & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\gamma} \partial \hat{c}} \\ \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{c} \partial \hat{\phi}_{1,0}} & \cdots & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{c} \partial \hat{\psi}_{1,p}} & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{c} \partial \hat{\gamma}} & \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{c}^2} \end{pmatrix}.$$

Cette matrice est symétrique puisque sous nos conditions :

$$\frac{\partial^2 F}{\partial a \partial b} = \frac{\partial^2 F}{\partial b \partial a}.$$

De plus, toutes les dérivées partielles de deuxième ordre dans cette matrice impliquant de dériver uniquement sur des paramètres linéaires sont nulles, c'est-à-dire toutes les dérivées secondes selon une combinaison des paramètres dans les ensembles ϕ_1 et ψ_1 . De plus, toutes les dérivées de deuxième ordre impliquant une dérivée par rapport à $\phi_{1,0}, \dots, \phi_{1,p}$, peu importe le deuxième paramètre de dérivation, sont nulles. Il y a ainsi $2(p+1) + 4$ valeurs non-nulles se devant d'être calculées dans la matrice hessienne : $2(p+1)$ valeurs pour les dérivées selon une combinaison de ψ_1 et d'un des deux paramètres non-linéaires, et 4 valeurs pour les dérivées selon deux paramètres linéaires, soit selon une combinaison de γ et c . Les éléments de la matrice sont donc :

$$\begin{aligned} \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,0} \partial \hat{\gamma}} &= \frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{\gamma}}, \\ \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,1} \partial \hat{\gamma}} &= y_{t-1} \frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{\gamma}}, \\ &\dots \\ \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,p} \partial \hat{\gamma}} &= y_{t-p} \frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{\gamma}}, \end{aligned} \tag{2.2.5}$$

avec

$$\frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{\gamma}} = \frac{(s_t - \hat{c}) \exp(-\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^2},$$

lorsque la fonction g est de forme logistique. De plus,

$$\begin{aligned}\frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,0} \partial \hat{c}} &= \frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{c}}, \\ \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,1} \partial \hat{c}} &= y_{t-1} \frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{c}}, \\ &\dots \\ \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\psi}_{1,p} \partial \hat{c}} &= y_{t-p} \frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{c}},\end{aligned}\tag{2.2.6}$$

avec

$$\frac{\partial g(s_t, \hat{\gamma}, \hat{c})}{\partial \hat{c}} = \frac{-\hat{\gamma} \exp(-\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^2},$$

lorsque la fonction g est de forme logistique. Finalement, les dérivées secondes impliquant deux paramètres non-linéaires sont également non-nulles :

$$\begin{aligned}\frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{\gamma}^2} &= \hat{\psi}'_1 \mathbf{x}_t \left(\frac{-(s_t - \hat{c})^2 \exp(-\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^2} + \frac{2(s_t - \hat{c})^2 \exp(-2\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^3} \right), \\ \frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{c}^2} &= \hat{\psi}'_1 \mathbf{x}_t \left(\frac{2\hat{\gamma}^2 \exp(-2\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^3} - \frac{\hat{\gamma}^2 \exp(-\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^2} \right),\end{aligned}\tag{2.2.7}$$

et

$$\frac{\partial^2 F(\mathbf{x}_t; \hat{\theta})}{\partial \hat{c} \partial \hat{\gamma}} = \hat{\psi}'_1 \mathbf{x}_t \left(\frac{(\hat{\gamma}(s_t - \hat{c}) - 1) \exp(-\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^2} - \frac{2\hat{\gamma}(s_t - \hat{c}) \exp(-2\hat{\gamma}(s_t - \hat{c}))}{[1 + \exp(-\hat{\gamma}(s_t - \hat{c}))]^3} \right).$$

Les calculs présentés ici s'appliquent lorsque la fonction g est de forme logistique.

Par la suite, le résidu est obtenu à partir du modèle STAR(p) ajusté, soit $\hat{\varepsilon}_t = y_t - F(\mathbf{x}_t; \hat{\theta})$, pour $t = 1, \dots, T$. À partir de ces valeurs calculées pour les données selon le modèle ajusté, les matrices \widehat{A}_T et \widehat{B}_T , et par conséquent l'estimé de la matrice C du théorème 2.2.1, sont aisément obtenues. Ceci permet d'obtenir une estimation de la variance des paramètres du modèle STAR(p) ajusté, et ainsi d'observer si ceux-ci sont significatifs ou non.

2.2.1. Problèmes numériques lors de l'estimation

Plusieurs problèmes numériques peuvent se présenter lors de l'estimation des paramètres d'un modèle $\text{STAR}(p)$ par optimisation non-linéaire. Ces problèmes sont élaborés par Teräsvirta (1994) et Dijk, Teräsvirta et Franses (2002). Ces problèmes sont également survenus lors de nos expériences empiriques, compliquant la mise en oeuvre d'études complètes de simulations.

Tout d'abord, il peut être difficile d'obtenir une bonne estimation du paramètre γ lorsque sa vraie valeur est élevée. En effet, si γ est élevé, le modèle $\text{STAR}(p)$ tend alors vers un modèle où la fonction g est indicatrice et où le changement entre les deux régimes se fait alors instantanément à $s_t = c$, tel que discuté à la section 1.3. Par conséquent, il est difficile d'obtenir une estimation précise de γ puisque γ détermine le niveau de lissage de la transition, et que la transition n'est pas suffisamment lisse pour ce faire.

De plus, si le modèle ajusté a une valeur de $\hat{\gamma}$ très élevée, les dérivées de la fonction g selon $\hat{\gamma}$ et \hat{c} utilisées pour le gradient peuvent avoir des valeurs grandement variables, puisqu'alors elles sont généralement très proches de 0 à l'exception de quelques points où elles sont très élevées. Ceci peut causer des problèmes lors de l'estimation de la matrice C intervenant dans le théorème 2.2.1, puisqu'alors on est en présence d'une matrice presque singulière pour la matrice A qui sera inversée. Ainsi, il est important de porter attention à l'estimateur de γ . Si cette valeur est trop élevée, alors le modèle $\text{STAR}(p)$ n'est peut-être pas le plus approprié pour les données, un modèle autorégressif à seuil limite de type TAR tel que présenté à la définition 1.3.1 pouvant être mieux puisque celui-ci impose la transition instantanée entre les deux régimes.

Un autre problème pouvant se présenter lors de l'estimation des paramètres d'un modèle $\text{STAR}(p)$ est que l'algorithme d'optimisation non-linéaire utilisé ne parvienne pas à converger et échoue donc à fournir les estimateurs des paramètres, ou alors qu'il ne parvienne pas à identifier des valeurs crédibles pour l'un des modèles $\text{AR}(p)$ formant le modèle $\text{STAR}(p)$. Ceci se produit lorsque la fonction g présentée aux définitions 1.3.2 et 1.3.3 est presque constante dans les données, soit qu'elle est constamment proche de 0 ou de 1 dans le jeu de données disponible. Par exemple, si la valeur de la fonction g est toujours entre 0 et 0,01 pour toutes les données disponibles à l'exception de quelques valeurs, il sera difficile pour l'algorithme de converger à des estimations de paramètres, puisqu'il manque

alors de données pour bien estimer la fonction de transition g et le deuxième modèle $AR(p)$ caractérisant le $STAR(p)$. Autrement dit, les estimations des paramètres linéaires ψ_1 ainsi que les paramètres non-linéaires γ et c du modèle $STAR(p)$ en (1.3.3) seront difficiles ou même impossibles à obtenir, ou alors elles seront peu fiables en raison du faible nombre de valeurs utilisées pour les estimer. Ainsi, lors de l'estimation par optimisation non-linéaire des paramètres, il est important de porter attention aux valeurs prises par la fonction g afin de s'assurer que l'algorithme puisse bien converger et que les estimations de paramètres obtenues sont fiables. Entre autres, observer un graphique de la valeur de la fonction g prise selon le temps permet de bien visualiser si les deux régimes sont suffisamment présents dans les données observées, et si la transition entre les deux est effectivement parfois lisse. Selon le temps t , on souhaite donc observer un graphique où la valeur de la fonction g varie dans le temps avec une bonne présence des valeurs extrêmes et des valeurs entre les deux extrêmes.

Finalement, l'optimisation non-linéaire de la fonction (2.2.1) nécessite en général des valeurs initiales devant être fournies à l'algorithme pour l'ensemble θ des paramètres. Il n'est pas toujours évident de déterminer ces valeurs, et celles-ci peuvent grandement influencer l'estimation des paramètres subséquente, ce qui devient problématique lorsque les valeurs initiales posées sont loin des vraies valeurs des paramètres. Ainsi, il est nécessaire de poser des valeurs initiales suffisamment proches des vraies valeurs afin d'obtenir une estimation adéquate des paramètres. Ces valeurs initiales peuvent être posées à partir du contexte des données et de l'observation de la série, toutefois cela est généralement impossible à effectuer en pratique. Certains paramètres sont plus complexes que d'autres. Dans notre contexte, ce sont les paramètres γ et c qui causent le plus de difficultés. Compte tenu de la nature de ces paramètres, il convient de procéder par recherche quadrillée afin de trouver des valeurs initiales satisfaisantes pouvant être fournies à l'algorithme. Cette stratégie a été adoptée par Dijk, Teräsvirta et Franses (2002), entre autres. C'est également l'approche que nous avons programmé. Nous la détaillons ici. L'estimation par recherche quadrillée consiste à pré-déterminer un espace de valeurs possible pour chacun des paramètres non-linéaires devant être estimés, soit γ et c dans le cas des modèles $STAR(p)$. Ensuite, un ensemble de valeurs est sélectionné dans cet espace, les valeurs sélectionnées ayant un espacement régulier entre elles. Par exemple, si l'on pense que la valeur de c se situe entre 1 et 3, et que l'on prend un espacement de 0,5 entre chaque valeur, l'ensemble de valeurs sélectionné pour le paramètre c sera $\zeta = \{1; 1,5; 2; 2,5; 3\}$. La même logique est alors appliquée pour le paramètre γ .

Les ensembles de valeurs possibles doivent être déterminés manuellement, et doivent se baser sur les données observées ainsi que les hypothèses du modèle $\text{STAR}(p)$. Ainsi, pour le paramètre γ , celui-ci est nécessairement supérieur à 0, tel que posé dans la définition du modèle et de la fonction g aux définitions 1.3.2 et 1.3.3, et par conséquent son ensemble de valeurs possibles sera également entièrement positif. Pour le paramètre c , celui-ci doit se situer dans l'étendue des valeurs de y_t dans la série observée, puisqu'il marque le point milieu du changement entre les deux régimes.

Une fois l'ensemble de valeurs possibles déterminé pour chacun des paramètres, une recherche quadrillée à deux dimensions peut être effectuée. Pour chaque paire (γ, c) des ensembles de valeurs pré-déterminés, l'argument $Q_T(\theta)$ en (2.2.1) est minimisé. Toutefois, comme les valeurs de γ et c sont posées, la minimisation de l'argument est alors simplement une optimisation linéaire pouvant être aisément effectuée par les méthodes conventionnelles. La fonction objective est ainsi minimisée pour chaque combinaison de valeurs pré-déterminées (γ, c) , et c'est la paire de valeurs minimisant le critère d'information d'Akaike, tel que défini précédemment, qui est choisie. Par la suite, les valeurs des paramètres linéaires sélectionnées correspondent à celles obtenues suite à l'optimisation linéaire avec la paire de valeurs sélectionnée pour les paramètres non-linéaires (γ, c) . Les valeurs ainsi obtenues servent alors comme valeurs initiales fournies à l'algorithme d'optimisation non-linéaire pour l'ensemble de paramètres θ .

Une fois la spécification du modèle et l'estimation des paramètres complétées, l'ajustement du modèle $\text{STAR}(p)$ est terminé. Il convient alors de procéder à l'évaluation du modèle. Ceci fait l'objet de la prochaine section.

2.3. Évaluation du modèle

L'évaluation du modèle est la dernière étape caractérisant l'ajustement des modèles $\text{STAR}(p)$. Cette étape consiste à observer le modèle précédemment ajusté et à vérifier les hypothèses sous lesquelles il a été ajusté, tel que proposé par Dijk, Teräsvirta et Franses (2002). Il convient tout d'abord d'examiner les régimes obtenus par l'ajustement du modèle $\text{STAR}(p)$, entre autres les valeurs des paramètres ainsi que les propriétés du squelette $F(\mathbf{x}_t, \hat{\theta})$ obtenu. Ceux-ci doivent être concordants avec le contexte dans lequel s'inscrivent les données. Par ailleurs, comme mentionné précédemment, il est pertinent d'observer les valeurs prises par la fonction g pour l'échantillon disponible afin d'éviter les problèmes d'estimation

mentionnés à la section 2.2.1. Finalement, il est nécessaire de tester les hypothèses faites lors de l'ajustement du modèle $\text{STAR}(p)$, soit celles d'absence d'autocorrélation des résidus, d'absence de non-linéarité restante et de constance des paramètres.

2.3.1. Absence d'autocorrélation des résidus

La première hypothèse testée est celle d'absence d'autocorrélation des résidus obtenus suite à l'ajustement du modèle $\text{STAR}(p)$. Ceci est effectué avec le test de Ljung-Box présenté par Ljung et Box (1978), un test portemanteau fréquemment utilisé dans l'analyse de séries chronologiques. Il permet de tester l'hypothèse nulle d'indépendance des erreurs du modèle ajusté contre l'hypothèse alternative que les erreurs soient corrélées entre elles. La présence d'autocorrélation n'est évidemment pas souhaitable puisque cela signifierait que le modèle ajusté ne capture pas bien le modèle sous-jacent aux données.

Le test de Ljung-Box permet d'observer l'autocorrélation en utilisant les résidus pour des délais prédéterminés. La statistique de test de Ljung-Box correspond à :

$$Q_M = T(T+2) \sum_{k=1}^M \frac{\hat{r}_k^2}{T-k}, \quad (2.3.1)$$

où T est le nombre de résidus, soit le nombre de valeurs disponibles pour la série observée ; M est le nombre de délais pour lequel on regarde l'autocorrélation entre les résidus ; et \hat{r}_k est l'autocorrélation dans les résidus avec un délai $k \in \{1, \dots, M\}$, définie par :

$$\hat{r}_k = \frac{\sum_{t=k+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^T \hat{\varepsilon}_t^2},$$

pour $k \in \{1, 2, \dots\}$. La statistique Q_M en (2.3.1) suit approximativement une loi χ_M^2 , et donc l'hypothèse d'absence d'autocorrélation des résidus est rejetée au niveau 5% si $Q_M > \chi_{0,05;M}^2$.

Il est à noter que les tests de Ljung-Box ont été étudiés en détail dans les modèles ARMA, et l'approximation de la loi par une variable χ^2 est valide précisément dans ces modèles. Par exemple, il est bien connu que dans l'ajustement des modèles $\text{ARMA}(p,q)$, la loi de Q_M est approximativement χ_{M-p-q}^2 . Ceci étant dit, Q_M est une forme quadratique dans les autocorrélations résiduelles, et comme mesure de distance, il demeure pertinent de la calculer pour plusieurs valeurs de M . Cependant, l'approximation par la loi χ^2 doit être utilisée avec précaution.

2.3.2. Absence de non-linéarité restante

Après avoir vérifié si les erreurs sont compatibles avec un bruit blanc, la deuxième hypothèse peut être testée. Celle-ci est que le modèle sous-jacent aux données contient seulement deux régimes et non plus. En effet, l'ajustement d'un modèle STAR commence par un modèle à deux régimes, et l'ajout de régimes supplémentaires dans le modèle doit être testé avant d'être effectué, par exemple si l'on voudrait avoir trois régimes ou plus dans le modèle. Avec le test de linéarité effectué lors de la spécification du modèle à la section 2.1, la linéarité a été testée et rejetée. Le test concluait donc que deux régimes, avec une transition entre les deux, représentait mieux les données qu'un seul régime. Ici, nous validons que la présence de deux régimes dans l'ajustement d'un modèle STAR(p) est suffisant pour bien modéliser les données. Ainsi, nous testons s'il est nécessaire d'inclure un troisième régime afin d'obtenir une représentation adéquate des données. Le test de cette hypothèse est caractérisé par les mêmes difficultés et problématiques rencontrées lors du test de linéarité élaboré à la section 2.1, comme mentionné par [Dijk, Teräsvirta et Franses \(2002\)](#). Il est donc nécessaire de procéder de nouveau à une estimation de la fonction $g_2(s_t, \gamma, c)$ par une approximation par série de Taylor, de la même manière que pour le test de linéarité.

Tel que défini à l'équation (1.3.4), le modèle STAR(p) à trois régimes est de la forme :

$$y_t = \phi_1' \mathbf{x}_t + \psi_1' \mathbf{x}_t g_1(s_{1,t}, \gamma_1, c_1) + \psi_2' \mathbf{x}_t g_2(s_{2,t}, \gamma_2, c_2) + \varepsilon_t.$$

La fonction $g_1(s_{1,t}, \gamma_1, c_1)$ est conservée telle quelle dans le test puisqu'elle est présente dans les deux modèles considérés et qu'elle n'a pas besoin d'être estimée vu que ceci a été fait lors de l'ajustement du modèle. Ainsi, avec un développement de Taylor de la fonction $g_2(s_{2,t}, \gamma_2, c_2)$ au troisième ordre tel que suggéré par [Dijk, Teräsvirta et Franses \(2002\)](#), en prenant $s_{2,t} = s_{1,t} = s_t$ tel que trouvé lors de l'ajustement du modèle, l'approximation devient :

$$y_t = \phi_1' \mathbf{x}_t + \psi_1' \mathbf{x}_t g_1(s_t, \gamma_1, c_1) + \beta_1' \mathbf{x}_t s_t + \beta_2' \mathbf{x}_t s_t^2 + \beta_3' \mathbf{x}_t s_t^3 + e_t. \quad (2.3.2)$$

Comme dans le cadre des tests de linéarité, si s_t est de forme autorégressive, soit $s_t = y_{t-d}$, il est important de retirer de l'équation les termes qui causeraient de la multicollinéarité. La statistique de test est ensuite calculée de manière similaire au test de linéarité, soit :

$$LM = T(SSR_0 - SSR_1)/SSR_0, \quad (2.3.3)$$

avec SSR_0 étant la somme des résidus au carré du modèle $STAR(p)$ à deux régimes ajusté, et SSR_1 étant la somme des résidus au carré du modèle présenté à l'équation (2.3.2). Cette statistique de test suit approximativement une loi χ^2 avec $3(p+1)$ degrés de liberté. Si l'hypothèse nulle de deux régimes est rejetée, alors cela signifie qu'il y a présence de non-linéarité restante. Ainsi, le modèle $STAR(p)$ à deux régimes représenterait moins bien les données que le modèle $STAR(p)$ à trois régimes. Il semble intuitivement clair qu'un grand nombre de données sera nécessaire afin de détecter la présence de plus de deux régimes, voir Dijk et Franses (1999).

2.3.3. Constance des paramètres

La dernière hypothèse se devant d'être testée dans le cadre de l'évaluation du modèle est celle de la constance des paramètres. En effet, il se doit d'être vérifié que les paramètres du modèle sont constants dans le modèle plutôt que variables et dépendants du temps t , soit qu'ils changeraient de façon lisse dans le temps. Ceci revient à tester le modèle $STAR(p)$ ajusté contre l'alternative d'un modèle $STAR(p)$ à plusieurs régimes, et avec $s_{2,t} = t$. Plus de détails sont disponibles dans Dijk, Teräsvirta et Franses (2002). L'approximation par série de Taylor qui permet de tester l'hypothèse est :

$$\begin{aligned}
y_t = & \phi_1' \mathbf{x}_t + \psi_1' \mathbf{x}_t g_1(s_{1,t}, \gamma_1, c_1) \\
& + \beta_1' \mathbf{x}_t t + \beta_2' \mathbf{x}_t t^2 + \beta_3' \mathbf{x}_t t^3 \\
& + (\beta_4' \mathbf{x}_t t + \beta_5' \mathbf{x}_t t^2 + \beta_6' \mathbf{x}_t t^3) g_1(s_{1,t}, \gamma_1, c_1) + e_t.
\end{aligned} \tag{2.3.4}$$

Comme toujours, la statistique de test correspond à celle présentée en (2.3.3), mais ici avec SSR_0 étant la somme des résidus au carré du modèle $STAR(p)$ à deux régimes ajusté, et SSR_1 étant la somme des résidus au carré du modèle présenté à l'équation (2.3.4). Cette statistique suit approximativement une loi χ^2 avec $6(p+1)$ degrés de liberté. Si l'hypothèse nulle de constance des paramètres est rejetée, alors le modèle le plus approprié pour les données en est un avec des paramètres changeants de façon lisse dans le temps plutôt que des paramètres constants ou fixes comme dans le modèle $STAR(p)$.

Des problèmes calculatoires peuvent se présenter lorsque la série observée possède un grand nombre d'observations T . Voir le chapitre 5 pour plus de détails.

Chapitre 3

Prévisions ponctuelles et intervalles de prévision

Dans ce chapitre, nous élaborons la théorie concernant les prévisions de valeurs futures dans le cadre des séries chronologiques ainsi que les intervalles de prévision en découlant. Nous définissons tout d'abord ce qu'est une prévision ponctuelle ainsi que les critères d'évaluation pour les prévisions effectuées, soit le biais et l'erreur quadratique moyenne. Ceux-ci permettent de témoigner de la performance des prévisions effectuées empiriquement. Nous présentons ensuite la théorie pour les prévisions pour les modèles autorégressifs linéaires $AR(p)$ et non-linéaires $STAR(p)$ à deux régimes présentés au chapitre 1. Les modèles linéaires classiques $AR(p)$ permettent d'effectuer des prévisions ponctuelles non-biaisées ainsi que d'obtenir des intervalles de prévision symétriques pour celles-ci. Pour les modèles $STAR(p)$, nous élaborons la prévision exacte et non-biaisée au temps $t + 1$, ainsi que les difficultés rencontrées pour l'obtention de prévisions non-biaisées à partir du temps de prévision $t + 2$. Nous présentons ainsi, pour les temps $t + 2$ et $t + 3$, les prévisions directes et exactes, ainsi que leurs difficultés d'application pour les modèles $STAR(p)$. Ensuite, nous présentons des méthodes de prévisions alternatives, soit les prévisions ponctuelles naïves, de Monte Carlo et de bootstrap. Nous élaborons le biais entraîné par les prévisions naïves, et présentons les méthodes de Monte Carlo et bootstrap comme alternatives à la prévision naïve afin d'obtenir une prévision sans biais. Finalement, nous élaborons les intervalles de prévision dans le cadre des modèles autorégressifs linéaires $AR(p)$, ainsi que les intervalles de prévision pouvant être obtenus à l'aide des méthodes de Monte Carlo et bootstrap dans le cadre des modèle $STAR(p)$.

3.1. Généralités

Dans plusieurs domaines d'applications des séries chronologiques, dont l'économie et le domaine de la finance, l'un des intérêts majeurs de l'ajustement d'un modèle à un jeu de données est de permettre la prévision de valeurs futures pour la série. L'ajustement d'un modèle autorégressif à une série chronologique permet de modéliser la structure sous-jacente aux données, ce qui facilite la prévision de valeurs futures pour la série. Pour une série dont

les valeurs jusqu'au temps t sont connues, il est désirable de prédire les valeurs futures que la série prendra aux temps suivants, soit $t + 1, t + 2, \dots$. L'ensemble des valeurs connues de la série, ou l'historique de la série, se définit comme l'ensemble $\Omega_t = \{y_t, \dots, y_1\}$, voir aussi le théorème 2.2.1 ; les valeurs futures sont définies comme y_{t+h} avec $h \geq 1$, et les prévisions ponctuelles de ces valeurs futures sont définies comme $\hat{y}_{t+h|t}$. Ainsi, lorsque le temps t est la dernière valeur disponible pour la série, la prévision ponctuelle optimale selon l'erreur quadratique moyenne de prévision au temps $t + h$, telle que définie par Shumway et Stoffer (2010), entre autres, est l'espérance conditionnelle :

$$\hat{y}_{t+h|t} = E(y_{t+h}|\Omega_t), \quad (3.1.1)$$

soit l'espérance de la vraie valeur y_{t+h} conditionnellement à l'historique de la série $\Omega_t = \{y_t, y_{t-1}, \dots, y_1\}$, connu jusqu'au temps t . L'erreur de prévision est alors définie comme :

$$e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}, \quad (3.1.2)$$

où y_{t+h} est inconnue et correspond à la vraie valeur de la série au temps $t + h$.

Les prévisions $\hat{y}_{t+h|t}$ en (3.1.1) peuvent être obtenues de différentes manières élaborées plus loin dans cette section, et celles-ci possèdent chacune des avantages et des inconvénients. De plus, elles mènent à des prévisions ponctuelles et des intervalles de prévision différents selon la méthode de prévision employée et le temps de la prévision observé. Par conséquent, pour comparer les prévisions obtenues selon les différentes méthodes, nous mesurons la performance de celles-ci selon deux critères, soit le biais et l'erreur quadratique moyenne.

La première méthode utilisée pour évaluer la performance d'une prévision est son biais. Ceci fait l'objet de la définition suivante.

Définition 3.1.1. *Le biais d'une prévision au temps $t + h$ correspond à :*

$$Biais(\hat{y}_{t+h|t}) = E(y_{t+h} - \hat{y}_{t+h|t}),$$

où $\hat{y}_{t+h|t}$ est une prévision au temps $t + h$, et y_{t+h} est la vraie valeur de la série au temps $t + h$.

Puisque y_{t+h} est une variable aléatoire, on remarque que par définition le biais d'une prévision est la valeur moyenne de l'erreur de prévision, où l'erreur de prévision d'horizon h est donnée dans l'expression (3.1.2). Ainsi, une prévision sera sans biais si $E(e_{t+h|t}) = 0$.

La deuxième méthode pour évaluer la performance d'une prévision est son erreur quadratique moyenne, qui fait l'objet de la définition suivante.

Définition 3.1.2. *L'erreur quadratique moyenne d'une prévision au temps $t+h$ correspond à :*

$$EQM(\hat{y}_{t+h|t}) = E \{ (\hat{y}_{t+h|t} - y_{t+h})^2 \},$$

où $\hat{y}_{t+h|t}$ est une prévision au temps $t+h$, et y_{t+h} est la vraie valeur de la série au temps $t+h$.

L'erreur quadratique moyenne présentée à la définition 3.1.2 a comme avantage qu'elle est toujours positive, contrairement au biais présenté à la définition 3.1.1 qui peut être négatif ou positif. Ainsi, le biais moyen d'un ensemble de prévisions peut être nul même si les prévisions sont toutes très éloignées de la moyenne. Par conséquent, il convient d'observer à la fois le biais et l'erreur quadratique moyenne afin d'avoir une bonne idée de la performance des prévisions.

