UNIVERSITÉ CHARLES DE GAULLE - LILLE III

T H È S E

en vue de l'obtention du grade de

Docteur Ès-Sciences

Discipline : Mathématiques Appliquées aux Sciences Économiques

présentée par

Ali AHMAD

sous le titre

Contribution à l'économétrie des séries temporelles à valeurs entières

Soutenue publiquement le 05 décembre 2016 devant le jury composé de

M. Christian FRANCQ	Professeur	Université Lille 3	Directeur.
M. Jean-Michel ZAKOÏAN	Professeur	Université Lille 3	Co-directeur.
Mme Esmeralda GONÇALVES	Professeure	Université de Coimbra	Rapporteur.
M. Célestin C. KOKONENDJI	Professeur	Université de Franche Comté	Examinateur.
M. Guy MÉLARD	Professeur	Université libre de Bruxelles	Examinateur.

Les pré-rapporteurs :

M. Konstantinos FOKIANOS	Professeur	Université de Chypre.
Mme Esmeralda GONÇALVES	Professeure	Université de Coimbra.

Remerciements

J'aimerais tout d'abord remercier chaleureusement mon directeur de thèse, Monsieur le professeur Christian FRANCQ, de m'avoir encadré, guidé, encouragé et conseillé. Sa compétence, sa rigueur scientifique et sa clairvoyance m'ont beaucoup appris. J'ai vraiment apprécié sa grande disponibilité et ses qualités humaines de patience, d'écoute et de compréhension tout au long de ce travail doctoral.

Je voudrais également remercier vivement, Monsieur le professeur Jean-Michel ZAKOÏAN, le co-directeur de ma thèse, pour ses encouragements, ses soutiens et surtout ses conseils précieux qui m'ont permis de mener à bien cette thèse.

Je tiens à remercier Monsieur le professeur Konstantinos FOKIANOS[1] et Madame le professeur Esmeralda GONÇALVES de l'honneur qu'ils me font d'avoir accepté d'être les rapporteurs de cette thèse. Je remercie également Messieurs les professeurs Célestin C. KOKONENDJI et Guy MÉLARD d'avoir accepté de siéger au sein de mon jury de thèse.

Je tiens à remercier tous les membres du laboratoire LEM de l'université Lille 3 qui m'ont permis d'effectuer ma thèse dans une ambiance sereine. De plus, j'aimerais remercier chaleureusement Madame Christiane FRANCQ d'avoir relu la partie en français de ce travail.

Je souhaite aussi remercier sincèrement mes amis fidèles et mes proches qui m'ont soutenu et encouragé dans toutes les phases de mes études en Syrie et en France.

Enfin, je dédie ce travail à ma famille en Syrie, à ma femme et à mon fils Jaafar, pour leur présence dans ma vie, leurs sacrifices et leurs soutiens.

1. I would like to thank Pr. Konstantinos Fokianos for agreeing to write a report on my PhD thesis and for his encouragements and his advises during our meeting in the Workshop "Count Data Modeling and Analysis" in Besançon-France 2014.

Résumé

Dans cette thèse, nous étudions des modèles de moyennes conditionnelles de séries temporelles à valeurs entières. Tout d'abord, nous proposons l'estimateur du quasi maximum de vraisemblance de Poisson (EQMVP) pour les paramètres de la moyenne conditionnelle. Nous montrons que, sous des conditions générales de régularité, cet estimateur est consistant et asymptotiquement normal pour une grande classe de modèles. Étant donné que les paramètres de la moyenne conditionnelle de certains modèles sont positivement contraints, comme par exemple dans les modèles INAR (INteger-valued AutoRegressive) et les modèles INGARCH (INteger-valued Generalized AutoRegressive Conditional Heteroscedastic), nous étudions la distribution asymptotique de l'EQMVP lorsque le paramètre est sur le bord de l'espace des paramètres. En tenant compte de cette dernière situation, nous déduisons deux versions modifiées du test de Wald pour la significativité des paramètres et pour la moyenne conditionnelle constante. Par la suite, nous accordons une attention particulière au problème de validation des modèles de séries temporelles à valeurs entières en proposant un test portmanteau pour l'adéquation de l'ajustement. Nous dérivons la distribution jointe de l'EQMVP et des autocovariances résiduelles empiriques. Puis, nous déduisons la distribution asymptotique des autocovariances résiduelles estimées, et aussi la statistique du test. Enfin, nous proposons l'EQMVP pour estimer équation-par-équation (EpE) les paramètres de la moyenne conditionnelle des séries temporelles multivariées à valeurs entières. Nous présentons les hypothèses de régularité sous lesquelles l'EQMVP-EpE est consistant et asymptotiquement normal, et appliquons les résultats obtenus à plusieurs modèles de séries temporelles multivariées à valeurs entières.

Mots clés. Bord de l'espace des paramètres, Consistance et normalité asymptotique, Modèles INAR, Modèles INGARCH, Estimateur du quasi maximum de vraisemblance de Poisson, Test portmanteau, Test d'adéquation, Séries temporelles multivariées à valeurs entières.

Abstract

The framework of this PhD dissertation is the conditional mean count time series models. We propose the Poisson quasi-maximum likelihood estimator (PQMLE) for the conditional mean parameters. We show that, under quite general regularity conditions, this estimator is consistent and asymptotically normal for a wide classe of count time series models. Since the conditional mean parameters of some models are positively constrained, as, for example, in the integer-valued autoregressive (INAR) and in the integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH), we study the asymptotic distribution of this estimator when the parameter lies at the boundary of the parameter space. We deduce a Wald-type test for the significance of the parameters and another Wald-type test for the constance of the conditional mean. Subsequently, we propose a robust and general goodness-of-fit test for the count time series models. We derive the joint distribution of the PQMLE and of the empirical residual autocovariances. Then, we deduce the asymptotic distribution of the estimated residual autocovariances and also of a portmanteau test. Finally, we propose the PQMLE for estimating, equation-by-equation (EbE), the conditional mean parameters of a multivariate time series of counts. By using slightly different assumptions from those given for PQMLE, we show the consistency and the asymptotic normality of this estimator for a considerable variety of multivariate count time series models.

Keywords. Boundary of the parameter space, Consistency and asymptotic normality, Integer-valued AR and GARCH models, Non-normal asymptotic distribution, Poisson quasi-maximum likelihood estimator, Portmanteau test, Goodness-of-fit, Multivariate time series of counts.

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Le contexte bibliographique et les contributions de la thèse

1.1 Introduction

Au cours des dernières années, les séries temporelles à valeurs entières ont suscité un intérêt croissant, avec des applications dans de nombreux domaines scientifiques, par exemple, l'économie et la finance (voir par exemple Blundell et al., 2002, Quoreshi, 2006, Jung & Tremayne, 2011a, Jung & Tremayne, 2011b et Christou & Fokianos, 2014), les sciences sociales (voir par exemple McCabe & Martin, 2005, Pedeli & Karlis, 2011 et Liu, 2012), la télécommunication (voir par exemple Lambert & Liu, 2006), le tourisme (voir par exemple Brännäs & Nordström, 2006, Garcia-Ferrer & Queralt, 1997 et Brännäs et al., 2002) et l'environnement (voir par exemple Thyregod et al., 1999, Cui & Lund, 2009, Boudreault & Charpentier, 2011 et Scotto et al., 2014). En général, les séries temporelles à valeurs entières surviennent quand on s'intéresse au nombre des occurrences d'un événement particulier dans un intervalle de temps spécifié. Plusieurs modèles ont été proposés dans la littérature pour modéliser ce type des séries. Ces modèles sont classifiés par Cox et al. (1981) en deux catégories principales : les modèles "parameter-driven" (pour lesquels le paramètre d'intérêt dépend d'un processus latent) et les modèles "observation-driven" (pour lesquels le paramètre d'intérêt ne dépend que des observations). Une des premières approches proposées est le modèle INAR(1) (INteger-valued AutoRegressive) introduit par McKenzie (1985) et Al-Osh & Alzaid (1987). Ce modèle utilise l'opérateur d'amincissement introduit par Steutel et al. (1979). Le modèle INAR appartient à la classe des modèles "parameter-driven" et représente un cas particulier de processus de branchement avec une immigration. La généralisation de ce modèle à l'ordre supérieur à un (*i.e.* INAR(p)) est établie par Alzaid & Al-Osh (1990) qui ont aussi étudié sa structure d'autocorrélation. L'étude par simulations de Al-Osh & Alzaid (1987) a montré que l'estimateur des moindres carrés conditionnels, et l'estimateur de Yule-Walker ne sont pas efficaces pour le modèle INAR(1) quand la distribution marginale de l'innovation est Poisson. Par ailleurs, l'estimateur du maximum de vraisemblance (EMV) est difficile à calculer, notamment lorsque l'ordre du modèle est grand. Ce problème a incité certains auteurs à proposer des techniques et algorithmes alternatifs pour estimer ce modèle (voir par exemple Pavlopoulos & Karlis, 2008, Drost et al., 2009 et Pedeli et al., 2014).

Un autre modèle populaire est le modèle INGARCH(p, q) (INteger-valued Generalized AutoRegressive Conditional Heteroscedastic) qui a été introduit par Ferland et al. (2006). Il est aussi appelé ACP (Autoregressive Conditional Poisson) dans Heinen (2003). Le modèle INGARCH(p, q) est classifié comme un modèle "observation-driven". Il est obtenu en supposant que la distribution conditionnelle de la série sachant le passé est Poisson avec un paramètre d'intensité d'une forme linéaire expliquant la dynamique du modèle. Ce modèle ne peut convenir que pour les séries sur-dispersées (*i.e.* lorsque la variance inconditionnelle est supérieure à l'espérance inconditionnelle). Christou & Fokianos (2014) ont étudié le modèle INGARCH(1,1) quand la distribution conditionnelle est binomiale négative. Heinen (2003) et Zhu (2012a) ont proposé d'utiliser respectivement la distribution double-Poisson de Efron (1986) et la distribution de Poisson-généralisée pour que ce modèle puisse être compatible avec la sous-dispersion (*i.e.* lorsque la variance inconditionnelle est inférieure à l'espérance inconditionnelle). Le lecteur est renvoyé à Kokonendji (2014) pour une revue générale des modèles de séries temporelles à valeurs entières sous et sur-dispersées. Gonçalves et al. (2015b) ont établi la stationnarité et l'ergodicité des modèles INGARCH(p, q) dont la distribution conditionnelle appartient à la famille des distributions de Poisson composées, qui comprend, comme cas particuliers, plusieurs distributions (par exemple la distribution de Poisson, la distribution binomiale négative et la distribution de Poisson généralisée). Fokianos & Tjøstheim (2011) ont introduit le modèle log-linéaire. Au contraire du modèle INGARCH dont les paramètres autorégressifs de la moyenne conditionnelle sont positivement contraints, les paramètres du modèle log-linéaire ne sont pas contraints à être positifs, ce qui rend le modèle capable de modéliser les séries ayant une autocorrélation négative. Par ailleurs ce modèle permet d'ajouter des variables explicatives exogènes à la moyenne conditionnelle. Motivé par le phénomène de zéro-inflation ($i.e.$ la proportion des zéros est plus grande que la proportion des zéros du modèle Poisson INGARCH) qui est commun dans les séries temporelles à valeurs entières, Zhu (2012c) a introduit les modèles zéro-inflation Poisson et zéro-inflation négatif binomial INGARCH. Récemment, Gonçalves et al. (2016) ont proposé le modèle INGARCH de Poisson composé avec une zéro-inflation qui représente une classe générale, et qui comprend les modèles zéro-inflation Poisson et zéro-inflation négatif binomial INGARCH comme cas particuliers. D'autres spécifications des modèles de moyennes conditionnelles sont disponibles dans la littérature (voir par exemple Gao et al., 2009, Fokianos & Tjøstheim, 2012, Zhu, 2012b, Christou, 2014 et Gonçalves et al., 2015a).

En ce qui concerne les tests d'adéquation de l'ajustement des modèles de séries temporelles à valeurs entières, la littérature est ténue. Notons cependant, la contribution de Zhu & Wang (2010), qui ont étudié un ensemble de tests pour l'adéquation de l'ajustement du modèle INARCH(p) de Poisson. Certains de ces tests ne prennent pas en compte l'influence des erreurs d'estimation des paramètres affectant la robustesse des tests, en particulier lorsque les valeurs des paramètres sont relativement grandes. D'autres tests dépendent de paramètres arbitraires. Neumann (2011) a proposé d'utiliser l'hypothèse de l'équi-dispersion conditionnelle du modèle INGARCH(p, q) de Poisson pour tester l'adéquation de la spécification du processus de l'intensité. En outre, Fokianos & Neumann (2013) ont présenté un test non paramétrique pour le modèle INGARCH(p, q) de Poisson. Récemment, Meintanis & Karlis (2014) ont proposé un test de qualité de l'ajustement pour la distribution de l'innovation du modèle INAR(1) de Poisson. Hudecová et al. (2015) ont introduit un test de qualité de l'ajustement fondé sur la fonction génératrice des probabilités pour le modèle INAR(p) de Poisson et le modèle INARCH(p) de Poisson. De plus, Schweer (2016) a introduit un test plus général dans lequel la fonction génératrice conjointe des probabilités empirique est considérée pour tester l'adéquation de l'ajustement d'une large classe de modèles de séries temporelles à valeurs entières, mais ce test dépend de paramètres arbitraires. En outre, la généralisation de la méthodologie de ce test à modèles d'ordre supérieur est difficile.

La littérature sur les séries temporelles multivariées à valeurs entières est moins développée. Une contribution importante est due à Franke & Subba Rao (1993) qui ont introduit le modèle M-INAR(1) (Multivariate INteger-valued AutoRegressive). Latour (1997) a étudié le modèle M-INAR quand l'ordre est supérieur à un. Récemment, Pedeli & Karlis (2011) ont introduit le modèle INAR bivarié d'ordre un dans lequel les innovations sont supposées suivre conjointement la distribution de Poisson bivariée ou la distribution binomiale négative bivariée. Ce modèle peut seulement modéliser les séries dont la corrélation croisée instantanée est positive. Un modèle INAR bivarié plus flexible a été introduit par Karlis & Pedeli (2013). Dans ce modèle, la distribution marginale de chaque innovation est Poisson univariée ou binomiale négative univariée, et les deux innovations sont supposées être connectées par une copule bivariée (voir Sklar, 1959 et Nelsen, 2006). Heinen & Rengifo (2007) ont introduit le modèle INGARCH multivarié avec copule. Ce modèle est obtenu en supposant que la distribution conditionnelle de chaque série est double-Poisson. La dépendance entre les séries est modélisée en utilisant une copule gaussienne multivariée. Ce modèle peut accueillir la corrélation croisée instantanée négative et positive entre les séries. De plus, il peut modéliser les séries sous-dispersées et les séries sur dispersées. Liu (2012) a proposé le modèle INGARCH(p, q) bivarié pour les séries sur-dispersées ayant une corrélation croisée instantanée positive. Le modèle est obtenu en supposant que la distribution conditionnelle est Poisson bivarié.

Cette thèse se situe dans la continuité des travaux de recherches cités précédemment. Son champ général de recherche est les modèles pour les moyennes conditionnelles de séries temporelles à valeurs entières. Les contributions principales de cette thèse sont réparties sur trois chapitres :

Dans le deuxième chapitre, nous proposons l'estimateur du quasi maximum de vraisemblance de Poisson (EQMVP) pour les paramètres de la moyenne conditionnelle des séries temporelles à valeurs entières. Cet estimateur a déjà été étudié, pour des modèles particuliers, par plusieurs auteurs, par exemple, Zeger (1988), Zeger &

Qaqish (1988), Kedem & Fokianos (2002, Chap. 4) and Christou & Fokianos (2014). Dans ce chapitre, nous présentons des conditions de régularité générales selon lesquelles l'EQMVP, écrit comme si la distribution conditionnelle de la série sachant le passé est Poisson, est consistent et asymptotiquement normal pour une grande classe de modèles. Nous considérons également le cas où les conditions de régularité mentionnées ci-dessus sont violées parce que le paramètre est sur le bord de l'espace des paramètres. En tenant compte sur cette dernière situation, nous proposons deux tests, le premier est une version modifiée du test de Wald pour la significativité des paramètres et le second est un test joint pour la moyenne conditionnelle constante. Nous appliquons nos résultats asymptotiques aux modèles particuliers des séries temporelles à valeurs entières, comme le modèle INGARCH(p, q), le modèle INAR(p) et le modèle log-linéaire(1,1). Les résultats asymptotiques sont également illustrés par des simulations de Monte Carlo et une application aux données financières. Ce chapitre est une version longue d'un article co-écrit avec Christian Francq, intitulé (**Poisson QMLE of count time series models**) et publié dans ***Journal of Time Series Analysis***.

Dans le troisième chapitre, nous proposons un test de type portmanteau robuste et générale pour les modèles des séries temporelles à valeurs entières. Le test représente un outil pour évaluer l'adéquation de l'ajustement des classes larges et importantes des modèles, par exemple, les modèles INGARCH (p, q), les modèles log-linéaire(1,1), les modèles non-linéaire(1,1) avec diverses distributions conditionnelles et les modèles INAR(p) avec différentes distributions marginales de l'innovations. Nous dérivons la distribution jointe de l'EQMVP et des autocovariances résiduelles empiriques. Puis, nous déduisons la distribution asymptotique des autocovariances résiduelles estimées, et aussi la statistique du test. Nous illustrons les propriétés du test à distance finie en utilisant des simulations de Monte Carlo et une application aux données financières.

Dans le quatrième chapitre, nous proposons une méthodologie générale et simple pour estimer les modèles de séries temporelles multivariées à valeurs entières. Nous utilisons l'EQMVP pour estimer les paramètres de la moyenne conditionnelle des séries multivariées équation par équation (EpE). Pour obtenir cet estimateur, il suffit de spécifier les moyennes conditionnelles des séries. Nous présentons des conditions générales de régularité selon lesquelles l'EQMVP-EpE est convergent et asymptotiquement normal. De plus, nous illustrons les résultats asymptotiques par des simulations de Monte Carlo et une application aux données réelles.

Le dernier chapitre présente la conclusion et les perspectives des recherches futures.

Nous présentons maintenant les résultats obtenus de façon plus détaillée.

1.2 Résultats du chapitre 2

Dans ce chapitre, nous définissons l'estimateur du quasi maximum de vraisemblance de Poisson en présentant les conditions de régularité générales sous lesquelles cet estimateur est consistant et asymptotiquement normal. De plus, nous discutons la situation quand le paramètre est sur le bord de l'espace des paramètres.

Distribution asymptotique de l'EQMVP

Supposons que $X_1, \ldots, X_n$ est une série à valeurs entières non négatives, de sorte que

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda(X_{t-1}, X_{t-2}, \ldots; \theta_0), \tag{1.2.1}$$

où $\mathcal{F}_{t-1}$ désigne le $\sigma-$algèbre engendré par $(X_u, u < t)$,

$$\lambda \text{ est une fonction mesurable à valeurs dans } (\underline{\omega}, +\infty) \text{ pour un } \underline{\omega} > 0 \tag{1.2.2}$$

et θ_0 est un paramètre inconnu appartenant à un espace des paramètres $\Theta \subset \mathbb{R}^d$. La distribution marginale est supposée avoir un moment légèrement supérieur à 1

$$EX_t^{1+\varepsilon} < \infty, \text{ pour un certain } \varepsilon > 0, \tag{1.2.3}$$

ce qui implique l'existence de la moyenne conditionnelle (1.2.1). Pour tout $\theta \in \Theta$, $x_0 \in \mathbb{R}^+$ et $t \geq 1$, nous avons

$$\lambda_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \ldots; \theta) \text{ et } \widetilde{\lambda}_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \ldots, X_1, x_0, x_0, \ldots; \theta).$$

Notons que $\widetilde{\lambda}_t(\theta)$ servira comme de proxy pour $\lambda_t(\theta)$. Il est obtenu en fixant à une valeur réelle positive x_0 les valeurs initiales inconnues $X_0, X_{-1}, \ldots$ impliquées dans

$\lambda_t(\theta)$. Cette valeur x_0 peut être un nombre entier fixe, par exemple $x_0 = 0$, ou une valeur fonction de θ, ou encore une valeur fonction des observations. Par exemple, lorsque $\lambda_t(\theta) = \omega + \alpha X_{t-1} + \beta\lambda_{t-1}(\theta)$ avec $\theta = (\omega, \alpha, \beta)$, nous pouvons prendre $\widetilde{\lambda}_t(\theta) = \omega + \alpha X_{t-1} + \beta\widetilde{\lambda}_{t-1}(\theta)$ avec $\widetilde{\lambda}_1(\theta) = \omega/(1-\beta)$ (ce qui correspond à $x_0 = 0$), ou avec $\widetilde{\lambda}_1(\theta) = \omega/(1-\alpha-\beta)$ (ce qui correspond à $x_0 = \omega/(1-\alpha-\beta)$), ou $\widetilde{\lambda}_1(\theta) = \overline{X}_5$ (la moyenne des jours de travail de la première semaine, pour les données quotidiennes). Il sera démontré que le choix de x_0 n'est pas asymptotiquement important. Cependant, nous avons besoin de supposer que p.s.

$$\lim_{t\to\infty} a_t = 0 \quad \text{et} \lim_{t\to\infty} X_t a_t = 0, \quad \text{où } a_t = \sup_{\theta\in\Theta} \left|\widetilde{\lambda}_t(\theta) - \lambda_t(\theta)\right|, \tag{1.2.4}$$

et

$$\widetilde{\lambda}_t(\theta) \geq \underline{\omega}, \quad \forall t \geq 1,\ \forall \theta \in \Theta. \tag{1.2.5}$$

Nous supposons aussi que

$$\theta \mapsto \lambda_t(\theta) \text{ est presque sûrement continue et } \Theta \text{ est un ensemble compact.} \tag{1.2.6}$$

L'estimateur du quasi maximum de vraisemblance de Poisson du paramètre θ_0 est défini comme une solution mesurable de

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \widetilde{L}_n(\theta), \quad \widetilde{L}_n(\theta) = \frac{1}{n}\sum_{t=s+1}^{n} \widetilde{\ell}_t(\theta), \tag{1.2.7}$$

où $\widetilde{\ell}_t(\theta) = -\widetilde{\lambda}_t(\theta) + X_t \log \widetilde{\lambda}_t(\theta)$ et s est un nombre entier. La valeur de s n'est pas asymptotiquement importante, mais elle peut affecter le comportement de l'EQMVP à distance finie en réduisant l'impact de la valeur initiale x_0.