Dans le cadre des modèles $STAR(p)$, il est pertinent d'étudier les prévisions ponctuelles ainsi que leurs intervalles de prévision dans un cadre comparatif. On verra que le calcul des prévisions ponctuelles est déjà compliqué dans le cadre des modèles non-linéaires, et les intervalles de prévision sont encore plus complexes. On note que des formules explicites existent dans les modèles ARMA pour le calcul de l'erreur quadratique moyenne, alors que dans le cas non-linéaire ce n'est pas le cas pour des horizons supérieurs à deux. Les modèles $STAR(p)$ présentés à la définition 1.3.4 étant plus complexes à ajuster, il convient de comparer leurs prévisions avec celles qui seraient obtenues par les modèles linéaires classiques, ceux-ci étant beaucoup plus simples à ajuster et évaluer. Comme les modèles $STAR(p)$ comportent une transition entre deux régimes, nous comparons donc leurs prévisions avec celles qui seraient obtenues par l'ajustement d'un modèle linéaire classique $AR(p)$ de la définition 1.2.1, en observant le biais et l'erreur quadratique moyenne des prévisions effectuées, afin de voir si l'ajustement d'un modèle $STAR(p)$ plutôt qu'un $AR(p)$ mène à une amélioration au niveau de ces deux méthodes d'évaluation des prévisions.

Le biais et l'erreur quadratique moyenne pour les prévisions sont calculés empiriquement à partir de simulations dans le chapitre 4, et à partir de vraies données dans le chapitre 5. Dans les deux cas, les valeurs futures y_{t+h} , $h \in \{1, 2, \dots\}$, sont disponibles et utilisées afin de comparer les prévisions effectuées à l'aide des méthodes de prévisions définies aux sections 3.2, 3.3 et 3.4, mais en pratique ces valeurs futures seraient inconnues.

Finalement, il est important de mentionner que les prévisions pour les modèles STAR(p) présentées dans ce chapitre sont effectuées à partir de l'ensemble de paramètre θ supposé connu. Toutefois, en pratique, les valeurs des paramètres sont inconnues et se doivent d'être estimées lors du processus d'ajustement d'un modèle STAR(p) présenté dans le chapitre 2, ceci menant à l'ensemble de paramètres estimé $\hat{\theta}$. Ces valeurs estimées entraînent donc de l'incertitude additionnelle dans les prévisions effectuées. Cependant, dans un contexte de séries chronologiques, si le modèle est correctement spécifié, il est bien connu que l'estimation aura peu d'impact sur la variance de l'erreur de prévision, particulièrement si la taille de la série est grande. De plus, l'estimation des paramètres en plus du calcul des prévisions sera illustré dans le prochain chapitre.

3.2. Prévisions ponctuelles pour les modèles linéaires

Pour le modèle linéaire classique AR(p) présenté à la définition 1.2.1, les prévisions ponctuelles sont élaborées entre autres par Shumway et Stoffer (2010) et Brockwell et Davis (2016). Dans cette section, on présume que le bruit blanc est fort, ainsi constitué de variables aléatoires indépendantes, de sorte que $E(\varepsilon_{t+h}|\Omega_t) = 0$, pour $h > 0$. La meilleure prévision ponctuelle telle qu'en (3.1.1) pour $h = 1$ est :

$$\begin{aligned}\hat{y}_{t+1|t} &= E(y_{t+1}|\Omega_t), \\ &= E(\phi_1 y_t + \phi_2 y_{t-1} + \dots + \phi_p y_{t-p+1} + \varepsilon_{t+1}|\Omega_t), \\ &= \phi_1 y_t + \phi_2 y_{t-1} + \dots + \phi_p y_{t-p+1},\end{aligned}\tag{3.2.1}$$

puisque ε_t est un bruit blanc fort, et que $E(\varepsilon_t) = 0$ tel qu'élaboré à la définition 1.1.3. La prévision ponctuelle est donc sans biais. De façon similaire, la prévision ponctuelle sans biais pour le temps $t + 2$ se base sur la prévision $\hat{y}_{t+1|t}$ obtenue pour le temps $t + 1$ en (3.2.1) :

$$\begin{aligned}\hat{y}_{t+2|t} &= E(y_{t+2}|\Omega_t), \\ &= E(\phi_1 y_{t+1} + \phi_2 y_t + \dots + \phi_p y_{t-p+2} + \varepsilon_{t+2}|\Omega_t), \\ &= \phi_1 \hat{y}_{t+1|t} + \phi_2 y_t + \dots + \phi_p y_{t-p+2}.\end{aligned}\tag{3.2.2}$$

Pour trouver la prévision ponctuelle optimale au temps $t + h$, il suffit donc de calculer celle-ci à partir des prévisions ponctuelles des temps passés, soit $\hat{y}_{t+h-1|t}, \dots, \hat{y}_{t+h-p|t}$. Ainsi, la prévision au temps $t + 3$ se base sur celles des temps $t + 2$ et $t + 1$ en (3.2.2) et (3.2.1) :

$$\begin{aligned}\hat{y}_{t+3|t} &= E(y_{t+3}|\Omega_t), \\ &= E(\phi_1 y_{t+2} + \phi_2 y_{t+1} + \phi_3 y_t + \dots + \phi_p y_{t-p+3} + \varepsilon_{t+3}|\Omega_t), \\ &= \phi_1 \hat{y}_{t+2|t} + \phi_2 \hat{y}_{t+1|t} + \phi_3 y_t + \dots + \phi_p y_{t-p+3}.\end{aligned}\tag{3.2.3}$$

Il est par la suite aisé d'obtenir les prévisions ponctuelles de façon récursive pour une série en se basant sur les prévisions obtenues aux temps précédents. Toutes les prévisions ponctuelles ainsi trouvées avec le modèle linéaire $AR(p)$ sont sans biais. Les intervalles de prévision associés sont élaborés à la section 3.4.

3.3. Prévisions ponctuelles pour les modèles $STAR(p)$

Le modèle $STAR(p)$ à deux régimes de la définition 1.3.4 prend la forme :

$$\begin{aligned} y_t &= \phi'_1 \mathbf{x}_t [1 - g(s_t, \gamma, c)] + \phi'_2 \mathbf{x}_t [g(s_t, \gamma, c)] + \varepsilon_t, \\ &= \phi'_1 \mathbf{x}_t + \psi'_1 \mathbf{x}_t g(s_t, \gamma, c) + \varepsilon_t, \\ &= F(\mathbf{x}_t; \theta) + \varepsilon_t, \end{aligned} \tag{3.3.1}$$

où $F(\mathbf{x}_t; \theta)$ est le squelette du modèle, avec $\mathbf{x}_t = (1, y_{t-1}, \dots, y_{t-p})'$; et θ est l'ensemble des paramètres du modèle, tel que défini au chapitre 2.

Lin et Granger (1994) et Teräsvirta, Tjøstheim et Granger (2010) élaborent la prévision générale au temps $t + 1$ pour les modèles autorégressifs non-linéaires, et Umer, Sevil et Sevil (2018) mentionnent cette prévision dans le cadre des modèles $STAR(p)$. Nous développons ici cette prévision dans le cadre spécifique des modèles $STAR(p)$ ayant comme variable de transition $s_t = y_{t-1}$. La meilleure prévision ponctuelle pour le temps $t + 1$, soit $h = 1$, prend alors la forme :

$$\begin{aligned} \hat{y}_{t+1|t} &= E(y_{t+1} | \Omega_t), \\ &= E(\phi'_1 \mathbf{x}_{t+1} + \psi'_1 \mathbf{x}_{t+1} g(y_t, \gamma, c) + \varepsilon_{t+1} | \Omega_t), \\ &= E(\phi'_1 \mathbf{x}_{t+1} + \psi'_1 \mathbf{x}_{t+1} g(y_t, \gamma, c) | \Omega_t) + E(\varepsilon_{t+1} | \Omega_t), \end{aligned} \tag{3.3.2}$$

où $\mathbf{x}_{t+1} = (1, y_t, y_{t-1}, \dots, y_{t-p+1})'$. Si la fonction de transition g est de forme logistique, on aura alors $g(y_t, \gamma, c) = [1 + \exp(-\gamma(y_t - c))]^{-1}$. Comme toutes ces valeurs font partie de l'ensemble Ω_t et sont donc connues, leur espérance conditionnelle l'est évidemment aussi. La prévision $\hat{y}_{t+1|t}$ dépend donc uniquement de $E(\varepsilon_{t+1} | \Omega_t)$. Comme $\{\varepsilon_t\}$ est un bruit blanc fort, $E(\varepsilon_{t+1} | \Omega_t) = 0$ et donc la prévision ponctuelle au temps $t + 1$ est sans biais et correspond à :

$$\hat{y}_{t+1|t} = \phi'_1 \mathbf{x}_{t+1} + \psi'_1 \mathbf{x}_{t+1} g(y_t, \gamma, c), \tag{3.3.3}$$

ce qui se calcule directement puisque toutes les valeurs sont connues, conditionnellement à Ω_t .

De manière similaire, lorsque la variable de transition s_t prend la forme y_{t-d} avec $d > 1$, la prévision ponctuelle optimale au temps $t + h$ est directement calculable et sans biais pour $h \in \{1, \dots, d\}$. Il n'y a donc pas de nouvelles difficultés, même dans ce cas.

Le calcul des prévisions ponctuelles pour des horizons supérieurs à un pour les modèles STAR(p) se complexifie lorsque la valeur utilisée comme variable de transition est inconnue et se doit tout d'abord d'être estimée. En effet, la présence d'une composante autorégressive dont la valeur est inconnue comme variable de transition dans la fonction non-linéaire g pose problème au niveau de la prévision ponctuelle. Par exemple, si le modèle a $s_t = y_{t-1}$ comme variable de transition, la prévision ponctuelle pour le temps $t + 2$ sera difficile à effectuer puisque la valeur de la variable de transition est alors $s_t = y_{t+1}$, valeur inconnue et qui se doit donc d'être estimée. L'estimation de cette valeur et son utilisation subséquente dans la fonction non-linéaire g rend la prévision au temps $t + 2$ problématique. En effet, si le modèle prend la forme :

$$y_t = \phi_1' \mathbf{x}_t + \psi_1' \mathbf{x}_t g(y_{t-1}, \gamma, c) + \varepsilon_t,$$

la prévision au temps $t + 2$ correspond à :

$$\hat{y}_{t+2|t} = E(y_{t+2} | \Omega_t).$$

Ceci est difficilement calculable lorsque l'erreur ε_{t+1} est inconnue, et par conséquent la valeur y_{t+1} et la fonction g non-linéaire aussi. Le même problème se présente pour les prévisions ponctuelles des temps qui suivent. Par exemple, pour la prévision au temps $t + 3$, celle-ci prend la forme :

$$\hat{y}_{t+3|t} = E(y_{t+3} | \Omega_t),$$

ce qui est également difficile à calculer compte tenu que l'erreur ε_{t+2} et par conséquent la valeur y_{t+2} sont toutes les deux inconnues.

Plusieurs méthodes peuvent donc être utilisées afin de calculer les prévisions aux temps $t + 2$ et $t + 3$, ainsi qu'aux temps suivants. Nous présentons ici cinq méthodes communes pour les modèles autorégressifs non-linéaires, élaborées entre autres par Brown et Mariano (1984), Lin et Granger (1994), Clements et Hendry (1998), Teräsvirta, Tjostheim et Granger (2010), Teräsvirta, Dijk et Medeiros (2005) et Franses, Dijk et Opschoor (2014). Nous mentionnons leurs avantages et inconvénients et les élaborons pour les prévisions aux temps $t + 2$ et $t + 3$ dans le cadre des modèles STAR(p) avec variable de transition endogène et $d = 1$, soit $s_t = y_{t-1}$. Ce choix de variable de transition permet de témoigner ou non de la présence d'un biais dès le temps $t + 2$, comme mentionné précédemment. Le tout

est facilement généralisable aux cas $s_t = y_{t-d}$ avec $d \in \{1, \dots, p\}$. On remarque que les sous-sections qui suivent n'abordent que les prévisions ponctuelles.

3.3.1. Prévisions directes

La prévision directe, élaborée entre autres par Teräsvirta, Tjostheim et Granger (2010) et Lin et Granger (1994), consiste à déterminer une fonction permettant de prédire $\hat{y}_{t+2|t}$ directement à partir de Ω_t , soit sans passer par la prévision $\hat{y}_{t+1|t}$. Ceci revient donc à étudier directement la relation entre y_{t+2} et y_t . Ainsi, un nouveau squelette $F_2(\mathbf{x}_t; \theta)$ pour le modèle doit être ajusté, et un nouvel ensemble de paramètres doit être trouvé pour ce modèle. Toutefois, l'ajustement d'un tel nouveau modèle implique que l'erreur n'est alors pas nécessairement un bruit blanc, ce qui rend l'estimation des paramètres du modèle et le calcul de la prévision difficiles. De plus, si l'on souhaite obtenir les prévisions pour les temps suivants, soit $t + 3$ par exemple, la modélisation doit être refaite en entier. La relation entre y_{t+3} et y_t est alors étudiée et un nouveau modèle $F_3(\mathbf{x}_t; \theta)$ est ajusté. Par conséquent, la prévision directe n'est pas utilisée en pratique.

3.3.2. Prévisions exactes

Les prévisions exactes, élaborées entre autres par Franses, Dijk et Opschoor (2014) et Teräsvirta, Dijk et Medeiros (2005), consistent à calculer la prévision ponctuelle en intégrant le squelette du modèle pour y_{t+2} sur la fonction de l'erreur ε_{t+1} :

$$\hat{y}_{t+2|t}^{(exacte)} = \int F(\hat{\mathbf{x}}_{t+2}; \theta) dD(\varepsilon_{t+1}), \quad (3.3.4)$$

où $\hat{\mathbf{x}}_{t+2} = (1, \hat{y}_{t+1|t} + \varepsilon_{t+1}, y_t, \dots, y_{t-p+1})'$ est, comme toujours, de dimension $(p + 1) \times 1$; et $D(\varepsilon_{t+1})$ est la distribution de l'erreur ε_{t+1} . Autrement dit, la prévision ponctuelle optimale est l'espérance conditionnelle, et les prévisions exactes s'attaquent aux calculs des intégrales en cause.

Deux difficultés se présentent lors de l'application des prévisions exactes. Tout d'abord, la distribution de l'erreur doit être connue puisque l'intégrale se fait sur celle-ci. Si elle est inconnue, elle doit être assumée, auquel cas l'hypothèse doit être testée. Par exemple, il peut être supposé que $\varepsilon_t \sim \mathcal{N}(0, \Omega_w)$, auquel cas la valeur de Ω_w est inconnue et doit être estimée à l'aide des résidus du modèle ajusté. Ensuite, une fois la distribution de ε_t connue, l'intégrale doit généralement être calculée numériquement en raison de sa complexité.

À partir du temps $t + 3$, les prévisions exactes se complexifient en raison de l'ajout d'une intégrale à chaque augmentation d'un temps pour la prévision. Par exemple, la prévision ponctuelle pour $t + 3$ implique l'intégrale double suivante :

$$\hat{y}_{t+3|t}^{(exacte)} = \int \int F(\hat{\mathbf{x}}_{t+3}; \theta) dD(\varepsilon_{t+1}) dD(\varepsilon_{t+2}), \quad (3.3.5)$$

où $\hat{\mathbf{x}}_{t+3} = (1, F(\hat{\mathbf{x}}_{t+2}; \theta) + \varepsilon_{t+2}, \hat{y}_{t+1|t} + \varepsilon_{t+1}, \dots, y_{t-p+2})'$. Ainsi, le calcul d'une prévision au temps $t + 10$ impliquerait neuf intégrales, ce qui est très complexe à calculer numériquement. Pour cette raison, les prévisions exactes ne sont pas fréquemment utilisées. Cependant, si l'horizon de prévision est petit (si l'on veut considérer $h \leq 3$, par exemple), cela demeure envisageable et nous considérons de telles prévisions dans le prochain chapitre.

3.3.3. Prévisions naïves

Les prévisions naïves sont les plus simples à effectuer en pratique et sont élaborées par plusieurs auteurs dont Clements et Hendry (1998) et Lin et Granger (1994). Elles impliquent de poser l'erreur comme étant nulle pour chaque temps de prévision, soit $\varepsilon_{t+i} = 0$ pour $i \in \{1, 2, \dots\}$, et de se baser sur les prévisions précédentes pour effectuer les prévisions aux temps suivants. Pour la prévision au temps $t + 2$, on pose donc $\varepsilon_{t+1} = 0$, et la prévision devient ainsi :

$$\hat{y}_{t+2|t}^{(naive)} = \phi_1' \hat{\mathbf{x}}_{t+2} + \psi_1' \hat{\mathbf{x}}_{t+2} g(\hat{y}_{t+1|t}, \gamma, c), \quad (3.3.6)$$

où $\hat{\mathbf{x}}_{t+2} = (1, \hat{y}_{t+1|t}, y_t, \dots, y_{t-p+1})'$; et $\hat{y}_{t+1|t}$ est la prévision naïve à $t + 1$, qui correspond à la prévision à $t + 1$ non-biaisée élaborée en (3.3.3).

Malgré sa facilité de calcul, la prévision en (3.3.6) n'est pas idéale puisqu'elle est biaisée pour y_{t+2} . En effet, si l'on prend la définition du biais présentée à la définition 3.1.1,

$$\begin{aligned} \text{Biais} \left(\hat{y}_{t+2|t}^{(naive)} \right) &= E \left\{ y_{t+2} - \hat{y}_{t+2|t}^{(naive)} \right\}, \\ &= E \left\{ y_{t+2} - F(\hat{\mathbf{x}}_{t+2}; \theta) \right\}, \\ &= E \left\{ \phi_1' \mathbf{x}_{t+2} + \psi_1' \mathbf{x}_{t+2} g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c) \right. \\ &\quad \left. - (\phi_1' \hat{\mathbf{x}}_{t+2} + \psi_1' \hat{\mathbf{x}}_{t+2} g(\hat{y}_{t+1|t}, \gamma, c)) \right\}, \\ &= E \left\{ \phi_1' \mathbf{x}_{t+2} - \phi_1' \hat{\mathbf{x}}_{t+2} \right\} \\ &\quad - E \left\{ \psi_1' \hat{\mathbf{x}}_{t+2} g(\hat{y}_{t+1|t}, \gamma, c) - \psi_1' \mathbf{x}_{t+2} g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c) \right\}. \end{aligned} \quad (3.3.7)$$

Le premier terme de cette équation est nul, puisque $\mathbf{x}_{t+2} = (1, y_{t+1}, y_t, \dots, y_{t-p+2})' = (1, \hat{y}_{t+1|t} + \varepsilon_{t+1}, y_t, \dots, y_{t-p+2})' = \hat{\mathbf{x}}_{t+2} + (0, \varepsilon_{t+1}, 0, \dots, 0)'$ et que $E\{\phi_{1,1}\varepsilon_{t+1}\} = 0$. Ainsi, le biais se réduit à :

$$\begin{aligned} & E \left\{ \boldsymbol{\psi}'_1 \hat{\mathbf{x}}_{t+2} g(\hat{y}_{t+1|t}, \gamma, c) - \boldsymbol{\psi}'_1 \mathbf{x}_{t+2} g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c) \right\} \\ &= E \left\{ \boldsymbol{\psi}'_1 \hat{\mathbf{x}}_{t+2} (g(\hat{y}_{t+1|t}, \gamma, c) - g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c)) - \psi_{1,1} \varepsilon_{t+1} g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c) \right\}, \quad (3.3.8) \\ &= \boldsymbol{\psi}'_1 \hat{\mathbf{x}}_{t+2} E \left\{ g(\hat{y}_{t+1|t}, \gamma, c) - g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c) \right\} - \psi_{1,1} E \left\{ \varepsilon_{t+1} g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c) \right\}. \end{aligned}$$

Si la fonction g est de forme logistique telle qu'à la définition 1.3.2, les deux espérances de cette équation ne sont pas nulles, ce qui entraîne un biais. En effet,

$$\begin{aligned} E \{g(y_{t+1}, \gamma, c)\} &= E \{g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c)\}, \\ &= E \left\{ \left[1 + \exp(-\gamma(\hat{y}_{t+1|t} + \varepsilon_{t+1} - c)) \right]^{-1} \right\}, \\ &= E \left\{ \left[1 + \exp(-\gamma(\hat{y}_{t+1|t} - c)) \exp(-\gamma\varepsilon_{t+1}) \right]^{-1} \right\}, \quad (3.3.9) \\ &\neq E \left\{ \left[1 + \exp(-\gamma(\hat{y}_{t+1|t} - c)) \right]^{-1} \right\} = E \{g(\hat{y}_{t+1|t}, \gamma, c)\}, \end{aligned}$$

et, de manière similaire, $E\{\varepsilon_{t+1}g(\hat{y}_{t+1|t} + \varepsilon_{t+1}, \gamma, c)\}$ n'est pas nul. La prévision $\hat{y}_{t+2|t}^{(naive)}$ en (3.3.6) est donc biaisée, et par conséquent son utilisation pour les prévisions des modèles STAR(p) entraîne un biais systématique qui devient de plus en plus important lorsque le temps de prévision augmente.

Pour la prévision au temps $t + 3$, celle-ci prend la forme :

$$\hat{y}_{t+3|t}^{(naive)} = \boldsymbol{\phi}'_1 \hat{\mathbf{x}}_{t+3} + \boldsymbol{\psi}'_1 \hat{\mathbf{x}}_{t+3} g(\hat{y}_{t+3|t}, \gamma, c), \quad (3.3.10)$$

où $\hat{\mathbf{x}}_{t+3} = (1, \hat{y}_{t+2|t}^{(naive)}, \hat{y}_{t+1|t}, y_t, \dots, y_{t-p+2})'$; et $\hat{y}_{t+2|t}^{(naive)}$ est la prévision naïve biaisée trouvée pour le temps $t+2$ en (3.3.6). Puisque la prévision se base sur celle biaisée du temps précédent et qu'elle est elle-même calculée avec $\varepsilon_{t+2} = 0$, la prévision au temps $t + 3$ est aussi biaisée. Il en sera de même pour toutes les prévisions naïves aux temps qui suivent. Les prévisions naïves sont donc faciles à calculer mais mènent à un biais systématique qui peut grandement fausser les prévisions, particulièrement lorsque les temps de prévisions sont éloignés. Elles ne sont donc malheureusement pas recommandées en pratique. Les problèmes de biais seront illustrés dans nos études de simulation présentées dans le prochain chapitre.

3.3.4. Prévisions de Monte Carlo

Élaborées par Franses, Dijk et Opschoor (2014), Brown et Mariano (1984) et Teräsvirta, Dijk et Medeiros (2005), entre autres, les prévisions de Monte Carlo permettent de calculer une prévision ponctuelle non-biaisée en simulant un grand nombre d'erreurs provenant d'une loi prédéterminée D . Si l'erreur du modèle est assumée bruit blanc Gaussien, la loi D sera $\mathcal{N}(0, \Omega_w)$. Un nombre N d'erreurs provenant de cette loi est donc simulé, et la moyenne de ces simulations correspond à la prévision ponctuelle au temps $t + 2$:

$$\begin{aligned}\hat{y}_{t+2|t}^{(MC)} &= N^{-1} \sum_{i=1}^N F(\hat{\mathbf{x}}_{t+2,i}; \theta), \\ &= N^{-1} \sum_{i=1}^N \{ \phi_1' \hat{\mathbf{x}}_{t+2,i} + \psi_1' \hat{\mathbf{x}}_{t+2,i} g(\hat{y}_{t+1|t} + \varepsilon_i, \gamma, c) \},\end{aligned}\tag{3.3.11}$$

où $\hat{\mathbf{x}}_{t+2,i} = (1, \hat{y}_{t+1|t} + \varepsilon_i, y_t, \dots, y_{t-p+1})'$, avec ε_i , $i \in \{1, \dots, N\}$, étant la variable aléatoire simulée de la loi D . Comme dans le cas des prévisions exactes, la forme de la fonction de distribution D doit être assumée, et sa variance Ω_w doit ainsi être estimée à partir des erreurs du modèle STAR(p) ajusté aux données.

De façon similaire, la prévision de Monte Carlo au temps $t + 3$ correspond à :

$$\begin{aligned}\hat{y}_{t+3|t}^{(MC)} &= N^{-1} \sum_{i=1}^N F(\hat{\mathbf{x}}_{t+3,i}; \theta), \\ &= N^{-1} \sum_{i=1}^N \{ \phi_1' \hat{\mathbf{x}}_{t+3,i} + \psi_1' \hat{\mathbf{x}}_{t+3,i} g(F(\hat{\mathbf{x}}_{t+2,i}; \theta) + \varepsilon_{2,i}, \gamma, c) \},\end{aligned}\tag{3.3.12}$$

où $\hat{\mathbf{x}}_{t+3,i} = (1, F(\hat{\mathbf{x}}_{t+2,i}; \theta) + \varepsilon_{2,i}, \hat{y}_{t+1|t} + \varepsilon_{1,i}, y_t, \dots, y_{t-p+2})'$ avec $\hat{\mathbf{x}}_{t+2,i} = (1, \hat{y}_{t+1|t} + \varepsilon_{1,i}, y_t, \dots, y_{t-p+1})'$. Les erreurs $\varepsilon_{1,i}$ et $\varepsilon_{2,i}$, $i \in \{1, \dots, N\}$, sont simulées d'une distribution bivariée et sont indépendantes l'une de l'autre. Chacune suit donc respectivement la fonction de distribution D avec variance Ω_w . Cette variance est encore estimée en se basant sur les résidus du modèle ajusté. Le même principe est appliqué pour les prévisions aux temps supérieurs à $t+3$. Ceci revient donc à faire en quelque sorte un bootstrap paramétrique.

Comme mentionné par Lin et Granger (1994), les prévisions de Monte Carlo requièrent donc d'assumer la distribution des erreurs comme c'est le cas pour les prévisions exactes élaborées à la section 3.3.2. Toutefois, les prévisions de Monte Carlo sont plus simples à

calculer, ce qui fait qu'elles sont souvent utilisées pour obtenir des prévisions ponctuelles dans le cadre des séries temporelles non-linéaires.

3.3.5. Prévisions par la méthode de bootstrap

La méthode de bootstrap est une autre façon de calculer une prévision ponctuelle non-biaisée en procédant par simulations et est élaborée par [Brown et Mariano \(1984\)](#), [Teräsvirta, Tjostheim et Granger \(2010\)](#) et [Lin et Granger \(1994\)](#), entre autres. Elle est similaire à la méthode de Monte Carlo présentée à la section 3.3.4 mais se différencie par le choix des erreurs ε_i . En effet, contrairement à la méthode de Monte Carlo, la forme de la distribution des erreurs D n'est pas supposée connue. Dans ce sens, c'est une méthode non-paramétrique. Les erreurs utilisées pour les simulations proviennent plutôt des résidus obtenus lors de l'ajustement du modèle $\text{STAR}(p)$. Ainsi, la prévision ponctuelle par la méthode de bootstrap au temps $t + 2$ est :

$$\begin{aligned}\hat{y}_{t+2|t}^{(BST)} &= N^{-1} \sum_{i=1}^N F(\hat{\mathbf{x}}_{t+2,i}; \theta), \\ &= N^{-1} \sum_{i=1}^N \{ \phi_1' \hat{\mathbf{x}}_{t+2,i} + \psi_1' \hat{\mathbf{x}}_{t+2,i} g(\hat{y}_{t+1|t} + e_i, \gamma, c) \},\end{aligned}\tag{3.3.13}$$

où $\hat{\mathbf{x}}_{t+2,i} = (1, \hat{y}_{t+1|t} + e_i, y_t, \dots, y_{t-p+1})'$, avec $e_i, i \in \{1, \dots, N\}$, étant tiré au hasard parmi l'ensemble des résidus obtenu après l'ajustement du modèle $\text{STAR}(p)$.

La méthode de bootstrap est ainsi intuitive pour la prévision au temps $t + 2$, toutefois cela n'est pas le cas pour les temps suivants. En effet, de façon intuitive la prévision $t + 3$ correspondrait à :

$$\hat{y}_{t+3|t}^{(BST)} = N^{-1} \sum_{i=1}^N F(\hat{\mathbf{x}}_{t+3,i}; \theta),$$

où $\hat{\mathbf{x}}_{t+3,i} = (1, F(\hat{\mathbf{x}}_{t+2,i}; \theta) + e_{2,i}, \hat{y}_{t+1|t} + e_{1,i}, y_t, \dots, y_{t-p+2})'$ avec $\hat{\mathbf{x}}_{t+2,i} = (1, \hat{y}_{t+1|t} + e_{1,i}, y_t, \dots, y_{t-p+1})'$; et avec les erreurs $e_{1,i}$ et $e_{2,i}, i \in \{1, \dots, N\}$, tirées de l'ensemble des résidus obtenu lors de l'ajustement du modèle $\text{STAR}(p)$. Toutefois, la prévision ponctuelle au temps $t + 3$ est, par sa définition, conditionnelle aux valeurs de la série observée, soit $\hat{y}_{t+3|t} = E(y_{t+3} | \Omega_t)$. Ainsi, effectuer deux tirages indépendants dans l'ensemble des erreurs pour $e_{1,i}$ et $e_{2,i}$ revient à effectuer une prévision non-conditionnelle, puisque la structure de la série originale ne serait alors pas conservée. Par conséquent, il est important de sélectionner une paire d'erreurs recréant la structure originale de la série, voir [McCullough \(1994\)](#) pour plus de détails. Ceci nous amène à sélectionner les erreurs $e_{1,i}$ et $e_{2,i}$ pour la

méthode de bootstrap de façon à ce qu'elles soient consécutives dans la série résiduelle, soit $(e_{1,i}, e_{2,i}) = (e_k, e_{k+1})$, où k est un temps tiré au hasard dans $\{1, \dots, T-1\}$, tel que mentionné par Teräsvirta, Tjostheim et Granger (2010). Ainsi, la prévision de bootstrap au temps $t+3$ correspond à :

$$\hat{y}_{t+3|t}^{(BST)} = N^{-1} \sum_{i=1}^N F(\hat{\mathbf{x}}_{t+3,i}; \theta), \quad (3.3.14)$$

où $\hat{\mathbf{x}}_{t+3,i}$ correspond à :

$$\hat{\mathbf{x}}_{t+3,i} = (1, F(\hat{\mathbf{x}}_{t+2,i}; \theta) + e_{k+1}, \hat{y}_{t+1|t} + e_k, y_t, \dots, y_{t-p+2})',$$

avec $\hat{\mathbf{x}}_{t+2,i} = (1, \hat{y}_{t+1|t} + e_k, y_t, \dots, y_{t-p+1})'$; et e_k et e_{k+1} sont des résidus subséquents provenant du modèle STAR(p) ajusté. La même logique s'applique pour les prévisions aux temps supérieurs à $t+3$.

En raison de l'utilisation de résidus se suivant dans la série originale comme erreurs dans la prévision de bootstrap, l'hypothèse d'indépendance entre les résidus est maintenant invalide, puisque les résidus utilisés dans le calcul de la prévision sont maintenant dépendants, tel que mentionné par Teräsvirta, Tjostheim et Granger (2010). En effet, le choix du résidu e_{k+1} est une conséquence directe du choix du résidu le précédant, soit e_k . La méthode de prévision de bootstrap n'est donc pas idéale dans le cadre des prévisions pour les temps supérieurs à $t+2$, et on lui préférera généralement la méthode de Monte Carlo lorsque la loi D est connue.

3.4. Intervalles de prévision

Les prévisions ponctuelles élaborées aux sections 3.2 et 3.3 ne permettent pas d'obtenir une idée de la variation de la future valeur de la série, par conséquent il est pertinent de calculer des intervalles de prévision afin d'observer l'étendue que peut prendre la future valeur. Pour les modèles autorégressifs linéaires tels qu'élaborés à la section 1.2, ces intervalles de prévision sont généralement calculés à partir de l'erreur de prévision telle que définie par Shumway et Stoffer (2010). Cette erreur de prévision, dans le cas des modèles AR(p), correspond à l'erreur quadratique moyenne. L'intervalle de prévision 95% pour le modèle AR(p) est ainsi :

$$\hat{y}_{t+h|t} \pm 1,96 \times \{EQM(\hat{y}_{t+h|t})\}^{1/2}, \quad (3.4.1)$$

pour $h \in \{1, 2, \dots\}$. Cet intervalle est symétrique et suppose que le processus et les

prévisions suivent une loi normale. On note en effet qu'il n'y a pas de théorème limite central justifiant une expression comme (3.4.1) : la normalité du processus doit être supposée.

Plus précisément, dans le cas d'un $AR(p)$, on utilise la représentation moyenne mobile du processus :

$$\begin{aligned}\phi(B)y_t &= \varepsilon_t, \\ y_t &= a(B)\varepsilon_t,\end{aligned}\tag{3.4.2}$$

où

$$a(B) = \phi^{-1}(B) = \sum_{j \geq 0} a_j B^j,$$

de sorte que $y_{t+h} = a(B)\varepsilon_{t+h}$ et l'erreur de prévision d'horizon h prend la forme :

$$e_{t+h} = \sum_{j=0}^{h-1} a_j \varepsilon_{t+h-j},$$

et

$$EQM(\hat{y}_{t+h|t}) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} a_j^2.$$

Ainsi, l'intervalle de prévision reposant sur l'hypothèse de normalité est de la forme :

$$\hat{y}_{t+h|t} \pm 1,96\sigma_\varepsilon c_h,$$

où

$$c_h = \left\{ \sum_{j=0}^{h-1} a_j^2 \right\}^{1/2}.$$

L'intervalle (3.4.1) pour les prévisions du modèle $AR(p)$ ne s'étend pas directement aux modèles non-linéaires élaborés à la section 1.3. En effet, dans le cadre des modèles linéaires, l'intervalle est symétrique autour de la prévision ponctuelle, ce qui est une conséquence directe de la normalité assumée des prévisions. Voir Franses, Dijk et Opschoor (2014). Toutefois, dans le cadre des modèles non-linéaires, un intervalle de prévision symétrique autour de la prévision ponctuelle n'est pas nécessairement approprié puisque les prévisions peuvent avoir une distribution asymétrique en raison de la non-linéarité du modèle.

Franses, Dijk et Opschoor (2014) présentent ainsi trois méthodes pour construire un intervalle de prévision pour les modèles non-linéaires généraux. Bien que les méthodes présentées soient en apparence complètes, peu est dit sur la mise en oeuvre précise de chacune

des méthodes. La première méthode consiste à définir un intervalle symétrique autour de la prévision ponctuelle, tel que cet intervalle couvre 95% de la densité. La deuxième méthode proposée est de sélectionner l'intervalle de prévision à partir des quantiles empiriques de la distribution obtenue. La troisième méthode consiste à sélectionner la région de la distribution ayant la plus haute densité. Dans le cadre des modèles linéaires, ces trois méthodes mènent à des intervalles identiques, contrairement aux modèles non-linéaires où les intervalles obtenus selon les trois méthodes peuvent grandement varier.

Dans le cadre de ce mémoire, nous présentons les intervalles de prévision à 95% à partir de la méthode des quantiles, soit la deuxième méthode présentée par Franses, Dijk et Opschoor (2014). Ainsi, le calcul des intervalles requiert une distribution empirique pour les prévisions afin d'en obtenir les quantiles. Par conséquent, les intervalles de prévision peuvent être calculés à partir des méthodes de Monte Carlo et de bootstrap uniquement telles que présentées à la section 3.3, puisque ce sont les seules méthodes élaborées pour les modèles STAR(p) menant à des distributions empiriques de prévisions ponctuelles.

Pour calculer l'intervalle de prévision par la méthode de Monte Carlo, il faut tout d'abord obtenir une distribution empirique pour la densité des prévisions obtenues par Monte Carlo. Pour ce faire, il est nécessaire d'ajouter une erreur ε_i aux prévisions calculées afin que l'intervalle de prévision obtenu soit pour la future valeur au temps $t+h$ et non pour la prévision au temps $t+h$. En théorie, si le modèle STAR est stationnaire, il devrait exister une représentation moyenne mobile et on pourrait considérer que l'erreur de prévision d'horizon h est telle que :

$$e_{t+h} \sim \mathcal{L}(0, \sigma_\varepsilon^2 c_h^2),$$

où $\mathcal{L}(0, \sigma_\varepsilon^2 c_h^2)$ désigne la loi de e_{t+h} avec variance $V(e_{t+h}) = \sigma_\varepsilon^2 c_h^2$. La constante c_h devrait augmenter en fonction de h , afin de refléter une incertitude à mesure que l'horizon augmente. Ainsi, pour la prévision au temps $t+2$, la distribution sera calculée à partir des prévisions obtenues par Monte Carlo tel que :

$$\hat{y}_{t+2|t}^{(MC)} = N^{-1} \sum_{i=1}^N F(\hat{\mathbf{x}}_{t+2,i}; \theta) + \varepsilon_i^{(2)}, \quad (3.4.3)$$

avec $\hat{\mathbf{x}}_{t+2,i} = (1, \hat{y}_{t+1|t} + \varepsilon_i^{(1)}, y_t, \dots, y_{t-p+1})'$, et $\varepsilon_i^{(2)} \sim \mathcal{L}(0, \sigma_\varepsilon^2 c_2^2)$. Ainsi, les intervalles de prévision calculés par la méthode de Monte Carlo correspondent à :

$$IP = [q_{\alpha/2}, q_{1-\alpha/2}], \quad (3.4.4)$$

où q_i correspond au quantile empirique de la distribution des prévisions obtenue selon la méthode de Monte Carlo. Dans le cas 95%, $\alpha = 0,05$.

Pour calculer l'intervalle par la méthode de bootstrap, la méthode est similaire à celle pour Monte Carlo, sauf qu'il est nécessaire de faire attention à la sélection des résidus étant subséquents dans la série résiduelle, comme c'est le cas pour les prévisions ponctuelles tel que mentionné à la section 3.3.5. Il est à noter que des intervalles de prévision bootstrap ont été étudiés dans Thombs et Schucany (1990) pour les $AR(p)$. Cependant, les méthodes s'étendent difficilement pour les processus non-linéaires car ils utilisent la représentation avant et arrière des $AR(p)$ (en anglais, *forward representation*, *backward representation*, respectivement). Des méthodes ont été étudiées récemment dans les modèles linéaires et non-linéaires dans Pan et Politis (2016). Plus précisément, Pan et Politis (2016) discutent comment appliquer le bootstrap mais se concentrent sur les modèles $TAR(p)$ tels que discutés dans la définition 1.3.1, et se limitent à l'horizon un compte tenu des complications numériques impliquées. Ainsi, par exemple, pour la prévision au temps $t + 2$, la distribution sera calculée à partir des prévisions obtenues par la méthode de bootstrap tel que les prévisions correspondent à :

$$\hat{y}_{t+2|t}^{(BST)} = N^{-1} \sum_{k=1}^N F(\hat{\mathbf{x}}_{t+2,k}; \theta) + e_{t+2}^{*(i)}, \quad (3.4.5)$$

où $\hat{\mathbf{x}}_{t+2,k} = (1, \hat{y}_{t+1|t} + e_k^{(i)}, y_t, \dots, y_{t-p+1})'$; et $e_k^{(i)}$ et $e_{k+1}^{(i)}$, $k \in \{1, \dots, T-1\}$, sont des résidus subséquents provenant du modèle $STAR(p)$ ajusté. Comme pour Monte Carlo, l'intervalle de prévision est celui de l'équation (3.4.4).

Avec la méthode bootstrap, il importe de déterminer la distribution bootstrap de l'erreur de prévision à l'horizon deux. L'algorithme 3.1 développé par Pan et Politis (2016), soit l'algorithme dit avant reposant sur les résidus, nécessite le calcul des résidus centrés, et la prévision à l'horizon deux des valeurs futures dans le modèle bootstrap. Basé sur les données observées, il est également possible de générer des observations futures utilisées afin de calculer des erreurs de prévision bootstrap à l'horizon deux. Plus de détails sont fournis dans Pan et Politis (2016).

Pour la méthode de Monte Carlo comme pour celle de bootstrap, le tout est facilement étendu aux temps suivants, ce qui permet d'obtenir des intervalles de prévision aux temps futurs pour les modèles $STAR(p)$.

La théorie élaborée dans ce chapitre est étudiée de façon numérique dans le prochain chapitre. Nous appliquons les méthodes de prévision étudiées dans le présent chapitre à des simulations Monte Carlo de modèles STAR(1) et STAR(2). Plus particulièrement, nous calculons les prévisions ponctuelles pour les modèles AR(p) élaborées à la section 3.2, en plus des prévisions ponctuelles exactes, naïves, de Monte Carlo et bootstrap pour les modèles STAR(p) élaborées à la section 3.3. Nous étudions ainsi, pour les temps de prévision $t + 1$, $t + 2$ et $t + 3$, les biais et erreurs quadratiques moyennes empiriques obtenus selon ces méthodes. De plus, nous calculons les intervalles de prévision des modèles AR(p) et ceux de Monte Carlo et bootstrap des modèles STAR(p) présentés à la section 3.4 pour ces mêmes temps de prévision, et nous étudions leurs taux d'inclusion ainsi que la longueur moyenne des intervalles obtenus selon ces différentes méthodes.

Chapitre 4

Étude par simulations : estimation et prévisions dans les modèles STAR

Dans les chapitres précédents, nous avons présenté les modèles non-linéaires $STAR(p)$. Nous avons discuté comment les estimer, en élaborant entre autres les difficultés d'estimation caractéristiques de ces modèles, et nous avons considéré l'élaboration d'un système de prévision afin de prédire des valeurs futures de la série ainsi que des intervalles de prévision pour ces valeurs.

Dans ce chapitre, nous effectuons des études empiriques. Nous considérons ainsi des expériences afin d'apprécier les propriétés des estimateurs discutés, ainsi que la qualité de l'approximation asymptotique dans le cadre des modèles non-linéaires STAR. De plus, nous vérifions si les estimateurs de variance des différents paramètres sont près des valeurs asymptotiques. Ceci a un grand intérêt pratique, compte tenu que dans le choix du modèle il est pertinent de vérifier si les coefficients obtenus sont statistiquement significatifs, et la construction d'intervalles de confiance pour les paramètres estimés est une préoccupation constante.

Dans un contexte de séries chronologiques, le calcul des prévisions est également fondamental, par conséquent nous effectuons donc une étude empirique des prévisions ponctuelles. Nous considérons également la construction d'intervalles de prévision. Notre étude inclut les modèles $STAR(1)$ et $STAR(2)$ à deux régimes. Des comparaisons entre les modèles $AR(p)$ et $STAR(p)$ sont effectuées. En ce qui a trait aux prévisions ponctuelles, nous comparons les biais des prévisions et les erreurs quadratiques moyennes lorsqu'un modèle $STAR(p)$ est correctement ajusté, plutôt qu'un modèle linéaire classique de type AR, sachant que le vrai modèle sous-jacent est un modèle non-linéaire $STAR(p)$. Pour les intervalles de prévision, nous étudions les taux de couverture des prévisions et la longueur moyenne des intervalles obtenus selon différentes méthodes, ainsi que la façon dont ils se comparent aux modèles

de type AR. La mise en oeuvre de nos expériences empiriques est abordée plus précisément dans la section suivante.

4.1. Description des simulations

Dans nos expériences empiriques, nous nous limitons aux modèles STAR(1) et STAR(2). Nous décrivons dans cette section les particularités des modèles étudiés. Dans cette étude empirique, le délai p , la forme de la fonction g et de la variable de transition s_t tels que décrits à la section 2.1 sont supposés connus lors de l'ajustement du modèle STAR(p) à la série, et la non-linéarité est assumée. Nous n'effectuons donc que l'estimation des paramètres et de leurs erreurs standards, en plus des prévisions ponctuelles et intervalles de prévision.

De plus, nous avons considéré des cas de modèles où la persistance est modérée ou faible. Ces choix ont été dictés par des raisons d'ordre pratique, nécessaires dans le cadre de cette étude par simulations. En effet, dans les cas de modèles où la persistance était plus élevée, nous avons obtenu des problèmes assez sévères au niveau de la convergence des algorithmes, ce qui fausse l'interprétation des résultats.

4.1.1. STAR(1)

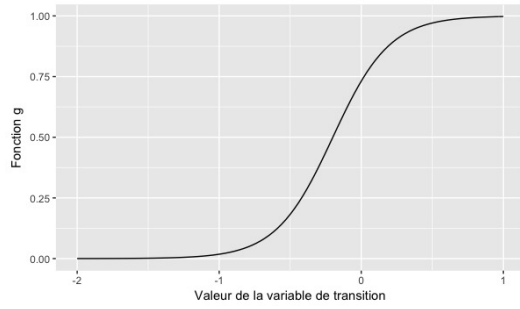
Nous simulons tout d'abord des modèles STAR(1) prenant la forme :

$$y_t = (\phi_{1,0} + \phi_{1,1}y_{t-1}) + (\psi_{1,0} + \psi_{1,1}y_{t-1}) [1 + \exp(-\gamma(y_{t-1} - c))]^{-1} + \varepsilon_t. \quad (4.1.1)$$

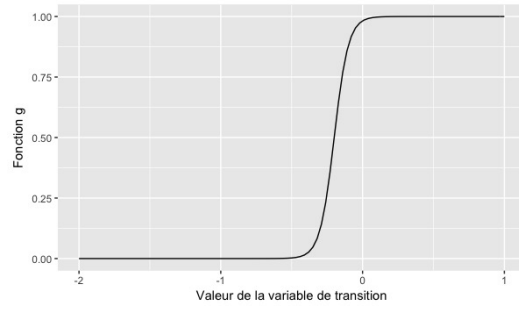
Le modèle est donc basé sur une fonction g de forme logistique. Les valeurs utilisées pour les paramètres de la composante linéaire du modèle sont $(\phi_{1,0}, \phi_{1,1}, \psi_{1,0}, \psi_{1,1})' = (-0,6; 0,5; 0,9; -0,2)'$. Nous étudions le modèle pour différentes valeurs de γ et c , soit $\gamma = 5$ et $\gamma = 20$, et $c = -0,2$ et $c = -0,5$. Lorsque $\gamma = 5$, les valeurs utilisées pour sa recherche quadrillée vont de 1 à 19 avec des écarts de 2 entre chaque, et lorsque $\gamma = 20$ les valeurs vont de 10 à 30 avec des écarts de 2 entre chaque. Pour le paramètre c , lorsqu'il vaut $-0,2$ les valeurs utilisées pour la recherche quadrillée vont de $-0,5$ à $0,2$ avec un écart de $0,1$ entre chaque, et lorsque c vaut $-0,5$ les valeurs vont de -1 à 0 avec un écart de $0,1$.

Comme nous utilisons différentes valeurs pour les paramètres non-linéaires, c'est la forme de la fonction g et son point milieu qui changent selon les différentes valeurs de ces paramètres du modèle STAR(1) observé. Ceci permet d'étudier les répercussions des variations de ces paramètres dans les séries obtenues et les modèles subséquemment ajustés. Les différentes formes prises par la fonction g selon les valeurs des paramètres γ et c sont montrées à la

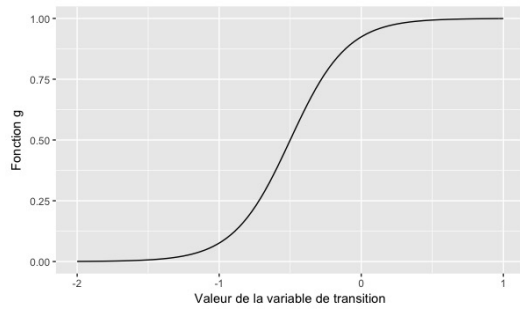
figure 4.1. Nous montrons également un exemple d'une série type pour ces différentes valeurs de paramètres non-linéaires à la figure 4.2.



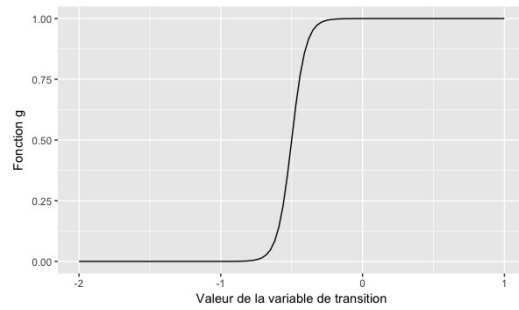
(a) $\gamma = 5, c = -0,2$



(b) $\gamma = 20, c = -0,2$

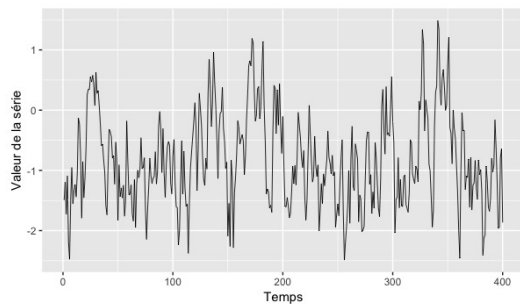


(c) $\gamma = 5, c = -0,5$

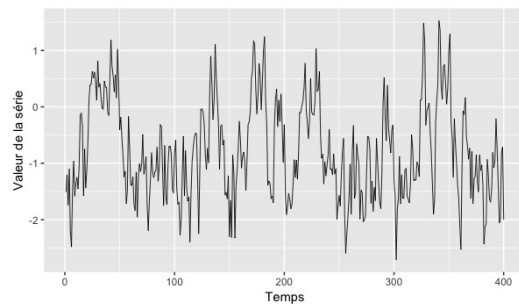


(d) $\gamma = 20, c = -0,5$

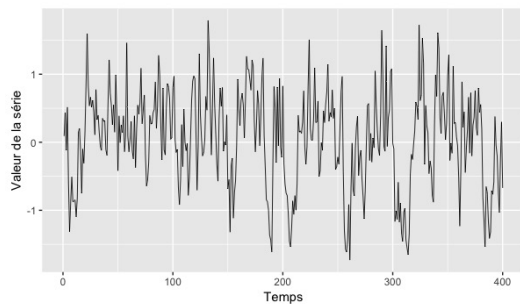
FIGURE 4.1. Forme prise par la fonction de transition selon γ et c .



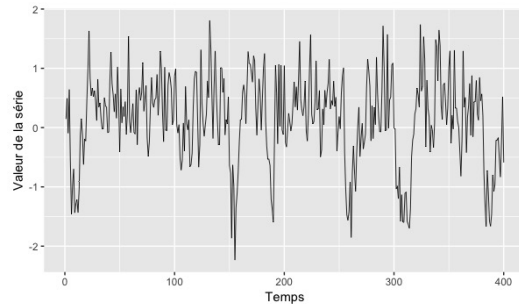
(a) $\gamma = 5, c = -0,2$



(b) $\gamma = 20, c = -0,2$



(c) $\gamma = 5, c = -0,5$



(d) $\gamma = 20, c = -0,5$

FIGURE 4.2. Exemple d'une série STAR(1) selon γ et c , avec $T = 400$.

4.1.2. STAR(2)

Nous simulons ensuite des modèles STAR(2) prenant la forme :

$$y_t = (\phi_{1,0} + \phi_{1,1}y_{t-1} + \phi_{1,2}y_{t-2}) + (\psi_{1,0} + \psi_{1,1}y_{t-1} + \psi_{1,2}y_{t-2}) [1 + \exp(-\gamma(y_{t-1} - c))]^{-1} + \varepsilon_t. \quad (4.1.2)$$

Le modèle est ici aussi basé sur une fonction g de forme logistique. Les valeurs utilisées pour les paramètres linéaires sont $(\phi_{1,0}, \phi_{1,1}, \phi_{1,2}, \psi_{1,0}, \psi_{1,1}, \psi_{1,2})' = (2, 0; -0, 1; -0, 5; -4, 0; 0, 4; 1, 1)'$. Nous étudions le modèle pour différentes valeurs de γ et c , soit pour les ensembles $\gamma = 2$ et $c = 1$, et $\gamma = 20$ et $c = 1, 2$. Lorsque $\gamma = 2$, les valeurs pour sa recherche quadrillée vont de 1 à 10 avec des écarts unitaires, et lorsque $\gamma = 20$ les valeurs vont de 12 à 28 avec des pas de deux. Pour le paramètre c , lorsqu'il vaut 1 les valeurs utilisées pour la recherche quadrillée vont de 0,1 à 3 avec un écart de 0,2 entre chacune des valeurs, et lorsque c vaut 1,2 les valeurs vont de 0,1 à 2 avec un écart de 0,2.

Les différentes formes prises par la fonction g selon les valeurs des paramètres γ et c sont illustrées à la figure 4.3. Nous montrons également un exemple d'une série type pour ces différentes valeurs de paramètres non-linéaires du modèle STAR(2) à la figure 4.4.

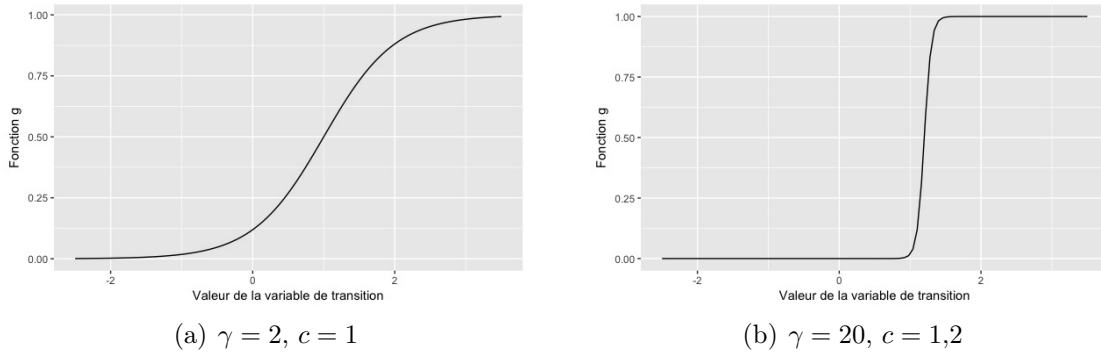


FIGURE 4.3. Forme prise par la fonction de transition selon γ et c .