La consistance de l'EQMVP

L'hypothèse essentielle pour la consistance de l'EQMVP est que la moyenne conditionnelle soit bien spécifiée. De toute façon, l'hypothèse d'identifiabilité suivante est également nécessaire :

$$\lambda_1(\theta) = \lambda_1(\theta_0) \text{ presque sûrement si et seulement si } \theta = \theta_0. \tag{1.2.8}$$

Théorème 1.2.1. *Nous supposons que (X_t) est une suite à valeurs entières non négatives, ergodique, strictement stationnaire et vérifiant (1.2.1)-(1.2.6) et (1.2.8). Donc, l'EQMVP défini dans (1.2.7) satisfait*

$$\widehat{\theta}_n \to \theta_0 \quad p.s. \qquad \text{quand } n \to \infty.$$

Remarque 1.2.1 (La moyenne conditionnelle linéaire). *Supposons que* Θ *est sous-ensemble compact de* $(0,\infty)\times[0,\infty)^{p+q}$, $\theta_0=(\omega_0,\alpha_{01},\ldots,\beta_{0p})$, *et*

$$\lambda_t(\theta_0)=\omega_0+\sum_{i=1}^{q}\alpha_{0i}X_{t-i}+\sum_{j=1}^{p}\beta_{0j}\lambda_{t-j}(\theta_0). \tag{1.2.9}$$

Supposons aussi que pour tout $\theta=(\omega,\alpha_1,\ldots,\beta_p)\in\Theta$, *nous avons* $\sum_{i=1}^{p}\beta_i<1$. *Notant que l'équation* (1.2.9) *est similaire à celle de la volatilité dans le modèle GARCH(p,q), il est alors facile de montrer (par les arguments utilisés pour montrer (7,30) dans Francq & Zakoian (2010), désigné ci-après par FZ), que* $a_t\le K\rho^t$. *Donc, la première condition dans* (1.2.4) *est directement vérifiée. Pour montrer la deuxième convergence de* (1.2.4), *il suffit d'utiliser le lemme de Borel-Cantelli et* $P(\rho^tX_t\ge\varepsilon)\le\rho^tEX_t/\varepsilon$, *pour* $\varepsilon>0$. *Alors, les conditions* (1.2.2) *et* (1.2.4)-(1.2.6) *sont satisfaites sans aucune contrainte additionnelle.*

Nous introduisons les polynômes $\mathcal{A}_\theta(z)=\sum_{i=1}^{q}\alpha_iz^i$ *et* $\mathcal{B}_\theta(z)=1-\sum_{i=1}^{p}\beta_iz^i$. *Comme dans le preuve de (b) Page 157 dans FZ, la condition d'identifiabilité (1.2.8) est satisfaite en supposant que la distribution conditionnelle de* X_t *n'est pas dégénérée et que*

$$\begin{aligned}&\text{si } p>0,\ \mathcal{A}_{\theta_0}(z)\ \text{et}\ \mathcal{B}_{\theta_0}(z)\ \text{n'ont pas de racine commune,}\\ &\text{les } \alpha_{0i}\neq0\ \text{sont no tous nuls, et}\ \beta_{0p}\neq0\ \text{si}\ \alpha_{0q}=0.\end{aligned} \tag{1.2.10}$$

Dans le cas $q=1$, *la condition* (1.2.10) *est simplement équivalente à supposer que* $\alpha_{01}>0$.

Les questions de stationnarité et d'ergodicité seront discutées pour des classes particulières des modèles des séries temporelles à valeurs entières non négatives dans la section 1.2.

La normalité asymptotique

Sous des conditions peu restrictives de régularité, l'EQMVP est asymptotiquement normal lorsque le paramètre appartient à l'intérieur de l'espace des paramètres. Dans le cas plus général où le paramètre est sur le bord de l'espace des paramètres, sa distribution asymptotique est la projection d'un vecteur aléatoire gaussien sur un cône convexe. Des estimateurs avec des distributions asymptotiques similaires ont été étudiées, par exemple, par Andrews (1999), Francq & Zakoïan (2009) et dans les références de ces articles.

Quand θ_0 appartient à l'intérieur de Θ

Nous avons besoin de supposer l'existence de

$$Var(X_t \mid \mathcal{F}_{t-1}) = E\left(X_t^2 \mid \mathcal{F}_{t-1}\right) - \lambda_t^2(\theta_0), \tag{1.2.11}$$

l'existence des dérivées continues d'ordre second pour $\lambda_t(\cdot)$ et $\widetilde{\lambda}_t(\cdot)$, ainsi que l'existence des matrices d'information

$$J = E\left(\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}\right), \quad I = E\left(\frac{Var(X_t \mid \mathcal{F}_{t-1})}{\lambda_t^2(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}\right). \tag{1.2.12}$$

Il est facile de montrer que ces matrices sont inversibles quand

$$c'\frac{\partial\lambda_t(\theta_0)}{\partial\theta} = 0 \quad p.s. \Rightarrow c = 0. \tag{1.2.13}$$

Nous supposons aussi qu'il y a un voisinage $V(\theta_0)$ de θ_0, tel que, pour tout $(i,j) \in \{1,\ldots,d\}$,

$$E \sup_{\theta\in V(\theta_0)} \left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta)\right| < \infty. \tag{1.2.14}$$

Nous supposons enfin que, p.s.,

$$b_t, b_tX_t \text{ et } a_td_tX_t \text{ sont d'ordre } O(t^{-\kappa}) \text{ pour un certain } \kappa > 1/2, \tag{1.2.15}$$

où

$$b_t = \sup_{\theta\in\Theta}\left\|\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta} - \frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|, \quad d_t = \sup_{\theta\in\Theta}\max\left\{\left\|\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|, \left\|\frac{1}{\widetilde{\lambda}_t(\theta)}\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta}\right\|\right\}.$$

Théorème 1.2.2. *Supposons que (X_t) satisfait les conditions de Théorème 1.2.1. Supposons aussi que (1.2.11)-(1.2.14) et (1.2.15). Si $\theta_0 \in \overset{\circ}{\Theta}$, où $\overset{\circ}{\Theta}$ désigne l'intérieur de Θ. Alors,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \Sigma := J^{-1}IJ^{-1}\right) \qquad \text{quand } n \to \infty.$$

Notons que quand la distribution conditionnelles de X_t sachant le passé est Poisson, nous avons $I = J$, et donc $\Sigma = J^{-1}$. Dans ce cas, nous revenons à la variance asymptotique obtenue par Ferland et al. (2006) pour MLE du modèle Poisson INGARCH.

Nous pouvons montrer que, sous les hypothèses de Théorème 1.2.2, la variance asymptotique de l'EQMVP peut être estimée par $\widehat{\Sigma} = \widehat{J}^{-1}\widehat{I}\widehat{J}^{-1}$ avec

$$\widehat{J} = \frac{1}{n}\sum_{t=s+1}^{n}\frac{1}{\widetilde{\lambda}_t(\widehat{\theta}_n)}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta'}, \tag{1.2.16}$$

$$\widehat{I} = \frac{1}{n}\sum_{t=s+1}^{n}\left(\frac{X_t}{\widetilde{\lambda}_t(\widehat{\theta}_n)}-1\right)^2\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta'}. \tag{1.2.17}$$

Remarque 1.2.2 (Des conditions alternatives pour (1.2.12) et (1.2.14))**.** *Notons que*

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta) = \left(\frac{X_t}{\lambda_t(\theta)}-1\right)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta) - \frac{X_t}{\lambda_t^2(\theta)}\frac{\partial}{\partial\theta_i}\lambda_t(\theta)\frac{\partial}{\partial\theta_j}\lambda_t(\theta). \tag{1.2.18}$$

En utilisant l'inégalité de Hölder et (1.2.3)*, nous pouvons assurer* (1.2.14) *en montrant que*

$$E\sup_{\theta\in V(\theta_0)}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta)\right| < \infty, \tag{1.2.19}$$

et

$$E\sup_{\theta\in V(\theta_0)}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial}{\partial\theta_i}\lambda_t(\theta)\right|^{r_0} < \infty, \qquad E\sup_{\theta\in V(\theta_0)}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta)\right|^{r_0} < \infty \tag{1.2.20}$$

pour tout $r_0 > 0$*. Nous constatons également que* (1.2.3) *et* (1.2.20) *impliquent l'existence de* J*. Pour l'existence de* I *une autre condition est nécessaire.*

En particulier, (1.2.12) *est obtenu sous les conditions* (1.2.20) *et*

$$E \mid Var(X_t \mid \mathcal{F}_{t-1}) \mid^{1+\varepsilon} < \infty, \ \textit{pour un certain } \varepsilon > 0, \tag{1.2.21}$$

Remarque 1.2.3 (La moyenne conditionnelle linéaire)**.** *Revenons au cadre de la Remarque 1.2.1. Parce que les racines des polynômes* $\mathcal{B}_\theta(z)$ *sont à l'extérieur du disque unitaire, nous avons* $\lambda_t(\theta) = \pi_0(\theta) + \sum_{k=1}^{\infty}\pi_k(\theta)X_{t-k}$*, où*

$$\sum_{k=1}^{\infty}\pi_k(\theta)z^k = \mathcal{B}_\theta^{-1}(z)\mathcal{A}_\theta(z) \qquad et \qquad \sup_{\theta\in\Theta}|\pi_k(\theta)| \le K\rho^k.$$

Nous avons également

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta) = \pi_0^{(i,j)}(\theta) + \sum_{k=1}^{\infty}\pi_k^{(i,j)}(\theta)X_{t-k} \qquad avec \qquad \sup_{\theta\in\Theta}|\pi_k^{(i,j)}(\theta)| \le K\rho^k.$$

Sous l'hypothèse de moment (1.2.3), la condition (1.2.19) est donc vérifiée, quelque soit le voisinage de $V(\theta_0)$ inclus dans Θ. Nous pouvons montrer (1.2.20) en utilisant les mêmes arguments utilisés pour montrer (7.54) dans FZ. Nous obtenons donc (1.2.14), et (1.2.12) sous (1.2.21).

Maintenant, notons que $b_t \leq K\rho^t$ et d_t admettent des moments de tous ordres, par des arguments déjà donnés.

Nous avons donc $E|t^\kappa b_t X_t| \leq Kt^\kappa \rho^t$ et $E|t^\kappa a_t d_t X_t| \leq Kt^\kappa \rho^t$, qui implique (1.2.15) en utilisant le lemme de Borel-Cantelli et l'inégalité de Markov. Finalement, en utilisant des arguments similaires à ceux utilisés pour montrer (b) Page 162 dans FZ, nous pouvons montrer que (1.2.13) est une conséquence de (1.2.10).

Quand θ_0 est sur le bord de Θ

Pour le calcul de l'EQMVP, il est nécessaire d'avoir $\widetilde{\lambda}_t(\theta) > 0$ presque sûrement, pour tout $\theta \in \Theta$. Pour cette raison, l'espace des paramètres Θ doit être contraint. Fréquemment, une ou plusieurs composantes de θ sont contraintes à être positives ou égales à zéro. Par exemple, lorsque nous avons un modèle INGARCH(1,1) de la forme $\lambda_t(\theta) = \omega + \alpha X_{t-1} + \beta\lambda_{t-1}(\theta)$, alors $\theta = (\omega, \alpha, \beta) \in \Theta \subset [\underline{\omega}, \infty) \times [0, \infty)^2$. Selon la méthodologie de Box-Jenkins, il est intéressant de tester si le modèle peut être simplifié. Pour le modèle INGARCH(1,1), il est intéressant de tester si la vraie valeur du paramètre $\theta_0 = (\omega_0, \alpha_0, 0)$, autrement dit si le PGD est un INARCH(1). Dans ce cas, le Théorème 1.2.2 n'est plus applicable parce que $\theta_0 \notin \overset{\circ}{\Theta}$. En outre, la distribution asymptotique de $\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)$ n'est plus gaussienne parce que les contraintes de positivité impliquent que $\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \geq 0$ avec une probabilité un quand $\beta_0 = 0$. Nous revenons maintenant au modèle général. La composante i du paramètre θ est dite positivement contrainte si la $i$ème section de Θ est de la forme $\left[0, \overline{\theta}_i\right]$ avec $\theta_i > 0$. Par exemple, pour le modèle linéaire (1.2.9), la première composante n'est pas positivement contrainte, mais les autres composantes le sont.

Soit $d_2 \in \{0, \dots, d\}$ est le nombre des composantes contraintes positivement de θ, et soit $d_1 = d - d_2$. Sans perte de généralité, supposons que ces composantes contraintes d_2 sont les dernières. Si une des dernières composantes d_2 est égale à zéro, Théorème 1.2.2 n'est plus applicable parce que θ_0 se situe sur le bord de Θ.

Nous supposons que $\Theta - \theta_0$ est assez grand pour contenir un hypercube de la forme $\prod_{i=1}^d [\underline{\theta}_i, \overline{\theta}_i]$, où pour tout $i \in \{1, \dots, d\}$, $\underline{\theta}_i = 0$ si $\theta_{0i} = 0$ avec $i > d_1$, $\underline{\theta}_i < 0$ autrement, et $\overline{\theta}_i > 0$. Sous cette hypothèse, nous avons

$$\lim_{n \to \infty} \sqrt{n}(\Theta - \theta_0) = \mathcal{C}, \tag{1.2.22}$$

où $\mathcal{C} = \prod_{i=1}^d \mathcal{C}_i$, où $\mathcal{C}_i = \mathbb{R}$ quand $i \leq d_1$ ou $\theta_{0i} > 0$ et $\mathcal{C}_i = [0, +\infty)$ quand $i > d_1$ et $\theta_{0i} = 0$. De façon similaire à (1.2.15), supposons que, p.s.,

$$\lim_{t \to \infty} c_t + X_t \left(a_t e_t + c_t + a_t d_t^2 + b_t d_t \right) = 0, \tag{1.2.23}$$

où

$$c_t = \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \widetilde{\lambda}_t(\theta)}{\partial \theta \partial \theta} - \frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta'} \right\|,$$
$$e_t = \sup_{\theta \in \Theta} \max \left\{ \left\| \frac{1}{\lambda_t(\theta)} \frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta'} \right\|, \left\| \frac{1}{\widetilde{\lambda}_t(\theta)} \frac{\partial^2 \widetilde{\lambda}_t(\theta)}{\partial \theta \partial \theta'} \right\| \right\}.$$

Notons que, dans le cadre des Remarques 1.2.1 et 1.2.3, la condition (1.2.23) est toujours satisfaite. Puisque J est définie positive, on peut considérer la norme $\|x\|_J^2 = x'Jx$ et le produit scalaire $\langle x, y \rangle_J = x'Jy$ pour $x, y \in \mathbb{R}^d$. Avec cette mesure, la projection d'un vecteur $Z \in \mathbb{R}^d$ sur le cône convexe $\mathcal{C}$ est définie par

$$Z^{\mathcal{C}} = \arg \inf_{C \in \mathcal{C}} \|C - Z\|_J$$

ou de manière équivalente par

$$Z^{\mathcal{C}} \in \mathcal{C} \quad \text{et} \quad \left\langle Z - Z^{\mathcal{C}}, C - Z^{\mathcal{C}} \right\rangle_J \leq 0, \quad \forall C \in \mathcal{C}. \tag{1.2.24}$$

Théorème 1.2.3. *Supposons que les conditions de Théorème 1.2.2 sont vérifiées (à l'exception que $\theta_0 \in \overset{\circ}{\Theta}$) et (1.2.22), (1.2.23). Alors, quand $n \to \infty$,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Z^{\mathcal{C}} = \arg \inf_{C \in \mathcal{C}} \|C - Z\|_J, \text{ où } Z \sim \mathcal{N}(0, \Sigma).$$

Comme application du Théorème 1.2.3, supposons que $d_2 > 0$ et considérons le problème de test

$$H_0 : \theta_{0d} = 0 \quad \text{contre} \quad H_1 : \theta_{0d} > 0.$$

Nous désignons par $\widehat{\theta}_{nd}$ la dernière composante de $\widehat{\theta}_n$ et nous désignons par $\chi_k^2(\underline{\alpha})$ le $\underline{\alpha}$-quantile de la loi khi-deux χ_k^2 avec degré de liberté k. Si nous supposons également que, sous l'hypothèse nulle, seulement la dernière composante de θ_0 est sur le bord (voir l'exemple 8.2 et la section 8.3.3 dans FZ), alors la région de rejet du test est

$$\left\{ \frac{n\widehat{\theta}_{nd}^2}{\widehat{\Sigma}(d,d)} \geq \chi_1^2(1-2\underline{\alpha}) \right\} \tag{1.2.25}$$

au niveau asymptotique $\underline{\alpha}$. Notons que lorsque la dernière composante n'est pas positivement contrainte (*i.e.* lorsque $d_2 = 0$), l'EQMVP a une distribution asymptotique normale et la valeur critique du test est $\chi_1^2(1-\underline{\alpha})$. Une autre application de Théorème 1.2.3 est donnée dans le corollaire suivant. Nous désignons par δ_0 la masse de Dirac en 0 et nous désignons par $p_0\delta_0 + \sum_{i=1}^q p_i\chi_{k_i}^2$ le mélange de δ_0 et de χ_{ki}^2- distributions, avec les poids de mélange $p_0, \ldots, p_q$.

Corollaire 1.2.1 (Un test pour la moyenne conditionnelle constante)**.** *Considérons un processus ergodique et strictement stationnaire (X_t) avec une moyenne conditionnelle linéaire de la forme $\lambda_t(\theta) = \omega + \sum_{i=1}^q \alpha_i X_{t-i}$. Supposons que la distribution conditionnelle de X_t ne dépend du passé que par $\lambda_t(\theta)$. Si $\theta_0 = (\omega_0, 0, \ldots, 0)' \in \Theta$, où Θ est un sous-ensemble compact de $(0,\infty) \times [0,\infty)^q$, et si $EX_t^4 < \infty$, alors la statistique*

$$S_n = n\sum_{i=1}^q \widehat{\alpha}_{ni}^2 \xrightarrow{d} \frac{1}{2^q}\delta_0 + \sum_{i=1}^q \binom{q}{i}\frac{1}{2^q}\chi_i^2 \qquad \textit{quand } n \to \infty. \tag{1.2.26}$$

La distribution asymptotique est appelée la distribution de Khi-bar-carré. La moyenne conditionnelle constante est rejetée au niveau asymptotique $\underline{\alpha}$ si $\{S_n > c_{q,\underline{\alpha}}\}$, où la valeur critique $c_{q,\underline{\alpha}}$ peut être trouvée dans le tableau 8.2 de FZ.

Applications aux modèles particuliers

Nous présentons maintenant certains exemples des modèles vérifiant les conditions de régularité générales requises pour les résultats asymptotiques.

Le modèle Poisson INGARCH

Le modèle Poisson INGARCH(p,q) est obtenu en supposant que la distribution conditionnelle de X_t sachant le passé est Poisson avec un paramètre d'intensité de la forme linéaire (1.2.9). Ce modèle peut seulement modéliser les séries temporelles à valeurs entières non négatives sur-dispersées. Sous la condition suivante

$$\sum_{i=1}^{q} \alpha_{0i} + \sum_{i=1}^{p} \beta_{0i} < 1, \tag{1.2.27}$$

Le modèle INGARCH(p, q) de Poisson est stationnaire et ergodique.

Le modèle binomial négatif INGARCH

Un autre modèle pour les séries sur-dispersées est le modèle binomial négatif INGARCH(p, q) qui est défini en supposant que la distribution conditionnelle de X_t sachant le passé est binomiale négative $\mathcal{NB}(r, p_t)$ avec des paramètres $r > 0$ et $p_t = r/(\lambda_t + r)$, où λ_t est de la forme (1.2.9). Le modèle binomial négatif INGARCH(p, q) est stationnaire et ergodique sous (1.2.27).

Le modèle double-Poisson INGARCH

La distribution double-Poisson avec les paramètres $\lambda > 0$ et $\gamma > 0$, est définie par la distribution de probabilité

$$P(X = x) = c(\lambda, \gamma) \left(\frac{e^{-x} x^{x}}{!X} \right) \left(\frac{e\lambda}{x} \right)^{\gamma} x, \qquad x = 0, 1, \ldots,$$

où $c(\lambda, \gamma)$ est une constante de normalisation. Nous utilisons ensuite la notation $X \sim \mathcal{DP}(\lambda, \gamma)$. Efron (1986) montre que la moyenne de la distribution $\mathcal{DP}(\lambda, \gamma)$ est λ, et que sa variance est approximativement égale à λ/γ. Le modèle double-Poisson INGARCH est défini en supposant que la distribution conditionnelle de X_t sachant le passé est $\mathcal{DP}(\lambda_t, \gamma)$ avec un paramètre λ_t de la forme (1.2.9). Ce modèle peut modéliser les séries sur-dispersées et les séries sous-dispersées. L'ergodicité de ce modèle n'est pas établie, mais nous pouvons conjecturer que, comme pour la distribution de Poisson, l'ergodicité est satisfaite sous (1.2.27).