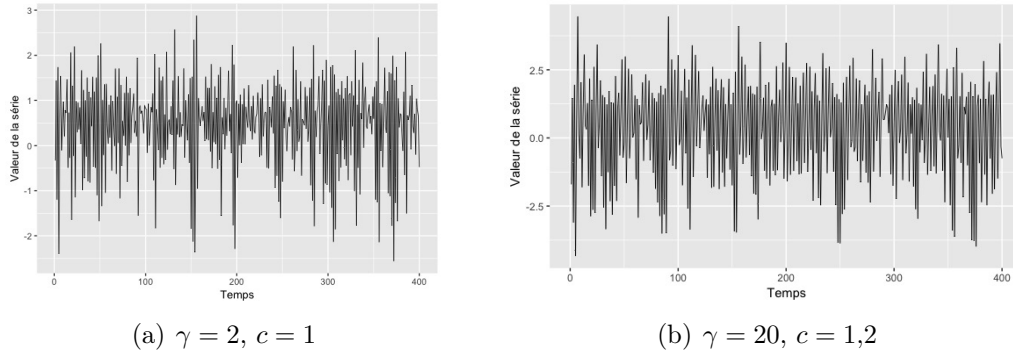


FIGURE 4.4. Exemple d'une série STAR(2) selon γ et c , avec $T = 400$.

4.2. Estimation dans les modèles STAR

Nous effectuons 1000 simulations Monte Carlo de séries pour chaque modèle STAR(p) présenté à la section 4.1. Pour chaque série simulée, nous effectuons l'étape d'estimation des paramètres tel que décrit à la section 2.2, soit d'estimer les paramètres par optimisation non-linéaire et leurs écart-types à partir de l'estimation de la matrice C . Nous obtenons ainsi la moyenne et l'écart-type Monte Carlo des estimations des paramètres, ainsi que la moyenne des écarts-types obtenus selon les estimés de la matrice C , le tout basé sur 1000 simulations de modèles.

Pour tous les modèles simulés, le processus d'erreur $\{\varepsilon_t\}$ est un bruit blanc Gaussien tel qu'à la définition 1.1.3, avec un écart-type de 0,5. L'étude est effectuée pour différentes tailles échantillonales dans les séries, soit $T = 400$, $T = 2000$ et $T = 4000$. Ceci permet d'étudier les estimations des paramètres et de leurs écart-types selon différentes longueurs de séries, particulièrement pour le paramètre non-linéaire γ , et par conséquent d'apprécier les propriétés asymptotiques de l'estimation non-linéaire. Il est important de préciser que le nombre T de valeurs dans la série ne correspond pas au nombre de valeurs simulées dans le code, mais plutôt au nombre de résidus disponibles pour effectuer l'estimation. En effet, nous retirons une centaine de valeurs initiales de la série afin d'éliminer l'effet d'initialisation du modèle, tel que recommandé par Lin et Granger (1994). Il y a également p valeurs utilisées dans l'ajustement pour lesquelles il est impossible de calculer un résidu, les valeurs précédentes dans la série n'étant pas disponibles. Ceci est expliqué par la nature autorégressive du modèle.

Les résultats des estimations de paramètres sont disponibles aux tableaux 4. I, 4. II, 4. III, 4. IV, 4. V et 4. VI.

TABLE 4. I. Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(1) avec $\gamma = 5$, $c = -0,2$.

Paramètre	Vraie valeur T	Moyenne MC			Erreur st. MC			Erreur st. estimée		
		400	2000	4000	400	2000	4000	400 ¹	2000	4000
$\phi_{1,0}$	-0,6	-0,80	-0,61	-0,61	1,37	0,08	0,05	8,74	0,08	0,06
$\phi_{1,1}$	0,5	0,44	0,49	0,50	0,32	0,05	0,03	2,20	0,05	0,03
$\psi_{1,0}$	0,9	10,46	0,93	0,91	220,96	0,16	0,11	156,36	0,17	0,11
$\psi_{1,1}$	-0,2	-9,60	-0,21	-0,21	259,76	0,11	0,07	162,67	0,11	0,07
γ	5,0	260,27	5,30	5,10	2390,49	1,72	0,81	32,98	1,68	0,84
c	-0,2	-0,12	-0,20	-0,20	0,49	0,07	0,05	1,98	0,06	0,04

¹ 55 matrices sur 1000 ont échoué à compléter l'estimation.

TABLE 4. II. Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(1) avec $\gamma = 5$, $c = -0,5$.

Paramètre	Vraie valeur T	Moyenne MC			Erreur st. MC			Erreur st. estimée		
		400	2000	4000	400	2000	4000	400 ¹	2000	4000
$\phi_{1,0}$	-0,6	-2,10	-0,63	-0,62	15,51	0,20	0,13	311,15	0,19	0,13
$\phi_{1,1}$	0,5	-0,05	0,48	0,49	6,23	0,11	0,07	116,81	0,11	0,07
$\psi_{1,0}$	0,9	0,66	0,94	0,92	61,64	0,23	0,15	314,34	0,22	0,14
$\psi_{1,1}$	-0,2	1,28	-0,19	-0,19	31,91	0,11	0,07	117,82	0,10	0,07
γ	5,0	192,58	5,34	5,11	1317,33	2,40	0,84	11,49	14,33	0,81
c	-0,5	-0,56	-0,50	-0,50	0,39	0,07	0,05	1,41	0,06	0,04

¹ 49 matrices sur 1000 ont échoué à compléter l'estimation.

TABLE 4. III. Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(1) avec $\gamma = 20$, $c = -0,2$.

Paramètre	Vraie valeur T	Moyenne MC			Erreur st. MC			Erreur st. estimée		
		400	2000	4000	400	2000	4000	400 ¹	2000 ²	4000
$\phi_{1,0}$	-0,6	-0,62	-0,61	-0,60	0,09	0,04	0,03	0,10	0,04	0,03
$\phi_{1,1}$	0,5	0,49	0,50	0,50	0,07	0,03	0,02	0,07	0,03	0,02
$\psi_{1,0}$	0,9	0,94	0,91	0,90	0,16	0,06	0,04	0,18	0,06	0,04
$\psi_{1,1}$	-0,2	-0,23	-0,20	-0,20	0,18	0,07	0,05	0,17	0,07	0,05
γ	20,0	311,79	24,13	20,83	1198,30	35,77	4,82	36,81	11,21	5,23
c	-0,2	-0,19	-0,20	-0,20	0,05	0,02	0,01	0,05	0,02	0,01

¹ 161 matrices sur 1000 ont échoué à compléter l'estimation.

² 5 matrices sur 1000 ont échoué à compléter l'estimation.

TABLE 4. IV. Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(1) avec $\gamma = 20$, $c = -0,5$.

Paramètre	Vraie valeur T	Moyenne MC			Erreur st. MC			Erreur st. estimée		
		400	2000	4000	400	2000	4000	400 ¹	2000 ²	4000
$\phi_{1,0}$	-0,6	-0,65	-0,61	-0,60	0,24	0,08	0,05	0,24	0,08	0,05
$\phi_{1,1}$	0,5	0,46	0,49	0,50	0,16	0,05	0,04	0,15	0,05	0,04
$\psi_{1,0}$	0,9	0,96	0,91	0,90	0,25	0,08	0,06	0,25	0,08	0,06
$\psi_{1,1}$	-0,2	-0,17	-0,19	-0,20	0,17	0,06	0,04	0,16	0,06	0,04
γ	20,0	354,80	22,44	20,94	2519,04	11,90	4,77	42,37	28,00	5,09
c	-0,5	-0,51	-0,50	-0,50	0,05	0,02	0,01	0,05	0,02	0,01

¹ 118 matrices sur 1000 ont échoué à compléter l'estimation.

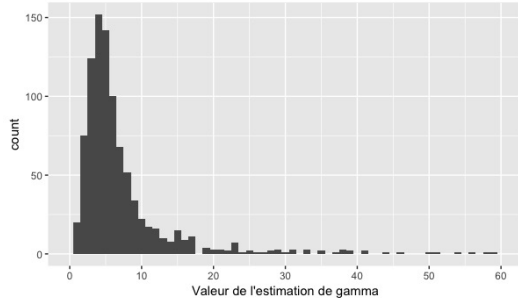
² 1 matrice sur 1000 a échoué à compléter l'estimation.

TABLE 4. V. Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(2) avec $\gamma = 2$, $c = 1$.

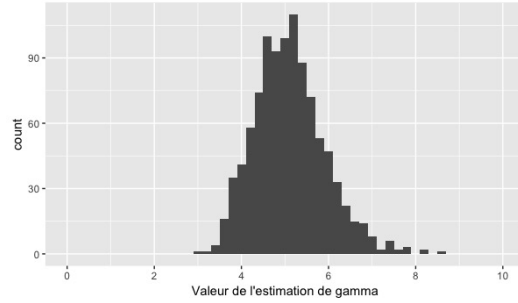
Paramètre	Vraie valeur T	Moyenne MC			Erreur st. MC			Erreur st. estimée		
		400	2000	4000	400	2000	4000	400	2000	4000
$\phi_{1,0}$	2,0	2,10	2,01	2,01	0,43	0,14	0,10	0,54	0,14	0,10
$\phi_{1,1}$	-0,1	-0,06	-0,09	-0,10	0,18	0,07	0,05	0,21	0,07	0,05
$\phi_{1,2}$	-0,5	-0,51	-0,50	-0,50	0,12	0,05	0,04	0,13	0,05	0,04
$\psi_{1,0}$	-4,0	-5,05	-4,11	-4,07	3,95	0,83	0,58	8,04	0,86	0,58
$\psi_{1,1}$	0,4	0,68	0,43	0,42	1,24	0,26	0,18	2,63	0,26	0,18
$\psi_{1,2}$	1,1	1,17	1,10	1,10	0,36	0,11	0,08	0,56	0,11	0,08
γ	2,0	2,07	2,02	2,00	0,68	0,20	0,14	0,63	0,20	0,14
c	1,0	1,05	1,01	1,00	0,29	0,10	0,07	0,36	0,10	0,07

TABLE 4. VI. Moyennes et erreurs standards selon les simulations de Monte Carlo, et erreurs standards moyennes selon l'estimé de la matrice C pour les paramètres du modèle STAR(2) avec $\gamma = 20$, $c = 1,2$.

Paramètre	Vraie valeur T	Moyenne MC			Erreur st. MC			Erreur st. estimée		
		400	2000	4000	400	2000	4000	400	2000	4000
$\phi_{1,0}$	2,0	2,00	2,00	2,00	0,04	0,02	0,01	0,04	0,02	0,01
$\phi_{1,1}$	-0,1	-0,10	-0,10	-0,10	0,03	0,01	0,01	0,03	0,01	0,01
$\phi_{1,2}$	-0,5	-0,50	-0,50	-0,50	0,02	0,01	0,01	0,02	0,01	0,01
$\psi_{1,0}$	-4,0	-4,03	-4,00	-4,00	0,18	0,08	0,05	0,18	0,08	0,06
$\psi_{1,1}$	0,4	0,41	0,40	0,40	0,08	0,04	0,02	0,08	0,04	0,03
$\psi_{1,2}$	1,1	1,09	1,10	1,10	0,04	0,02	0,01	0,04	0,02	0,01
γ	20,0	20,12	20,03	20,01	2,04	0,91	0,65	2,02	0,88	0,62
c	1,2	1,20	1,20	1,20	0,01	0,00	0,00	0,01	0,00	0,00

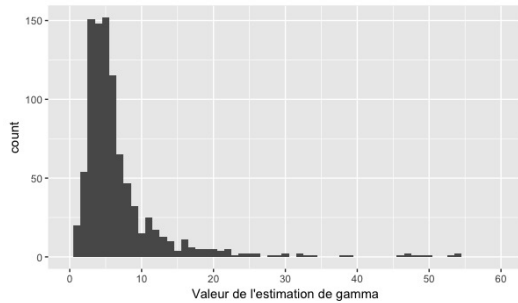


(a) $T = 400$

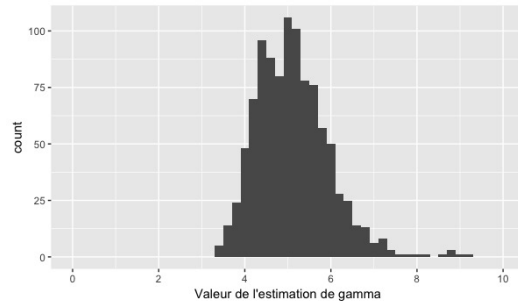


(b) $T = 4000$

FIGURE 4.5. Distributions empiriques de $\hat{\gamma}$ pour le modèle STAR(1) avec $\gamma = 5$, $c = -0,2$ selon le nombre de valeurs dans la série. Pour $T = 400$, 69 valeurs sont entre 60 et 59000 et ne sont pas illustrées.

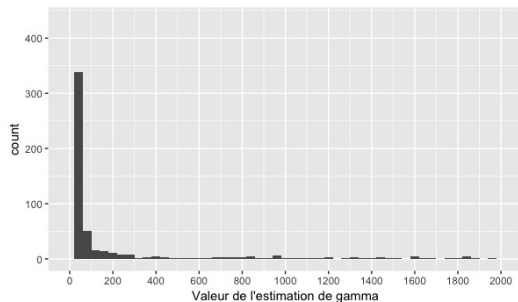


(a) $T = 400$

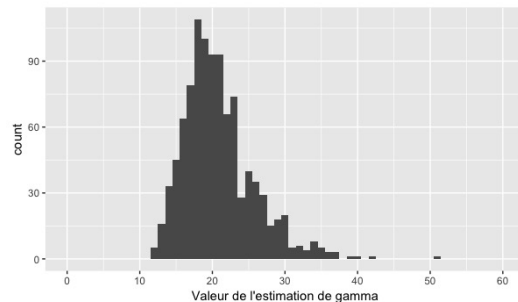


(b) $T = 4000$

FIGURE 4.6. Distributions empiriques de $\hat{\gamma}$ pour le modèle STAR(1) avec $\gamma = 5$, $c = -0,5$ selon le nombre de valeurs dans la série. Pour $T = 400$, 62 valeurs sont entre 60 et 26000 et ne sont pas illustrées.



(a) $T = 400$



(b) $T = 4000$

FIGURE 4.7. Distributions empiriques de $\hat{\gamma}$ pour le modèle STAR(1) avec $\gamma = 20$, $c = -0,2$ selon le nombre de valeurs dans la série. Pour $T = 400$, 38 valeurs sont entre 2000 et 16000 et ne sont pas illustrées.

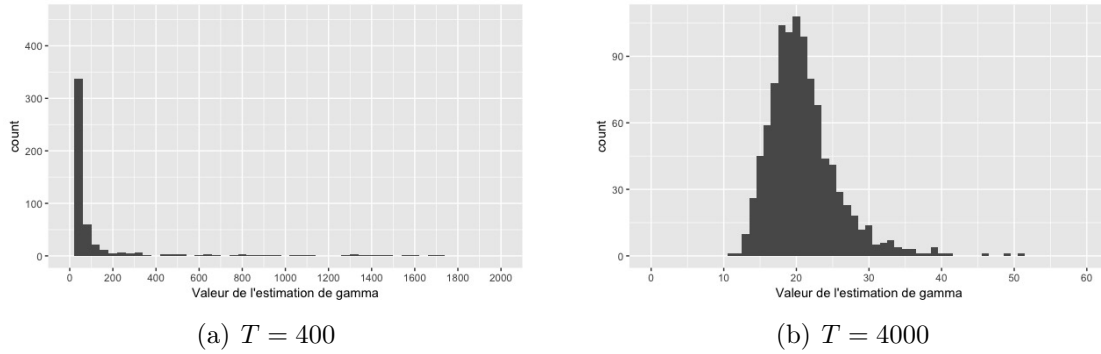


FIGURE 4.8. Distributions empiriques de $\hat{\gamma}$ pour le modèle STAR(1) avec $\gamma = 20$, $c = -0,5$ selon le nombre de valeurs dans la série. Pour $T = 400$, 45 valeurs sont entre 2000 et 71000 et ne sont pas illustrées.

4.3. Discussion des résultats de simulation sur l'estimation

Les estimations des paramètres et de leurs erreurs standards selon les tailles échantillonnelles de la série sont montrées aux tableaux 4. I, 4. II, 4. III et 4. IV pour le modèle STAR(1), et aux tableaux 4. V et 4. VI pour le modèle STAR(2).

Nous observons tout d'abord la moyenne des estimateurs des paramètres obtenus par les simulations de Monte Carlo. On remarque que, pour tous les modèles et valeurs de paramètres non-linéaires, les estimations des paramètres convergent vers les vraies valeurs lorsque la taille échantillonnelle de la série T augmente, ce qui est concordant avec le théorème 2.2.1. Ainsi, lorsque la série a un grand nombre de valeurs, soit lorsque $T = 4000$, les moyennes Monte Carlo des paramètres sont toutes très proches des vraies valeurs des paramètres du modèle, et ce pour tous les modèles étudiés. Dans une moindre mesure, c'est également le cas lorsque la série comporte $T = 2000$ valeurs.

Toutefois, lorsque la série comporte un petit nombre de valeurs, soit que $T = 400$, les estimations des paramètres selon la moyenne Monte Carlo sont parfois très éloignées des vraies valeurs, et c'est particulièrement le cas pour les modèles STAR(1) étudiés. Ainsi, on remarque aux tableaux 4. III et 4. IV que lorsque le modèle est un STAR(1) et que $\gamma = 20$, la moyenne de $\hat{\gamma}$ est grandement supérieure à la vraie valeur de γ , soit de 311,79 lorsque $c = -0,2$ et de 354,80 lorsque $c = -0,5$. Ceci est dû au fait que, comme γ est plus élevé, une petite taille échantillonnelle pour la série signifie un petit nombre de valeurs situées dans la transition entre les régimes, puisque la transition est moins lisse et s'effectue donc plus rapidement dû à la grande valeur de γ . Ainsi, il est plus difficile d'obtenir une estimation exacte de γ à partir d'un très petit nombre de valeurs dans les transitions entre les régimes.

Toujours avec $T = 400$, lorsque le modèle étudié est un STAR(1) et que $\gamma = 5$, soit aux tableaux 4. I et 4. II, on observe que, en plus du γ , les paramètres linéaires ont également une moyenne Monte Carlo très éloignée de leurs vraies valeurs. Cela est dû à la valeur de c et à la transition lisse entre les deux modèles. En effet, si l'on observe les graphiques des séries à la figure 4.2, on remarque que le deuxième régime est peu représenté dans l'évolution de la série, ce qui mène à des difficultés d'estimations lorsque le nombre de valeurs n'est pas assez suffisant dans un régime ou dans la transition, comme mentionné à la section 2.2.1. Toutefois, ce problème n'est présent que lorsque $T = 400$, et lorsque T augmente le nombre de valeurs dans le deuxième régime et dans les transitions est alors suffisamment élevé pour obtenir une estimation des régimes plus proche des vraies valeurs, ce qui est concordant avec la théorie asymptotique de l'estimation du modèle.

Ces incertitudes au niveau de l'estimation des paramètres par les moyennes de Monte Carlo se reflètent dans les erreurs standards de Monte Carlo. En effet, les modèles comportant des estimations moyennes de paramètres éloignées des vraies valeurs pour un certain T possèdent des erreurs standards Monte Carlo énormes dues aux difficultés d'estimation du modèle, tel que mentionné dans le paragraphe précédent. De plus, lorsque T augmente les erreurs standards de Monte Carlo diminuent dans tous les cas, ce qui est concordant avec le fait que les paramètres des modèles sous-jacents sont estimés avec plus de certitude lorsque le nombre de valeurs dans la série est plus grand. Autrement, nous utilisons les erreurs standards de Monte Carlo comme point de comparaison pour les erreurs standards estimées par les estimés de la matrice C .

Nous pouvons maintenant observer les erreurs standards obtenues selon les estimés des matrices C . Les valeurs présentes dans les tableaux 4. I, 4. II, 4. III, 4. IV, 4. V et 4. VI correspondent aux moyennes Monte Carlo des estimés des erreurs standards obtenues selon les estimés de la matrice C . Tout d'abord, on peut noter le fait que, pour certains modèles, la matrice C ne converge pas toujours pour toutes les simulations. En effet, les modèles STAR(1) avec un petit nombre de valeurs dans les séries ne parviennent pas toujours à calculer l'estimé de la matrice C , avec un maximum de taux d'échec de la convergence de 16,1% pour le modèle STAR(1) avec $\gamma = 20$, $c = -0,2$ et $T = 400$.

Cette non-convergence est principalement due au paramètre non-linéaire γ étant grandement surestimé dans certains modèles et à un problème calculatoire du logiciel utilisé associé à ces grandes valeurs. En effet, certaines séries simulées mènent à un estimé de

paramètre γ très grand lorsque le nombre de valeurs dans la série est petit, comme montré aux figures 4.5, 4.6, 4.7 et 4.8 où l'on peut observer les distributions empiriques de $\hat{\gamma}$ pour les modèles STAR(1). Le fait que l'estimé se trompe autant et atteint une si grande valeur pour le paramètre γ cause le logiciel à ne pas pouvoir calculer l'estimé de la matrice. En effet, l'exponentielle du paramètre γ est calculée dans les dérivées utilisées pour le calcul de la matrice C , tel qu'élaboré à la section 2.2. Lorsque $\hat{\gamma}$ prend une grande valeur, l'exponentielle de cette valeur surpasse les valeurs acceptées par le logiciel, ce qui fait que la valeur est considérée comme infinie. L'inversion des matrices subséquentes échoue donc en raison de cette valeur non-numérique présente dans la matrice. Ainsi, les non-convergences sont dues aux problèmes calculatoires de la matrice lorsque $\hat{\gamma}$ a une valeur élevée. Lorsque γ est estimé par une valeur plus basse et plus proche de sa vraie valeur, la matrice C converge de façon satisfaisante, et c'est le cas pour tous les modèles lorsque $T = 4000$ en raison de l'estimation plus précise des paramètres du modèle, tel que mentionné précédemment.

Ainsi, si l'on fait fi des matrices C échouant à être estimées en raison des valeurs trop grandes de $\hat{\gamma}$, on peut tout d'abord observer que lorsque T augmente, l'estimé de la matrice C converge vers les valeurs des erreurs standards obtenues par Monte Carlo, ce qui concorde avec le théorème de la convergence asymptotique de l'estimation non-linéaire des paramètres élaboré au théorème 2.2.1.

Ensuite, on remarque que, lorsque le modèle STAR possède des difficultés d'estimation pour les paramètres linéaires et non-linéaires du modèle, comme c'est le cas pour les modèles STAR(1) avec $\gamma = 5$ et $T = 400$ aux tableaux 4. I et 4. II, les erreurs standards associées obtenues par les estimés de la matrice C sont énormes. Ceci n'est pas surprenant sachant que la moyenne Monte Carlo est très loin des vraies valeurs pour la plupart des paramètres linéaires. Toutefois, pour ces modèles, le paramètre γ a une erreur standard beaucoup plus raisonnable que celles des paramètres linéaires. Ceci est dû au fait que les matrices C associées aux grandes valeurs de $\hat{\gamma}$ ne font pas parties de la moyenne des matrices puisqu'elles ne sont pas calculables, comme mentionné précédemment. Ainsi, l'erreur associée à $\hat{\gamma}$ par l'estimé de la matrice C est beaucoup plus petite que celle obtenue par Monte Carlo, qui elle prend en compte les grandes valeurs de $\hat{\gamma}$. Lorsque $T = 2000$, et encore plus lorsque $T = 4000$, le modèle et ses paramètres sont mieux estimés et ce phénomène de grandes erreurs standards n'est plus observable, et alors les erreurs standards obtenues par l'estimé de la matrice C sont très proches de celles obtenues par Monte Carlo.

Pour les modèles STAR(1) avec $\gamma = 20$, les paramètres linéaires sont mieux estimés par le modèle, ce qui est observable par le fait que les erreurs standards de Monte Carlo et de la matrice C sont beaucoup plus raisonnables peu importe le nombre de valeurs T dans la série. Ainsi, pour ces modèles, le seul paramètre dont l'erreur standard est élevée est γ , mais celle-ci converge tout de même lorsque T augmente. Pour $T = 400$, il existe toutefois une grande différence entre les erreurs standards de Monte Carlo et celles de l'estimé de la matrice C . Ceci est dû aux matrices non-calculables pour les grandes valeurs de $\hat{\gamma}$ qui ne sont ainsi pas dans l'erreur standard de la matrice C mais qui sont dans l'erreur standard Monte Carlo, comme c'était le cas pour les modèles STAR(1) avec $\gamma = 5$.

Finalement, les modèles STAR(2) ne possèdent pas de matrices C non-calculables comme c'est le cas pour les modèles STAR(1), puisque leurs paramètres sont beaucoup mieux estimés même quand T est petit. Ainsi, il est déjà possible de voir une concordance des erreurs standards de Monte Carlo et de la matrice C lorsque $T = 2000$, plutôt que pour $T = 4000$ dans les modèles STAR(1). Ceci démontre bien les propriétés asymptotiques du théorème 2.2.1 dans le cadre des modèles STAR(p). Il n'en demeure pas moins que les résultats pour $T = 400$ sont décevants. Il faut retenir que les modèles STAR exigent des tailles échantillonnelles importantes.

4.4. Prévisions ponctuelles et intervalles de prévision

Suite à l'estimation des paramètres de chaque série, nous calculons les prévisions ponctuelles pour le modèle STAR ajusté par les méthodes exactes, naïves, de Monte Carlo et de bootstrap telles qu'élaborées à la section 3.3, en plus de calculer les prévisions ponctuelles qui seraient obtenues si le modèle ajusté était de type AR tel qu'élaboré à la section 3.2. Nous ajustons deux types de modèles AR : un modèle AR(p) avec p correspondant à celui du modèle STAR(p) ajusté, soit un ou deux, et un modèle AR(p) avec p étant sélectionné par le critère AIC comme à l'équation (2.1.1) pour chaque série simulée. De plus, nous calculons les intervalles de prévision et leurs longueurs moyennes par les méthodes de Monte Carlo et notre méthode de type bootstrap élaborés à la section 3.4, en plus de ceux qui seraient obtenus par l'ajustement des modèles AR(p). On note que dans l'application de la méthode Monte Carlo, un terme d'erreur doit être simulé. Idéalement, l'erreur d'horizon h devrait être de loi $e_{t+h} \sim \mathcal{L}(0, \sigma_\varepsilon^2 c_h^2)$. Ici, par simplicité et compte tenu que les horizons considérés sont petits, on pose $c_1 = c_2 = c_3 = 1$. De plus, notre bootstrap est une variante de la méthode de Monte Carlo, puisque nous ajoutons des résidus (donc des erreurs de prévision d'horizon un) dans l'expression (3.4.5). Nous appelons tout de même cette variante

une méthode de type bootstrap car elle repose moins sur l'hypothèse de normalité que la méthode Monte Carlo.

Nous simulons également les vraies valeurs de la série pour les temps de prévisions étudiés, ce qui nous permet de calculer le biais moyen et l'erreur quadratique moyenne pour chaque prévision ponctuelle, en plus d'observer si la vraie valeur de la série se situe dans les intervalles de prévision obtenus. Pour la méthode de bootstrap, nous centrons les résidus du modèle ajusté avant l'application de la méthode développée afin d'être concordants avec la théorie de la section 3.4.

Les résultats pour les prévisions ponctuelles sont présentés aux tableaux 4. VII, 4. VIII, 4. IX, 4. X, 4. XI et 4. XII, et les résultats pour les intervalles de prévision sont présentés aux tableaux 4. XIII, 4. XIV, 4. XV, 4. XVI, 4. XVII et 4. XVIII.

TABLE 4. VII. Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 5$ et $c = -0,2$.

	Temps de prév. T	t+1			t+2			t+3		
		400	2000	4000	400	2000	4000	400	2000	4000
Biais moy.	AR(1)	-0,015	0,022	0,002	0,007	0,031	0,004	-0,015	0,010	0,001
	AR(p)	-0,018	0,025	0,002	0,004	0,037	0,003	-0,018	0,018	0,000
	STAR(1) Exacte	-0,018	0,020	0,006	0,004	0,030	0,009	-0,017	0,009	0,006
	STAR(1) Naïve	-	-	-	0,010	0,038	0,019	-0,002	0,031	0,030
	STAR(1) MC	-	-	-	0,004	0,029	0,009	-0,018	0,009	0,005
	STAR(1) BS	-	-	-	0,005	0,029	0,008	-0,017	0,009	0,005
EQM moy.	AR(1)	0,259	0,268	0,284	0,432	0,392	0,446	0,514	0,480	0,522
	AR(p)	0,266	0,266	0,284	0,454	0,401	0,446	0,554	0,495	0,533
	STAR(1) Exacte	0,244	0,247	0,266	0,406	0,370	0,418	0,489	0,466	0,499
	STAR(1) Naïve	-	-	-	0,416	0,379	0,422	0,522	0,503	0,525
	STAR(1) MC	-	-	-	0,406	0,371	0,418	0,489	0,466	0,498
	STAR(1) BS	-	-	-	0,406	0,370	0,418	0,489	0,466	0,500

TABLE 4. VIII. Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 5$ et $c = -0,5$.

	Temps de prév. T	t+1			t+2			t+3		
		400	2000	4000	400	2000	4000	400	2000	4000
Biais moy.	AR(1)	-0,012	0,010	0,008	0,010	0,015	0,006	-0,008	-0,001	-0,004
	AR(p)	-0,019	0,012	0,012	0,002	0,018	0,011	-0,016	0,004	0,002
	STAR(1) Exacte	-0,017	0,019	0,006	0,006	0,025	0,004	-0,012	0,010	-0,006
	STAR(1) Naïve	-	-	-	-0,030	-0,009	-0,033	-0,079	-0,054	-0,076
	STAR(1) MC	-	-	-	0,006	0,025	0,005	-0,012	0,010	-0,005
	STAR(1) BS	-	-	-	0,006	0,025	0,004	-0,012	0,010	-0,006
EQM moy.	AR(1)	0,277	0,291	0,311	0,416	0,393	0,439	0,492	0,475	0,509
	AR(p)	0,281	0,290	0,304	0,432	0,401	0,426	0,524	0,485	0,503
	STAR(1) Exacte	0,242	0,247	0,267	0,382	0,350	0,387	0,455	0,430	0,457
	STAR(1) Naïve	-	-	-	0,394	0,353	0,387	0,490	0,453	0,473
	STAR(1) MC	-	-	-	0,382	0,350	0,388	0,455	0,431	0,456
	STAR(1) BS	-	-	-	0,382	0,350	0,387	0,454	0,432	0,458

TABLE 4. IX. Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 20$ et $c = -0,2$.

	Temps de prév. T	t+1			t+2			t+3		
		400	2000	4000	400	2000	4000	400	2000	4000
Biais moy.	AR(1)	-0,007	0,020	-0,004	0,022	0,028	-0,006	-0,005	0,017	-0,011
	AR(p)	-0,016	0,021	-0,003	0,009	0,031	-0,005	-0,019	0,021	-0,010
	STAR(1) Exacte	-0,020	0,021	0,006	0,008	0,029	0,005	-0,018	0,019	0,001
	STAR(1) Naïve	-	-	-	0,002	0,031	0,008	-0,026	0,021	0,005
	STAR(1) MC	-	-	-	0,008	0,028	0,006	-0,020	0,018	0,000
	STAR(1) BS	-	-	-	0,009	0,028	0,005	-0,019	0,017	0,000
EQM moy.	AR(1)	0,275	0,290	0,307	0,450	0,387	0,463	0,535	0,483	0,554
	AR(p)	0,286	0,287	0,302	0,456	0,394	0,459	0,555	0,505	0,561
	STAR(1) Exacte	0,243	0,248	0,266	0,396	0,345	0,401	0,484	0,437	0,489
	STAR(1) Naïve	-	-	-	0,407	0,355	0,405	0,513	0,462	0,507
	STAR(1) MC	-	-	-	0,396	0,345	0,401	0,482	0,437	0,488
	STAR(1) BS	-	-	-	0,396	0,345	0,401	0,484	0,437	0,490

TABLE 4. X. Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 20$ et $c = -0,5$.

	Temps de prév. T	t+1			t+2			t+3		
		400	2000	4000	400	2000	4000	400	2000	4000
Biais moy.	AR(1)	-0,004	0,010	0,017	0,016	0,020	0,021	-0,003	-0,003	0,000
	AR(p)	-0,014	0,012	0,017	-0,001	0,024	0,021	-0,024	0,003	-0,001
	STAR(1) Exacte	-0,016	0,020	0,006	0,003	0,033	0,009	-0,019	0,006	-0,017
	STAR(1) Naïve	-	-	-	-0,009	0,027	-0,003	-0,034	0,003	-0,031
	STAR(1) MC	-	-	-	0,004	0,033	0,009	-0,014	0,011	-0,012
	STAR(1) BS	-	-	-	0,004	0,032	0,008	-0,014	0,011	-0,012
EQM moy.	AR(1)	0,297	0,324	0,332	0,441	0,435	0,466	0,518	0,510	0,551
	AR(p)	0,297	0,316	0,319	0,451	0,438	0,444	0,544	0,504	0,528
	STAR(1) Exacte	0,242	0,249	0,267	0,377	0,351	0,387	0,453	0,419	0,474
	STAR(1) Naïve	-	-	-	0,388	0,358	0,395	0,485	0,443	0,497
	STAR(1) MC	-	-	-	0,378	0,351	0,387	0,452	0,419	0,473
	STAR(1) BS	-	-	-	0,377	0,350	0,387	0,452	0,420	0,474

TABLE 4. XI. Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(2), AR(p) et STAR(2) avec $\gamma = 2$ et $c = 1$.

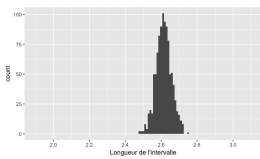
	Temps de prév. T	t+1			t+2			t+3		
		400	2000	4000	400	2000	4000	400	2000	4000
Biais moy.	AR(2)	0,009	0,005	-0,006	0,035	-0,004	0,004	-0,036	-0,052	-0,022
	AR(p)	0,009	0,006	-0,006	0,034	-0,003	0,003	-0,036	-0,050	-0,023
	STAR(2) Exacte	-0,020	0,020	0,005	0,051	-0,009	-0,002	-0,053	-0,042	-0,017
	STAR(2) Naïve	-	-	-	0,074	0,011	0,015	-0,163	-0,137	-0,114
	STAR(2) MC	-	-	-	0,052	-0,009	-0,002	-0,051	-0,042	-0,017
	STAR(2) BS	-	-	-	0,050	-0,008	-0,001	-0,052	-0,042	-0,018
EQM moy.	AR(2)	0,409	0,458	0,458	0,688	0,705	0,718	0,849	0,883	0,897
	AR(p)	0,409	0,459	0,459	0,689	0,703	0,713	0,848	0,881	0,900
	STAR(2) Exacte	0,241	0,248	0,268	0,584	0,557	0,629	0,745	0,778	0,827
	STAR(2) Naïve	-	-	-	0,600	0,552	0,640	0,804	0,815	0,872
	STAR(2) MC	-	-	-	0,586	0,556	0,628	0,744	0,778	0,827
	STAR(2) BS	-	-	-	0,584	0,560	0,629	0,747	0,779	0,829

TABLE 4. XII. Biais et erreur quadratique moyenne pour les prévisions ponctuelles selon les modèles AR(2), AR(p) et STAR(2) avec $\gamma = 20$ et $c = 1,2$.

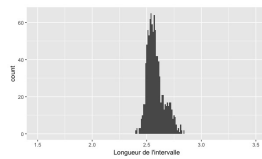
	Temps de prév. T	t+1			t+2			t+3		
		400	2000	4000	400	2000	4000	400	2000	4000
Biais moy.	AR(2)	0,025	0,050	0,037	0,070	-0,079	-0,001	-0,076	-0,065	-0,080
	AR(p)	0,022	0,037	0,016	0,097	-0,078	0,029	-0,078	-0,068	-0,069
	STAR(2) Exacte	-0,017	0,021	0,005	0,095	-0,063	0,033	-0,096	-0,084	-0,089
	STAR(2) Naïve	-	-	-	0,292	0,088	0,266	-0,049	0,020	-0,053
	STAR(2) MC	-	-	-	0,095	-0,063	0,033	-0,092	-0,083	-0,086
	STAR(2) BS	-	-	-	0,095	-0,060	0,034	-0,094	-0,084	-0,087
EQM moy.	AR(2)	2,475	2,555	2,424	3,200	2,915	3,374	3,296	3,491	3,251
	AR(p)	2,318	2,307	2,227	3,019	2,600	3,237	3,115	3,309	3,217
	STAR(2) Exacte	0,245	0,247	0,266	2,197	1,823	2,419	2,276	2,411	2,513
	STAR(2) Naïve	-	-	-	3,051	2,129	3,326	2,903	3,108	3,536
	STAR(2) MC	-	-	-	2,203	1,824	2,426	2,274	2,431	2,505
	STAR(2) BS	-	-	-	2,196	1,827	2,411	2,294	2,421	2,518

TABLE 4. XIII. Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 5$ et $c = -0,2$.

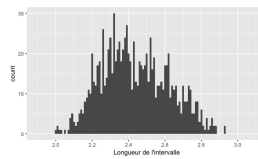
Niv.	Modèle	Temps de prév. T	t+1			t+2			t+3		
			400	2000	4000	400	2000	4000	400	2000	4000
80	AR(1)	Taux d'incl. (%)	81,8	79,3	77,8	80,6	81,7	80,3	80,6	82,7	80,7
		Long. moy. de l'IP	1,3	1,3	1,3	1,7	1,7	1,7	1,9	1,9	1,9
	AR(p)	Taux d'incl. (%)	81,6	79,1	78,2	79,6	80,4	79,5	80,9	81,6	80,1
		Long. moy. de l'IP	1,3	1,3	1,3	1,7	1,7	1,7	1,9	1,9	1,9
	STAR(1) MC	Taux d'incl. (%)	80,4	79,4	77,8	78,0	80,6	77,9	81,2	81,0	79,8
		Long. moy. de l'IP	1,3	1,3	1,3	1,6	1,6	1,6	1,8	1,8	1,8
STAR(1) BS	Taux d'incl. (%)	81,0	79,2	77,8	77,5	81,1	78,6	80,7	80,7	79,3	
	Long. moy. de l'IP	1,3	1,3	1,3	1,6	1,6	1,6	1,8	1,8	1,8	
90	AR(1)	Taux d'incl. (%)	90,0	91,3	89,0	89,6	90,4	90,0	91,0	91,8	91,5
		Long. moy. de l'IP	1,7	1,7	1,7	2,2	2,2	2,2	2,4	2,5	2,5
	AR(p)	Taux d'incl. (%)	90,2	90,9	89,0	89,7	89,7	89,8	90,6	91,0	90,4
		Long. moy. de l'IP	1,7	1,7	1,7	2,2	2,2	2,2	2,5	2,4	2,4
	STAR(1) MC	Taux d'incl. (%)	88,9	90,2	88,2	89,4	90,5	88,9	90,4	91,0	90,1
		Long. moy. de l'IP	1,6	1,6	1,6	2,0	2,0	2,1	2,3	2,3	2,3
STAR(1) BS	Taux d'incl. (%)	89,6	89,7	88,1	88,9	90,7	88,8	89,9	90,3	90,1	
	Long. moy. de l'IP	1,6	1,6	1,6	2,0	2,0	2,1	2,3	2,3	2,3	
95	AR(1)	Taux d'incl. (%)	95,1	95,4	94,7	95,1	96,7	94,9	96,0	97,0	95,1
		Long. moy. de l'IP	2,0	2,0	2,0	2,6	2,6	2,6	2,9	2,9	2,9
	AR(p)	Taux d'incl. (%)	95,3	95,6	94,9	94,2	96,0	94,5	95,5	96,2	95,1
		Long. moy. de l'IP	2,1	2,0	2,0	2,6	2,6	2,6	3,0	2,9	2,9
	STAR(1) MC	Taux d'incl. (%)	94,1	95,7	94,4	94,4	95,7	93,7	95,1	96,2	94,4
		Long. moy. de l'IP	1,9	1,9	1,9	2,4	2,4	2,5	2,7	2,7	2,8
STAR(1) BS	Taux d'incl. (%)	93,8	95,9	94,3	94,7	95,5	93,8	94,5	95,5	94,3	
	Long. moy. de l'IP	1,9	1,9	1,9	2,4	2,4	2,5	2,7	2,7	2,8	
99	AR(1)	Taux d'incl. (%)	98,5	99,3	99,1	98,9	99,8	98,7	99,4	99,7	99,4
		Long. moy. de l'IP	2,7	2,7	2,7	3,4	3,4	3,4	3,8	3,9	3,9
	AR(p)	Taux d'incl. (%)	98,6	99,3	99,2	98,8	99,5	98,4	99,0	99,2	99,1
		Long. moy. de l'IP	2,7	2,7	2,7	3,4	3,4	3,4	3,9	3,8	3,8
	STAR(1) MC	Taux d'incl. (%)	98,3	99,1	98,9	98,6	99,6	98,1	98,8	99,2	98,7
		Long. moy. de l'IP	2,5	2,5	2,5	3,2	3,2	3,2	3,5	3,5	3,6
STAR(1) BS	Taux d'incl. (%)	98,4	99,1	98,7	98,2	99,2	98,0	98,7	99,2	99,1	
	Long. moy. de l'IP	2,5	2,5	2,5	3,2	3,2	3,2	3,5	3,5	3,6	



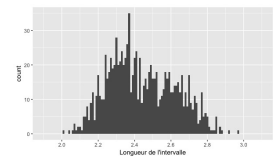
(a) AR(1)



(b) AR(p)



(c) STAR(1) Monte Carlo



(d) STAR(1) bootstrap

FIGURE 4.9. Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(1) avec $T = 2000$, $\gamma = 5$, $c = -0,2$.

TABLE 4. XIV. Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 5$ et $c = -0,5$.

Niv.	Modèle	Temps de prév. T	t+1			t+2			t+3		
			400	2000	4000	400	2000	4000	400	2000	4000
80	AR(1)	Taux d'incl. (%)	80,5	77,1	77,2	81,4	82,5	81,4	81,0	82,4	82,4
		Long. moy. de l'IP	1,4	1,4	1,4	1,7	1,7	1,7	1,9	1,9	1,9
	AR(p)	Taux d'incl. (%)	80,8	77,3	78,0	81,3	80,3	80,2	81,3	80,6	81,1
		Long. moy. de l'IP	1,4	1,4	1,4	1,7	1,7	1,7	1,9	1,8	1,8
	STAR(1) MC	Taux d'incl. (%)	81,0	79,2	78,0	79,0	82,5	79,3	79,2	80,6	80,2
		Long. moy. de l'IP	1,3	1,3	1,3	1,5	1,5	1,5	1,7	1,7	1,7
STAR(1) BS	Taux d'incl. (%)	81,6	79,0	77,8	79,4	82,0	78,9	79,7	80,8	80,4	
	Long. moy. de l'IP	1,3	1,3	1,3	1,5	1,5	1,5	1,7	1,7	1,7	
90	AR(1)	Taux d'incl. (%)	90,8	89,5	88,0	90,9	92,6	89,9	90,7	92,3	90,1
		Long. moy. de l'IP	1,7	1,7	1,7	2,2	2,2	2,2	2,4	2,4	2,4
	AR(p)	Taux d'incl. (%)	90,1	89,1	88,3	90,0	91,5	89,4	91,0	90,2	90,0
		Long. moy. de l'IP	1,8	1,7	1,7	2,2	2,1	2,1	2,4	2,4	2,3
	STAR(1) MC	Taux d'incl. (%)	90,4	90,1	88,2	88,8	91,4	88,6	89,1	90,6	90,5
		Long. moy. de l'IP	1,6	1,6	1,6	2,0	2,0	2,0	2,2	2,2	2,2
STAR(1) BS	Taux d'incl. (%)	89,9	89,8	88,5	89,1	91,2	88,6	88,3	90,9	90,2	
	Long. moy. de l'IP	1,6	1,6	1,6	2,0	2,0	2,0	2,2	2,2	2,2	
95	AR(1)	Taux d'incl. (%)	94,9	95,5	93,6	95,4	96,2	95,3	95,6	96,8	95,1
		Long. moy. de l'IP	2,1	2,1	2,1	2,6	2,6	2,6	2,9	2,9	2,9
	AR(p)	Taux d'incl. (%)	94,6	95,6	93,8	94,3	95,7	94,4	95,0	95,3	94,6
		Long. moy. de l'IP	2,1	2,1	2,1	2,6	2,5	2,5	2,9	2,8	2,8
	STAR(1) MC	Taux d'incl. (%)	94,4	95,7	94,1	94,2	96,0	94,2	94,8	95,7	94,6
		Long. moy. de l'IP	1,9	1,9	1,9	2,4	2,4	2,4	2,6	2,6	2,6
STAR(1) BS	Taux d'incl. (%)	94,8	96,4	94,1	94,2	95,4	94,4	94,7	95,9	94,5	
	Long. moy. de l'IP	1,9	1,9	1,9	2,4	2,4	2,4	2,6	2,7	2,6	
99	AR(1)	Taux d'incl. (%)	98,6	99,1	99,0	99,5	99,5	98,8	99,1	99,6	99,0
		Long. moy. de l'IP	2,7	2,7	2,7	3,4	3,5	3,5	3,8	3,8	3,8
	AR(p)	Taux d'incl. (%)	98,8	99,4	99,3	99,0	99,2	98,7	98,7	99,6	98,5
		Long. moy. de l'IP	2,7	2,7	2,7	3,4	3,3	3,3	3,8	3,7	3,7
	STAR(1) MC	Taux d'incl. (%)	98,3	99,1	98,8	99,0	99,5	98,9	98,7	99,3	98,4
		Long. moy. de l'IP	2,5	2,5	2,5	3,1	3,1	3,1	3,4	3,5	3,5
STAR(1) BS	Taux d'incl. (%)	98,5	99,2	98,6	98,7	99,1	98,5	98,2	99,4	98,7	
	Long. moy. de l'IP	2,5	2,5	2,5	3,1	3,1	3,1	3,4	3,5	3,5	

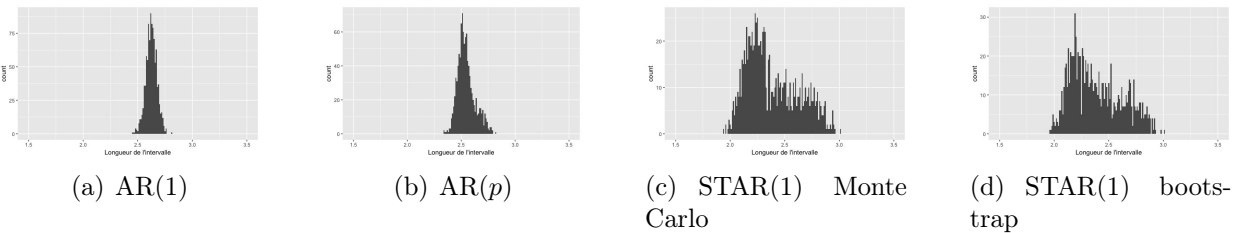


FIGURE 4.10. Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(1) avec $T = 2000$, $\gamma = 5$, $c = -0,5$.

TABLE 4. XV. Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 20$ et $c = -0,2$.

Niv.	Modèle	Temps de prév. T	t+1			t+2			t+3		
			400	2000	4000	400	2000	4000	400	2000	4000
80	AR(1)	Taux d'incl. (%)	81,8	78,9	77,0	80,3	83,4	80,4	81,6	84,3	81,2
		Long. moy. de l'IP	1,4	1,4	1,4	1,8	1,8	1,8	2,0	2,0	2,0
	AR(p)	Taux d'incl. (%)	81,3	79,3	77,6	80,1	81,3	79,7	81,4	81,8	79,5
		Long. moy. de l'IP	1,4	1,4	1,4	1,7	1,7	1,7	1,9	1,9	1,9
	STAR(1) MC	Taux d'incl. (%)	80,4	79,4	78,1	78,8	80,6	78,2	79,8	80,8	79,6
		Long. moy. de l'IP	1,3	1,3	1,3	1,5	1,5	1,5	1,7	1,7	1,7
STAR(1) BS	Taux d'incl. (%)	80,7	79,2	77,7	78,3	81,4	77,8	80,4	80,6	79,5	
	Long. moy. de l'IP	1,3	1,3	1,3	1,5	1,5	1,6	1,7	1,7	1,7	
90	AR(1)	Taux d'incl. (%)	90,1	89,7	89,2	90,2	92,8	90,0	91,5	93,5	91,1
		Long. moy. de l'IP	1,8	1,8	1,8	2,2	2,3	2,3	2,5	2,5	2,5
	AR(p)	Taux d'incl. (%)	90,2	89,3	89,1	90,5	91,4	87,8	90,9	90,8	89,1
		Long. moy. de l'IP	1,8	1,8	1,8	2,2	2,2	2,2	2,5	2,4	2,4
	STAR(1) MC	Taux d'incl. (%)	90,2	90,5	88,2	88,4	91,1	88,6	90,1	91,5	89,4
		Long. moy. de l'IP	1,6	1,6	1,6	2,0	2,0	2,0	2,2	2,2	2,3
STAR(1) BS	Taux d'incl. (%)	90,0	90,0	88,2	88,3	91,0	88,3	88,7	91,4	89,8	
	Long. moy. de l'IP	1,6	1,6	1,6	2,0	2,0	2,0	2,2	2,2	2,3	
95	AR(1)	Taux d'incl. (%)	95,2	94,7	94,5	95,6	97,7	95,5	95,7	97,6	95,4
		Long. moy. de l'IP	2,1	2,1	2,1	2,7	2,7	2,7	3,0	3,0	3,0
	AR(p)	Taux d'incl. (%)	94,9	95,5	94,8	94,2	96,7	94,2	95,3	96,4	94,5
		Long. moy. de l'IP	2,1	2,1	2,1	2,7	2,6	2,6	3,0	2,9	2,9
	STAR(1) MC	Taux d'incl. (%)	94,4	95,8	94,1	94,5	95,8	93,7	94,9	96,5	94,2
		Long. moy. de l'IP	1,9	1,9	1,9	2,4	2,4	2,4	2,7	2,7	2,7
STAR(1) BS	Taux d'incl. (%)	93,8	96,1	94,2	94,1	95,7	93,9	94,3	96,3	93,8	
	Long. moy. de l'IP	1,9	1,9	1,9	2,4	2,4	2,4	2,7	2,7	2,7	
99	AR(1)	Taux d'incl. (%)	98,6	99,4	99,5	99,2	99,9	99,3	99,3	99,6	99,1
		Long. moy. de l'IP	2,8	2,8	2,8	3,5	3,6	3,6	3,9	4,0	4,0
	AR(p)	Taux d'incl. (%)	98,7	99,2	99,0	98,8	99,5	98,7	98,1	99,2	98,7
		Long. moy. de l'IP	2,8	2,8	2,8	3,5	3,4	3,4	3,9	3,8	3,8
	STAR(1) MC	Taux d'incl. (%)	98,0	99,0	98,9	98,4	99,5	98,2	98,5	99,3	98,8
		Long. moy. de l'IP	2,5	2,5	2,5	3,2	3,2	3,2	3,5	3,6	3,6
STAR(1) BS	Taux d'incl. (%)	98,4	99,1	98,7	98,0	99,3	98,1	98,4	99,1	98,8	
	Long. moy. de l'IP	2,5	2,5	2,5	3,2	3,2	3,2	3,5	3,6	3,6	

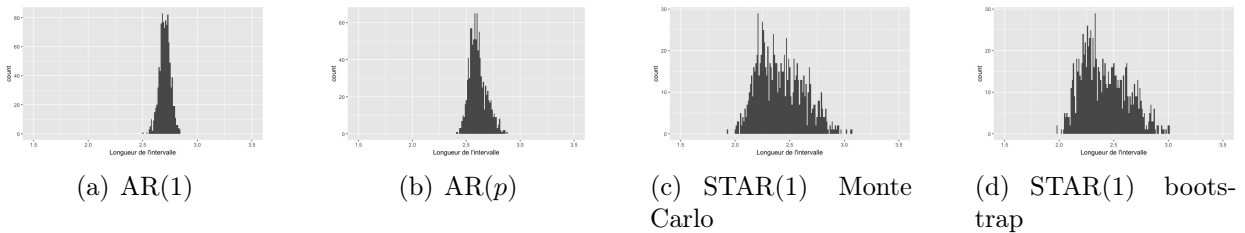


FIGURE 4.11. Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(1) avec $T = 2000$, $\gamma = 20$, $c = -0,2$.

TABLE 4. XVI. Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(1), AR(p) et STAR(1) avec $\gamma = 20$ et $c = -0,5$.