Le modèle Poisson-généralisé INGARCH

Zhu (2012a) a présenté le modèle Poisson-généralisé INGARCH, qui peut modéliser les séries sur-dispersées et les séries sous-dispersées. La distribution de Poisson-généralisé (GP) a deux paramètres $\lambda^* > 0$ et $\kappa < 1$. Lorsque $\kappa \geq 0$, une variable X suivant cette distribution, $X \sim \mathcal{GP}(\lambda^*, \kappa)$, satisfait

$$P(X = x) = \lambda^* (\lambda^* + \kappa x)^{x-1} \frac{e^{-(\lambda^* + \kappa x)}}{x!}, \qquad x = 0, 1, \ldots.$$

La définition peut être étendue à certaines valeurs de $\kappa < 0$ (voir Zhu, 2012a). Nous avons $EX = \lambda^*/(1-\kappa)$. Donc, le modèle Poisson-généralisé INGARCH pourrait être défini en supposant que la distribution conditionnelle de X_t sachant le passé est $\mathcal{GP}(\lambda_t(1-\kappa), \kappa)$ avec λ_t de la forme (1.2.9). Ce modèle est stationnaire et ergodique sous (1.2.27).

Le modèle log-linéaire

Les coefficients des modèles précédents sont positivement contraints, ce qui entraîne des difficultés statistiques lorsqu'un coefficient est égal à zéro (voir Théorème 1.2.3) et entraîne aussi des difficultés pour ajouter des variables explicatives exogènes à λ_t. Un autre inconvénient est que les autocovariances $\mathrm{Cov}(X_t, X_{t-h})$ sont non négatives à tout retard h (voir Christou & Fokianos (2014) pour l'expression explicite de ces autocovariances pour les modèles de premier ordre). Pour faire face à ces problèmes, Fokianos & Tjøstheim (2011) ont proposé un modèle dans lequel la distribution conditionnelle de X_t sachant le passé est Poisson de paramètre d'intensité $\lambda_t = e^{\upsilon_t}$, où

$$\upsilon_t = \omega_0 + \alpha_0 \log(X_{t-1} + 1) + \beta_0 \upsilon_{t-1}. \tag{1.2.28}$$

Cela conduit au modèle Poisson log-linéaire de premier ordre. De toute évidence, nous pouvons modifier la distribution conditionnelle pour définir, par exemple, un modèle binomial négatif log-linéaire. Parce que le processus υ_t n'est pas contraint à être positif, les coefficients ω_0, α_0 et β_0 en (1.2.28) peuvent être négatifs. Sous la condition

$$|\alpha_0 + \beta_0| \vee |\alpha_0| \vee |\beta_0| < 1, \tag{1.2.29}$$

Douc et al. (2013) ont montré que le modèle Poisson log-linéaire avec une log-intensité (1.2.28) admet une solution stationnaire et ergodique.

Le modèle INAR

Un des modèles les plus populaires des séries temporelles à valeurs entières non négatives est le processus autorégressif à valeur entière (INAR). Contrairement aux modèles précédents, le modèle INAR est "parameter-driven". Le modèle INAR(1) définit X_t comme une convolution d'une distribution binomiale $\mathcal{B}(X_{t-1}, \alpha_0)$ (avec la convention $\mathcal{B}(X_{t-1}, \alpha_0) = 0$ quand $X_{t-1} = 0$) et de la distribution de l'innovation ϵ_t dont le support est à valeurs entières non négatives. Nous pouvons interpréter $\mathcal{B}(X_{t-1}, \alpha_0)$ comme le nombre de survivants de la population X_{t-1} et ϵ_t comme le nombre de nouveaux arrivants, qui est supposé être indépendant de X_{t-1}. Avec ce modèle, nous avons (1.2.1) (voir Grunwald et al., 2000) avec $\lambda_t = \omega_0 + \alpha_0 X_{t-1}$ et $\omega_0 = E\epsilon_1$, évidemment sous l'hypothèse que l'espérance existe. Dans ce cas, et lorsque la séquence (ϵ_t) est *iid* et $\alpha_0 < 1$, (X_t) est toujours stationnaire et $EX_1 = E\epsilon_1/(1-\alpha_0)$.

Illustrations numériques

La première partie de cette section examine le comportement de l'EQMVP à distance finie par des expériences de Monte Carlo basées sur 1000 réplications indépendantes. Les simulations sont réalisées sur les modèles INGARCH et les modèles log-linéaires avec diverses distributions conditionnelles et le modèle INAR avec plusieurs distributions marginales des innovations. Ces expériences montrent que, pour tous les modèles considérés, les moyennes des paramètres estimés par EQMVP sont très proches de leurs valeurs théoriques. De plus, les erreurs standards estimées basées sur la théorie asymptotique $\sqrt{\widehat{\Sigma}_{(i,i)}/n}$, où $\widehat{\Sigma}$ est définie dans (1.2.16) et (1.2.17), et les erreurs moyennes quadratiques empiriques obtenues par les simulations sont très similaires et diminuent quand les tailles des échantillons augmentent. La deuxième partie présente une étude de simulation pour le test de nullité des coefficients (1.2.25) et le test de moyenne conditionnelle constante (1.2.26). Les expériences ont pour but d'évaluer la taille et la puissance des deux tests pour les modèles dans lesquels les paramètres sont positivement contraints (*i.e.* les modèles INGARCH et les modèles INAR). Les résultats montrent que les tailles des tests sont très proches de leurs niveaux asymptotiques. De plus, les puissances des tests augmentent quand la taille d'échantillon augmente. Finalement, nous présentons une application aux données financières. Ces illustrations numériques montrent que la théorie asymptotique fournit de bonnes approximations.

1.3 Résultats du chapitre 3

Dans ce chapitre, nous présentons un test portmanteau pour l'adéquation de l'ajustement des modèles des séries temporelles à valeurs entières non négatives.

Le modèle et les hypothèses

Supposons que $\{X_t \in \mathbb{N}\}$ est une série temporelle à valeurs entières non négatives, telle que

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda_t(\theta_0) = \lambda(X_{t-1}, X_{t-2}, \ldots; \theta_0), \tag{1.3.1}$$

où $\mathcal{F}_{t-1}$ désigne $\sigma-$algèbre engendré par $(X_u, u < t)$,

$$\lambda \text{ est une fonction mesurable à valeurs dans } (\underline{\omega}, +\infty) \text{ pour certains } \underline{\omega} > 0 \tag{1.3.2}$$

et θ_0 est un paramètre inconnu appartenant à un espace des paramètres $\Theta \subset \mathbb{R}^d$. Nous supposons que

$$\begin{aligned} &\text{le support de la distribution conditionnelle de } X_t \text{ sachant le passé} \\ &\text{a au moins trois valeurs différentes.} \end{aligned} \tag{1.3.3}$$

Pour tout $\theta \in \Theta$, $x_0 \in \mathbb{R}^+$ et $t \geq 1$, nous avons

$$\lambda_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \ldots; \theta) \text{ et } \widetilde{\lambda}_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \ldots, X_1, x_0, x_0, \ldots; \theta).$$

La variable $\widetilde{\lambda}_t(\theta)$ servira comme un proxy pour $\lambda_t(\theta)$. Il est obtenu en fixant à une valeurs réelle positive x_0 les valeurs initiales inconnues $X_0, X_{-1}, \ldots$ impliquées dans $\lambda_t(\theta)$. Nous supposons que le processus $\widetilde{\lambda}_t(\theta)$ vérifie l'hypothèse suivante

$$\widetilde{\lambda}_t(\theta) \geq \underline{\omega} \text{ pour un certain } \underline{\omega} > 0, \forall t > 1 \text{ et } \theta \in \Theta. \tag{1.3.4}$$

Nous supposons aussi que le moment d'ordre quatre de la distribution marginale de X_t existe

$$EX_t^4 < \infty. \tag{1.3.5}$$

Nous définissons les résidus comme

$$\epsilon_t(\theta_0) = X_t - \lambda_t(\theta_0). \tag{1.3.6}$$

Notons que l'hypothèse (1.3.5) implique que $E(\epsilon_t(\theta_0)^4) < \infty$ ainsi que $E(\lambda_t(\theta_0)^4) < \infty$. Si la moyenne conditionnelle est correctement spécifiée, sous l'hypothèse de la stationnarité de la série X_t, nous pouvons montrer que $\epsilon_t(\theta_0)$ est une suite de bruit blanc non-corrélé, où

$$E(\epsilon_t(\theta_0)) = E\left(E\left(X_t - \lambda_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) = 0$$

et

$$\begin{aligned}
var(\epsilon_t(\theta_0)) &= E\left(Var\left(\epsilon_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) + Var\left(E\left(\epsilon_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) \\
&= E\left(Var\left(\epsilon_t(\theta_0)|\mathcal{F}_{t-1}\right)\right), \text{ depuis que } Var\left(E\left(\epsilon_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) = 0 \\
&= E\left(E\left(\epsilon_t(\theta_0)^2|\mathcal{F}_{t-1}\right)\right) \\
&= E\left(E\left((X_t - \lambda_t(\theta_0))^2|\mathcal{F}_{t-1}\right)\right) \\
&= E\left(Var\left(X_t|\mathcal{F}_{t-1}\right)\right). \qquad (1.3.7)
\end{aligned}$$

De plus, pour $h > 0$, nous avons

$$Cov(\epsilon_t(\theta_0)\epsilon_{t+h}(\theta_0)) = E\left(\epsilon_t(\theta_0)E\left(\epsilon_{t+h}(\theta_0)|\mathcal{F}_{t+h-1}\right)\right) = 0.$$

La formulation présentée ci-dessus contient comme cas particuliers plusieurs modèles des séries temporelles à valeurs entières non négatives, par exemple, le modèle INGARCH(p,q) (voir par exemple Heinen, 2003, Ferland et al., 2006, Zhu, 2012a et Christou & Fokianos, 2014), le modèle log-linéaire (voir Fokianos & Tjøstheim, 2011), le modèle non-linéaire (voir Fokianos & Tjøstheim, 2012) et le modèle INAR(p) (voir par exemple Alzaid & Al-Osh, 1990). Selon Ahmad & Francq (2016), tous les modèles mentionnés ci-dessus pourraient être estimés par EQMVP, qui est défini comme toute solution mesurable du problème de maximisation suivant

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \widetilde{L}_n(\theta), \quad \widetilde{L}_n(\theta) = \frac{1}{n}\sum_{t=s+1}^{n} \widetilde{\ell}_t(\theta), \qquad (1.3.8)$$

où $\widetilde{\ell}_t(\theta) = -\widetilde{\lambda}_t(\theta) + X_t \log \widetilde{\lambda}_t(\theta)$. La valeur de s n'est pas asymptotiquement importante, mais elle peut affecter le comportement de l'EQMVP à distance finie en réduisant l'impact de la valeur initiale x_0. La matrice variance-covariance asymptotique de $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ est désignée par $\Sigma := J^{-1}IJ^{-1}$, où

$$J = E\left(\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}\right), \quad I = E\left(\frac{Var(X_t|\mathcal{F}_{t-1})}{\lambda_t^2(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}\right). \qquad (1.3.9)$$

Le test Portmanteau

Comme déjà mentionné, ce test est basé sur les fonctions des autocovariances des résidus. Tout d'abord, les résidus pris en compte dans le test sont définis comme suit

$$\widetilde{\epsilon}_t(\widehat{\theta}_n) = X_t - \widetilde{\lambda}_t(\widehat{\theta}_n).$$

Nous désignons par $\widehat{\gamma}_m = (\widehat{\gamma}_{(1)}, .., \widehat{\gamma}_{(m)})'$ le vecteur des autocovariances des résidus, où, pour $m < n$ et $h \in \{1, .., m\}$, ses composantes sont données par

$$\widehat{\gamma}_{(h)} = \frac{1}{n} \sum_{t=h+1}^{n} \widetilde{\epsilon}_t(\widehat{\theta}_n)\widetilde{\epsilon}_{t-h}(\widehat{\theta}_n).$$

Nous définissons la matrice $\widehat{\Sigma}_{\gamma_m}$ de dimension $m \times m$ dont les composantes pour $(h, l) \in \{1, .., m\}$, sont données par

$$\widehat{\Sigma}_{\gamma_m}(h, l) = \frac{1}{n} \sum_{t=max(h,l)+1}^{n} \widetilde{\epsilon}_t^2(\widehat{\theta}_n)\widetilde{\epsilon}_{t-h}(\widehat{\theta}_n)\widetilde{\epsilon}_{t-l}(\widehat{\theta}_n).$$

Soit $\widehat{\Sigma}$ un estimateur consistant de la matrice Σ définie dans Section 1.3

$$\widehat{\Sigma} = \widehat{J}^{-1}\widehat{I}\widehat{J}^{-1},$$

où

$$\widehat{J} = \frac{1}{n} \sum_{t=s+1}^{n} \frac{1}{\widetilde{\lambda}_t(\widehat{\theta}_n)} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta'} \quad \text{et} \quad \widehat{I} = \frac{1}{n} \sum_{t=s+1}^{n} \left(\frac{X_t}{\widetilde{\lambda}_t(\widehat{\theta}_n)} - 1 \right)^2 \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta'}.$$

Soit d le nombre de paramètres du modèle. Nous définissons la matrice $\widehat{C}_m$ de dimension $m \times d$ et la matrice $\widehat{\Sigma}_{\theta,\gamma_m}$ de dimension $d \times m$ dont les composantes, pour $1 \leq k \leq d$ et $1 \leq h \leq m$, sont respectivement obtenues par

$$\widehat{C}_m(h, k) = -\frac{1}{n} \sum_{t=h+1}^{n} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta_k} \widetilde{\epsilon}_{t-h}(\widehat{\theta}_n)$$

et

$$\widehat{\Sigma}_{\theta,\gamma_m}(k, h) = \widehat{J}^{-1} \frac{1}{n} \sum_{t=h+1}^{n} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial \theta_k} \frac{\widetilde{\epsilon}_t^2(\widehat{\theta}_n)}{\widetilde{\lambda}_t(\widehat{\theta}_n)} \widetilde{\epsilon}_{t-h}(\widehat{\theta}_n).$$

Maintenant, nous définissons la matrice $\widehat{\Sigma}_{\widehat{\gamma}_m}$ comme

$$\widehat{\Sigma}_{\widehat{\gamma}_m} = \widehat{\Sigma}_{\gamma_m} + \widehat{C}_m \widehat{\Sigma} \widehat{C}'_m + \widehat{C}_m \widehat{\Sigma}_{\theta,\gamma_m} + \widehat{\Sigma}'_{\theta,\gamma_m} \widehat{C}'_m$$

et nous supposons que

$$\text{la matrice } \widehat{\Sigma}_{\widehat{\gamma}_m} \text{ est inversible.} \tag{1.3.10}$$

Pour montrer que les effets des valeurs initiales sur la statistique $\widehat{\gamma}_m$ sont asymptotiquement négligeables, nous avons besoin des convergences suivantes

$$\lim_{t\to\infty} \lambda_t(\theta_0)a_{t-h} = 0, \ \lim_{t\to\infty} \widetilde{\lambda}_t(\theta_0)a_t = 0 \text{ et } \lim_{t\to\infty} \frac{\partial \widetilde{\lambda}_t(\theta_0)}{\partial\theta} a_{t-h} = 0,$$
$$\text{où } a_t = \sup_{\theta\in\Theta} \left|\widetilde{\lambda}_t(\theta) - \lambda_t(\theta)\right|. \tag{1.3.11}$$

Nous supposons aussi que, pour tout $\theta \in \Theta$, $(i, j) \in \{1, .., d\}$ et $h \in \{1, .., m\}$, il y a un voisinage $V(\theta_0)$ de θ_0, tel que

$$E \sup_{\theta\in V(\theta_0)} \left\| \frac{\partial\lambda_{t-h}(\theta_0)}{\partial\theta_i} \frac{\partial\lambda_t(\theta_0)}{\partial\theta_j} \right\| < \infty \text{ et } E \sup_{\theta\in V(\theta_0)} \left\| \frac{\partial^2\lambda(\theta)}{\partial\theta_i\theta_j} \epsilon_{t-h} \right\| < \infty. \tag{1.3.12}$$

Théorème 1.3.1. *Sous* (1.3.1)-(1.3.5), (1.3.10)-(1.3.12) *et les autres hypothèses de régularité requises pour l'EQMVP, nous avons*

$$n\widehat{\gamma}_m' \widehat{\Sigma}^{-1}_{\widehat{\gamma}_m} \widehat{\gamma}_m \xrightarrow{\mathcal{L}} \chi^2_m.$$

L'adéquation du modèle est rejetée au niveau asymptotique $\underline{\alpha}$ quand

$$n\widehat{\gamma}_m' \widehat{\Sigma}^{-1}_{\widehat{\gamma}_m} \widehat{\gamma}_m > \chi^2_m(1 - \underline{\alpha}).$$

Illustrations numériques

Dans cette section, nous étudions le comportement du test à distance finie en utilisant des expériences de Monte Carlo. Les expériences ont pour but d'évaluer la taille et la puissance du test pour une grande collection de modèles (INGARCH, log-linéaire, non-linéaire, INAR) aux différents niveaux asymptotiques et avec différentes tailles des échantillons. Comme prévu, les résultats des simulations ont confirmé la théorie proposée dans ce chapitre. Pour tous les modèles, les tailles sont très proches de leurs niveaux asymptotiques, surtout quand la taille d'échantillon est assez grande. De plus, les résultats ont montré que la puissance du test augmente quand la taille d'échantillon augmente. Finalement, nous présentons une application aux données financières. Ces illustrations numériques montrent que la théorie asymptotique s'applique avec succès en échantillons finis.

1.4 Résultats du chapitre 4

Dans ce chapitre, nous proposons l'EQMVP pour estimer les paramètres de la moyenne conditionnelle de séries temporelles multivariées à valeurs entières non négatives équation-par-équation (EbE).

Le modèle et les résultats principaux

Supposons que X_t est un vecteur de dimension k de séries temporelles à valeurs dans $\mathbb{N}^k$, supposons aussi que la moyenne conditionnelle de la $d$ème composante $(X_{d,t})$ de ce vecteur sachant le passé est donnée par

$$E\left(X_{d,t} \mid \mathcal{F}_{t-1}\right) = \lambda_d\left(X_{t-1}, X_{t-2}, \ldots; \theta_{0,d}\right) = \lambda_{d,t}, \tag{1.4.1}$$

où $\mathcal{F}_{t-1}$ désigne le $\sigma-$algèbre engendré par $(X_u, u < t)$, $\theta_{0,d}$ est un paramètre inconnu appartenant à un espace de paramètres Θ_d, et pour tout $(x_u,\ u < t) \in \mathbb{N}^k$ et $\theta_d \in \Theta_d$,

$$\lambda_d(x_{t-1}, x_{t-2}, \ldots; \theta_{0,d}) \text{ est évaluée dans } (\underline{\omega}, +\infty) \text{ pour un certain } \underline{\omega} > 0. \tag{1.4.2}$$

De plus, nous supposons que la distribution marginale a un moment légèrement supérieur à 1

$$EX_{d,t}^{1+\varepsilon} < \infty,\ \text{pour un } \varepsilon > 0,\ \text{et pour } \ d = 1, .., k, \tag{1.4.3}$$

ce qui implique l'existence de la moyenne conditionnelle. Pour tout $\theta_d \in \Theta_d$ et $t \geq 1$, soit

$$\widetilde{\lambda}_{d,t}(\theta_d) = \lambda_d(X_{t-1}, X_{t-1}, \ldots, X_1, \widetilde{X}_0, \widetilde{X}_{-1}, \ldots; \theta_d) \text{ et } \lambda_{d,t}(\theta_d) = \lambda_d(X_{t-1}, X_{t-1}, \ldots; \theta_d),$$

où $\widetilde{X}_i$, for $i \leq 0$, sont des valeurs initiales arbitraires. Notons que $\widetilde{\lambda}_{d,t}(\theta_d)$ servira comme de proxy pour $\lambda_{d,t}(\theta_d)$. Il est obtenu en fixant aux valeurs entières non négatives $\widetilde{X}_0,\ \widetilde{X}_{-1}, \ldots$ les valeurs initiales inconnues $X_0, X_{-1}, \ldots$ impliquées dans $\lambda_t(\theta)$. Nous supposons que

$$\widetilde{\lambda}_{d,t}(\theta_d) > \underline{\omega},\ \forall t \geq 1 \text{ et } \theta_d \in \Theta. \tag{1.4.4}$$

Nous supposons aussi que

$$\theta_d \mapsto \lambda_{d,t}(\theta_d) \text{ est presque sûrement continue et } \Theta_d \text{ est compact.} \tag{1.4.5}$$

L'EQMVP-EpE est défini comme toute solution du problème de maximisation suivant

$$\widehat{\theta}_{n,d} = \arg\max_{\theta_d \in \Theta_d} \widetilde{L}_{n,d}(\theta_d), \quad \widetilde{L}_{n,d}(\theta_d) = \frac{1}{n}\sum_{t=s+1}^{n} \widetilde{\ell}_{d,t}(\theta_d), \tag{1.4.6}$$

où $\widetilde{\ell}_t(\theta) = -\widetilde{\lambda}_t(\theta) + X_t \log \widetilde{\lambda}_t(\theta)$. La valeur de s n'est pas asymptotiquement importante, mais elle peut affecter le comportement de l'EQMVP à distance finie en réduisant l'impact de la valeur initiale x_0. Soit aussi

$$L_{n,d}(\theta_d) = \frac{1}{n}\sum_{t=s+1}^{n} \ell_{d,t}(\theta_d), \text{ où } \ell_{d,t}(\theta_d) = -\lambda_{d,t}(\theta_d) + X_{d,t}\log\lambda_{d,t}(\theta_d).$$

Autres hypothèses techniques

Pour la consistance de EQMVP-EpE, nous supposons que

A1 Nous avons $\theta_{0,d} \in \Theta_d$, où Θ_d est compact.