Niv.	Modèle	Temps de prév. T	t+1			t+2			t+3		
			400	2000	4000	400	2000	4000	400	2000	4000
80	AR(1)	Taux d'incl. (%)	80,9	78,1	77,1	81,5	83,2	81,2	82,0	83,4	83,5
		Long. moy. de l'IP	1,4	1,4	1,4	1,8	1,8	1,8	2,0	2,0	2,0
	AR(p)	Taux d'incl. (%)	80,9	77,7	77,7	81,5	80,3	79,2	81,0	82,2	81,7
		Long. moy. de l'IP	1,4	1,4	1,4	1,7	1,7	1,7	1,9	1,9	1,9
	STAR(1) MC	Taux d'incl. (%)	80,8	78,8	77,8	78,9	81,2	77,1	79,4	80,6	79,1
		Long. moy. de l'IP	1,3	1,3	1,3	1,5	1,5	1,5	1,6	1,7	1,7
STAR(1) BS	Taux d'incl. (%)	81,0	78,7	77,6	78,5	81,5	77,4	80,2	80,9	79,2	
	Long. moy. de l'IP	1,3	1,3	1,3	1,5	1,5	1,5	1,6	1,7	1,7	
90	AR(1)	Taux d'incl. (%)	90,6	88,9	88,8	92,0	91,6	90,8	91,5	92,5	91,2
		Long. moy. de l'IP	1,8	1,8	1,8	2,3	2,3	2,3	2,5	2,6	2,6
	AR(p)	Taux d'incl. (%)	90,6	89,6	89,1	89,9	89,9	89,2	89,2	90,7	89,7
		Long. moy. de l'IP	1,8	1,8	1,8	2,2	2,2	2,2	2,5	2,4	2,4
	STAR(1) MC	Taux d'incl. (%)	89,7	90,0	88,4	90,5	91,0	88,2	88,8	91,6	89,9
		Long. moy. de l'IP	1,6	1,6	1,6	1,9	2,0	2,0	2,2	2,2	2,2
STAR(1) BS	Taux d'incl. (%)	89,5	89,8	88,3	89,6	91,0	88,4	88,4	91,2	89,4	
	Long. moy. de l'IP	1,6	1,6	1,6	1,9	2,0	2,0	2,2	2,2	2,2	
95	AR(1)	Taux d'incl. (%)	95,2	95,1	94,3	95,6	96,2	96,4	97,1	96,9	95,8
		Long. moy. de l'IP	2,2	2,2	2,2	2,7	2,8	2,8	3,0	3,1	3,1
	AR(p)	Taux d'incl. (%)	94,8	95,0	94,1	94,3	94,7	94,2	94,6	95,5	94,4
		Long. moy. de l'IP	2,2	2,2	2,2	2,6	2,6	2,6	2,9	2,9	2,8
	STAR(1) MC	Taux d'incl. (%)	94,6	95,5	94,3	93,9	95,3	93,7	94,7	95,7	94,8
		Long. moy. de l'IP	1,9	1,9	1,9	2,4	2,4	2,4	2,6	2,7	2,7
STAR(1) BS	Taux d'incl. (%)	93,9	96,1	94,1	94,3	95,3	94,0	94,8	95,5	94,1	
	Long. moy. de l'IP	1,9	1,9	1,9	2,4	2,4	2,4	2,6	2,7	2,7	
99	AR(1)	Taux d'incl. (%)	98,7	99,2	99,1	99,5	99,5	99,3	99,4	99,6	99,3
		Long. moy. de l'IP	2,8	2,9	2,9	3,6	3,6	3,6	4,0	4,0	4,0
	AR(p)	Taux d'incl. (%)	98,6	99,2	99,0	98,8	98,6	99,2	98,7	98,8	98,1
		Long. moy. de l'IP	2,8	2,8	2,8	3,5	3,4	3,4	3,8	3,8	3,7
	STAR(1) MC	Taux d'incl. (%)	98,2	98,9	98,7	98,5	99,2	99,1	98,9	99,1	98,3
		Long. moy. de l'IP	2,5	2,5	2,5	3,2	3,2	3,2	3,5	3,5	3,5
STAR(1) BS	Taux d'incl. (%)	98,4	99,3	98,6	98,5	98,8	98,7	98,3	99,3	98,6	
	Long. moy. de l'IP	2,5	2,5	2,5	3,2	3,2	3,2	3,5	3,5	3,5	

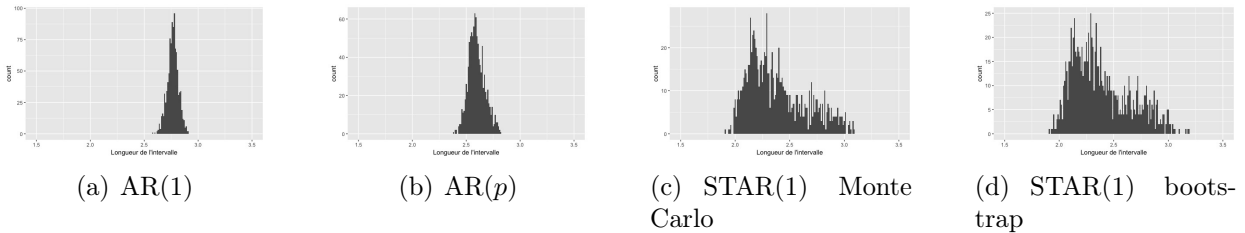


FIGURE 4.12. Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(1) avec $T = 2000$, $\gamma = 20$, $c = -0,5$.

TABLE 4. XVII. Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(2), AR(p) et STAR(2) avec $\gamma = 2$ et $c = 1$.

Niv.	Modèle	Temps de prév. T	t+1			t+2			t+3		
			400	2000	4000	400	2000	4000	400	2000	4000
80	AR(2)	Taux d'incl. (%)	<i>83,2</i>	78,2	79,9	80,5	80,1	78,1	79,5	78,7	77,6
		Long. moy. de l'IP	1,7	1,7	1,7	2,1	2,1	2,1	2,3	2,3	2,3
	AR(p)	Taux d'incl. (%)	<i>83,6</i>	78,8	79,4	80,7	80,2	78,4	80,0	79,0	77,9
		Long. moy. de l'IP	1,7	1,7	1,7	2,1	2,1	2,1	2,3	2,3	2,3
	STAR(2) MC	Taux d'incl. (%)	80,8	78,9	<i>77,3</i>	81,6	81,0	<i>76,4</i>	80,5	81,2	77,7
		Long. moy. de l'IP	1,3	1,3	1,3	1,9	1,9	1,9	2,2	2,2	2,2
STAR(2) BS	Taux d'incl. (%)	80,1	79,1	<i>77,5</i>	81,3	80,9	<i>76,9</i>	79,5	80,8	77,6	
	Long. moy. de l'IP	1,3	1,3	1,3	1,9	1,9	1,9	2,2	2,2	2,2	
90	AR(2)	Taux d'incl. (%)	91,7	89,3	90,0	90,3	89,8	89,7	88,7	88,2	88,5
		Long. moy. de l'IP	2,2	2,2	2,2	2,7	2,7	2,7	3,0	3,0	3,0
	AR(p)	Taux d'incl. (%)	91,6	89,7	89,8	90,5	89,9	89,7	88,6	88,6	88,7
		Long. moy. de l'IP	2,2	2,2	2,2	2,7	2,7	2,7	3,0	3,0	3,0
	STAR(2) MC	Taux d'incl. (%)	90,0	90,1	88,3	90,4	90,6	88,7	89,5	91,1	88,5
		Long. moy. de l'IP	1,6	1,6	1,6	2,4	2,4	2,4	2,8	2,8	2,9
STAR(2) BS	Taux d'incl. (%)	90,2	89,7	<i>88,1</i>	90,6	90,4	<i>88,1</i>	89,1	90,5	88,4	
	Long. moy. de l'IP	1,6	1,6	1,6	2,4	2,4	2,4	2,8	2,8	2,8	
95	AR(2)	Taux d'incl. (%)	95,1	95,1	94,6	94,5	94,9	95,1	94,1	94,0	94,1
		Long. moy. de l'IP	2,6	2,6	2,6	3,2	3,2	3,2	3,5	3,5	3,5
	AR(p)	Taux d'incl. (%)	95,3	95,0	94,7	94,6	94,9	95,1	94,3	94,0	94,4
		Long. moy. de l'IP	2,6	2,6	2,6	3,2	3,2	3,2	3,5	3,6	3,6
	STAR(2) MC	Taux d'incl. (%)	95,0	95,8	94,3	95,1	95,3	<i>93,3</i>	94,4	95,0	<i>93,4</i>
		Long. moy. de l'IP	1,9	1,9	1,9	2,9	2,8	2,9	3,3	3,3	3,3
STAR(2) BS	Taux d'incl. (%)	94,4	95,4	93,8	95,0	94,7	93,9	94,4	94,3	94,0	
	Long. moy. de l'IP	1,9	1,9	1,9	2,9	2,8	2,8	3,3	3,3	3,3	
99	AR(2)	Taux d'incl. (%)	98,8	98,5	98,5	98,7	98,6	<i>98,3</i>	98,5	99,0	98,8
		Long. moy. de l'IP	3,4	3,4	3,4	4,2	4,2	4,2	4,6	4,6	4,6
	AR(p)	Taux d'incl. (%)	98,8	98,5	98,5	98,4	98,5	98,4	98,6	99,0	99,0
		Long. moy. de l'IP	3,4	3,4	3,4	4,2	4,2	4,2	4,7	4,7	4,7
	STAR(2) MC	Taux d'incl. (%)	<i>98,0</i>	99,1	99,0	<i>98,2</i>	98,9	98,5	<i>98,2</i>	98,8	98,5
		Long. moy. de l'IP	2,5	2,5	2,5	3,7	3,6	3,6	4,1	4,2	4,2
STAR(2) BS	Taux d'incl. (%)	<i>98,2</i>	99,2	98,7	<i>98,1</i>	99,0	98,6	<i>98,1</i>	99,1	98,6	
	Long. moy. de l'IP	2,5	2,5	2,5	3,6	3,6	3,6	4,1	4,2	4,2	

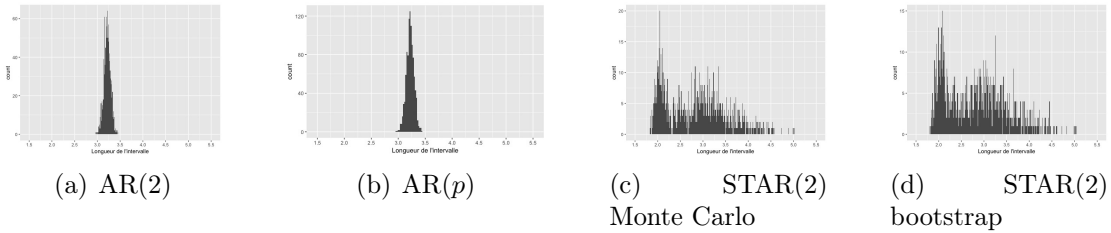


FIGURE 4.13. Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(2) avec $T = 2000$, $\gamma = 2$, $c = 1$.

TABLE 4. XVIII. Taux d'inclusion et longueurs moyennes des intervalles de prévision selon les modèles AR(2), AR(p) et STAR(2) avec $\gamma = 20$ et $c = 1,2$.

Niv.	Modèle	Temps de prév. T	t+1			t+2			t+3		
			400	2000	4000	400	2000	4000	400	2000	4000
80	AR(2)	Taux d'incl. (%)	76,6	76,4	78,1	78,3	79,0	75,8	78,9	76,0	77,4
		Long. moy. de l'IP	4,0	4,0	4,0	4,5	4,5	4,5	4,6	4,6	4,6
	AR(p)	Taux d'incl. (%)	78,3	77,9	80,1	80,0	82,9	79,1	79,9	77,7	78,0
		Long. moy. de l'IP	3,8	3,8	3,8	4,3	4,3	4,3	4,6	4,6	4,6
	STAR(2) MC	Taux d'incl. (%)	80,7	78,7	77,6	80,5	81,2	77,9	79,6	78,3	79,1
		Long. moy. de l'IP	1,3	1,3	1,3	2,9	2,9	3,1	3,5	3,6	3,5
STAR(2) BS	Taux d'incl. (%)	80,8	79,1	77,7	80,0	80,4	78,5	79,7	78,7	79,8	
	Long. moy. de l'IP	1,3	1,3	1,3	2,9	2,9	3,1	3,5	3,6	3,5	
90	AR(2)	Taux d'incl. (%)	89,1	86,6	87,6	90,0	91,0	89,3	89,5	87,6	89,7
		Long. moy. de l'IP	5,1	5,1	5,1	5,7	5,8	5,8	5,9	5,9	5,9
	AR(p)	Taux d'incl. (%)	89,3	87,2	88,7	90,7	91,8	88,7	89,9	87,8	89,0
		Long. moy. de l'IP	4,9	4,9	4,9	5,6	5,5	5,6	5,9	5,9	5,9
	STAR(2) MC	Taux d'incl. (%)	88,7	90,3	88,3	90,5	89,9	87,8	90,8	88,9	89,1
		Long. moy. de l'IP	1,6	1,6	1,6	3,5	3,4	3,6	4,1	4,2	4,1
STAR(2) BS	Taux d'incl. (%)	89,4	89,5	88,5	90,4	90,3	88,1	90,4	89,2	88,8	
	Long. moy. de l'IP	1,6	1,6	1,6	3,5	3,4	3,6	4,1	4,2	4,1	
95	AR(2)	Taux d'incl. (%)	94,6	92,7	94,6	94,7	96,7	95,0	96,2	95,2	96,2
		Long. moy. de l'IP	6,0	6,1	6,1	6,8	6,9	6,9	7,0	7,1	7,1
	AR(p)	Taux d'incl. (%)	94,1	93,4	94,6	94,9	96,3	93,6	95,6	94,4	95,4
		Long. moy. de l'IP	5,9	5,9	5,9	6,6	6,6	6,6	7,0	7,0	7,0
	STAR(2) MC	Taux d'incl. (%)	94,6	95,6	94,2	94,9	96,2	93,4	95,2	94,8	94,8
		Long. moy. de l'IP	1,9	1,9	1,9	3,9	3,8	4,0	4,6	4,6	4,6
STAR(2) BS	Taux d'incl. (%)	94,2	96,2	94,2	95,0	96,2	93,5	94,8	94,0	94,7	
	Long. moy. de l'IP	1,9	1,9	1,9	3,9	3,8	4,0	4,6	4,6	4,6	
99	AR(2)	Taux d'incl. (%)	98,5	99,6	99,1	99,0	99,8	98,9	99,8	99,8	100,0
		Long. moy. de l'IP	7,9	8,0	8,0	9,0	9,0	9,0	9,3	9,3	9,3
	AR(p)	Taux d'incl. (%)	97,9	99,1	98,8	97,9	98,9	97,7	99,6	99,7	99,8
		Long. moy. de l'IP	7,7	7,7	7,7	8,7	8,7	8,7	9,2	9,2	9,2
	STAR(2) MC	Taux d'incl. (%)	97,9	99,0	98,9	98,0	98,7	97,9	98,3	98,8	98,9
		Long. moy. de l'IP	2,5	2,5	2,5	4,6	4,5	4,7	5,4	5,4	5,4
STAR(2) BS	Taux d'incl. (%)	98,2	99,1	98,6	97,8	98,8	98,5	98,4	98,8	98,8	
	Long. moy. de l'IP	2,5	2,5	2,5	4,6	4,5	4,7	5,4	5,4	5,4	

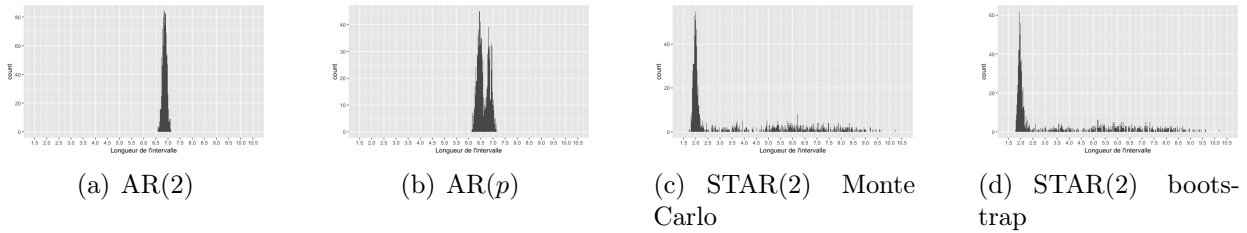


FIGURE 4.14. Distributions empiriques des longueurs d'intervalles de prévision 95% au temps $t + 2$ pour le modèle STAR(2) avec $T = 2000$, $\gamma = 20$, $c = 1,2$.

4.5. Discussion des résultats de simulation sur les prévisions

4.5.1. Prévisions ponctuelles

Les biais et erreurs quadratiques moyennes des prévisions ponctuelles pour les modèles STAR(1) sont présentés aux tableaux 4. VII, 4. VIII, 4. IX et 4. X, et aux tableaux 4. XI et 4. XII pour les modèles STAR(2).

Nous concentrons notre étude des prévisions ponctuelles sur deux questions principales. En premier lieu, nous étudions les modèles STAR par rapport aux modèles linéaires classiques AR, et nous observons si ceux-ci mènent à des meilleures prévisions ponctuelles, tant au niveau du biais que de l'erreur quadratique moyenne. Le modèle STAR(p) qui est supposé stationnaire admet en effet une représentation de type linéaire AR(p), et nous souhaitons étudier si cette représentation mène à de meilleures prévisions. Nous appliquons deux modèles AR(p) : un modèle avec AR(p) où p prend la même valeur que le modèle STAR(p), soit un ou deux, et un modèle AR(p) où la valeur de p varie selon les séries simulées et est sélectionnée selon le critère AIC. En effet, il est possible que la meilleure représentation AR(p) contre le STAR(p) soit celle ayant la même valeur de p , toutefois cela n'est pas certain, et donc il convient aussi de comparer le modèle STAR à un modèle AR(p) avec p choisi avec les données. Ceci permet donc d'étudier les caractéristiques prévisionnelles des modèles STAR(p) en comparaison aux modèles AR(p). En deuxième lieu, nous comparons les performances des différentes méthodes de prévisions pour les modèles STAR(p) entre elles, et nous observons si elles diffèrent les unes des autres au niveau du biais et de l'erreur quadratique moyenne.

Nous commençons par étudier le biais et l'erreur quadratique moyenne pour le temps de prévision $t + 1$. Pour le modèle STAR, les prévisions exactes sont calculées, et les prévisions naïves, de Monte Carlo et bootstrap ne sont pas calculées puisque toutes les valeurs précédentes de la série sont connues à ce temps, comme expliqué au chapitre 3. Ainsi, nous comparons cette prévision du modèle STAR(p) exacte avec celles obtenues par les modèles AR(p). Pour le biais, il est impossible de conclure puisque celui-ci est parfois plus petit pour le modèle STAR et parfois pour les modèles AR, dépendamment du modèle étudié et de la taille échantillonnale de la série. On peut donc considérer que le biais n'est pas mieux pour le modèle STAR que pour le AR au temps $t + 1$. Toutefois, cela n'est surprenant. En effet, le modèle AR est aussi sans biais pour ces prévisions, ce qui fait que même si elles sont éloignées de la vraie valeur, leur moyenne tend tout de même vers la vraie valeur.

Pour l'erreur quadratique moyenne, on remarque par contre que, pour tous les modèles et tailles échantillonnales étudiés, l'erreur quadratique moyenne est constamment plus petite pour le modèle STAR(p) que pour les modèles AR(p) au temps $t + 1$, et ce peu importe le p du modèle AR(p). Selon les modèles, cette différence peut être relativement minime ou plutôt importante. Par exemple, pour le modèle STAR(1) avec $\gamma = 5$, $c = -0,2$ et $T = 400$, la différence entre les erreurs quadratiques moyennes du STAR(1) et AR(1) est de 0,015, et de 0,022 entre le STAR(1) et AR(p), alors que pour le modèle STAR(2) avec $\gamma = 20$, $c = 1,2$ et $T = 2000$, cette différence est de 2,308 entre le STAR(2) et AR(2), et de 2,073 entre le STAR(2) et AR(p). Ainsi, la différence semble plus ou moins majeure selon le modèle étudié et le comportement de la série, mais globalement, l'erreur quadratique moyenne est toujours plus petite pour le modèle STAR(p) que pour les modèles AR(p), le tout au temps $t + 1$.

Par la suite, nous observons les temps $t + 2$ et $t + 3$ conjointement. Nous commençons par étudier le biais selon les différentes méthodes de prévisions du modèle STAR(p), puis nous comparons celles-ci avec les modèles AR(p). Ainsi, nous observons tout d'abord les biais aux temps $t + 2$ et $t + 3$ entraînés par les méthodes de prévisions exactes, naïves, de Monte Carlo et bootstrap. Globalement, on remarque que les méthodes de prévisions exactes, de Monte Carlo et bootstrap mènent à des valeurs très similaires dans tous les cas de modèles et de tailles échantillonnales pour les séries. Ainsi, il ne semble pas y avoir de différences majeures au niveau du biais empirique selon la méthode de prévision sans biais employée pour effectuer les prévisions, bien que les biais de ces prévisions ne soient pas empiriquement nuls comme il serait attendu selon la théorie. Pour les modèles STAR(p) uniquement, la seule différence remarquable au niveau du biais est par conséquent pour la méthode de prévision naïve. Comme mentionné au chapitre 3, cette méthode est biaisée à partir du temps $t + 2$. Ainsi, le biais moyen amené par cette méthode est, sauf exceptions, généralement plus grand que celui des méthodes exactes, de Monte Carlo et de bootstrap. Toutefois, il existe certains cas où le biais empirique moyen est moindre pour les prévisions naïves, par exemple pour la prévision à $t + 3$ du modèle STAR(2) avec $\gamma = 20$ et $c = 1,2$ au tableau 4. XII, et pour la prévision $t + 2$ du modèle STAR(1) avec $\gamma = 5$, $c = -0,5$ et $T = 2000$ au tableau 4. VIII. Par contre, il est difficile de conclure pour les cas spécifiques où le biais moyen obtenu par la méthode naïve peut être plus petit que celui des autres méthodes, puisque les biais empiriques plus petits semblent être dus au hasard et n'ont pas de comportements remarquables, et que dans la plupart des cas la méthode naïve présente un plus grand biais que les autres méthodes pour le modèle STAR.

Nous pouvons maintenant regarder les biais obtenus selon les modèles $AR(p)$ et comment ceux-ci se comparent à ceux obtenus par les méthodes de prévision du modèle $STAR(p)$. Généralement, on observe que le biais empirique entraîné par les modèles $AR(p)$ est plus petit en valeur absolue que celui entraîné par les prévisions naïves pour le modèle $STAR(p)$, mais qu'il est plus grand que ceux entraînés par les méthodes de prévisions exacte, de Monte Carlo et bootstrap. Toutefois, cette tendance observable n'est pas généralisée, et il y a un nombre non-négligeable de modèles où les prévisions par $AR(p)$ sont plus petites que celles de toutes les méthodes de prévision du modèle $STAR$. De plus, il y a certains cas où le modèle $AR(p)$ mène à des prévisions plus biaisées que toutes les méthodes de prévisions pour le modèle $STAR(p)$. Ainsi, il est difficile de conclure que le modèle $STAR(p)$ mène à un meilleur ou pire biais que le modèle $AR(p)$ en raison de l'absence de comportement généralisé pour les résultats, puisque cela semble varier autant selon le modèle que la taille échantillonnale de la série. Toutefois, on peut supposer qu'en général, le modèle $AR(p)$ est mieux au niveau du biais que la méthode de prévision naïve, mais pas que les méthodes de prévision exactes, de Monte Carlo et bootstrap.

Par la suite, nous observons l'erreur quadratique moyenne obtenue selon les différentes méthodes de prévisions pour le modèle $STAR(p)$. On remarque en observant les résultats que l'erreur quadratique moyenne obtenue selon la méthode de prévision naïve est systématiquement plus grande que celle obtenue par les méthodes de prévisions exactes, de Monte Carlo et bootstrap. Une seule exception existe, au tableau 4. XI, où la prévision naïve a une EQM très légèrement inférieure à celles des autres méthodes de prévisions. Autrement, la prévision naïve mène toujours à une plus grande EQM, et dans certains cas cette différence peut être très importante, par exemple au tableau 4. XII où les différences entre la méthode naïve et les autres méthodes pour le modèle $STAR(p)$ sont grandes pour tous les temps et tailles échantillonales de la série. De plus, on remarque que les méthodes de prévisions exactes, de Monte Carlo et bootstrap ne semblent pas avoir de grandes différences entre elles au niveau de l'erreur quadratique moyenne qu'elles entraînent. En effet, leurs valeurs sont très semblables dans pratiquement tous les cas, et lorsque des différences minimales existent, elles ne semblent pas avoir un comportement précis qui permettrait de conclure qu'une de ces trois méthodes est meilleure qu'une autre au niveau de l'erreur quadratique moyenne.

Finalement, nous comparons les erreurs quadratiques moyennes obtenues selon les modèles $AR(p)$ avec celles obtenues par le modèle $STAR(p)$. Globalement, on remarque que les modèles $AR(p)$ mènent à une erreur quadratique moyenne plus élevée que toutes les méthodes

de prévisions pour le modèle STAR(p), sauf pour quelques rares exceptions où la méthode de prévision naïve a une erreur quadratique moyenne plus grande que les modèles AR(p). Ainsi, les modèles AR(p) donnent des erreurs quadratiques moyennes généralement proches l'une de l'autre, systématiquement plus grandes que les méthodes de prévisions exacte, de Monte Carlo et bootstrap pour le modèle STAR(p), et parfois cette différence peut être relativement importante. Le tout signifie que, bien que le modèle AR(p) ne mène pas nécessairement à une prévision plus biaisée, tel qu'élaboré précédemment, les prévisions obtenues par ce modèle sont en moyenne plus éloignées de la vraie valeur que les prévisions pouvant être obtenues par le modèle STAR(p). Par conséquent, l'ajustement d'un modèle STAR(p) aux données plutôt qu'un modèle linéaire classique AR(p) mène à des prévisions plus exactes globalement.

4.5.2. Intervalles de prévision

Les résultats pour les intervalles de prévision empiriques sont présentés aux tableaux 4. XIII, 4. XIV, 4. XV et 4. XVI pour le modèle STAR(1), et aux tableaux 4. XVII et 4. XVIII pour le modèle STAR(2). Ces tableaux comprennent les taux d'inclusion des vraies valeurs dans les intervalles calculés ainsi que la longueur moyenne des intervalles, le tout selon les intervalles obtenus selon les modèles AR(p) et les intervalles obtenus par Monte Carlo et bootstrap selon le modèle STAR(p). Si les taux d'inclusion sont statistiquement différents à 5% du taux visé, ils sont indiqués en italiques dans les tableaux.

Nous commençons par étudier si les taux d'inclusion des vraies valeurs correspondent aux taux d'inclusion visés pour les différents niveaux. Dans le cas des modèles AR(p), les intervalles calculés sont théoriquement sans biais pour le niveau visé, tel qu'élaboré au chapitre 3. Ainsi, on remarque que les taux d'inclusion pour le modèle AR(p) avec p déterminé selon le critère AIC sont presque toujours statistiquement non-différents du taux visé à 95%. Pour le modèle AR(p) avec p équivalant à celui du modèle STAR(p), soit un et deux, les taux sont plus fréquemment statistiquement différents des taux visés. Ceci est sans doute dû au fait que, dans certain cas, l'imposition de la valeur de p ne permet pas de bien représenter la série, et donc de bien prédire les intervalles des futures valeurs. Pour les modèles STAR(p), on observe que les taux d'inclusion sont parfois plus petits que les taux visés, bien que généralement ils ne soient pas statistiquement différents du taux visé. Ainsi, bien que les intervalles de Monte Carlo et de notre méthode de type bootstrap soient possiblement biaisés, il est difficile de conclure que ce biais est marqué à la vue de nos résultats empiriques.