A2 Le processus $(X_{d,t})$ est stationnaire et ergodique.

A3 $\lambda_{d,1}(\theta_d) = \lambda_{d,1}(\theta_{0,d})$ presque sûrement si et seulement si $\theta_d = \theta_{0,d}$.

Les deux hypothèses suivantes sont pour montrer que les valeurs initiales n'affectent pas les propriétés asymptotiques de l'EQMVP-EpE.

A4 Nous avons p.s. $\lim_{t\to\infty} a_{d,t} = 0 \quad$ et $\lim_{t\to\infty} X_{d,t}a_{d,t} = 0$,

où $a_{d,t} = \sup_{\theta_d\in\Theta_d}\left|\widetilde{\lambda}_{d,t}(\theta_d) - \lambda_{d.t}(\theta_d)\right|$.

Pour la normalité asymptotique, nous avons besoin des hypothèses suivantes

A5 $\theta_{0,d} \in \overset{\circ}{\Theta}_d$, où $\overset{\circ}{\Theta}_d$ désigne l'intérieur de Θ_d.

L'hypothèse suivante est introduite pour gérer les valeurs initiales

A6 $b_{d,t}, b_{d,t}X_{d,t}$ et $a_{d,t}g_{d,t}X_{d,t}$ sont d'ordre $O(t^{-\kappa})$ pour certains $\kappa > 1/2$,

où

$$\begin{aligned} b_{d,t} &= \sup_{\theta_d\in\Theta_d}\left\|\frac{\partial\widetilde{\lambda}_{d,t}(\theta_d)}{\partial\theta_d} - \frac{\partial\lambda_{d,t}(\theta_d)}{\partial\theta_d}\right\|. \\ g_{d,t} &= \sup_{\theta_d\in\Theta_d}\max\left\{\left\|\frac{1}{\lambda_{d,t}(\theta_d)}\frac{\partial\lambda_{d,t}(\theta_d)}{\partial\theta_d}\right\|, \left\|\frac{1}{\widetilde{\lambda}_{d,t}(\theta_d)}\frac{\partial\widetilde{\lambda}_{d,t}(\theta_d)}{\partial\theta_d}\right\|\right\} \end{aligned}$$

A7 L'existence de la variance conditionnelle de $X_{d,t}$ sachant le passé, telle que

$$E \mid Var\left(X_{d,t} \mid \mathcal{F}_{t-1}\right) \mid^{1+\varepsilon} < \infty, \text{ pour certain } \varepsilon > 0,$$

A8 $\lambda_{d,t}$, $\widetilde{\lambda}_{d,t}$ admettent continûment des dérivées du second ordre. Les matrices

$$J_{dd} = E\left(\frac{1}{\lambda_{d,t}(\theta_{0,d})}\frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta_d}\frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta'_d}\right)$$

et

$$I_{dd} = E\left(\frac{Var\left(X_{d,t} \mid \mathcal{F}_{t-1}\right)}{\lambda^2_{d,t}(\theta_{0,d})}\frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta_d}\frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta'_d}\right)$$

existent et J_{dd} est inversible.

Théorème 1.4.1. *Soit* $(X_{d,t})$ *un processus stationnaire et ergodique défini dans* (1.4.1)*, satisfaisant* (1.4.2)-(1.4.3), (1.4.5), (1.4.4) *et les hypothèses* ***A1-A4****. Soit* $\widehat{\theta}_{n,d}$ *l'EQMVP-EpE. Alors*

$$\widehat{\theta}_{n,d} \to \theta_{0,d} \quad p.s. \qquad \text{quand } n \to \infty.$$

Remarque 1.4.1 (Le modèle linéaire multivarié)**.** *Une des spécifications les plus communes du modèle général* (1.4.1) *est le linéaire ou le soi-disant modèle INGARCH*(p,q) *multivarié qui admet la représentation suivante*

$$\Lambda_t(\theta_0) = \Omega_0 + \sum_{i=1}^{q} A_{0,i}X_{t-i} + \sum_{j=1}^{p} B_{0,j}\Lambda_{t-j}(\theta_0), \tag{1.4.7}$$

où $\Lambda_t(\theta_0) = (\lambda_{1,t}(\theta_{0,1}),\dots,\lambda_{k,t}(\theta_{0,k}))'$, $\Omega_0 = (\omega_{0,1},\dots,\omega_{0,k})'$ *avec des coefficients strictement positifs et les matrices*

$$A_{0,i} = \begin{pmatrix} \alpha_{0,11,i} & \alpha_{0,12,i} & \cdots & \alpha_{0,1k,i} \\ \alpha_{0,21,i} & \alpha_{0,22,i} & \cdots & \alpha_{0,2k,i} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{0,k1,i} & \alpha_{0,k2,i} & \cdots & \alpha_{0,kk,i} \end{pmatrix}, \qquad B_{0,j} = \begin{pmatrix} \beta_{0,11,j} & \beta_{0,12,j} & \cdots & \beta_{0,1k,j} \\ \beta_{0,21,j} & \beta_{0,22,j} & \cdots & \beta_{0,2k,j} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{0,k1,j} & \beta_{0,k2,j} & \cdots & \beta_{0,kk,j} \end{pmatrix}$$

de dimension $k \times k$ *avec des coefficients positifs. Afin de faciliter les procédures d'estimation et faire des hypothèses plus explicites, nous supposons que*

$$\text{les matrices } B_{0,j}, \text{ pour } j = 1,\dots,p, \text{ sont diagonales.} \tag{1.4.8}$$

Sous (1.4.8)*, nous pouvons écrire la dème équation de ce modèle comme suit*

$$\lambda_{d,t}(\theta_{0,d}) = \omega_d + \sum_{i=1}^{q} \alpha_{0,d1,i}X_{1,t-i} + \cdots + \sum_{i=1}^{q} \alpha_{0,dk,i}X_{k,t-i} + \sum_{j=1}^{p} \beta_{0,dd,j}\lambda_{d,t-j}. \tag{1.4.9}$$

Nous pouvons montrer l'hypothèse d'identifiabilté ***A3*** *en utilisant* ***A2****, les hypothèses* (1.4.1)-(1.4.3) *et les conditions suivantes :*

(a) *Pour tout $c \in \{1,\dots,k\}$, il n'y a pas de racines communes entre $\mathcal{A}_\theta^{(dc)}(z)$ et $\mathcal{B}_\theta^{(d)}(z)$, où*

$$\mathcal{A}_\theta^{(dc)}(z) = \sum_{i=1}^{q} \alpha_{0,dc,i} z^i \quad et \quad \mathcal{B}_\theta^{(d)}(z) = 1 - \sum_{j=1}^{p} \beta_{0,dd,j} z^j.$$

(b) *Si $p > 0$, $\sum_{j=1}^{p} \beta_{0,dd,j} < 1$. Donc, $\mathcal{B}_\theta^{(d)}(z)$ est inversible.*

(c) *$Var(X_t)$ est une matrice définie positive.*

(d) *Au moins un $\alpha_{0,dc,i}$ est strictement positif, pour $c = 1,\dots,k$ et $i = 1,..,q$.*

Sous la condition (b), nous pouvons réécrire l'équation (1.4.9) en forme vectorielle autorégressive, telle que

$$\underline{\lambda}_{d,t}(\theta_{0,d}) = \underline{c}_{d,t}(\theta_{0,d}) + B^* \underline{\lambda}_{d,t-1}(\theta_{0,d}), \tag{1.4.10}$$

où $\underline{\lambda}_{d,t}(\theta_{0,d}) = (\lambda_{d,t},\dots,\lambda_{d,t-p+1})'$,

$$\underline{c}_{d,t}(\theta_{0,d}) = \begin{pmatrix} \omega_d + \sum_{i=1}^{q} \alpha_{0,d1,i} X_{1,t-i} + \cdots + \sum_{i=1}^{q} \alpha_{0,dk,i} X_{k,t-i} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

et

$$B^* = \begin{pmatrix} \beta_{0,d,1} & \beta_{0,d,2} & \cdots & \beta_{0,d,p} \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}$$

Donc, nous pouvons montrer ***A4*** *en utilisant des arguments similaires à ceux de Ahmad & Francq (2016) et à ceux utilisés pour montrer (7.30) dans Francq & Zakoian (2010).*

Théorème 1.4.2. *Supposons que $(X_{d,t})$ satisfait les conditions de Théorème 1.4.1. Sous les hypothèses* ***A5-A8****, nous avons*

$$\sqrt{n}(\widehat{\theta}_{n,d} - \theta_{0,d}) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{dd} := J_{dd}^{-1} I_{dd} J_{dd}^{-1}\right) \qquad quand\ n \to \infty.$$

Nous pouvons montrer que, sous les hypothèses de Théorème 1.4.2, la matrice Σ_{dd} peut être estimée de manière convergente par

$$\widehat{\Sigma}_{dd} = \widehat{J}_{dd}^{-1} \widehat{I}_{dd} \widehat{J}_{dd}^{-1}, \tag{1.4.11}$$

où

$$\widehat{J}_{dd} = \frac{1}{n}\sum_{t=s+1}^{n}\frac{1}{\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}\frac{\partial\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial\theta_d}\frac{\partial\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial\theta'_d}, \tag{1.4.12}$$

$$\widehat{I}_{dd} = \frac{1}{n}\sum_{t=s+1}^{n}\left(\frac{X_{d,t}}{\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}-1\right)^2\frac{\partial\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial\theta_d}\frac{\partial\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial\theta'_d}. \tag{1.4.13}$$

Remarque 1.4.2 (Le modèle linéaire multivarié)**.** *Sous l'hypothèse d'existence de moments* (1.4.3)*, nous pouvons montrer que **A6** et **A8** sont satisfaites par des arguments similaires à ceux utilisés dans Remarque 2.3 de Ahmad & Francq (2016). L'inversibilité de la matrice J_{dd} peut être démontrée en utilisant* (1.4.3)*,* (1.4.2) *et la condition (c) de Remarque 1.4.1.*

Nous désignons par θ_0 le vecteur des paramètres de toutes les équations du modèle. Même si les composantes de θ_0 sont estimées équation-par-équation, les composantes de $\widehat{\theta}_n$ ne sont pas nécessairement indépendantes asymptotiquement en présence d'une autocorrélation entre les séries. Nous supposons que

A9 Pour (c,d) $\in$ $\{1,...k\}$, nous avons p.s. $\lim_{n\to\infty}\sup_{\theta_d\in\Theta_d}\left\|\frac{\partial}{\partial\theta_d}\widetilde{L}_{n,d}(\theta_{0,d})-\frac{\partial}{\partial\theta_d}L_{n,d}(\theta)\right\|h_{c,t}=0$,
où $h_{c,t}=\sup_{\theta_c\in\Theta_c}\max\left\{\left\|\frac{\partial\lambda_{c,t}(\theta_c)}{\partial\theta_c}\right\|,\left\|\frac{\partial\widetilde{\lambda}_{c,t}(\theta_c)}{\partial\theta_c}\right\|\right\}$.

Théorème 1.4.3. *Supposons que, pour $d\in\{1,\dots,k\}$, $(X_{d,t})$ satisfait les conditions des Théorèmes 1.4.1 et 1.4.2. Sous **A9**, nous avons*

$$\begin{pmatrix}\sqrt{n}(\widehat{\theta}_{n,1}-\theta_{0,1})\\ \vdots\\ \sqrt{n}(\widehat{\theta}_{n,k}-\theta_{0,k})\end{pmatrix}\xrightarrow{d}\mathcal{N}(0,\Sigma_\theta) \qquad \textit{quand } n\to\infty,$$

où

$$\Sigma_\theta = \begin{pmatrix}\Sigma_{11} & \cdots & \Sigma_{1k}\\ \vdots & \ddots & \vdots\\ \Sigma'_{k1} & \cdots & \Sigma_{kk}\end{pmatrix}. \tag{1.4.14}$$

Pour (c,d) $\in\{1,\dots,k\}$, les sous-matrices de Σ_θ sont données par

$$\Sigma_{dc} = J_{dd}^{-1}I_{dc}J_{cc}^{-1\prime}. \tag{1.4.15}$$

La matrice I_{dc} est obtenue par

$$I_{dc} = \lim_{n\to\infty}\frac{1}{n}\sum_{t=s+1}^{n}\frac{\epsilon_{d,t}(\theta_{0,d})}{\lambda_{d,t}(\theta_{0,d})}\frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta_d}\frac{\epsilon_{c,t}(\theta_{0,c})}{\lambda_{c,t}(\theta_{0,c})}\frac{\partial\lambda_{c,t}(\theta_{0,c})}{\partial\theta'_c}, \tag{1.4.16}$$

où $\epsilon_{d,t}(\theta_{0,d}) = X_{d,t} - \lambda_{d,t}(\theta_{0,d})$ et $\epsilon_{c,t}(\theta_{0,c}) = X_{c,t} - \lambda_{c,t}(\theta_{0,c})$, et elle peut être estimée de manière convergente par

$$\widehat{I}_{dc} = \frac{1}{n}\sum_{t=s+1}^{n} \frac{\widetilde{\epsilon}_{d,t}(\widehat{\theta}_d)}{\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})} \frac{\partial\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial\widehat{\theta}_d} \frac{\widetilde{\epsilon}_{c,t}(\widehat{\theta}_{n,c})}{\widetilde{\lambda}_{c,t}(\widehat{\theta}_{n,c})} \frac{\partial\widetilde{\lambda}_{c,t}(\widehat{\theta}_{n,c})}{\partial\widehat{\theta}'_c}. \tag{1.4.17}$$

Applications aux modèles particuliers

Nous présentons maintenant certains exemples de modèles vérifiant les hypothèses des Théorèmes 1.4.1, 1.4.2 et 1.4.3.

Le modèle INAR multivarié

Le modèle INAR multivarié (M-INAR) est introduit par Franke & Subba Rao (1993). Ce modèle définit X_t comme un vecteur évalué dans $\mathbb{N}^k$, tel que

$$X_t = \sum_{i=1}^{q} A_{0,i} \circ X_{t-1} + E_t, \tag{1.4.18}$$

où $A_{0,i}$ est une matrice de dimension $k \times k$ avec les valeurs $\alpha_{0,dc,i} \in \{0,1\}$ pour $(d,c) \in \{1,...,k\}$. La matrice des opérateurs d'amincissement $A_{0,i}\circ$ [1] est définie comme une matrice ordinaire au niveau de la multiplication matricielle et conserve, en même temps, les propriétés de l'opérateur d'amincissement (voir Steutel et al., 1979). Le terme des innovations $E_t = (\epsilon_{1,t}, ..., \epsilon_{k,t})'$ est un vecteur aléatoire *iid* à valeurs dans $\mathbb{N}^k$ dont les moyennes ω_d et les variances $\sigma^2_{\epsilon_{d,t}}$ de ses composantes sont finies. Le processus M-INAR(1) (*i.e.* $q = 1$) est stationnaire et ergodique si $\rho(A_0) < 1$, où $\rho(A_0)$ est le rayon spectral de la matrice A_0, et si $\|E_t\| < \infty$. Pour le modèle INAR(1) bivarié ($k = 2$), il y a plusieurs choix de distributions pour le terme des innovations, par exemple, la distribution de Poisson bivariée et la distribution binomiale négative bivariée. Dans ces deux cas, le modèle ne peut pas modéliser les séries ayant une corrélation croisée négative. Une approche alternative a été proposée par Karlis & Pedeli (2013) pour introduire une corrélation croisée négative entre les séries. Ce modèle est défini en supposant que les innovations du modèle INAR(1) bivarié ont des distributions marginales de Poisson ou de binomiale négative, et les

1. L'opérateur d'amincissement, appliqué à une variable à valeurs entières non négatives X, donne $\alpha \circ X = \sum_{i=1}^{X} y_i$ avec la convention $\alpha \circ X = 0$ lorsque $X = 0$, et où y_i's sont une variable aléatoire *iid* de Bernoulli indépendante de X et a une probabilité de succès α.

deux innovations sont connectées par une copule bivariée de Frank ou une copule gaussienne bivariée. En prenant l'espérance conditionnelle du modèle M-INAR, nous allons obtenir un modèle similaire à celui défini dans (1.4.7), où $E(X_t|\mathcal{F}_{t-1}) = \Lambda_t = \Omega + \sum_{i=1}^{q} A_{0,i}X_{t-i}$ et $E(E_t) = \Omega$.

Le modèle INGARCH de Poisson bivarié

Le modèle INGARCH de Poisson bivarié est défini comme (1.4.7) en supposant que la distribution conditionnelle de X_t sachant le passé est Poisson bivarié

$$X_t|\mathcal{F}_{t-1} \rightsquigarrow \mathcal{BP}(\lambda_{1.t}, \lambda_{2.t}, \phi), \tag{1.4.19}$$

où $\Lambda_t = (\lambda_{1.t}, \lambda_{2.t})'$ satisfait (1.4.7). La covariance conditionnelle entre les séries est ϕ. Selon la proposition 4.3.1 de Liu (2012), le modèle INGARCH de Poisson bivarié est stationnaire et ergodique si

$$2^{1-1/r}\sum_{i=1}^{q}\|A_{0,i}\|_r + \sum_{j=1}^{p}\|B_{o,j}\|_r < 1, \tag{1.4.20}$$

où, pour $0 \leq r \leq \infty$, $\|g\|_r$ est la r-*norm* de la matrice carrée g. La limitation de ce modèle est qu'il peut uniquement accueillir la corrélation croisée positive entre les séries.

Le modèle INGARCH multivarié avec copule

Une autre approche alternative qui peut accueillir la corrélation croisée négative entre les séries, est le modèle INGARCH multivarié avec une copule. Ce modèle est étudié par Heinen & Rengifo (2007) et est défini par (1.4.7), en supposant que la distribution conditionnelle de $X_{d,t}$ sachant le passé est double-Poisson $\mathcal{DP}(\lambda_{d,t}, \gamma_d)$. Les composantes du vecteur X_t sont supposées corrélées par une copule gaussienne multivariée. Pour faire face aux difficultés liées à l'utilisation des copules avec les distributions discrètes, Heinen & Rengifo (2007) a proposé d'utiliser l'extension continue des données de comptage (voir Denuit & Lambert, 2005).

Illustrations numériques

Dans la première partie, nous étudions le comportement de l'EQMVP-EpE à distance finie par des expériences de Monte Carlo basées sur 1000 réplications indépendantes. Les simulations sont réalisées sur les modèles INAR bivariés et les modèles INGARCH bivariés. Ces expériences montrent que, pour tous les modèles considérés, les moyennes des paramètres estimés par EQMVP-EpE sont très proches de leurs valeurs théoriques. De plus, les erreurs standards estimées basées sur la théorie asymptotique $\sqrt{\widehat{\Sigma}_{dd(i,i)}/n}$, où $\widehat{\Sigma}_{dd}$ est définie dans (1.4.11), et les erreurs moyennes quadratiques empiriques obtenues par les simulations sont très similaires et diminuent quand les tailles des échantillons augmentent. La deuxième partie de cette section présente une comparaison à distance finie entre EQMVP-EpE et MLE pour les modèles INAGRCH quand la distribution conditionnelle est Poisson bivarié et quand elle est binomiale négative bivariée. Les résultats montrent que MLE est plus efficace que EQMVP-EpE quand la taille d'échantillon est petite, mais la différence entre les erreurs standards des deux estimateurs diminue avec l'augmentation de la taille d'échantillon. Finalement, nous présentons une application aux données réelles.

Références

Ahmad, A. & Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis*, 37, 291–314.

Al-Osh, M. A. & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8, 261–275.

Alzaid, A. & Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 27, 314–324.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67, 1341–1383.

Blundell, R., Griffith, R., & Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics*, 108, 113–131.

Boudreault, M. & Charpentier, A. (2011). Multivariate integer-valued autoregressive models applied to earthquake counts. *arXiv :1112.0929*.

Brännäs, K., Hellström, J., & Nordström, J. (2002). A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting*, 18, 19–30.

Brännäs, K. & Nordström, J. (2006). Tourist accommodation effects of festivals. *Tourism Economics*, 12, 291–302.

Christou, V. (2014). *Statistical theory for mixed poisson time series models.* PhD thesis, University of Cyprus.

Christou, V. & Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35, 55–78.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., & Lauritzen, S. L. (1981). Statistical analysis of time series : some recent developments. *Scandinavian Journal of Statistics*, 8, 93–115.

Cui, Y. & Lund, R. (2009). A new look at time series of counts. *Biometrika*, 96, 781–792.

Denuit, M. & Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93, 40–57.

Douc, R., Doukhan, P., & Moulines, E. (2013). Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Processes and their Applications*, 123, 2620–2647.

Drost, F. C., Akker, R. v. d., & Werker, B. J. (2009). Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 71, 467–485.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81, 709–721.

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27, 923–942.

Fokianos, K. & Neumann, M. (2013). A goodness-of-fit test for poisson count processes. *Electronic Journal of Statistics*, 7, 793–819.

Fokianos, K. & Tjøstheim, D. (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102, 563–578.

Fokianos, K. & Tjøstheim, D. (2012). Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64, 1205–1225.

Francq, C. & Zakoïan, J.-M. (2009). Testing the nullity of GARCH coefficients : correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association*, 104, 313–324.

Francq, C. & Zakoian, J.-M. (2010). *GARCH models : structure, statistical inference and financial applications*. Chichester : John Wiley & Sons.

Franke, J. & Subba Rao, T. (1993). *Multivariate first order integer valued autoregressions*. Technical report, University of Kaiserslautern, Department of Mathematics.

Gao, J., King, M., Lu, Z., Tjøstheim, D., et al. (2009). Specification testing in nonlinear and nonstationary time series autoregression. *The Annals of Statistics*, 37, 3893–3928.

Garcia-Ferrer, A. & Queralt, R. A. (1997). A note on forecasting international tourism demand in spain. *International journal of forecasting*, 13, 539–549.

Gonçalves, E., Lopes, N. M., & Silva, F. (2015a). A new approach to integer-valued time series modeling : The neyman type-A INGARCH model. *Lithuanian Mathematical Journal*, 55, 231–242.

Gonçalves, E., Mendes-Lopes, N., & Silva, F. (2015b). Infinitely divisible distributions in integer-valued Garch models. *Journal of Time Series Analysis*, 36, 503–527.

Gonçalves, E., Mendes-Lopes, N., & Silva, F. (2016). Zero-inflated compound poisson distributions in integer-valued GARCH models. *Statistics*, 50, 558–578.

Grunwald, G. K., Hyndman, R. J., Tedesco, L., & Tweedie, R. L. (2000). Theory & methods : Non-gaussian conditional linear AR(1) models. *Australian & New Zealand Journal of Statistics*, 42, 479–495.

Heinen, A. (2003). Modelling time series count data : an autoregressive conditional poisson model. *Discussion Paper, vol. 2003/62. Center for operations research and econometrics (CORE), Catholic University of Louvain, Belgium.*

Heinen, A. & Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14, 564–583.

Hudecová, Š., Hušková, M., & Meintanis, S. G. (2015). Tests for time series of counts based on the probability-generating function. *Statistics*, 49, 316–337.

Jung, R. C. & Tremayne, A. (2011a). Convolution-closed models for count time series with applications. *Journal of Time Series Analysis*, 32, 268–280.

Jung, R. C. & Tremayne, A. (2011b). Useful models for time series of counts or simply wrong ones ? *AStA Advances in Statistical Analysis*, 95, 59–91.

Karlis, D. & Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. *Communications in Statistics-Theory and Methods*, 42, 723–740.

Kedem, B. & Fokianos, K. (2002). *Regression models for time series analysis*. Hoboken, New Jersey : John Wiley & Sons.

Kokonendji, C. C. (2014). Over-and underdispersion models. *The Wiley Encyclopedia of Clinical Trials-Methods and Applications of Statistics in Clinical Trials*, 2, 506–526.

Lambert, D. & Liu, C. (2006). Adaptive thresholds : monitoring streams of network counts. *Journal of the American Statistical Association*, 101, 78–88.

Latour, A. (1997). The multivariate GINAR(p) process. *Advances in Applied Probability*, 29, 228–248.

Liu, H. (2012). *Some models for time series of counts*. PhD thesis, University of Colombia.

McCabe, B. & Martin, G. M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting*, 21, 315–330.

McKenzie, E. (1985). Some simple models for discrete variate time series1. *Water Resources Bulletin*, 21, 645–650.

Meintanis, S. G. & Karlis, D. (2014). Validation tests for the innovation distribution in INAR time series models. *Computational Statistics*, 29, 1221–1241.

Nelsen, R. (2006). *An introduction to copulas, 2nd*. New York : Springer Science Business Media.

Neumann, M. (2011). Absolute regularity and ergodicity of poisson count processes. *Bernoulli*, 17, 1268–1284.

Pavlopoulos, H. & Karlis, D. (2008). INAR(1) modeling of overdispersed count series with an environmental application. *Environmetrics*, 19, 369–393.

Pedeli, X., Davison, A. C., & Fokianos, K. (2014). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association*, (published online DOI :10.1080/01621459.2014.983230).

Pedeli, X. & Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical modelling*, 11, 325–349.

Quoreshi, A. S. (2006). Bivariate time series modeling of financial count data. *Communications in Statistics-Theory and Methods*, 35, 1343–1358.

Schweer, S. (2016). A goodness-of-fit test for integer-valued autoregressive processes. *Journal of Time Series Analysis*, 37, 77–98.

Scotto, M. G., Weiß, C. H., Silva, M. E., & Pereira, I. (2014). Bivariate binomial autoregressive models. *Journal of Multivariate Analysis*, 125, 233–251.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

Steutel, F., van Harn, K., et al. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7, 893–899.

Thyregod, P., Carstensen, J., Madsen, H., & Arnbjerg-Nielsen, K. (1999). Integer valued autoregressive models for tipping bucket rainfall measurements. *Environmetrics*, 10, 395–411.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, 75, 621–629.

Zeger, S. L. & Qaqish, B. (1988). Markov regression models for time series : a quasi-likelihood approach. *Biometrics*, 44, 1019–1031.

Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications*, 389, 58–71.

Zhu, F. (2012b). Modeling time series of counts with COM-poisson INGARCH models. *Mathematical and Computer Modelling*, 56, 191–203.

Zhu, F. (2012c). Zero-inflated poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference*, 142, 826–839.

Zhu, F. & Wang, D. (2010). Diagnostic checking integer-valued ARCH(p) models using conditional residual autocorrelations. *Computational Statistics & Data Analysis*, 54, 496–508.

Chapitre 2

Poisson QMLE of count time series models

Abstract. Regularity conditions are given for the consistency of the Poisson quasi-maximum likelihood estimator of the conditional mean parameter of a count time series model. The asymptotic distribution of the estimator is studied when the parameter belongs to the interior of the parameter space and when it lies at the boundary. Tests for the significance of the parameters and for constant conditional mean are deduced. Applications to specific integer-valued autoregressive (INAR) and integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH) models are considered. Numerical illustrations, Monte Carlo simulations and real data series, are provided.

Keywords. Boundary of the parameter space, Consistency and asymptotic normality, Integer-valued AR and GARCH models, Non-normal asymptotic distribution, Poisson quasi-maximum likelihood estimator, Time series of counts.

2.1 Introduction

The literature on time series of counts is becoming increasingly abundant, with applications in numerous domains (see *e.g.* the monographs by Christou (2014) and Liu (2012), and the references therein). It is common to assume a conditional Poisson distribution with the intensity parameter depending on the past values. This leads to models that are quite tractable [1], but extremely constrained, since their conditional variance and conditional mean coincide. Many extensions and alternative conditional distributions have been proposed in the literature, for example, Heinen (2003), Zhu (2011, 2012a,b,c) and Christou & Fokianos (2014), but either the conditional distribution remains relatively constrained or it contains extra parameters that are difficult to estimate and interpret.

In the present chapter we adopt a semi-parametric approach, in which only the conditional mean is specified. Since the works of Wedderburn (1974), White (1982), McCullagh (1983) and Gourieroux et al. (1984), it is known that certain maximum likelihood estimators (MLEs) can be consistent and asymptotically normal (CAN) for the parameters of the conditional mean and variance, even if the actual conditional distribution is not the one which is assumed by the MLE. In particular, the Gaussian quasi-maximum likelihood estimator (QMLE), in which the conditional mean and variance parameters are estimated by maximizing a pseudo-likelihood written as if the condition mean were Gaussian, is the method of choice for estimating autoregressive moving average-generalized autoregressive conditional heteroscedasticity (ARMA-GARCH)-type models. For time series of counts, the Poisson QMLE (PQMLE) can be employed to identify the conditional mean. This estimator has been studied by several authors for particular count time series models, for example, Zeger (1988), Zeger & Qaqish (1988), Kedem & Fokianos (2002, Chap. 4) and Christou & Fokianos (2014). In this chapter, we consider a quite general framework and give regularity conditions under which the PQMLE is CAN. We also consider the case where the aforementioned regularity conditions are violated because the parameter stands at the boundary of the parameter space. In that case the asymptotic distribu-

1. even if the probabilistic structure, in particular the ergodicity, of these models is not easy to derive (see Tjøstheim (2012, 2016) and the references therein)

tion is not Gaussian. This situation must be considered for testing the nullity of some conditional mean parameters. For important classes of time series of counts, such as the integer-valued GARCH (INGARCH) models, the significance test statistics are not asymptotically distributed as a standard chi-square, but as chi-bar-square. The general results are applied to specific models, namely the integer-valued autoregressive (INAR), the INGARCH and the log-linear models, with different specifications of the conditional distribution. Thus, the main contribution of the present chapter is threefold. Firstly, the asymptotic theory of the PQMLE is developed for estimating the conditional mean parameters without having to specify entirely the conditional distribution. We complete the aforementioned existing results by considering a very general parametric form for the conditional mean, and by providing applications to linear or log-linear models. Second, the asymptotic distribution of the estimator is obtained without positivity constraint on the coefficients, which is new for count time series. Third, Wald-type significance tests are proposed. Due to boundary effects, the asymptotic distribution of these tests is not standard, but they can however be easily implemented and are obviously useful to model identification. These theoretical results are illustrated by Monte Carlo simulations and applications on financial data. The chapter is organized as follows. Section 2.2 contains the main results concerning the asymptotic behavior of the PQMLE and of the related significance tests. Section 2.3 applies the general results to particular observation-driven and parameter-driven models (according to the nomenclature introduced by Cox et al. (1981)). Section 2.4 studies the finite sample properties of the PQMLE and of the significance tests, via a set of Monte Carlo experiments. In Section 2.5, we use the PQMLE to fit INGARCH(p, q) models on daily series of the number of transactions of stocks. The proofs are collected in Section 2.6, and Section 2.7 concludes.

2.2 Asymptotic distribution of the Poisson QMLE

Assume that we have observations $X_1, \dots, X_n$ of a times series valued in $\mathbb{N}$, such that

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda(X_{t-1}, X_{t-2}, \dots; \theta_0), \tag{2.2.1}$$

where $\mathcal{F}_{t-1}$ denotes the $\sigma-$field generated by $(X_u, u < t)$,

$$\lambda \text{ is a measurable function valued in } (\underline{\omega}, +\infty) \text{ for some } \underline{\omega} > 0 \tag{2.2.2}$$

and θ_0 is an unknown parameter belonging to some parameter space $\Theta \subset \mathbb{R}^d$. The marginal distribution is assumed to have a moment slightly greater than 1

$$EX_t^{1+\varepsilon} < \infty, \text{ for some } \varepsilon > 0, \tag{2.2.3}$$

which entails the existence of the conditional mean (2.2.1). For all $\theta \in \Theta$, $x_0 \in \mathbb{R}^+$ and $t \geq 1$, let

$$\lambda_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \dots; \theta) \text{ and } \widetilde{\lambda}_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \dots, X_1, x_0, x_0, \dots; \theta).$$

Note that $\widetilde{\lambda}_t(\theta)$ will serve as a proxy for $\lambda_t(\theta)$. It is obtained by setting to some positive real value x_0 the unknown initial values $X_0, X_{-1}, \dots$ involved in $\lambda_t(\theta)$. This value x_0 can either be a fixed integer, for instance $x_0 = 0$, or a value depending on θ, or a value depending on the observations. For example, when $\lambda_t(\theta) = \omega + \alpha X_{t-1} + \beta\lambda_{t-1}(\theta)$ with $\theta = (\omega, \alpha, \beta)$, one can take $\widetilde{\lambda}_t(\theta) = \omega + \alpha X_{t-1} + \beta\widetilde{\lambda}_{t-1}(\theta)$ with $\widetilde{\lambda}_1(\theta) = \omega/(1-\beta)$ (which corresponds to $x_0 = 0$), or with $\widetilde{\lambda}_1(\theta) = \omega/(1-\alpha-\beta)$ (which corresponds to $x_0 = \omega/(1-\alpha-\beta)$), or with $\widetilde{\lambda}_1(\theta) = \overline{X}_5$ (the average of the working days of the first week, for daily data). It will be shown that the choice of x_0 is asymptotically unimportant, provided we have a.s.

$$\lim_{t\to\infty} a_t = 0 \quad \text{ and } \lim_{t\to\infty} X_t a_t = 0, \quad \text{ where } a_t = \sup_{\theta\in\Theta} \left|\widetilde{\lambda}_t(\theta) - \lambda_t(\theta)\right|, \tag{2.2.4}$$

and

$$\widetilde{\lambda}_t(\theta) \geq \underline{\omega}, \quad \forall t \geq 1, \ \forall \theta \in \Theta. \tag{2.2.5}$$

Assuming that

$$\theta \mapsto \lambda_t(\theta) \text{ is almost surely continuous and } \Theta \text{ is a compact set,} \tag{2.2.6}$$

then a PQMLE of θ_0 is defined as any measurable solution of

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \widetilde{L}_n(\theta), \quad \widetilde{L}_n(\theta) = \frac{1}{n}\sum_{t=s+1}^{n} \widetilde{\ell}_t(\theta), \tag{2.2.7}$$

where $\widetilde{\ell}_t(\theta) = -\widetilde{\lambda}_t(\theta) + X_t \log \widetilde{\lambda}_t(\theta)$ and s is an integer. The value of s is asymptotically unimportant, but it can affect the finite sample behavior of the PQMLE by reducing the impact of the initial value x_0. Note that $\widehat{\theta}_n$ is equal to the maximum likelihood estimator (MLE) of θ_0 if the conditional distribution of X_t is Poisson with parameter $\lambda_t(\theta_0)$. Since we do not make any specific assumption on the conditional distribution of X_t, the estimator is called "quasi" MLE (QMLE). The reader is referred to Gourieroux et al. (1984) for a general reference on QMLE.

2.2.1 Consistency of the PQMLE

As shown by the following theorem, the essential assumption required for the consistency of the PQMLE is that the conditional mean be well specified. Obviously, the following identifiability assumption is also required :

$$\lambda_1(\theta) = \lambda_1(\theta_0) \text{ almost surely if and only if } \theta = \theta_0. \tag{2.2.8}$$

Theorem 2.2.1. *Assume that (X_t) is an ergodic strictly stationary sequence valued in $\mathbb{N}$, satisfying (2.2.1)-(2.2.6) and (2.2.8). Then the PQMLE defined by (2.2.7) satisfies*

$$\widehat{\theta}_n \to \theta_0 \quad a.s. \qquad as\ n \to \infty.$$

In the sequel, K and ρ denote generic constants, or random variables depending on $\{X_u, u \leq 0\}$, such that $K > 0$ and $\rho \in (0,1)$. It is often assumed that $\lambda_t(\theta)$ is a linear function of the past values. In that case, the regularity conditions become much simpler.

Remark 2.2.1 (linear conditional mean)**.** *Assume that Θ is a compact subset of $(0,\infty) \times [0,\infty)^{p+q}$, that $\theta_0 = (\omega_0, \alpha_{01}, \dots, \beta_{0p})$, and that*

$$\lambda_t(\theta_0) = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} X_{t-i} + \sum_{j=1}^{p} \beta_{0j} \lambda_{t-j}(\theta_0). \tag{2.2.9}$$

Assume also that for all $\theta = (\omega, \alpha_1, \dots, \beta_p) \in \Theta$, we have $\sum_{i=1}^{p} \beta_i < 1$. This condition ensures that $\lambda_t(\theta_0)$ is well defined for all $\theta \in \Theta$. Noting that the equation (2.2.9) is similar to that satisfied by the volatility in a GARCH(p, q) model, it is easy to show (by for instance the arguments used to show (7.30) in Francq & Zakoian (2010), denoted hereafter as FZ) that $a_t \leq K\rho^t$. The first condition in (2.2.4) directly follows. To show the second convergence of (2.2.4), it suffices to use the Borel-Cantelli lemma and $P(\rho^t X_t \geq \varepsilon) \leq \rho^t EX_t/\varepsilon$, for $\varepsilon > 0$. The conditions (2.2.2) and (2.2.4)-(2.2.6) are thus satisfied without any additional constraint. Let the polynomials $\mathcal{A}_\theta(z) = \sum_{i=1}^{q} \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{i=1}^{p} \beta_i z^i$. As in the proof of (b) Page 157 in FZ, the identifiability condition (2.2.8) is satisfied by assuming that the conditional distribution of X_t is not degenerated and

$$\begin{aligned} &\text{if } p > 0,\ \mathcal{A}_{\theta_0}(z) \text{ and } \mathcal{B}_{\theta_0}(z) \text{ have no common root,} \\ &\text{at least one } \alpha_{0i} \neq 0 \text{ for } i = 1, \dots, q, \text{ and } \beta_{0p} \neq 0 \text{ if } \alpha_{0q} = 0. \end{aligned} \tag{2.2.10}$$

In the case $q = 1$, the conditions (2.2.10) simply amount to assuming $\alpha_{01} > 0$.

The stationarity and ergodicity issues will be discussed for particular classes of count models in Section 2.3.

2.2.2 Asymptotic distribution

As expected, under mild regularity conditions, the Poisson QMLE turns out to be asymptotically normal when the parameter belongs to the interior of the parameter space. In the more general situation where the parameter may lie at the boundary of the parameter space, its asymptotic distribution is the projection of a Gaussian random vector on a convex cone. Estimators with similar asymptotic distributions have been studied by, for example, Andrews (1999), Francq & Zakoïan (2009b) and the references therein.

2.2.2.1 When θ_0 belongs to the interior of Θ

The asymptotic normality of the PQMLE has been studied by several authors (see in particular Christou & Fokianos (2014) and the other references cited in the introduction). The general framework of a conditional mean of the form (2.2.1) has however never been considered. Fokianos & Tjøstheim (2012) considered the MLE for a first-order Poisson nonlinear model. Here, we give conditions for the asymptotic normality of the PQMLE when the true conditional distribution is unspecified. We need to assume the existence of

$$Var(X_t \mid \mathcal{F}_{t-1}) = E\left(X_t^2 \mid \mathcal{F}_{t-1}\right) - \lambda_t^2(\theta_0), \tag{2.2.11}$$

the existence of continuous second-order derivatives for $\lambda_t(\cdot)$ and $\widetilde{\lambda}_t(\cdot)$, as well as the existence of the information matrices

$$J = E\left(\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}\right), \quad I = E\left(\frac{Var(X_t \mid \mathcal{F}_{t-1})}{\lambda_t^2(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}\right). \tag{2.2.12}$$

It is easy to see that the matrix J is invertible when

$$c'\frac{\partial\lambda_t(\theta_0)}{\partial\theta} = 0 \quad a.s. \Rightarrow c = 0. \tag{2.2.13}$$

We also assume that there exists a neighborhood $V(\theta_0)$ of θ_0 such that, for all $(i,j) \in \{1,\ldots,d\}$,

$$E \sup_{\theta\in V(\theta_0)}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta)\right| < \infty. \tag{2.2.14}$$

Assume also that, a.s.,

$$b_t, b_tX_t \text{ and } a_td_tX_t \text{ are of order } O(t^{-\kappa}) \text{ for some } \kappa > 1/2, \tag{2.2.15}$$

where

$$b_t = \sup_{\theta\in\Theta}\left\|\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta} - \frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|, \quad d_t = \sup_{\theta\in\Theta}\max\left\{\left\|\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|, \left\|\frac{1}{\widetilde{\lambda}_t(\theta)}\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta}\right\|\right\}.$$

Theorem 2.2.2. *Assume that (X_t) satisfies the conditions of Theorem 2.2.1. Assume also (2.2.11)-(2.2.15). If $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ, then*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \Sigma := J^{-1}IJ^{-1}\right) \qquad as\ n \to \infty.$$

Note that when the distribution of X_t conditional to its past is Poisson, we have $I = J$, and thus $\Sigma = J^{-1}$. In that case we retrieve the asymptotic variance obtained by Ferland et al. (2006) for the MLE of a Poisson INGARCH model.

It can be shown that, under the assumptions of Theorem 2.2.2, the asymptotic variance of the PQMLE can be consistently estimated by $\widehat{\Sigma} = \widehat{J}^{-1}\widehat{I}\widehat{J}^{-1}$ with

$$\widehat{J} = \frac{1}{n}\sum_{t=s+1}^{n} \frac{1}{\widetilde{\lambda}_t(\widehat{\theta}_n)} \frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta'}, \tag{2.2.16}$$

$$\widehat{I} = \frac{1}{n}\sum_{t=s+1}^{n} \left(\frac{X_t}{\widetilde{\lambda}_t(\widehat{\theta}_n)} - 1\right)^2 \frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta'}. \tag{2.2.17}$$

Remark 2.2.2 (alternative conditions to (2.2.12) and (2.2.14))**.** *. Note that*

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta) = \left(\frac{X_t}{\lambda_t(\theta)} - 1\right)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta) - \frac{X_t}{\lambda_t^2(\theta)}\frac{\partial}{\partial\theta_i}\lambda_t(\theta)\frac{\partial}{\partial\theta_j}\lambda_t(\theta). \tag{2.2.18}$$

Using Hölder's inequality and (2.2.3)*, one can thus obtain* (2.2.14) *by showing that*

$$E\sup_{\theta\in V(\theta_0)}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta)\right| < \infty, \tag{2.2.19}$$

and

$$E\sup_{\theta\in V(\theta_0)}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial}{\partial\theta_i}\lambda_t(\theta)\right|^{r_0} < \infty, \qquad E\sup_{\theta\in V(\theta_0)}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta)\right|^{r_0} < \infty \tag{2.2.20}$$

for all $r_0 > 0$. Note also that (2.2.3) *and* (2.2.20) *entail the existence of J. For the existence of I an extra assumption is needed. In particular,* (2.2.12) *is obtained under the conditions* (2.2.20) *and*

$$E\mid Var\,(X_t \mid \mathcal{F}_{t-1})\mid^{1+\varepsilon} < \infty, \; \textit{for some } \varepsilon > 0, \tag{2.2.21}$$

Remark 2.2.3 (linear conditional mean)**.** *Let us come back to the case of the linear conditional mean* (2.2.9)*. Because the roots of the polynomial $\mathcal{B}_\theta(z)$ are outside the unit disk, we have $\lambda_t(\theta) = \pi_0(\theta) + \sum_{k=1}^{\infty}\pi_k(\theta)X_{t-k}$, where*

$$\sum_{k=1}^{\infty}\pi_k(\theta)z^k = \mathcal{B}_\theta^{-1}(z)\mathcal{A}_\theta(z) \quad \textit{and} \quad \sup_{\theta\in\Theta}|\pi_k(\theta)| \le K\rho^k.$$

Letting $\pi_k^{(i,j)}(\theta) = \partial^2\pi_k(\theta)/\partial\theta_i\partial\theta_j$, we also have

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_t(\theta) = \pi_0^{(i,j)}(\theta) + \sum_{k=1}^{\infty}\pi_k^{(i,j)}(\theta)X_{t-k} \quad \textit{with} \quad \sup_{\theta\in\Theta}|\pi_k^{(i,j)}(\theta)| \le K\rho^k.$$

Under the moment assumption (2.2.3)*, the condition* (2.2.19) *is thus satisfied, whatever the neighborhood $V(\theta_0)$ included in Θ. One can show* (2.2.20) *by the arguments used to prove (7.54) in FZ. We thus obtain* (2.2.14)*, and* (2.2.12) *under* (2.2.21)*.*

Now, note that $b_t \leq K\rho^t$ and that d_t admits moments at any order, by arguments already given. We thus have $E|t^\kappa b_t X_t| \leq K t^\kappa \rho^t$ and $E|t^\kappa a_t d_t X_t| \leq K t^\kappa \rho^t$, which entails (2.2.15) by the Borel-Cantelli lemma and the Markov inequality.