Par la suite, bien que les taux d'inclusion varient selon les différentes méthodes, nous les considérons tous comme semblables pour l'analyse qui suit. Cela permet de comparer les longueurs moyennes des intervalles de prévision selon les différentes méthodes étudiées. Nous commençons par comparer les méthodes de prévision pour le modèle $\text{STAR}(p)$ entre elles, soit les intervalles obtenus par les méthodes de Monte Carlo et notre méthode de type bootstrap. On remarque que, pour un même modèle, niveau d'intervalle et taille échantillonnale de la série, les longueurs moyennes des intervalles de prévision par Monte Carlo et bootstrap semblent toujours mener à des valeurs très similaires, avec aucune de ces méthodes ne menant constamment à une plus petite ou plus grande valeur que l'autre. Ainsi, les méthodes de Monte Carlo et bootstrap pour les intervalles de prévision semblent mener à des longueurs d'intervalles constamment très semblables.

On peut maintenant comparer les longueurs moyennes des intervalles de prévision pour les modèles $\text{AR}(p)$ contre celles obtenues pour le modèle $\text{STAR}(p)$ par Monte Carlo et bootstrap. On remarque tout d'abord que le modèle $\text{AR}(p)$ avec p sélectionné selon le critère AIC mène généralement à des intervalles moyens plus petits que le modèle $\text{AR}(p)$ avec le p étant le même que le modèle $\text{STAR}(p)$, sauf pour quelques exceptions lorsque $T = 400$. Cela n'est pas surprenant, puisque le p sélectionné spécifiquement pour chaque série permet de mieux représenter chaque série et donc d'obtenir de meilleures prévisions. Toutefois, peu importe le p dans le modèle $\text{AR}(p)$, celui-ci mène systématiquement à une longueur d'intervalle moyenne plus grande que le modèle $\text{STAR}(p)$, et ce pour tous les modèles étudiés. De plus, cette différence est toujours présente, mais selon le modèle elle peut être minime ou relativement importante. Par exemple, pour le temps de prévision $t + 1$ du modèle $\text{STAR}(1)$ avec $\gamma = 5$, $c = -0,2$ et $T = 4000$ au niveau 80% au tableau 4. XIII, la longueur moyenne de l'intervalle pour le modèle $\text{AR}(1)$ est de 1,3, et de 1,3 pour le modèle $\text{STAR}(1)$. La longueur d'intervalle moyenne du modèle $\text{AR}(1)$ est ici 4,2% plus élevée que celle du modèle $\text{STAR}(1)$. À l'opposé, pour le modèle $\text{STAR}(2)$ avec $\gamma = 20$, $c = 1,2$ et $T = 4000$ au niveau 80% au tableau 4. XVIII, la longueur moyenne de l'intervalle pour le modèle $\text{AR}(2)$ est de 4,0, contre 1,3 pour le modèle $\text{STAR}(2)$. La longueur d'intervalle moyenne du modèle $\text{AR}(2)$ est donc 212% plus élevée que celle du modèle $\text{STAR}(2)$. Généralement, il est donc avantageux au niveau de l'intervalle de prévision d'utiliser les modèles $\text{STAR}(p)$ plutôt que $\text{AR}(p)$ lorsque le modèle sous-jacent est un modèle $\text{STAR}(p)$, puisque cela ne mène pas nécessairement à une différence au niveau de la performance des prévisions ponctuelles, tel que discuté dans la section 4.5.1, mais que cela peut grandement réduire l'intervalle de prévision obtenu pour les différents temps de prévisions.

Finalement, nous pouvons observer les distributions empiriques des longueurs des intervalles de prévision pour les différents modèles et méthodes de prévision. Les figures 4.9, 4.10, 4.11, 4.12, 4.13 et 4.14 montrent ces distributions pour les intervalles 95% pour le temps de prévision $t + 2$ avec $T = 2000$. Pour un modèle spécifié, les graphiques des intervalles pour les deux modèles $AR(p)$ et les modèles $STAR(p)$ avec Monte Carlo et bootstrap comportent la même échelle pour l'axe horizontal afin de rendre la visualisation plus intuitive.

En observant ces graphiques, on remarque que les modèles $AR(p)$ mènent toujours à des distributions empiriques en forme de pic et très concentrées autour d'une certaine valeur. Ainsi, les intervalles obtenus par les modèles $AR(p)$ ont tous des longueurs similaires pour les différentes séries simulées à partir d'un même modèle. Pour les modèles $STAR(p)$ avec méthodes de prévisions Monte Carlo et bootstrap, les distributions sont beaucoup plus irrégulières et dispersées sur une plus grande étendue de valeurs, et se ressemblent grandement pour les deux méthodes. De plus, les distributions sont généralement asymétriques avec deux modes présents dans les données. Selon les modèles, ces deux modes sont plus ou moins remarquables, par exemple la figure 4.13 où les deux modes sont très distinguables, contre la figure 4.11 où ils le sont très peu. En particulier, les modèles $STAR(2)$ étudiés ont une très grande étendue de valeurs pour les intervalles, mais avec un pic très marqué à une petite longueur d'intervalle. La présence de ces deux modes dans les distributions empiriques est due à la particularité du modèle $STAR(p)$ de comprendre deux régimes présents en alternance. Ainsi, selon la valeur de la fonction g au dernier temps connu, l'intervalle prédit sera plus précis puisque le régime sous-jacent présentement en fonction représente alors mieux les données qu'un modèle général $AR(p)$, et que le modèle $STAR(p)$ parvient à représenter ce régime sous-jacent. Cette plus grande précision dans les intervalles de prévision est ce qui cause les deux modes de la distribution : si le régime en fonction est peu variable, l'intervalle prédit sera généralement petit, et si le régime en fonction est très variable, l'intervalle de prévision sera beaucoup plus grand. C'est ce phénomène qui permet de simultanément parvenir à un intervalle moyen plus petit, tel que vu dans les tableaux précédemment, et d'obtenir des valeurs d'intervalles qui peuvent être beaucoup plus importantes que ce qui est observable pour les modèles $AR(p)$.

Chapitre 5

Analyse de données réelles

Dans ce chapitre, nous appliquons la théorie des modèles $\text{STAR}(p)$ élaborée aux chapitres 1, 2 et 3 sur un jeu de données réel. Les données utilisées proviennent de Yahoo Finance et représentent les rendements quotidiens du fonds négocié en bourse SPDR (pour *Standard & Poor's Depository Receipts*) suivant l'indice boursier S&P 500 (*SPY*) entre le 1er janvier 2010 et le 31 décembre 2018. Dans un premier lieu, nous spécifions le modèle en déterminant le délai p et en testant les données pour la linéarité, en plus de sélectionner la variable de transition s_t et la forme de la fonction de transition g . Par la suite, nous procédons à l'estimation des paramètres du modèle $\text{STAR}(p)$ par optimisation non-linéaire, en portant attention aux difficultés calculatoires mentionnées à la section 2.2.1. Notamment, nous procédons par recherche quadrillée pour trouver les valeurs initiales fournies à l'algorithme. Ensuite, nous évaluons le modèle en testant les hypothèses faites lors de l'ajustement. Finalement, nous étudions les prévisions pouvant être faites pour cette série de données pour les temps $t + 1, \dots, t + 10$, en appliquant le modèle $\text{AR}(p)$ et le modèle ajusté $\text{STAR}(p)$ aux données. Puis, nous comparons les prévisions ponctuelles obtenues par ces modèles à l'aide de mesures du biais et de l'erreur quadratique moyenne. Entre autres, pour le modèle $\text{STAR}(p)$, nous effectuons les prévisions naïves, de Monte Carlo et par la méthode de bootstrap. De plus, nous calculons les intervalles de prévision obtenus par le modèle $\text{AR}(p)$ ainsi que ceux pouvant être obtenus par le modèle $\text{STAR}(p)$ via les méthodes de Monte Carlo et notre méthode de type bootstrap. Nous observons alors si la vraie valeur de la série au temps futur est incluse dans l'intervalle ou non selon le modèle ajusté et la méthode de prévision employée. Finalement, nous discutons de notre analyse et de nos résultats en les comparant à ceux disponibles dans la littérature, en élaborant autant au niveau des difficultés présentes lors de l'estimation du modèle qu'au niveau de la performance des différentes méthodes de prévisions ponctuelles et des intervalles de prévision obtenus. Le code R utilisé pour cette section est disponible sur demande.

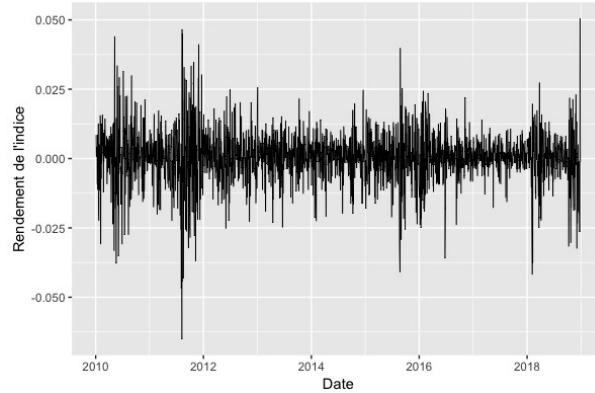


FIGURE 5.1. Rendement de l'indice SPDR S&P 500 (SPY) selon la date

5.1. Description du jeu de données

Le jeu de données utilisé est une série temporelle obtenue sur le site web de Yahoo Finance. La série comporte les rendements quotidiens du fonds négocié en bourse SPDR (pour *Standard & Poor's Depository Receipts*) suivant l'indice boursier S&P 500 (*SPY*) entre le 1er janvier 2010 et le 31 décembre 2018. Les rendements sont disponibles pour tous les jours d'ouverture des marchés durant cette période, ce qui donne un total de 2263 valeurs disponibles dans la série temporelle. La première valeur disponible est celle du 4 janvier 2010, et la dernière est celle du 28 décembre 2018. La série est illustrée à la figure 5.1. Les données ont été téléchargées en mai 2019 et sont disponibles sur demande.

5.2. Ajustement d'un modèle $\text{STAR}(p)$

L'ajustement d'un modèle $\text{STAR}(p)$ s'effectue tel que décrit dans le chapitre 2. Nous commençons ainsi par la spécification du modèle, suivi de l'estimation des paramètres, pour terminer par l'évaluation du modèle.

5.2.1. Spécification du modèle

5.2.1.1. Sélection du délai p

La première étape consiste à sélectionner le délai p approprié pour ce jeu de données. Pour ce faire, nous ajustons un modèle $\text{AR}(p)$ tel qu'à la définition 1.2.1, et ce pour $p \in \{1, \dots, 20\}$, puis nous observons la valeur du critère AIC pour chaque modèle. Ces valeurs observées sont présentées au tableau 5. I. Le modèle $\text{AR}(5)$ étant celui qui présente le plus petit AIC, avec une valeur de -14 692,4, nous sélectionnons donc un délai de $p = 5$ pour ce jeu de données, ce qui mène par la suite à l'ajustement d'un modèle $\text{STAR}(5)$.

TABLE 5. I. Valeur du critère AIC obtenue selon le délai p dans le modèle $AR(p)$

Délai p	AIC	Délai p	AIC
1	-14 684,0	11	-14 687,6
2	-14 682,0	12	-14 685,6
3	-14 684,0	13	-14 683,6
4	-14 682,4	14	-14 688,7
5	-14 692,4	15	-14 689,5
6	-14 690,9	16	-14 688,1
7	-14 689,1	17	-14 689,1
8	-14 690,3	18	-14 687,9
9	-14 689,6	19	-14 690,2
10	-14 689,0	20	-14 689,6

5.2.1.2. *Test de linéarité, choix de la variable de transition s_t et de la fonction de transition g*

La deuxième étape consiste à tester les données pour la linéarité avant de procéder à l'ajustement d'un modèle STAR(5). Tout d'abord, nous ajustons un modèle linéaire classique avec un délai de 5 sur les données, et l'on calcule la somme des résidus au carré pour ce modèle. On obtient ainsi $SSR_0 = 0,1994$. Il est important de mentionner ici que l'ajustement de ce modèle et des subséquents mènent à 2258 valeurs résiduelles plutôt que 2263, la régression linéaire ne pouvant être effectuée pour les cinq premières valeurs de la série en raison du manque de valeurs précédentes dans la série pour calculer ces résidus.

Par la suite, nous ajustons un modèle avec un développement de Taylor d'ordre un de la fonction logistique g sur les données. Comme la variable de transition s_t n'est pas encore définie, on ajuste cinq modèles, soit un pour chaque possibilité considérée pour s_t . Ici, nous considérons les valeurs autorégressives seulement comme possibilités pour la variable de transition, soit $s_t = y_{t-1}$, $s_t = y_{t-2}$, $s_t = y_{t-3}$, $s_t = y_{t-4}$ ou $s_t = y_{t-5}$. Les modèles prennent donc la forme :

$$y_t = \beta'_0 \mathbf{x}_t + \beta'_1 \mathbf{x}_t s_t + e_t, \quad (5.2.1)$$

où $\beta_i = (\beta_{i,0}, \beta_{i,1}, \beta_{i,2}, \beta_{i,3}, \beta_{i,4}, \beta_{i,5})'$ sont les paramètres de la régression linéaire ajustée ; $\mathbf{x}_t = (1, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5})'$; et s_t est soit y_{t-1} , y_{t-2} , y_{t-3} , y_{t-4} ou y_{t-5} . Comme mentionné précédemment, on retire $\beta_{1,0} s_t$ du modèle pour éviter la multicollinéarité. Nous effectuons donc le test de linéarité pour chacune des possibilités de s_t . Par exemple, pour la variable de transition $s_t = y_{t-1}$, le modèle linéaire ajusté sur les données prend la forme suivante :

$$\begin{aligned}
y_t = & \beta_{0,0} + \beta_{0,1}y_{t-1} + \beta_{0,2}y_{t-2} + \beta_{0,3}y_{t-3} + \beta_{0,4}y_{t-4} + \beta_{0,5}y_{t-5} \\
& + \beta_{1,1}y_{t-1}^2 + \beta_{1,2}y_{t-1}y_{t-2} + \beta_{1,3}y_{t-1}y_{t-3} + \beta_{1,4}y_{t-1}y_{t-4} + \beta_{1,5}y_{t-1}y_{t-5} + e_t,
\end{aligned} \tag{5.2.2}$$

et $SSR_1^{(y_{t-1})}$ correspond à la somme des résidus au carré pour ce modèle. Le degré de liberté pour le test est de $p + 1 = 6$ puisque la taille échantillonnale T de la série est grande. Les résultats obtenus sont présentés au tableau 5. II.

TABLE 5. II. Valeur- p du test de linéarité selon le choix de la variable de transition s_t

Choix de s_t	$SSR_1^{(s_t)}$	Valeur- p
y_{t-1}	0,1981	0,0214
y_{t-2}	0,1987	0,2945
y_{t-3}	0,1966	<0,0001
y_{t-4}	0,1984	0,1017
y_{t-5}	0,1970	0,0001

Le choix de $s_t = y_{t-3}$ rejetant la linéarité au niveau 5% et donnant la plus petite valeur- p , c'est cette composante qui est choisie comme variable de transition s_t pour le modèle STAR(5).

En observant les données, nous pensons que le choix de la fonction g de forme logistique est raisonnable. Nous ajusterons donc le modèle avec cette fonction, et si celle-ci s'avère inappropriée à l'étape de l'évaluation du modèle, alors nous ajusterons le modèle avec la fonction g de forme exponentielle. Le modèle STAR(5) ajusté aux sections subséquentes est donc de la forme suivante :

$$\begin{aligned}
y_t = & \phi_1' \mathbf{x}_t + \psi_1' \mathbf{x}_t g(y_{t-3}, \gamma, c) + \varepsilon_t \\
= & \phi_{1,0} + \phi_{1,1}y_{t-1} + \phi_{1,2}y_{t-2} + \phi_{1,3}y_{t-3} + \phi_{1,4}y_{t-4} + \phi_{1,5}y_{t-5} \\
& + (\psi_{1,0} + \psi_{1,1}y_{t-1} + \psi_{1,2}y_{t-2} + \psi_{1,3}y_{t-3} + \psi_{1,4}y_{t-4} + \psi_{1,5}y_{t-5}) \\
& \times [1 + \exp(-\gamma(y_{t-3} - c))]^{-1} + \varepsilon_t.
\end{aligned} \tag{5.2.3}$$

Maintenant que l'étape d'ajustement du modèle est complétée et que le modèle à estimer est posé tel qu'à l'équation (5.2.3), il convient de procéder à l'estimation des paramètres tel qu'élaboré à la section 2.2.

5.2.2. Estimation des paramètres

Il est maintenant possible d'estimer les paramètres du modèle STAR(5) avec variable de transition $s_t = y_{t-3}$ et fonction de transition g de forme logistique tel que présenté à l'équation (5.2.3). L'estimation se fait par optimisation non-linéaire telle qu'élaborée à la section 2.2, par conséquent il convient tout d'abord de déterminer les valeurs initiales fournies à l'algorithme à l'aide d'une recherche quadrillée. Pour le paramètre γ , les valeurs considérées vont de 1 à 2000 avec un écart de 2 entre chaque, soit l'ensemble $\{1; 3; \dots; 1997; 1999\}$. Pour le paramètre c , nous considérons les valeurs générales prises par la série y_t , soit entre -0,03 et 0,03, avec un écart de 0,005 entre chaque valeur, soit l'ensemble $\{-0,03; -0,025; \dots; 0,025; 0,03\}$. Suite à la recherche quadrillée, les valeurs obtenues minimisant le critère AIC sont 957 et -0,03 pour γ et c respectivement. Les valeurs initiales pour les paramètres linéaires sont subséquentement sélectionnées par régression linéaire à partir de ces valeurs de γ et c . L'optimisation non-linéaire est ainsi effectuée à partir de ces valeurs initiales.

Suite à l'estimation par optimisation non-linéaire et au calcul de la matrice de variances-covariances, le modèle obtenu est le suivant :

$$y_t = \hat{\phi}'_1 \mathbf{x}_t + \hat{\psi}'_1 \mathbf{x}_t g(y_{t-3}, \hat{\gamma}, \hat{c}) + e_t, \quad (5.2.4)$$

avec les valeurs des paramètres estimés ainsi que leurs erreurs standards basées sur l'estimé de la matrice C étant présentées au tableau 5. III pour le modèle initial.

On remarque au tableau 5. III que les estimés des paramètres $\phi_{1,4}$, $\psi_{1,4}$ et γ sont statistiquement non-significatifs. Nous retirons ainsi les paramètres linéaires $\phi_{1,4}$ et $\psi_{1,4}$ du modèle, toutefois nous conservons γ dans le modèle même si son erreur standard est très grande. En effet, cette grande erreur est due à l'incertitude associée aux grandes valeurs de γ , tel que discuté au chapitre 2. De plus, le fait que γ est non-significatif ne peut être considéré comme une preuve de non-linéarité faible pour le modèle, voir Dijk, Teräsvirta et Franses (2002).

TABLE 5. III. Estimation des paramètres du modèle STAR(5) ainsi que leurs intervalles de confiance 95%

Paramètre	Estimé	Err. std.	IC 95% Inf.	IC 95% Sup.
$\phi_{1,0}$	-0,035	0,018	-0,070	-0,001
$\phi_{1,1}$	-0,226	0,084	-0,391	-0,061
$\phi_{1,2}$	-0,686	0,308	-1,290	-0,083
$\phi_{1,3}$	-1,135	0,444	-2,006	-0,264
$\phi_{1,4}$	0,049	0,221	-0,385	0,483
$\phi_{1,5}$	-0,482	0,186	-0,846	-0,118
$\psi_{1,0}$	0,036	0,018	0,001	0,071
$\psi_{1,1}$	0,205	0,090	0,029	0,380
$\psi_{1,2}$	0,698	0,311	0,088	1,308
$\psi_{1,3}$	1,131	0,442	0,264	1,998
$\psi_{1,4}$	-0,064	0,223	-0,501	0,374
$\psi_{1,5}$	0,436	0,189	0,066	0,806
γ	957,000	2257,465	-3467,630	5381,631
c	-0,030	0,003	-0,035	-0,025

Le modèle STAR(5) final ajusté aux données est donc :

$$\begin{aligned}
 y_t = & -0,035 - 0,226y_{t-1} - 0,686y_{t-2} - 1,135y_{t-3} - 0,482y_{t-5} \\
 & + (0,036 + 0,205y_{t-1} + 0,698y_{t-2} + 1,131y_{t-3} + 0,436y_{t-5}) \\
 & \times [1 + \exp \{-957,000(y_{t-3} + 0,030)\}]^{-1} + e_t.
 \end{aligned} \tag{5.2.5}$$

On remarque que les régimes possèdent des valeurs similaires mais opposées pour les paramètres linéaires influençant les mêmes valeurs autorégressives. Par exemple, $\phi_{1,0} = -0,035$ et $\psi_{1,0} = 0,036$. Ceci signifie que, lorsque la fonction g est de 1, les coefficients du régime en fonction sont presque nuls, et donc que le régime entraîne peu de variation dans la série. Pour observer ce phénomène, nous pouvons décortiquer les deux régimes présents dans le modèle, en observant le comportement du modèle selon les valeurs extrêmes prises par la fonction g , soit 0 et 1. Si la fonction g vaut 0, le régime en fonction est :

$$y_t = -0,035 - 0,226y_{t-1} - 0,686y_{t-2} - 1,135y_{t-3} - 0,482y_{t-5} + e_t, \tag{5.2.6}$$

et si la fonction g est de 1, le régime en fonction est alors :

$$y_t = -0,022y_{t-1} + 0,012y_{t-2} - 0,004y_{t-3} - 0,046y_{t-5} + e_t. \tag{5.2.7}$$

Ceci confirme que l'un des régimes dans le modèle amène de grandes variations dans la série, alors que l'autre régime est caractérisé par une évolution beaucoup plus constante de la série dans le temps, soit avec de très faibles variations.

Par la suite, pour avoir une bonne idée du modèle ajusté et de son application sur les données, nous pouvons observer différents graphiques. Entre autres, la figure 5.2 présente l'évolution de la fonction g dans le temps. Elle montre que les deux régimes sont bien présents en alternance dans les données, et qu'il y a un nombre non-négligeable de points présents dans les transitions entre les régimes. Ceci confirme que le modèle STAR(5) représente adéquatement les données et que les estimés des paramètres non-linéaires de la fonction g , soit γ et c , sont relativement fiables en raison d'un nombre suffisant de valeurs dans la série permettant de les estimer, tel que mentionné à la section 2.2.1. De plus, on observe que la fonction g n'est pas souvent nulle. Comme la fonction g nulle est associée à un régime avec de fortes perturbations, soit celui présenté à l'équation (5.2.6), cela signifie que la série est généralement stable, avec quelques périodes marquées par de plus fortes perturbations. Ceci est également observable à la figure 5.1.

On peut également regarder le graphique de la fonction g selon la valeur de la variable de transition s_t , lequel est présenté à la figure 5.3, et qui permet de voir quelle est la valeur prise par la fonction g lorsque la variable de transition a une certaine valeur au temps t . Ce graphique montre ainsi que la fonction de transition g est quelque peu lisse. Évidemment, la fonction g vaut 0,5 lorsque $s_t = c$, puisque le paramètre c marque le point milieu de transition entre les deux régimes, comme mentionné à la section 1.3.

Finalement, nous pouvons observer le graphique de densité des résidus du modèle ajusté de l'équation (5.2.5) présenté à la figure 5.4. Entre autres, nous remarquons que les résidus sont regroupés autour de 0 mais ne semblent pas suivre exactement une densité normale. Ceci sera discuté lors de l'application des prévisions de Monte Carlo effectuées à la section 5.3.

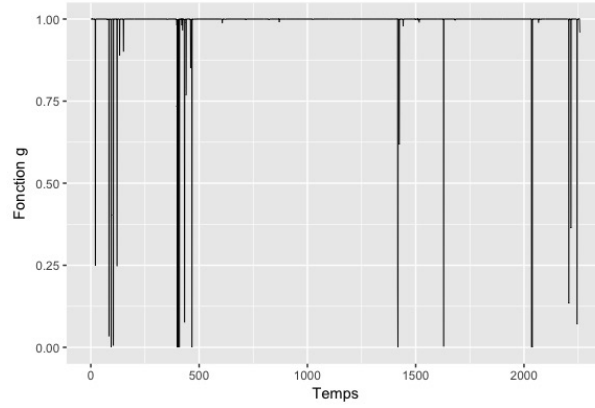


FIGURE 5.2. Valeur prise par la fonction g selon le temps pour le jeu de données *SPY*

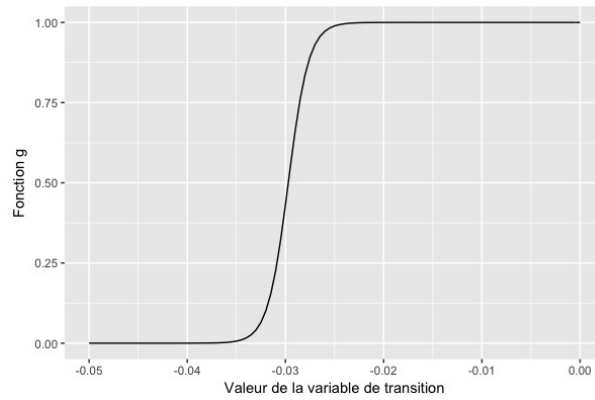


FIGURE 5.3. Valeur prise par la fonction g selon la valeur s_t de la série pour le jeu de données *SPY*

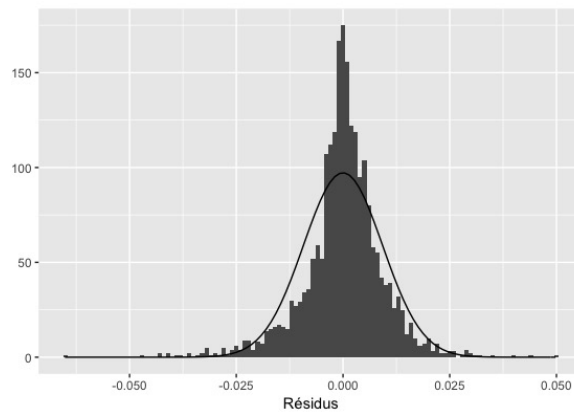


FIGURE 5.4. Densité des résidus du modèle STAR(5) pour le modèle ajusté

5.2.3. Évaluation du modèle

Le modèle STAR(5) étant maintenant ajusté, on procède à l'évaluation de celui-ci. Cette étape consiste à valider les hypothèses faites lors de l'ajustement du modèle, soit les hypothèses présentées à la section 2.3.

5.2.3.1. Absence d'autocorrélation des résidus

Les valeurs- p du test de Ljung-Box sont toutes supérieures à 5% pour les délais de 1 à 19. La valeur- p est très légèrement inférieure à 5% pour le délai 20. Par conséquent, il ne semble pas y avoir de forte autocorrélation présente entre les résidus du modèle. Les valeurs- p sont présentées dans le tableau 5. IV.