Finally, note that (2.2.13) is a consequence of (2.2.10), by the arguments used to show (b), Page 162 in FZ.

2.2.2.2 When θ_0 belongs to the boundary of Θ

For the computation of the PQMLE it is obviously necessary to have $\widetilde{\lambda}_t(\theta) > 0$ almost surely, for any $\theta \in \Theta$. For that reason, the parameter space Θ must be constrained. Very often, one or several components of θ are constrained to be positive or equal to zero. For example, when we have an INGARCH(1,1) model of the form $\lambda_t(\theta) = \omega + \alpha X_{t-1} + \beta\lambda_{t-1}(\theta)$ then $\theta = (\omega, \alpha, \beta) \in \Theta \subset [\underline{\omega}, \infty) \times [0, \infty)^2$. Following the celebrated Box-Jenkins time series methodology, it is often interesting to test if the model can be simplified. For the INGARCH(1,1) example, it is of interest to test if the true parameter is of the form $\theta_0 = (\omega_0, \alpha_0, 0)$, that is, if the DGP is an INARCH(1). Theorem 2.2.2 does not apply because, in that situation, $\theta_0 \notin \overset{\circ}{\Theta}$. Moreover the asymptotic distribution of $\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)$ is not Gaussian (see Figure 2.12 in Appendix) because the positivity constraints entail that $\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \geq 0$ with probability one when $\beta_0 = 0$.

We now come back to the general model. The component i of the parameter θ is said to be positively constrained if the ith section of Θ is of the form $[0, \overline{\theta}_i]$ with $\overline{\theta}_i > 0$. For example, for the linear model (2.2.9), the first component is not positively constrained, but the other components are. Let $d_2 \in \{0, \dots, d\}$ be the number of positively constrained components of θ, and let $d_1 = d - d_2$. Without loss of generality, assume that these d_2 constrained components are the last ones. If one of the last d_2 components of θ_0 is equal to zero, then Theorem 2.2.2 does not apply because θ_0 belongs to the boundary of Θ. We assume that $\Theta - \theta_0$ is large enough to contain a hypercube of the form $\prod_{i=1}^d [\underline{\theta}_i, \overline{\theta}_i]$ where, for all $i \in \{1, \dots, d\}$, $\underline{\theta}_i = 0$ if $\theta_{0i} = 0$ with $i > d_1$, $\underline{\theta}_i < 0$ otherwise, and $\overline{\theta}_i > 0$. Under this assumption we have

$$\lim_{n\to\infty} \sqrt{n}(\Theta - \theta_0) = \mathcal{C}, \tag{2.2.22}$$

where $\mathcal{C} = \prod_{i=1}^d \mathcal{C}_i$, in which $\mathcal{C}_i = \mathbb{R}$ when $i \leq d_1$ or $\theta_{0i} > 0$ and $\mathcal{C}_i = [0, +\infty)$ when

$i > d_1$ and $\theta_{0i} = 0$.

Similarly to (2.2.15), assume that, a.s.,

$$\lim_{t\to\infty} c_t + X_t \left(a_t e_t + c_t + a_t d_t^2 + b_t d_t \right) = 0, \tag{2.2.23}$$

where

$$c_t = \sup_{\theta\in\Theta} \left\| \frac{\partial^2 \widetilde{\lambda}_t(\theta)}{\partial\theta\partial\theta} - \frac{\partial^2 \lambda_t(\theta)}{\partial\theta\partial\theta'} \right\|,$$
$$e_t = \sup_{\theta\in\Theta} \max \left\{ \left\| \frac{1}{\lambda_t(\theta)} \frac{\partial^2 \lambda_t(\theta)}{\partial\theta\partial\theta'} \right\|, \left\| \frac{1}{\widetilde{\lambda}_t(\theta)} \frac{\partial^2 \widetilde{\lambda}_t(\theta)}{\partial\theta\partial\theta'} \right\| \right\}.$$

Note that for the linear model (2.2.9), the condition (2.2.23) is always satisfied. Since J is positive definite, one can consider the norm $\|x\|_J^2 = x'Jx$ and the scalar product $\langle x, y\rangle_J = x'Jy$ for $x, y \in \mathbb{R}^d$. With this metric, the projection of a vector $Z \in \mathbb{R}^d$ on the convex cone $\mathcal{C}$ is defined by

$$Z^{\mathcal{C}} = \arg \inf_{C\in\mathcal{C}} \|C - Z\|_J$$

or equivalently by

$$Z^{\mathcal{C}} \in \mathcal{C} \quad \text{and} \quad \left\langle Z - Z^{\mathcal{C}}, C - Z^{\mathcal{C}} \right\rangle_J \leq 0, \quad \forall C \in \mathcal{C}. \tag{2.2.24}$$

Theorem 2.2.3. *Assume the conditions of Theorem 2.2.2 (except that $\theta_0 \in \overset{\circ}{\Theta}$) and (2.2.22), (2.2.23). Then, as $n \to \infty$,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Z^{\mathcal{C}} = \arg \inf_{C\in\mathcal{C}} \|C - Z\|_J, \ \text{where } Z \sim \mathcal{N}(0, \Sigma).$$

Note that, when $\theta_0 \in \overset{\circ}{\Theta}$ we have $\mathcal{C} = \mathbb{R}^d$ and $Z^{\mathcal{C}} = Z$. In that case, we retrieve the CAN of the PQMLE, as stated in Theorem 2.2.2. When $\theta_0 \notin \overset{\circ}{\Theta}$, the conditions required for the existence of the information matrices I and J can however be more demanding in terms of moments of X_t.

Remark 2.2.4 (alternative to (2.2.12) and (2.2.14))**.** *When θ does not belong to the interior of Θ, the conditions (2.2.20) generally impose restrictive moment conditions on the observed process. For example, in the linear case of the linear model (2.2.9), they may impose the existence of $EX_t^{r_0}$. By (2.2.18) and Hölder's inequality, (2.2.14) can be obtained by showing (2.2.19), (2.2.20) for $r_0 = 3$ and $EX_t^3 < \infty$. In the linear model (2.2.9), this is equivalent to $EX_t^3 < \infty$. To obtain (2.2.12), one can impose the existence of I and the additional moment condition $E \mid Var(X_t \mid \mathcal{F}_{t-1}) \mid^3 < \infty$. Alternatively, one can impose $E \mid Var(X_t \mid \mathcal{F}_{t-1}) \mid^2 < \infty$ and $r_0 = 4$.*

Note that the matrices I and J are still estimated by (2.2.16) and (2.2.17). As an application of Theorem 2.2.3, let us assume $d_2 > 0$ and consider the testing problem

$$H_0 : \theta_{0d} = 0 \quad \text{against} \quad H_1 : \theta_{0d} > 0.$$

Denote by $\widehat{\theta}_{nd}$ the last component of $\widehat{\theta}_n$ and denote by $\chi^2_k(\underline{\alpha})$ the $\underline{\alpha}$-quantile of a chi-squared distribution with k degrees of freedom χ^2_k. If one also assume that, under the null, only the last component of θ_0 is at the boundary (see Example 8.2 and Section 8.3.3 in FZ), then the test of rejection region

$$\left\{ \frac{n\widehat{\theta}^2_{nd}}{\widehat{\Sigma}(d,d)} \geq \chi^2_1(1-2\underline{\alpha}) \right\} \tag{2.2.25}$$

has the asymptotic level $\underline{\alpha}$. Note that when the last component is not positively constrained (*i.e.* when $d_2 = 0$) the PQMLE has a normal asymptotic distribution and the critical value of the test is $\chi^2_1(1-\underline{\alpha})$. Another application of Theorem 2.2.3 is given in the following corollary. Denote by δ_0 the Dirac mass at 0 and denote by $p_0\delta_0 + \sum_{i=1}^q p_i\chi^2_{k_i}$ the mixture of δ_0 and of $\chi^2_{k_i}$-distributions, with the mixture weights $p_0, \ldots, p_q$.

Corollary 2.2.1 (testing for constant conditional mean)**.** . *Consider an ergodic strictly stationary process (X_t) with a linear conditional mean of the form $\lambda_t(\theta) = \omega + \sum_{i=1}^q \alpha_i X_{t-i}$. Assume that the conditional distribution of X_t depends on the past only through $\lambda_t(\theta)$. If $\theta_0 = (\omega_0, 0, \ldots, 0)' \in \Theta$, where Θ is a compact subset of $(0,\infty) \times [0,\infty)^q$, and if $EX_t^4 < \infty$, then the statistic*

$$S_n = n\sum_{i=1}^q \widehat{\alpha}^2_{ni} \xrightarrow{d} \frac{1}{2^q}\delta_0 + \sum_{i=1}^q \binom{q}{i}\frac{1}{2^q}\chi^2_i \qquad \text{as } n \to \infty, \tag{2.2.26}$$

where $\widehat{\alpha}_{ni} = \widehat{\theta}_{n,i+1}$.

The asymptotic distribution is known as a chi-bar-square distribution, and has been tabulated. The constant conditional mean assumption is rejected at the asymptotic level $\underline{\alpha}$ if $\{S_n > c_{q,\underline{\alpha}}\}$, where the critical value $c_{q,\underline{\alpha}}$ can be found in Table 8.2 of FZ.

2.3 Application to particular models

We now show that the regularity conditions required for the asymptotic results of the previous section can be made explicit for the most popular classes of observation-driven and parameter-driven models for time series of counts (see Cox et al. (1981) for the distinction between observation-driven and parameter-driven models).

2.3.1 Poisson INGARCH model

One of the most natural count time series model is the Poisson INGARCH model, which has been studied by Ferland et al. (2006). This model is also called Autoregressive Conditional Poisson in Heinen (2003). The INGARCH(p, q) model is obtained by assuming that the conditional distribution of X_t given its past values is Poisson with intensity parameter of the linear form (2.2.9). Ferland et al. (2006) showed that there exists a stationary process (X_t) satisfying the INGARCH model, with second-order moments, under the assumption

$$\sum_{i=1}^{q} \alpha_{0i} + \sum_{i=1}^{p} \beta_{0i} < 1. \tag{2.3.1}$$

As shown in Tjøstheim (2012, 2016), the ergodicity of the stationary solution is a difficult issue. Fokianos et al. (2009) showed that this model can be approximated by an ergodic process and applied this result to the likelihood inference. By using different techniques and different frameworks encompassing the Poisson INARCH model, Neumann (2011), Liu (2012), Davis & Liu (2016), Christou & Fokianos (2014) and Gonçalves et al. (2015) showed the ergodicity. Under (2.3.1) and the assumptions of Remark 2.2.1 on the linear model (2.2.9), Theorem 2.2.1 thus establishes the strong consistency of the PQMLE. Since (2.3.1) also entails the existence of moments of any order (see Christou & Fokianos, 2014), the condition (2.2.21) is obviously verified, and Remark 2.2.3 entails that the conclusion of Theorem 2.2.2 holds true when θ_0 belongs to the interior of the parameter space. Davis & Liu (2016) studied the MLE for a wide class of integer-valued models. Their conditions for the CAN of the MLE coincide with ours in the Poisson INGARCH case. This was quite expected because the PQMLE is actually the MLE in the framework of this section. Similarly, the regularity conditions required for Theorem 2.2.3 and Corollary 2.2.1 are satisfied. To our knowledge, the asymptotic behaviour of the MLE had never been studied for count time series models with parameter at the boundary of the parameter space.

2.3.2 Negative binomial INGARCH model

As an alternative to the conditional Poisson distribution, Zhu (2011) and Christou & Fokianos (2014) considered the negative binomial distribution $\mathcal{NB}(r, p_t)$ with parameters $r > 0$ and $p_t = r/(\lambda_t + r)$ where λ_t is, for instance, of the form (2.2.9). We still have $E(X_t \mid \mathcal{F}_{t-1}) = \lambda_t$, but the conditional variance $\lambda_t + \lambda_t^2/r$ is larger than the conditional variance of the Poisson case, which reflects the (conditional) overdispersion that is suspected to be present on real series (see Christou & Fokianos, 2014 and Gonçalves et al., 2015). From Proposition 3.4.1 in Liu (2012), the condition (2.3.1) entails the existence of an ergodic and strictly stationary solution (X_t). In the case $(p, q) = (1, 1)$, it can be shown (see Christou & Fokianos, 2014), that the stationary solution is such that $EX_t^2 < \infty$ if and only if

$$(\alpha_0 + \beta_0)^2 + \frac{\alpha_0^2}{r} < 1, \tag{2.3.2}$$

writing α_0 and β_0 instead of α_{01} and β_{01}. Always in the case $(p, q) = (1, 1)$, it can be shown (see the appendix), that $EX_t^4 < \infty$ if and only if

$$(\alpha_0 + \beta_0)^4 + \frac{6\alpha_0^2(\alpha_0 + \beta_0)^2}{r} + \frac{\alpha_0^3(11\alpha_0 + 8\beta_0)}{r^2} + \frac{6\alpha_0^4}{r^3} < 1. \tag{2.3.3}$$

The conditions ensuring the existence of EX_t^2 are much more complicated for general orders p and q (see Theorem 2 in Zhu, 2011). The regularity conditions required for Theorems 2.2.1 and 2.2.2 are thus explicit, at least in the case $(p, q) = (1, 1)$. Christou & Fokianos (2014) had already noted that the Poisson QMLE is consistent in the case of a negative binomial conditional distribution. To our knowledge the result stated in Theorem 2.2.3 and Corollary 2.2.1 are however new.

Motivated by the zero inflation phenomenon (*i.e.* the fact that, for numerous count time series, the proportion of zeros is greater than the proportion of zeros in a Poisson INGARCH model), Zhu (2012c) proposed zero-inflated Poisson and negative binomial INGARCH models. Since the conditional mean of these models remains of the form (2.2.9), the PQMLE can directly be used to estimate the parameters of λ_t, with or without zero inflation in the conditional distribution.

2.3.3 Double-Poisson INGARCH model (DACP model)

Count time series often exhibit over-dispersion, that is, the variance larger than the mean, but the opposite phenomenon may be encountered. The Poisson and negative binomial INGARCH models can not take into account the under-dispersion. To tackle the problem, Heinen (2003) proposes a model based on the double-Poisson distribution of Efron (1986). This distribution, which has two parameters $\lambda > 0$ and $\gamma > 0$, is defined by the probability mass function

$$P(X = x) = c(\lambda, \gamma) \left(\frac{e^{-x} x^x}{x!} \right) \left(\frac{e\lambda}{x} \right)^{\gamma x}, \qquad x = 0, 1, \ldots.$$

where $c(\lambda, \gamma)$ is a normalization constant. We then use the notation $X \sim \mathcal{DP}(\lambda, \gamma)$. Efron (1986) shows that the mean of the $\mathcal{DP}(\lambda, \gamma)$ distribution is λ, and that its variance is approximately equal to λ/γ. The double-Poisson INGARCH model is defined by assuming that the conditional distribution of X_t given its past values is $\mathcal{DP}(\lambda_t, \gamma)$ with a parameter λ_t of the form (2.2.9). For $(p, q) = (1, 1)$, according to Propositions 3.1 and 3.2 in Heinen (2003), the condition (2.3.1) entails the existence of a stationary solution (X_t) such that

$$E(X_t) = \frac{\omega_0}{1 - \alpha_0 - \beta_0}, \qquad \mathrm{Var}(X_t) = \frac{1 - (\alpha_0 + \beta_0)^2 + \alpha_0^2}{1 - (\alpha_0 + \beta_0)^2} \frac{E[X_t]}{\gamma}. \tag{2.3.4}$$

The ergodicity has not been established, but one can conjecture that, as for the Poisson distribution, the ergodicity holds under the same conditions. In view of Remark 2.2.1, the consistency result of Theorem 2.2.1 thus holds true in the case $(p, q) = (1, 1)$ when $\alpha_{01} + \beta_{01} < 1$ and $\alpha_{01} > 0$. Similarly, the conditions required for Theorems 2.2.2 and 2.2.3 are explicit in the INGARCH(1,1) case.

2.3.4 Generalized Poisson INGARCH model

Zhu (2012a) introduced the Generalized-Poisson (GP) INGARCH model, which can also account for both overdispersion and underdispersion. The GP distribution has two parameters $\lambda^* > 0$ and $\kappa < 1$. When $\kappa \geq 0$, a variable X following this distribution, $X \sim \mathcal{GP}(\lambda^*, \kappa)$, satisfies

$$P(X = x) = \lambda^* (\lambda^* + \kappa x)^{x-1} \frac{e^{-(\lambda^* + \kappa x)}}{x!}, \qquad x = 0, 1, \ldots.$$

The definition can be extended to some values of $\kappa < 0$ (see Zhu, 2012a). Since $EX = \lambda^*/(1-\kappa)$, the GP INGARCH model is defined by assuming that the conditional distribution of X_t is $\mathcal{GP}(\lambda_t(1-\kappa), \kappa)$ with λ_t of the form (2.2.9). By Theorem 1 and Remark 2 in Lee et al. (2015), who considered a zero-inflated GP AR model, strict and second order stationarity, as well as ergodicity, hold under (2.3.1) in the case $(p,q) = (1,1)$. Since the GP distribution belongs to the wide class of the compound Poisson distributions studied by Gonçalves et al. (2015), the latter reference shows that the general GP INGARCH(p,q) model is strictly stationary and ergodic if and only if (2.3.1) holds. By Theorem 2.1, the PQMLE of this model is thus consistent under (2.3.1) and (2.2.10).

2.3.5 COM Poisson INGARCH model

Zhu (2012b) proposed the Conway-Maxwell (COM) Poisson INGARCH model as an alternative to the GP and DP INGARCH models, and showed that the GP and COM-Poisson models often perform better than other competitors, in particular the DP INGARCH model.

A variable X of COM-Poisson distribution with parameters $\mu > 0$ and $\nu > 0$ is such that

$$P(X = x) = \left(\frac{\mu^x}{x!}\right)^\nu \frac{1}{S(\mu,\nu)}, \qquad x = 0, 1, \dots,$$

where $S(\mu,\nu) = \left(\sum_{k=0}^\infty \mu^k/k!\right)^\nu$. The COM-Poisson INGARCH model is defined by assuming that the conditional distribution of X_t given its past values is COM-Poisson with parameters μ_t and ν, where

$$\mu_t = a_0 + \sum_{i=1}^q a_i X_{t-i} + \sum_{j=1}^p b_j \mu_{t-j},$$

with positive constraints on the coefficients. Zhu (2012b) has shown that the expectation of the COM-Poisson distribution can be approximated by

$$\lambda := \sum_{k=1}^\infty \frac{k\mu^{k\nu}}{(k!)^\nu S(\mu,\nu)} \simeq \mu - \frac{\nu-1}{2\nu}.$$

In view of this result and of Theorem 2.2.1, one can thus expect that the PQMLE converges to a parameter close to

$$\left(a_0 - (1 - \sum_{j=1}^p b_j)\frac{\nu-1}{2\nu}, a_1, \dots, a_q, b_1, \dots, b_p\right).$$

2.3.6 Log-linear models

One drawback of the previous models is that their coefficients are positively constrained, which entails statistical difficulties when a coefficient is equal to zero (see Theorem 2.2.3) and makes difficult to add exogenous explanatory variables to λ_t. Another drawback is that the autocovariances $\mathrm{cov}(X_t, X_{t-h})$ are nonnegative at any lag h (see Christou & Fokianos (2014) for the explicit expression of these autocovariances for first-order models). To tackle these problems, Fokianos & Tjøstheim (2011) proposed a model in which the conditional distribution of X_t given its past values is Poisson with intensity parameter $\lambda_t = e^{\upsilon_t}$, where

$$\upsilon_t = \omega_0 + \alpha_0 \log(X_{t-1} + 1) + \beta_0 \upsilon_{t-1}. \tag{2.3.5}$$

This leads to the first-order Poisson log-linear model. Obviously, one can change the conditional distribution to define, for instance, a negative binomial log-linear model. Because the process υ_t is not constrained to be positive, the coefficients ω_0, α_0 and β_0 in (2.3.5) can be negative. Under the condition

$$|\alpha_0 + \beta_0| \vee |\alpha_0| \vee |\beta_0| < 1, \tag{2.3.6}$$

Douc et al. (2013) showed that the Poisson log-linear model with the log-intensity (2.3.5) admits a stationary and ergodic solution. Note however that (2.3.6) is not imposed *a priori* because the PQMLE does not assume that the conditional distribution is Poisson. For log-linear models, the existence of the lower bound $\underline{\omega} > 0$ in (2.2.2) and (2.2.5) is a restrictive assumption. It is satisfied when Θ is a compact set and, for instance, when the distribution of X_t is bounded or when the coefficients α and β are non negative on Θ. In Fokianos & Tjøstheim (2011), Pages 576-577, it is shown that, in the Poisson case and under a condition similar to (2.3.6), $Ee^{r_0 \upsilon_t} < \infty$ for $r_0 = 1$. Their arguments extend to show the inequality for any $r_0 > 0$. The moment conditions (2.2.3) and (2.2.21) then follow, for any $\varepsilon > 0$. With the notation $\upsilon_t(\theta) = \log \lambda_t(\theta)$ and $\widetilde{\upsilon}_t(\theta) = \log \widetilde{\lambda}_t(\theta)$, a Taylor expansion yields

$$\left|\widetilde{\lambda}_t(\theta) - \lambda_t(\theta)\right| = e^{\sum_{i=0}^{t-1} \beta^i \{\omega + \alpha \log(X_{t-i-1}+1)\}} |\beta|^t \left|\upsilon_0(\theta) - \widetilde{\upsilon}_0(\theta)\right| e^{\beta^t \upsilon^*},$$

where v^* is between $v_0(\theta)$ and $\widetilde{v}_0(\theta)$. Using $\sup_t E\log(X_t+1) < \infty$ and the additional assumption $\sup_{\theta\in\Theta}|\beta| < 1$, Lemma 7.1 in Francq et al. (2013) shows that

$$\limsup_{t\to\infty}\sup_{\theta\in\Theta}\frac{\log\left|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)\right|}{t} < 0,$$

which entails (2.2.4) (by arguments given in Remark 2.2.1). One can show (2.2.15) similarly. The identifiability conditions (2.2.8) and (2.2.13) are satisfied by assuming that the conditional distribution of X_t is not degenerated and $\alpha_0 \neq 0$. By the moment condition (2.2.3), the compactness of Θ and the condition $\sup_{\theta\in\Theta}|\beta| < 1$, the variables $v_t(\theta)$, $\partial\log\lambda_t(\theta)/\partial\theta_i = \partial v_t(\theta)/\partial\theta_i$ and $\partial^2 v_t(\theta)/\partial\theta_i\partial\theta_j$ admit moments of any order uniformly in $\theta\in\Theta$. Together with (2.2.3) and (2.2.21), this entails (2.2.12). The assumptions required for Theorem 2.2.1 and Theorem 2.2.2 have thus been discussed. Note that for the log-linear models, Theorem 2.2.3 is not relevant because there is no boundary effect when the parameter space Θ is well chosen. In particular, because the parameters are not constrained to be nonnegative, a value of θ_0 such as $\beta_0 = 0$ does not lie at the boundary of Θ.

2.3.7 INAR

One of the most popular count time series model is the integer-valued autoregressive process. Contrary to the previous models, the INAR is parameter-driven. Since it is not obvious to compute the MLE of a parameter-driven model, several alternative techniques and algorithms have been proposed in the literature (see *e.g.* Drost et al., 2009, Pavlopoulos & Karlis, 2008 and Pedeli et al., 2014). The PQMLE model seems also particularly attractive in this framework. The INAR(1) model defines X_t as the convolution of a binomial distribution $\mathcal{B}(X_{t-1},\alpha_0)$ (with the convention $\mathcal{B}(X_{t-1},\alpha_0) = 0$ when $X_{t-1} = 0$) with a distribution ϵ_t on the integers. One can interpret $\mathcal{B}(X_{t-1},\alpha_0)$ as the number of survivors from the population X_{t-1} and ϵ_t as the number of new arrivals, which is assumed to be independent of X_{t-1}. With this model we have (2.2.1) (see Grunwald et al., 2000) with $\lambda_t = \omega_0 + \alpha_0 X_{t-1}$ and $\omega_0 = E\epsilon_1$, obviously under the assumption that the expectation exists. In this case, and when the sequence (ϵ_t) is iid and $\alpha_0 < 1$, (X_t) is always stationary and $EX_1 = E\epsilon_1/(1-\alpha_0)$. If $E\epsilon_1 \neq 0$, one can choose Θ such that (2.2.5) holds true. It is easy to see that the identifiability condition (2.2.8) is satisfied when the conditional distribution of X_t is not degenerated, which is the case when $\alpha_0 \neq 0$ or $\mathrm{Var}(\epsilon_1) > 0$. Now, note that $Var(X_t \mid \mathcal{F}_{t-1}) = \alpha_0(1-\alpha_0)X_{t-1} + \mathrm{Var}(\epsilon_1)$. Therefore (2.2.11) is satisfied. The information matrices I and J in (2.2.12) exist because we have $|Var(X_t \mid \mathcal{F}_{t-1})/\lambda_t(\theta_0)| \leq c_0$, $\left\| \frac{1}{\lambda_t(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right\| \leq c_0 + c_1 X_1$ for some constants c_0 and c_1. We show (2.2.13) by the argument used to show (2.2.8). The second-order derivatives of $\lambda_t(\theta)$ being equal to zero, (2.2.14) is easily verified. Since $a_t = b_t = c_t = d_t = e_t = 0$ for $t \geq 2$, the conditions (2.2.4), (2.2.15) and (2.2.23) are trivially satisfied.

2.4 Numerical illustrations

The first part of this section examines the finite sample behaviour of the PQMLE. The second part presents a simulation study concerning the test of nullity of one coefficient and the test of constant conditional mean. All the results of this section are based on $N = 1000$ independent replications of Monte Carlo simulations of different sample sizes n. For each simulation, the first 100 observations are omitted, so that the process approaches its stationary regime.

2.4.1 Finite sample behaviour of the PQMLE

The PQML function $\widetilde{L}_n$, defined in (2.2.7), is optimized numerically, using the PORT routine (implemented by the function `nlminb()` of R).

The first Monte Carlo experiments concern the INAR(1) model. When the innovation ϵ_t follows a Poisson $\mathcal{P}(\lambda)$ distribution, then the conditional mean is $\lambda_t = \omega_0 + \alpha_0 X_{t-1}$ with $\omega_0 = \lambda$. When ϵ_t follows the geometric distribution $\mathcal{G}(p)$ with parameter $p \in (0,1)$, then $\omega_0 = (1-p)/p$. When $\epsilon_t \sim \mathcal{NB}(r,p)$ then $\omega_0 = rp/(1-p)$. We also simulated INGARCH(1,1) and log-linear(1,1) models, with Poisson, double-Poisson and negative binomial conditional distributions. Note that, in view of (2.3.4), the mean and variance of the DP INGARCH model are respectively

$$EX_t = \frac{2}{1-0.3-0.6} = 20 \qquad \text{and} \qquad \mathrm{Var}X_t = \frac{1-0.81+0.09}{1-0.81}\frac{20}{2} = 14.737.$$

The DP INGARCH model is thus underdispersed. We have checked empirically that for the DP log-linear model, the simulated series are also underdispersed. For the other models, the simulated series are overdispersed.

The means of the estimated values of θ_0 are displayed in bold in the rows "PQMLE" of Table 2.1. This table also gives four different estimators of the root-mean-square deviation $\sqrt{E\left(\widehat{\theta}_n - \theta_0\right)^2}$: the empirical standard error (ESE), the estimated standard error based on the asymptotic theory (ASE), the theoretical standard error based on the asymptotic theory (TSE), and the Poisson standard error based on the asymptotic theory assuming a Poisson conditional distribution (PSE). The ESE is equal to the root mean square error of estimation over the N replications. The ASE of the estimator of the i-th parameter is equal to the empirical mean of the estimated standard errors $\sqrt{\widehat{\Sigma}(i,i)/n}$, where $\widehat{\Sigma}$ is obtained from (2.2.16) and (2.2.17). The TSE is defined like the ASE, except that $\widehat{\Sigma}$ is replaced by Σ computed from a very large simulation ($n = 5000$). The PSE is equal to the empirical mean of $\sqrt{\widehat{J}^{-1}(i,i)/n}$ (noting that $\Sigma = J^{-1}$ when the conditional distribution is Poisson). The ESE offers the best view of the finite sample standard error of the PQMLE but, on real data series, only ASE and PSE are computable.

It is known that, for estimating a mean parameter λ, any quasi MLE based on a linear exponential family is consistent even if the true density function does not belong to the exponential family (see Gourieroux et al. 1984). Examples of such exponential families are the Poisson distributions $\{\mathcal{P}(\lambda), \lambda > 0\}$ or the negative binomial distributions $\{\mathcal{NB}(r, r/\lambda + r), \lambda > 0\}$, with r fixed. This leads to view the MLE based on the negative binomial distribution as a quasi maximum likelihood estimator. We thus call it the NB QMLE. In Table 2.1, the rows "NB-QMLE" give the means of the estimates obtained by NB QMLE, and the empirical standard errors are given in the rows below.

Table 2.1 shows that, for all the models (overdispersed or underdispersed) and for the two estimators, the means of the estimated parameters are satisfactorily close to their theoretical values, especially for large sample sizes. Moreover, for the PQMLE, the first three estimations of the standard deviations, the ESE, ASE and TSE, are very similar. The ASE and TSE are close because Σ is well estimated by (2.2.16)–(2.2.17). The closeness between ESE and ASE means that the asymptotic theory

Table 2.1 – Finite sample behaviour of the PQMLE

INAR(1)							
		$\epsilon_t \sim \mathcal{P}(2)$		$\epsilon_t \sim \mathcal{G}(0.5)$		$\epsilon_t \sim \mathcal{NB}(2, 0.5)$	
n		ω_0=2	α_0=0.9	ω_0=1	α_0=0.9	ω_0=2	α_0=0.9
500	PQMLE	**2.159**	**0.892**	**1.075**	**0.892**	**2.173**	**0.892**
	ESE	0.412	0.021	0.232	0.023	0.442	0.022
	ASE	0.404	0.020	0.225	0.022	0.425	0.020
	TSE	0.409	0.020	0.229	0.022	0.424	0.021
	PSE	0.895	0.046	0.340	0.037	0.713	0.037
	NB QMLE	**2.159**	**0.892**	**1.077**	**0.892**	**2.159**	**0.892**
	ESE	0.460	0.023	0.251	0.024	0.474	0.023
1000	PQMLE	**2.066**	**0.897**	**1.048**	**0.895**	**2.070**	**0.896**
	ESE	0.286	0.014	0.169	0.016	0.287	0.014
	ASE	0.283	0.014	0.159	0.016	0.294	0.015
	TSE	0.285	0.014	0.161	0.016	0.297	0.015
	PSE	0.620	0.032	0.236	0.025	0.497	0.026
	NB QMLE	**2.081**	**0.896**	**1.038**	**0.896**	**2.067**	**0.897**
	ESE	0.300	0.015	0.165	0.017	0.313	0.016

INGARCH(1,1)										
		$\mathcal{P}(\lambda_t)$			$\mathcal{DP}(\lambda_t, 2)$			$\mathcal{NB}(3, p_t)$		
n		ω_0= 2	α_0= 0.3	β_0= 0.6	ω_0= 2	α_0= 0.3	β_0= 0.6	ω_0= 2	α_0= 0.3	β_0= 0.6
500	PQMLE	**2.229**	**0.296**	**0.592**	**2.249**	**0.295**	**0.592**	**2.239**	**0.292**	**0.595**
	ESE	0.703	0.039	0.059	0.732	0.037	0.057	0.758	0.045	0.064
	ASE	0.653	0.038	0.057	0.660	0.038	0.057	0.643	0.047	0.064
	TSE	0.619	0.038	0.056	0.617	0.038	0.055	0.667	0.047	0.063
	PSE	0.658	0.038	0.058	0.939	0.054	0.082	0.445	0.027	0.041
	NB QMLE	**2.273**	**0.298**	**0.588**	**2.245**	**0.299**	**0.589**	**2.232**	**0.296**	**0.591**
	ESE	0.739	0.039	0.059	0.732	0.0389	0.060	0.728	0.043	0.063
1000	PQMLE	**2.134**	**0.298**	**0.595**	**2.104**	**0.298**	**0.597**	**2.168**	**0.298**	**0.596**
	ESE	0.476	0.026	0.040	0.460	0.027	0.040	0.496	0.033	0.046
	ASE	0.444	0.027	0.040	0.440	0.027	0.040	0.481	0.032	0.045
	TSE	0.438	0.027	0.039	0.430	0.027	0.039	0.468	0.033	0.044
	PSE	0.446	0.027	0.040	0.626	0.038	0.015	0.313	0.019	0.056
	NB QMLE	**2.125**	**0.300**	**0.594**	**2.116**	**0.299**	**0.592**	**2.119**	**0.298**	**0.596**
	ESE	0.471	0.026	0.039	0.452	0.028	0.040	0.482	0.030	0.043

Log-linear(1,1)										
		$\mathcal{P}(\lambda_t)$			$\mathcal{DP}(\lambda_t, 2)$			$\mathcal{NB}(3, p_t)$		
n		ω_0= 2	α_0= 0.3	β_0= -0.6	ω_0= 2	α_0= 0.3	β_0= -0.6	ω_0= 2	α_0= 0.3	β_0= -0.6
500	PQMLE	**1.946**	**0.304**	**-0.570**	**1.909**	**0.308**	**-0.588**	**1.976**	**0.301**	**-0.596**
	ESE	0.214	0.048	0.116	0.236	0.050	0.131	0.217	0.052	0.116
	ASE	0.206	0.049	0.108	0.217	0.051	0.112	0.208	0.050	0.106
	TSE	0.205	0.050	0.105	0.223	0.053	0.111	0.177	0.051	0.084
	PSE	0.208	0.049	0.105	0.309	0.072	0.159	0.140	0.072	0.159
	NB QMLE	**1.939**	**0.293**	**-0.555**	**1.978**	**0.303**	**-0.589**	**1.987**	**0.303**	**-0.596**
	ESE	0.386	0.071	0.284	0.238	0.055	0.126	0.234	0.054	0.121
1000	PQMLE	**1.975**	**0.303**	**-0.587**	**1.958**	**0.302**	**-0.591**	**1.990**	**0.300**	**-0.598**
	ESE	0.150	0.035	0.076	0.154	0.034	0.083	0.155	0.037	0.079
	ASE	0.146	0.035	0.073	0.155	0.036	0.079	0.148	0.036	0.075
	TSE	0.146	0.035	0.073	0.192	0.037	0.100	0.136	0.036	0.067
	PSE	0.146	0.035	0.074	0.219	0.051	0.111	0.099	0.024	0.050
	NB QMLE	**1.989**	**0.298**	**-0.591**	**1.996**	**0.301**	**-0.596**	**1.989**	**0.301**	**-0.595**
	ESE	0.211	0.043	0.146	0.166	0.038	0.092	0.155	0.035	0.080

provides a reliable view on the actual standard error of the PQMLE. As expected, the standard errors decrease as the sample sizes increase. It is important to note that the PSE is different from the other estimators when the conditional distribution is not Poisson. The fact that PSE may be more than twice smaller or larger than the ESE demonstrates that, for a valid inference based on the PQMLE, it is crucial to rely on the asymptotic variance $\Sigma = J^{-1}IJ^{-1}$ instead of $\Sigma = J^{-1}$. We can thus draw the conclusion that ASE is a much more robust estimator of the PQMLE standard deviation than PSE. Concerning the comparison of the performance of the two estimators, we do not observe important difference, except for the Poisson log-linear(1,1) model, where the NB QMLE seems less efficient than the PQMLE, which is not surprising because the PQMLE is actually the MLE in this case. More surprisingly, the NB QMLE does not outperform the PQMLE when the conditional distribution is negative binomial.

Figures 2.1 displays the boxplot and histogram of the $N = 1000$ values of the PQMLE (centred and reduced) for simulations of length $n = 3000$ of an INAR(1) with $\epsilon_1 \sim \mathcal{NB}(3, 0.6)$ and $\alpha_0 = 0.9$. In agreement with Theorem 2.2.2, the empirical distribution of the estimator resembles the standard Gaussian law. Other simulation experiments, not presented here for sake of conciseness, reveal similar behaviors for other models and other values of parameters, provided they are sufficiently far from zero. Indeed, in accordance with Theorem 2.2.3, the empirical distribution of the PQMLE does not resemble the Gaussian when the parameter is close to the boundary of the parameter space. Table 2.2 gives the p-values of the Kolmogorov-Smirnov test of normality for the N values of the PQMLE, computed on simulations of size $n = 3000$ of each of the models considered in Table 2.1. The normality assumption is never rejected.

Table 2.2 – p-values of the Kolmogorov-Smirnov test of normality of the PQMLE

	INAR(1)			INGARCH(1,1)			Log-linear(1,1)		
	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$	$\mathcal{P}(\lambda_t)$	$\mathcal{DP}(\lambda_t,0.5)$	$\mathcal{NB}(3,p_t)$	$\mathcal{P}(\lambda_t)$	$\mathcal{DP}(\lambda_t,0.5)$	$\mathcal{NB}(3,p_t)$
$\widehat{\omega}$	0.244	0.395	0.302	0.396	0.080	0.257	0.936	0.961	0.385
$\widehat{\alpha}$	0.318	0.449	0.768	0.542	0.707	0.841	0.756	0.487	0.584
$\widehat{\beta}$				0.848	0.384	0.851	0.658	0.890	0.969

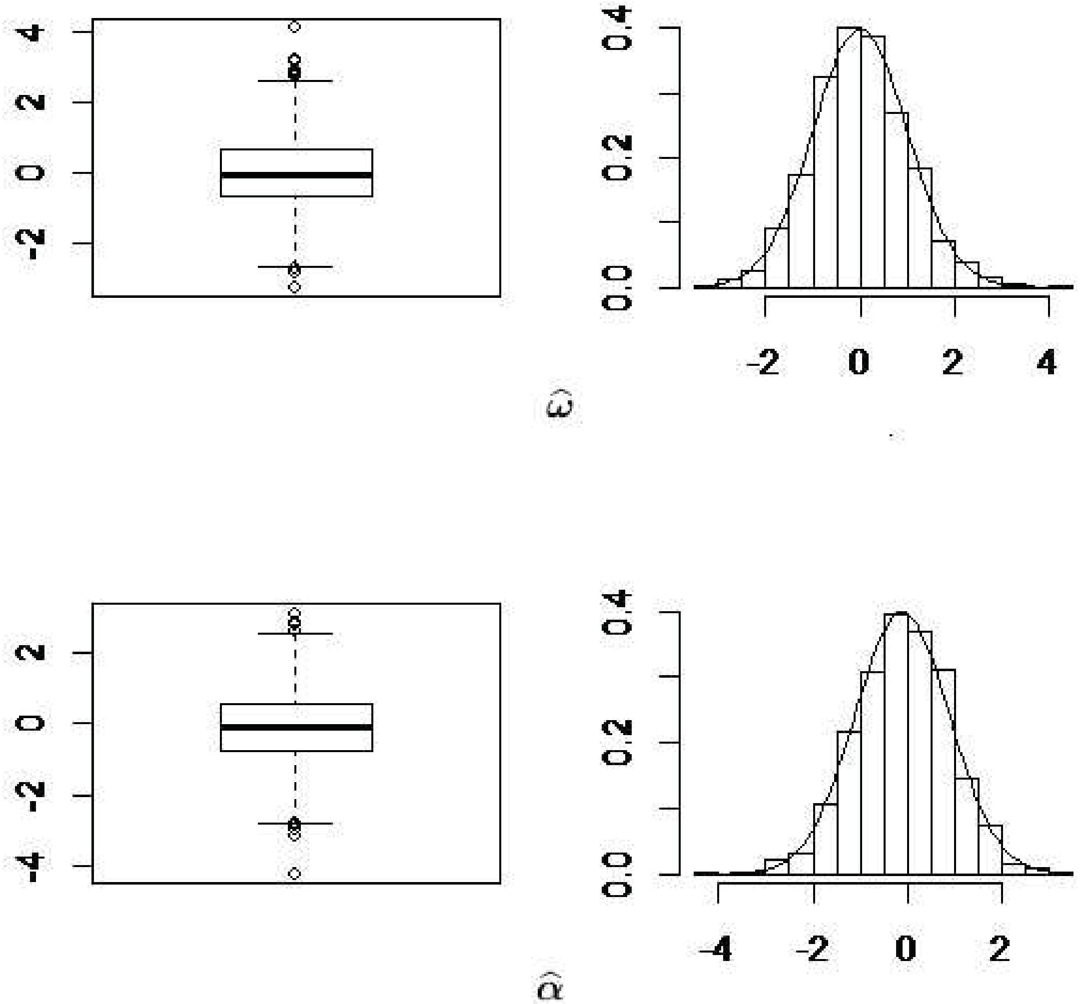

Figure 2.1 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}$, $\widehat{\alpha})$ for an INAR(1) model with negative binomial innovation. Superimposed is the standard normal density function.

2.4.2 Significance tests based on the PQMLE

We now report a Monte Carlo experiment for examining the performance of two adequacy tests for the conditional mean : the test that one coefficient is equal to zero, and the test of constant conditional mean. The simulation is implemented to obtain the sizes and the powers of the tests for different sample sizes. The tests are carried out at asymptotic level $\underline{\alpha} = 5\%$.

2.4.2.1 Empirical behavior of the tests under the null

For the test of nullity of one coefficient, we consider the INGARCH(1,2) models with (ω_0, α_{01},α_{02}, β_0) =(2, 0.2, 0, 0.4) and three different conditional distributions, a Poisson, a negative binomial and a double-Poisson. In addition, we consider the INAR(2) models with (α_{01},α_{02}) =(0.2, 0) when the marginal distribution of the innovation is Poisson, negative binomial and geometric. On each of the $N = 1000$ simulations, we fit the INGARCH(1,2) and INAR(2) models by PQMLE, and carry out the test of $H_0 : \alpha_{02} = 0$ against $H_1 : \alpha_{02} > 0$ for three nominal levels $\underline{\alpha} = \{1\%, 5\%, 10\%\}$. The null is rejected for large values of the test statistic $n\widehat{\alpha}_{n2}^2/\widehat{\Sigma}(3,3)$. As the parameters of the two models are positively constrained, according to (2.2.25), we use the critical value $\chi_1^2(1 - 2\underline{\alpha})$. The relative rejection frequencies are shown in Table 2.3. Recall that, over $N = 1000$ independent replications of a test, the relative rejection frequencies should vary respectively between $[0.4\%, 1.6\%]$ when $\underline{\alpha} = 1\%$, $[3.6\%, 6.4\%]$ when $\underline{\alpha} = 5\%$ and between $[8.1\%, 11.9\%]$ when $\underline{\alpha} = 10\%$ with probability 95%. For the sample size $n = 1000$, the empirical sizes of the tests are thus in perfect agreement with the nominal levels.