TABLE 5. IV. Valeur- p obtenue selon le délai m pour le test de Ljung-Box

Délai m	Valeur- p	Délai m	Valeur- p
1	0,892	11	0,626
2	0,959	12	0,707
3	0,989	13	0,771
4	0,959	14	0,187
5	0,979	15	0,150
6	0,984	16	0,155
7	0,844	17	0,059
8	0,602	18	0,078
9	0,678	19	0,056
10	0,718	20	0,042

5.2.3.2. Absence de non-linéarité restante

Pour observer s'il y a présence de non-linéarité restante suite à l'ajustement du modèle STAR(5) à deux régimes, nous ajustons le modèle suivant :

$$y_t = \phi_1' \mathbf{x}_t + \psi_1' \mathbf{x}_t g_1(y_{t-3}, \gamma_1, c_1) + \beta_1' \mathbf{x}_t y_{t-3} + \beta_2' \mathbf{x}_t y_{t-3}^2 + \beta_3' \mathbf{x}_t y_{t-3}^3 + e_t,$$

où, comme nous utilisons maintenant le modèle réduit, $\phi_1 = (\phi_{1,0}, \phi_{1,1}, \phi_{1,2}, \phi_{1,3}, \phi_{1,5})'$, $\psi_1 = (\psi_{1,0}, \psi_{1,1}, \psi_{1,2}, \psi_{1,3}, \psi_{1,5})'$, et $\mathbf{x}_t = (1, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-5})'$. Comme la première partie de ce modèle a déjà été ajustée, cela revient à ajuster la deuxième partie du modèle aux résidus obtenus par l'ajustement du modèle STAR(5) :

$$e_t^{STAR(5)} = \beta_1' \mathbf{x}_t y_{t-3} + \beta_2' \mathbf{x}_t y_{t-3}^2 + \beta_3' \mathbf{x}_t y_{t-3}^3 + e_t. \quad (5.2.8)$$

Nous retirons les termes $\beta_{2,0}y_{t-3}^2$ et $\beta_{3,0}y_{t-3}^3$ du modèle ci-haut afin d'éviter la multicolinéarité. On calcule la somme au carré des résidus de ce modèle, ce qui correspond à $SSR_1 = 0,1929$. Le test présenté à la section 2.3.2 est ainsi effectué avec $3(p+1) = 18$ degrés de liberté et la valeur- p obtenue est de 0,924, ce qui n'est pas significatif au niveau 5%. Ainsi, il ne semble pas y avoir de non-linéarité restante après l'ajustement du modèle STAR(5). On considère donc ce modèle à deux régimes comme satisfaisant pour le jeu de données *SPY*.

5.2.3.3. Constance des paramètres

Le test de l'hypothèse de constance des paramètres présenté à la section 2.3.3 consiste ici à ajuster le modèle :

$$e_t^{STAR(5)} = \beta'_1 \mathbf{x}_t t + \beta'_2 \mathbf{x}_t t^2 + \beta'_3 \mathbf{x}_t t^3 + (\beta'_4 \mathbf{x}_t t + \beta'_5 \mathbf{x}_t t^2 + \beta'_6 \mathbf{x}_t t^3) g_1(s_{1,t}, \gamma_1, c_1) + e_t, \quad (5.2.9)$$

où, encore une fois, $\mathbf{x}_t = (1, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-5})'$. Toutefois, en raison du grand nombre de valeurs dans la série et des limites informatiques, les termes comprenant t^3 ne peuvent être ajustés dans le modèle puisque cela entraîne une erreur au niveau programmatore. Par conséquent, nous divisons tous les temps t par 1000, ce qui rend le calcul possible et n'affecte ultimement pas la somme des résidus au carré. On ajuste donc le modèle de l'équation (5.2.9), et la somme des résidus au carré obtenue est de $SSR_1 = 0,1905$. Avec un degré de liberté de 36, la valeur- p obtenue est de 0,367, et donc on ne rejette pas l'hypothèse de constance des paramètres du modèle.

Toutes les hypothèses du modèles sont donc testées, et l'on peut procéder à la prévision de valeurs futures pour le jeu de données basée sur le modèle ajusté obtenu.

5.3. Prévisions de valeurs futures

Nous commençons par obtenir les vraies valeurs futures de la série. Ceci nous permettra par la suite de calculer les biais et erreurs quadratiques moyennes des prévisions effectuées, et de voir si les vraies valeurs sont comprises dans les intervalles de prévision obtenus. Les valeurs futures utilisées sont celles du rendement quotidien de l'indice SPDR S&P 500 (*SPY*) pour les dix premiers jours ouvrables du mois de janvier 2019, soit les dix valeurs qui suivent directement la série analysée. Les valeurs futures vont donc du 2 janvier 2019 au 15 janvier 2019 inclusivement, avec aucune valeur pour les 5, 6, 12 et 13 janvier 2019. Nous concentrons l'étude sur dix temps de prévisions futurs seulement puisque ceci correspond à deux semaines en jours ouvrables. De plus, cela devrait permettre d'étudier des prévisions suffisamment lointaines dans le temps pour voir des différences éventuelles entre les méthodes de prévision puisque le modèle étudié est un STAR(5).

Nous appliquons les différentes méthodes de prévision élaborées au chapitre 3, soit les prévisions naïves, de Monte Carlo et notre méthode de type bootstrap pour le modèle STAR(5), en plus des prévisions obtenues avec le modèle linéaire classique AR(5). Nous observons des mesures de biais et d'erreur quadratique moyenne ainsi que les intervalles de prévision lorsqu'applicable. Comme dans l'étude de simulation, les erreurs de prévision d'horizon un sont utilisées. Les prévisions exactes ne sont pas effectuées en raison de leur complexité calculatoire aux temps supérieurs. Les résultats sont présentés aux tableaux 5. V et 5. VI.

TABLE 5. V. Biais moyen et mesure d'EQM des prévisions ponctuelles pour dix temps de prévisions du rendement quotidien de l'indice *SPY*

	Biais	EQM
	$\times 10^{-2}$	$\times 10^{-3}$
AR(5)	1,203	0,259
STAR(5) Naïve	1,186	0,248
STAR(5) MC	1,186	0,249
STAR(5) BS	1,186	0,249

TABLE 5. VI. Inclusion des vraies valeurs dans les intervalles de prévision pour dix temps de prévisions du rendement quotidien de l'indice *SPY*

		Temps de prévision									
		1	2	3	4	5	6	7	8	9	10
Incl. dans l'IP	AR(5)	O	N	N	O	O	O	O	O	O	O
	STAR(5) MC	O	N	N	O	O	O	O	O	O	O
	STAR(5) BS	O	N	N	O	O	O	O	O	O	O

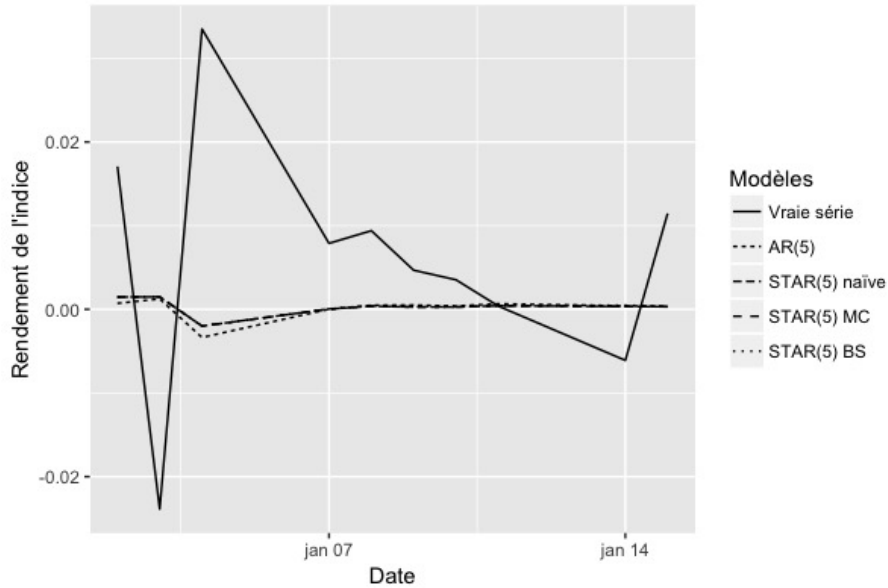


FIGURE 5.5. Prévisions ponctuelles selon les différentes méthodes étudiées et vraies valeurs de la série

La figure 5.5 montre les prévisions ponctuelles selon les différentes méthodes de prévisions, en plus des vraies valeurs de la série. Les figures en 5.6 montrent les prévisions ponctuelles et intervalles de prévision pour les méthodes de prévision des modèles AR(5), STAR(5) Monte Carlo et STAR(5) bootstrap, en plus des vraies valeurs de la série.

On remarque au tableau 5. V que le modèle AR(5) est celui donnant le biais moyen des dix prévisions ponctuelles le plus grand, avec un biais moyen de $1,203 \times 10^{-2}$. Les trois méthodes de prévision pour le modèle STAR(5), pour leur part, donnent toutes des biais moyens presque égaux, soit d'environ $1,186 \times 10^{-2}$. Par conséquent, au niveau du biais moyen, le modèle STAR(5) est plus performant que le modèle AR(5), et il ne semble pas y avoir de différences entre les méthodes employées pour les prévisions du modèle STAR(5).

Pour la mesure d'erreur quadratique moyenne, c'est encore le modèle AR(5) qui est le moins performant, avec une mesure d'EQM de $0,259 \times 10^{-3}$. Le modèle STAR(5) mène pour

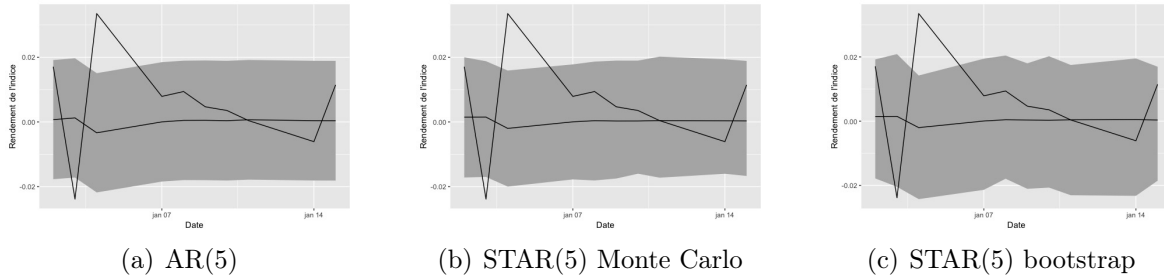


FIGURE 5.6. Prévisions ponctuelles et intervalles de prévision obtenus selon les modèles AR(5) et STAR(5)

sa part à une mesure d'EQM de $0,248 \times 10^{-3}$ pour la méthode naïve, et de $0,249 \times 10^{-3}$ pour la méthode de Monte Carlo et notre méthode de type bootstrap. Ainsi, globalement le modèle STAR(5) est mieux au niveau des prévisions ponctuelles que le modèle linéaire classique AR(5) dans le cadre de ce jeu de données.

Par la suite, nous pouvons observer les intervalles de prévision calculés selon les modèles AR(5) et STAR(5). Comme montré au tableau 5. V, les inclusions des vraies valeurs dans les intervalles de prévision sont identiques pour toutes les méthodes étudiées, soit que les prévisions aux temps 2 et 3 sont les seules non incluses dans les intervalles de prévision. À la figure 5.6, nous observons que le modèle AR(5) mène rapidement à des intervalles de prévision constants dans le temps, alors que le modèle STAR(5), via les méthodes de Monte Carlo et notre méthode de type bootstrap, mène à des intervalles plus variables. Par ailleurs, la méthode de prévision de type bootstrap mène dans le cas de ce jeu de données à des intervalles de prévision très variables dans le temps, quoi que cela ne fait pas de différence au niveau de l'inclusion des vraies valeurs dans les intervalles. La méthode de Monte Carlo mène à des intervalles légèrement plus constants dans le temps, bien que celle-ci se base sur une distribution normale des résidus, ce qui n'est pas tout à fait le cas en réalité, comme montré à la figure 5.4. Ainsi, il est difficile de conclure que les intervalles de prévision obtenus sont meilleurs pour les modèles STAR(p) que pour les modèles AR(p) dans le cadre de ce jeu de données. De plus, les intervalles de prévision plus grands sont approximatifs, puisqu'ils reposent sur les erreurs de prévision d'horizon un.

5.4. Comparaison avec la littérature

5.4.1. Estimation des modèles STAR(p)

Dans la littérature, l'ajustement d'un modèle STAR(p) à des données réelles ainsi que l'estimation des paramètres et de leurs erreurs standards associées sont élaborés de différentes façons par les auteurs en raison de la complexité des modèles STAR(p). Tout d'abord, selon le jeu de données étudié, certaines restrictions peuvent être imposées au niveau des paramètres linéaires du modèle. Entre autres, Teräsvirta (1994) estime des modèles STAR(p) sur des séries comportant respectivement environ une centaine de valeurs, et il impose des restrictions au niveau des paramètres, par exemple $\psi_{1,0} = 0$, lorsqu'il considère que cela est supporté par les données. Cela lui permet d'obtenir des estimations plus précises pour le reste des paramètres. Dans notre cas, les données de la série *SPY* ne permettent pas de faire de telles suppositions.

Dijk, Teräsvirta et Franses (2002) ajustent pour leur part un modèle STAR(p) au taux de chômage mensuel des États-Unis, et se basent comme nous sur le critère AIC afin de déterminer le délai p . Toutefois, ils décident d'augmenter ce délai suite à l'observation de l'autocorrélation des résidus du modèle ajusté. De plus, ils obtiennent une valeur de $\hat{\gamma}$ associée à une grande erreur standard, comme c'est le cas pour nos données, et ils déterminent que cette non-significativité du paramètre n'est pas un signe de linéarité de la série, ce qui concorde avec notre analyse. De plus, leurs paramètres linéaires sont statistiquement significatifs pour la plupart, comme c'est le cas pour notre jeu de données. Ceci n'est pas le cas de Umer, Sevil et Sevil (2018), dont les modèles STAR(p) ajustés au rendement mensuel d'un indice ont des paramètres presque tous non-significatifs. Comme mentionné par Teräsvirta (1994), ceci peut être un signe que certains paramètres linéaires dans le modèle sont redondants. Dans notre cas, notre jeu de données mène toutefois à des estimations de paramètres presque toutes significatives.

Finalement, Tsay (2010) applique pour sa part un modèle STAR(p) aux rendements mensuels des actions d'une compagnie. Il ajuste un modèle STAR(2) pour fins de comparaison avec un modèle ARCH(2). Toutefois, nous avons testé ce jeu de données, et si le choix du délai p avait été effectué avec le critère AIC comme c'est généralement le cas, le modèle ajusté aurait été un STAR(3). Nous avons tenté d'ajuster ce modèle aux données sans succès, l'optimisation non-linéaire par recherche quadrillée menant à un modèle où seules quelques valeurs sont disponibles pour estimer un des deux régimes, la presque totalité des valeurs

restantes ayant 1 comme valeur de fonction g associée. Ainsi, l'estimation des paramètres ne parvenait pas à converger dans le cas d'un STAR(3) ajusté via la recherche quadrillée. De plus, Tsay (2010) n'inclut pas l'estimation du paramètre γ dans la modélisation du modèle STAR(2) aux données. Plutôt, la valeur de ce paramètre est arbitrairement fixée à 1000, ce qui élimine la plupart des difficultés associées à l'estimation des modèles STAR(p), notamment la difficulté d'estimation précise pour le paramètre de transition γ , que nous avons abordée à la section 2.2.1.

5.4.2. Prévisions ponctuelles pour les modèles STAR(p)

Dans la littérature, plusieurs auteurs s'attardent également aux prévisions ponctuelles pour les modèles STAR(p). Tout d'abord, Dijk, Teräsvirta et Franses (2002) effectuent des prévisions ponctuelles par la méthode de bootstrap et étudient entre autres l'erreur quadratique moyenne. Ils trouvent que les prévisions semblent être biaisées pour les deux modèles étudiés, soit AR(p) et STAR(p), mais que l'erreur quadratique moyenne semble être légèrement plus petite pour le modèle STAR(p), particulièrement lorsque l'analyse des prévisions est divisée selon le régime en vigueur dans la série. Les conclusions pour l'erreur quadratique moyenne sont donc concordantes avec celles obtenues pour notre jeu de données, et également pour le biais notre conclusion est similaire.

Umer, Sevil et Sevil (2018) trouvent pour leur part que l'utilisation d'un modèle STAR(p) avec méthode de prévision de bootstrap pour leur jeu de données ne mène pas à une plus petite erreur de prévision, et donc que le modèle STAR(p) ne semble pas mener à de meilleures prévisions que le modèle AR(p). C'est également les conclusions de Teräsvirta et Anderson (1992) qui analysent l'indice de la production industrielle pour plusieurs pays. Ils trouvent que, au niveau de l'erreur quadratique moyenne des prévisions ponctuelles, le modèle STAR(p) ne semble pas donner de meilleures ou pires prévisions ponctuelles que le modèle AR(p). Ainsi, pour leur jeu de données l'utilisation d'un modèle STAR(p) ne semble pas amener d'avantages ou d'inconvénients par rapport à l'utilisation d'un modèle AR(p) au niveau des prévisions ponctuelles. Teräsvirta, Dijk et Medeiros (2005) analysent pour leur part plusieurs variables macroéconomiques mensuelles pour les pays du G7. Ils trouvent que le modèle qui performe le mieux, parmi entre autres les modèles AR(p) et STAR(p), varie selon la variable analysée, le temps de prévision et le pays à l'étude. Ainsi, aucun des modèles ne semble amener un avantage par rapport à un autre de façon systématique.

Par conséquent, nos conclusions au niveau de la différence entre les performances des prévisions ponctuelles des modèles $STAR(p)$ et $AR(p)$ pour notre jeu de données sont légèrement différentes de certaines conclusions de la littérature, puisque pour nos données le biais et l'EQM semblent plus petits pour le modèle STAR que le modèle AR. Il faut avouer que les gains pour le STAR étaient petits. Les intervalles de prévision ne sont pour leur part pas effectués par aucun auteur pour les modèles $STAR(p)$ avec un vrai jeu de données, et donc il nous n'est pas possible de comparer nos conclusions avec celles de la littérature. Bien que nos méthodes pour calculer les intervalles de prévision étaient approximatives pour les horizons lointains, un pas de plus dans cette direction demeure une contribution du mémoire.

Finalement, les modèles STAR considérés dans ce mémoire portaient sur la modélisation d'une série chronologique observée, et nous avons étudié les propriétés statistiques des estimateurs des modèles STAR dans ce cadre. En pratique, il peut être argumenté que les modèles STAR peuvent être utilisés afin de modéliser la variance conditionnelle d'une série chronologique financière, plutôt que la série en soi. Bien que cela soit au-delà des objectifs de ce mémoire, il serait intéressant de dégager les propriétés des estimateurs en cause, et de vérifier, minimalement par simulations, les propriétés empiriques des estimateurs considérés. Le calcul des prévisions dans une modélisation de la variance conditionnelle avec des modèles STAR est certainement un sujet qui mérite d'être étudié et qui pourrait être une suite à ce travail. En fait, il est souhaité que les résultats présentés dans ce mémoire susciteront un intérêt pour les modèles STAR dans ce contexte. Nous laissons ces considérations pour des travaux de recherche futurs.

Conclusion

Dans ce mémoire, nous avons étudié la théorie et l'aspect empirique de l'ajustement et des méthodes prévisionnelles pour les modèles autorégressifs à transition lisse. En premier lieu, nous avons tout d'abord présenté la théorie entourant les séries temporelles linéaires et non-linéaires en plus des conditions de stationnarité de celles-ci. Nous avons ensuite élaboré le processus d'ajustement des modèles STAR(p) ainsi que les difficultés pratiques s'y rattachant, suivi des méthodes de prévisions ponctuelles et intervalles de prévision pour les modèles autorégressifs linéaires et les modèles STAR(p).

En deuxième lieu, nous avons effectué des simulations de Monte Carlo pour plusieurs modèles STAR(p) afin d'étudier empiriquement le processus d'ajustement du modèle ainsi que la performance des prévisions ponctuelles et intervalles de prévision, le tout dans un but comparatif du modèle STAR(p) avec le modèle AR(p) et des méthodes de prévisions du modèle STAR(p) entre elles. Les simulations ont été effectuées pour trois tailles échantillonnales dans les séries temporelles, soit $T = 400$, $T = 2000$ et $T = 4000$, permettant de comparer les estimations des paramètres et des matrices de variances-covariances. Nous trouvons que les estimations des paramètres sont plus imprécises et que les matrices sont parfois non-calculables lorsque les séries comportent un petit nombre de valeurs T . Au niveau des prévisions ponctuelles, nous observons que le biais est généralement plus petit en valeur absolue lorsque la méthode de prévision est exacte, de Monte Carlo ou bootstrap, suivi des prévisions par le modèle AR(p), et avec la méthode de prévision naïve donnant généralement le plus grand biais en valeur absolue. L'erreur quadratique moyenne est pour sa part toujours plus petite pour les méthodes de prévisions exactes, de Monte Carlo et de bootstrap du modèle STAR(p), et parfois plus petite pour la méthode de prévision naïve que pour le modèle AR(p). Globalement, les prévisions obtenues par les méthodes exactes, de Monte Carlo et de bootstrap sont donc des prévisions ponctuelles plus exactes lorsque le modèle sous-jacent aux données est un STAR(p). Pour les intervalles de prévision, nous trouvons que les intervalles obtenus par les méthodes de Monte Carlo et bootstrap sont toujours plus petits en moyenne, et parfois de façon importante, que ceux obtenus par l'ajustement d'un modèle AR(p), puisque ceux-ci parviennent à mieux représenter les deux

régimes présents dans la série. Une direction de recherche poursuivant ce qui a été fait dans ce mémoire porte sur des intervalles de prévision valides à des horizons supérieurs à un, par les méthodes de Monte Carlo et de bootstrap. Les travaux de [Pan et Politis \(2016\)](#) devraient constituer le point de départ.

En troisième lieu, nous avons appliqué la théorie étudiée à un vrai jeu de données comportant les rendements quotidiens de l'indice *SPY*. Nous avons effectué le processus d'ajustement en commençant par la spécification du modèle, ce qui a mené à un choix de modèle STAR(5), puis en effectuant l'estimation des paramètres, dont certains étaient non-significatifs, et en observant les régimes obtenus, puis nous avons terminé par l'évaluation du modèle. Par la suite, nous avons calculé les prévisions ponctuelles et intervalles de prévision selon les différentes méthodes présentées au chapitre 3, et avons conclu que l'ajustement d'un STAR(p), peu importe la méthode utilisée, ne semble pas mener des gains importants, en termes de meilleures prévisions et intervalles qu'un modèle AR(p), dans le cadre de ce jeu de données réelles. Nous avons finalement discuté de nos résultats et de la façon dont ils se comparent à ceux disponibles dans la littérature.

Références

- Akaike, H. (1974). "A new look at the statistical model identification". Dans *Selected Papers of Hirotugu Akaike*, Springer, New York, 215-222.
- Brockwell, P. J. et Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. 3e édition. Springer Texts in Statistics. Springer-Verlag, New York.
- Brown, B. W. et Mariano, R. S. (1984). "Residual-based procedures for prediction and estimation in a nonlinear simultaneous system". *Econometrica : Journal of the Econometric Society* **52**, 321-343.
- Clements, M. et Hendry, D. (1998). *Forecasting Economic Time Series*. Cambridge University Press, Cambridge.
- De Gooijer, J. G. (2017). *Elements of Nonlinear Time Series Analysis and Forecasting*. Springer Series in Statistics. Springer International Publishing.
- Dijk, D. V. et Franses, P. H. (1999). "Modeling multiple regimes in the business cycle". *Macroeconomic Dynamics* **3**, 311-340.
- Dijk, D. V., Teräsvirta, T. et Franses, P. H. (2002). "Smooth transition autoregressive models —a survey of recent developments". *Econometric Reviews* **21**, 1-47.
- Franses, P. H., Dijk, D. V. et Opschoor, A. (2014). *Time Series Models for Business and Economic Forecasting*. 2e édition. Cambridge University Press, Cambridge.
- Granger, C. W. et Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford University Press, New York.
- Hannan, E.J. et Deistler, M. (2012). *The Statistical Theory of Linear Systems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.
- Kheifets, I. L. et Saikkonen, P. J. (2018). "Stationarity and ergodicity of vector STAR models". *arXiv preprint*, arXiv :1805.11311.
- Lin, J. L. et Granger, C. W. (1994). "Forecasting from non-linear models in practice". *Journal of Forecasting* **13**, 1-9.

- Ljung, G. M. et Box, G. E. (1978). "On a measure of lack of fit in time series models". *Biometrika* **65**, 297-303.
- Luukkonen, R., Saikkonen, P. et Teräsvirta, T. (1988). "Testing linearity against smooth transition autoregressive models". *Biometrika* **75**, 491-499.
- McCullough, B. D. (1994). "Bootstrapping forecast intervals : an application to AR(p) models". *Journal of Forecasting* **13**, 51-66.
- Pan, L. et Politis, D. N. (2016). "Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions". *Journal of Statistical Planning and Inference* **177**, 1-27.
- Pötscher, B. M. et Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models : Asymptotic Theory*. Springer-Verlag, Berlin.
- Shumway, R. H. et Stoffer, D. S. (2010). *Time Series Analysis and its Applications : with R Examples*. 3e édition. Springer Texts in Statistics. Springer, New York.
- Teräsvirta, T. (1994). "Specification, estimation, and evaluation of smooth transition autoregressive models". *Journal of the American Statistical Association* **89**, 208-218.
- Teräsvirta, T. et Anderson, H. (1992). "Characterizing nonlinearities in business cycles using smooth transition autoregressive models". *Journal of Applied Econometrics* **7**, S119-S136.
- Teräsvirta, T., Dijk, D. V. et Medeiros, M. C. (2005). "Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series : a re-examination". *International Journal of Forecasting* **21**, 755-774.
- Teräsvirta, T., Tjøstheim, D. et Granger, C. W. (2010). *Modelling Nonlinear Economic Time Series*. Advanced Texts in Econometrics. Oxford University Press, New York.
- Teräsvirta, T. et Yang, Y. (2014). "Specification, estimation and evaluation of vector smooth transition autoregressive models with applications". CREATES Research Papers 2014-08, Department of Economics and Business Economics, Aarhus University.
- Thombs, L. A. et Schucany, W. R. (1990). "Bootstrap prediction intervals for autoregression". *Journal of the American Statistical Association* **85**, 486-492.
- Tsay, R. S. (1986). "Nonlinearity tests for time series". *Biometrika* **73**, 461-466.
- Tsay, R. S. (1989). "Testing and modeling threshold autoregressive processes". *Journal of the American Statistical Association* **84**, 231-240.

- Tsay, R. S. (2010). *Analysis of Financial Time Series*. 3e édition. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, New Jersey.
- Umer, U. M., Sevil, T. et Sevil, G. (2018). "Forecasting performance of smooth transition autoregressive (STAR) model on travel and leisure stock index". *The Journal of Finance and Data Science* 4, 90-100.
- Wooldridge, J. M. (1994). "Estimation and inference for dependent processes". *Handbook of Econometrics* 4, 2639-2738.