Table 2.3 – Size of the test of nullity of α_{02} (in %)

	INGARCH(1,2)			INAR(2)		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 3)$	$\mathcal{DP}(\lambda_t, 0.5)$	$\mathcal{P}(2)$	$\mathcal{NB}(2, 0.5)$	$\mathcal{G}(0.5)$
			$\underline{\alpha} = 1\%$			
100	0.5	1.8	1.5	0.6	0.7	0.5
300	1.3	1.4	1.3	0.8	1.1	0.7
1000	1.1	0.9	0.9	1.3	1.3	1.2
			$\underline{\alpha} = 5\%$			
100	3.2	5.3	5.9	4.0	4.6	3.1
300	4.7	5.9	5.7	3.7	4.7	4.9
1000	5.6	5.1	4.9	5.6	4.7	5.4
			$\underline{\alpha} = 10\%$			
100	6.7	9.6	11.6	7.8	8.3	7.4
300	9.1	9.8	11.9	8.8	8.6	8.7
1000	9.8	10.4	9.8	10.1	8.9	9.7

For the test of constant conditional mean, we simulate INARCH(3) with (ω_0, α_{01}, α_{02}, α_{03})= (2, 0, 0, 0) and INAR(3) models with (α_{01}, α_{02}, α_{03})= (0, 0, 0). We

Table 2.4 – Size of the test of constant conditional mean (in %)

	INARCH(3)			INAR(3)		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 3)$	$\mathcal{DP}(\lambda_t, 0.5)$	$\mathcal{P}(2)$	$\mathcal{NB}(2, 0.5)$	$\mathcal{G}(0.5)$
			$\underline{\alpha} = 1\%$			
100	0.3	0.6	0.7	0.4	1.2	1.2
300	0.9	0.9	0.6	0.6	0.7	0.8
1000	0.8	1.0	0.9	1.1	1.4	0.9
			$\underline{\alpha} = 5\%$			
100	3.3	3.6	3.5	3.4	5.3	5.5
300	4.3	4.4	5.5	4.4	3.8	5.9
1000	4.9	4.5	5.2	5.2	5.6	5.1
			$\underline{\alpha} = 10\%$			
100	7.7	7.9	7.5	7.8	8.9	8.7
300	8.6	8.7	9.4	8.4	8.6	10.8
1000	9.8	9.2	10.2	8.8	9.0	10.2

then carry out the test of

$$H_0 : \alpha_{01} = \alpha_{02} = \alpha_{03} = 0 \quad \text{against} \quad H_1 : \text{ at least one } \alpha_{0i} > 0 \text{ for } i = 1, 2, 3.$$

In view of Corollary 2.2.1, the null is rejected when the statistic $S_n = n(\widehat{\alpha}_{n1}^2 + \widehat{\alpha}_{n2}^2 + \widehat{\alpha}_{n3}^2)$ exceeds the $\underline{\alpha}$-quantile of the chi-bar-square distribution. When $q = 3$ and for $\underline{\alpha} = \{1\%, 5\%, 10\%\}$, the $\underline{\alpha}$-quantiles are respectively $c_{3,1\%} = 8.75$, $c_{3,5\%} = 5.43$ and $c_{3,10\%} = 4.01$. The relative frequencies of rejection are given in Table 2.4. We can note that, the observed relative rejection frequencies of all the tests are not significantly different from their theoretical levels.

2.4.2.2 Empirical behavior of the tests under the alternative

To study the power of the tests, we now simulate INGARCH(1,2) processes with $(\omega_0,\ \alpha_{01},\ \beta_0)$= (2, 0.2, 0.4) and INAR(2) processes with $\alpha_{01} = 0.2$. For both models $\alpha_{02} \in \{0.05,\ 0.1,\ 0.3\}$. We carry out the test of nullity of the coefficient α_{02}. Table 2.5 shows that the test works as expected : the power increases as the sample size increases and as the value of α_{02} increases.

For the test of constant conditional mean, we simulate INARCH(3) model with $(\omega_0,\ \alpha_{01},\ \alpha_{02})$= (2, 0, 0) and INAR(3) with $\alpha_{01} = \alpha_{02} = 0$. For the two models $\alpha_{03} \in \{0.05,\ 0.1,\ 0.4\}$. Table 2.6 shows that this test also works reasonably well.

Table 2.5 – Power of the test of nullity of α_{02} (in %)

		INGARCH(1,2)			INAR(2)		
n	α_{02}	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(3,p_t)$	$\mathcal{DP}(\lambda_t,0.5)$	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$
100	$\alpha_{02}=0.05$	10.2	9.3	10.6	10.5	8.9	10.6
	$\alpha_{02}=0.1$	18.2	16.9	19.8	21.9	20.5	22.4
	$\alpha_{02}=0.3$	64.9	59.7	69.5	84.9	85.3	84.5
300	$\alpha_{02}=0.05$	14.5	17.9	16.6	20.0	17.2	17.6
	$\alpha_{02}=0.1$	30.4	32.1	33.4	50.3	46.6	43.8
	$\alpha_{02}=0.3$	95.4	89.0	92.6	100.0	99.9	99.8
1000	$\alpha_{02}=0.05$	22.6	23.3	23.4	44.1	40.7	42.4
	$\alpha_{02}=0.1$	61.8	56.5	58.2	92.0	88.5	90.1
	$\alpha_{02}=0.3$	100.0	100.0	100.0	100.0	100.0	100.0

Table 2.6 – Power of the test of constant conditional mean (in %)

		INARCH(3)			INAR(3)		
n	θ_0	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(3,p_t)$	$\mathcal{DP}(\lambda_t,0.5)$	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$
100	$\theta_0=(2,0,0,0.05)$	5.5	7.7	11.0	6.7	9.0	8.8
	$\theta_0=(2,0,0,0.1)$	13.3	13.1	13.8	14.5	14.7	13.9
	$\theta_0=(2,0,0,0.4)$	92.6	90.1	91.1	94.5	92.7	92.5
300	$\theta_0=(2,0,0,0.05)$	12.6	13.3	14.2	14.5	11.7	11.1
	$\theta_0=(2,0,0,0.1)$	30.6	32.9	36.2	32.7	33.7	33.2
	$\theta_0=(2,0,0,0.4)$	100.0	100.0	100.0	100.0	100.0	100.0
1000	$\theta_0=(2,0,0,0.05)$	29.9	29	24.5	30.3	26.9	31.3
	$\theta_0=(2,0,0,0.1)$	83.3	81.1	76.0	83.5	82.5	81.2
	$\theta_0=(2,0,0,0.4)$	100.0	100.0	100.0	100.0	100.0	100.0

A way to visualize the power of a test is to plot the function of the relative rejection frequencies (RRF)

$$RRF(z) = \frac{1}{N}\sum_{j=1}^{N} I(p_j < z), \quad z \in [0,1],$$

where p_j denotes the observed p-value for the j-th replication of the test, and $I(p_j < z)$ is an indicator function that takes the value 1 if its argument is true and 0 otherwise. Figure 2.2 displays the RRF functions of the test of nullity of one coefficient (2.2.25) and of the test of constant conditional mean (2.2.26), for different sample sizes n. The first test is applied with the null $H_0 : \alpha_{02} = 0$ on simulations of the INGARCH(1,2) process with (ω_0, α_{01}, α_{02}, β_0)= (2, 0.3, 0.2, 0.4) and the conditional distribution $\mathcal{NB}(3, p_t)$. The second test is applied to the INARCH(3) process with the conditional distribution $\mathcal{NB}(3, p_t)$ and (ω_0, α_{01}, α_{02}, α_{03})= (2, 0, 0, 0.1). In Figure 2.2, the more concave the shape of a curve is, the better the corresponding test is in terms of power. Note that, for the first test, $RRF(z)$ does not reach 1 when $z = 1$. This is due to the fact that when the test statistic takes the value zero (which appears with non zero probability, even under the alternative) the p-value is equal to 1 (*i.e.* the probability that a chi-bar-square distribution be positive or equal to zero).

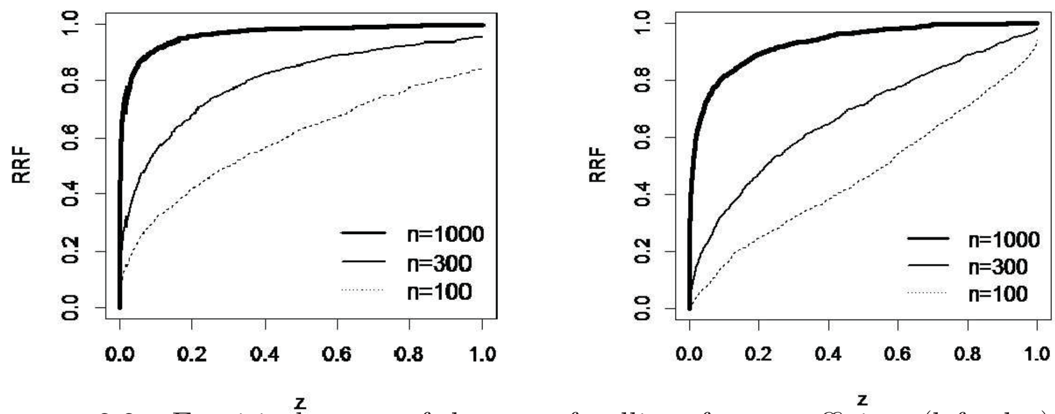

Figure 2.2 – Empirical power of the test of nullity of one coefficient (left plot) and of the test of constant conditional mean (right plot), measured by the function of the relative rejection frequencies (RRF)

2.5 Real data application

In this section, we report an application of the PQMLE to financial time series data. The data set is obtained from the QUANDL search engine and it contains the daily number of trades of 6 stocks listed in the NYSE Euronext group, namely CR.FONC.MONACO, Siraga, Technofirst, Siparex Croissance, Proximedia and Achmea (see Figure 2.3). The size of the series varies from 1006 to 3633. Table 2.7 shows that the series are overdispersed (their empirical variances are larger than their means), with the exception of the PROXIMEDIA stock which is underdispersed. For each series, we fitted INGARCH(1,3), INGARCH(1,2) and INGARCH(1,1) models. The estimated parameters are shown in Table 2.8. This table also gives, into parentheses, the p-values of the test (2.2.25) of nullity of the corresponding coefficient. The p-values that are less than 0.05 are displayed in bold. Recall that we used the function `nlminb()` of R to minimize the opposite of the log-likelihood function. The values of minimised function are provided in the rows "-Log-Like". To illustrate the table, take the example of the daily number of transactions of the SIRAGA stock. For the full INGARCH(1,3) model with 5 parameters, the parameters $\widehat{\alpha}_2$ and $\widehat{\alpha}_3$ are not statistically significant. Among the 3 models with 4 parameters, the INGARCH(1,2) model and the constrained INGARCH(1,3) models (assuming $\alpha_{02} = 0$ or $\alpha_{01} = 0$), the first two models are the same and are prefered to the third because the likelihood is larger, but they contain a non significant parameter. Among the 3 models with 3 parameters, the INGARCH(1,1) has the largest likelihood and all its parameters are significant. Therefore, a conditional mean of the form $\lambda_t = 0.201 + 0.258X_{t-1} + 0.687\lambda_{t-1}$ is retained for this series. In general, the INGARCH(1,1) representation seems the most appropriate for the majority of the series, except maybe for the TECHNOFIRST series which apparently could be better modeled by a INGARCH(1,3) with the constraint $\alpha_2 = 0$. Figure 2.13 (see Appendix) shows that the residuals of the selected models can be considered as white noises. It is interesting to note that, for all the series, the sum of the estimated values of the α and β coefficients is close to 0.9, which indicates a strong persistence in the dynamics. This is in accordance with the clusters of high values that are observed on the series plotted in Figure 2.3.

Table 2.7 – The dispersion of the data

	C.F.M	SIRAGA	TECHNOFIRST	SIPAREX	PROXIMEDIA	ACHMEA
Mean	2.226	3.589	4.132	10.019	1.736	23.788
Variance	2.963	17.578	19.562	129.730	1.586	234.854

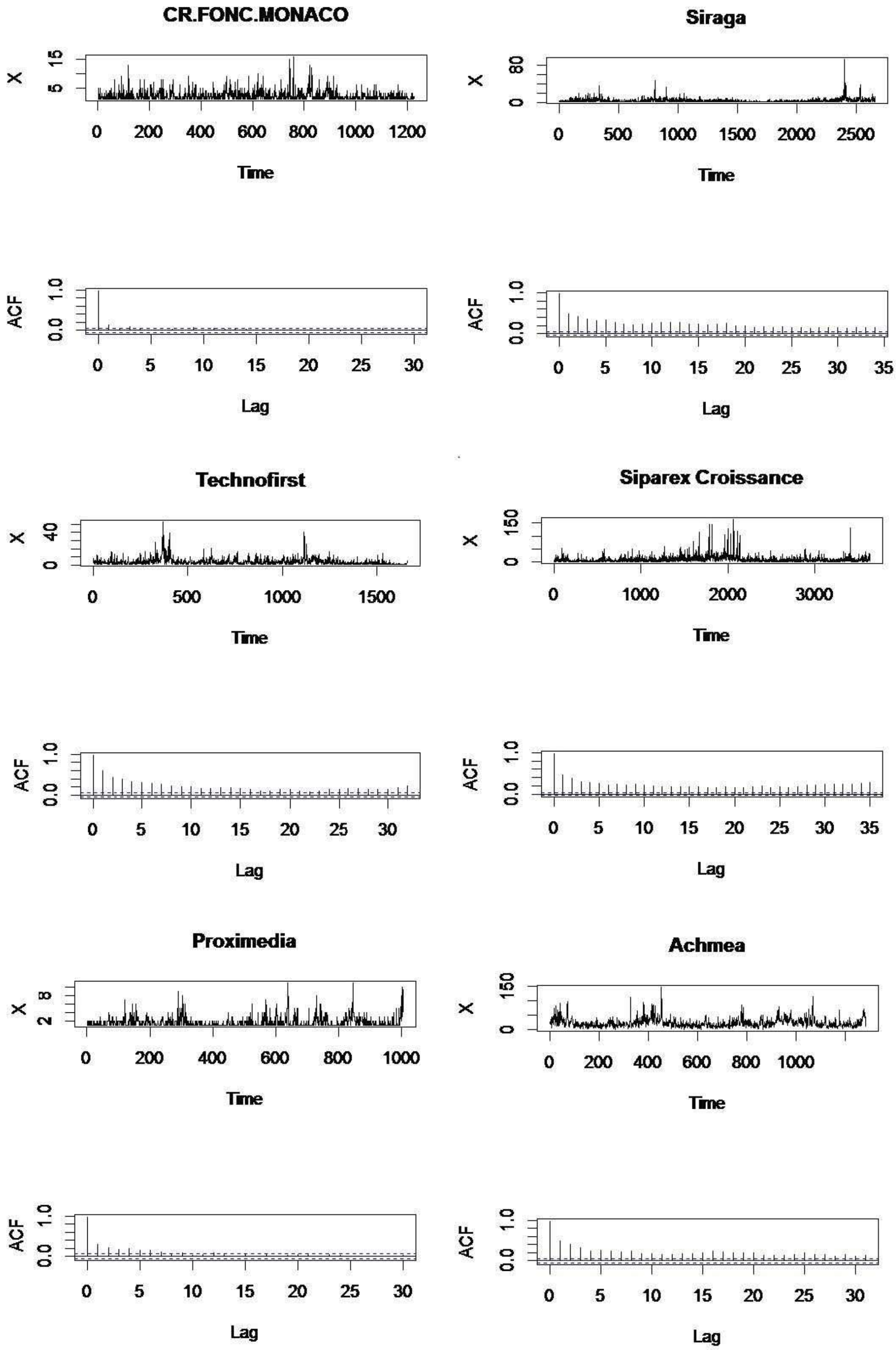

Figure 2.3 – The time series and their sample autocorrelation functions

Table 2.8 – PQMLE of INGARCH models for the expected number of transactions (and p-value of the test of nullity of the coefficient)

	C.F.M	SIRAGA	TECHNOFIRST	SIPAREX	PROXIMEDIA	ACHMEA
$\widehat{\omega}$	**1.466**	**0.201**	**0.929**	**0.802**	**0.158**	2.842
$\widehat{\alpha}_1$	**0.123 (0.000)**	**0.258 (0.000)**	**0.396 (0.000)**	**0.288 (0.000)**	**0.183 (0.000)**	**0.326 (0.000)**
$\widehat{\alpha}_2$	0.042 (0.359)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
$\widehat{\alpha}_3$	**0.101 (0.006)**	0.000 (1.000)	**0.104 (0.007)**	0.000 (1.000)	0.009 (0.857)	0.000 (1.000)
$\widehat{\beta}$	0.076 (0.767)	**0.687 (0.000)**	**0.275 (0.046)**	**0.632 (0.000)**	**0.718 (0.000)**	**0.554 (0.000)**
-Log-Lik	0.421	-1.559	-2.307	-14.582	0.717	-52.960
$\widehat{\omega}$	**0.292**	**0.201**	**0.567**	**0.802**	**0.148**	**2.842**
$\widehat{\alpha}_1$	**0.090 (0.008)**	**0.258 (0.000)**	**0.343 (0.000)**	**0.288 (0.000)**	**0.184(0.000)**	**0.326 (0.000)**
$\widehat{\alpha}_2$	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
$\widehat{\beta}$	**0.779 (0.000)**	**0.687 (0.000)**	**0.519 (0.000)**	**0.632 (0.000)**	**0.733 (0.000)**	**0.554 (0.000)**
-Log-Lik	0.422	-1.559	-2.303	-14.582	0.717	-52.960
$\widehat{\omega}$	**1.213**	**0.201**	**0.929**	**0.802**	**0.158**	**2.842**
$\widehat{\alpha}_1$	**0.129 (0.000)**	**0.258 (0.000)**	**0.396 (0.000)**	**0.288 (0.000)**	**0.183 (0.000)**	**0.326 (0.000)**
$\widehat{\alpha}_3$	**0.093 (0.008)**	0.000 (1.000)	**0.104 (0.002)**	0.000 (1.000)	0.009 (0.849)	0.000 (1.000)
$\widehat{\beta}$	0.233 (0.176)	**0.687 (0.000)**	**0.275 (0.000)**	**0.632 (0.000)**	**0.718 (0.000)**	**0.554 (0.000)**
-Log-Lik	0421	-1.559	-2.307	-14.582	0.717	-52.960
$\widehat{\omega}$	**1.519**	**0.139**	**0.228**	**0.481**	**0.156**	**2.331**
$\widehat{\alpha}_2$	**0.067 (0.040)**	**0.178 (0.000)**	**0.165 (0.000)**	**0.179 (0.000)**	**0.152 (0.000)**	**0.224 (0.000)**
$\widehat{\alpha}_3$	**0.111 (0.006)**	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)	0.016 (0.781)	0.000 (1.000)
$\widehat{\beta}$	0.140 (0.524)	**0.783 (0.000)**	**0.779 (0.000)**	**0.773 (0.000)**	**0.738 (0.000)**	**0.678 (0.000)**
-Log-Lik	0.429	-1.490	-2.166	-14.332	0.729	-52.627
$\widehat{\omega}$	**0.290**	**0.201**	**0.568**	**0.802**	**0.148**	**2.842**
$\widehat{\alpha}_1$	**0.090 (0.000)**	**0.258 (0.000)**	**0.343 (0.000)**	**0.288 (0.000)**	**0.184 (0.000)**	**0.326 (0.000)**
$\widehat{\beta}$	**0.779 (0.000)**	**0.687 (0.000)**	**0.519 (0.000)**	**0.632 (0.000)**	**0.733 (0.000)**	**0.554 (0.000)**
-Log-Lik	0.422	-1.559	-2.303	-14.582	0.717	-52.960
$\widehat{\omega}$	**0.164**	**0.139**	**0.228**	**0.481**	**0.155**	**2.330**
$\widehat{\alpha}_2$	**0.052 (0.004)**	**0.178 (0.000)**	**0.165 (0.000)**	**0.179 (0.000)**	**0.161 (0.000)**	**0.224 (0.000)**
$\widehat{\beta}$	**0.875 (0.000)**	**0.783 (0.000)**	**0.779 (0.000)**	**0.773 (0.000)**	**0.752 (0.000)**	**0.678 (0.000)**
-Log-Lik	0.429	-1.498	-2.166	-14.333	0.729	-52.627
$\widehat{\omega}$	**1.555**	**0.084**	**0.198**	**0.131**	**0.183**	**0.624**
$\widehat{\alpha}_3$	**0.131 (0.000)**	**0.117 (0.000)**	**0.138 (0.000)**	**0.072 (0.000)**	**0.157 (0.000)**	**0.088 (0.000)**
$\widehat{\beta}$	0.171 (0.396)	**0.859 (0.000)**	**0.814 (0.000)**	**0.915 (0.000)**	**0.740 (0.000)**	**0.886 (0.000)**
-Log-Lik	0.432	-1.442	-2.118	-14.197	0.737	-52.453

2.6 Proofs

Proof of Theorem 2.2.1. Let $\ell_t(\theta)$ and $L_n(\theta)$ be the random variables obtained by replacing $\widetilde{\lambda}_t(\theta)$ by $\lambda_t(\theta)$ in $\widetilde{\ell}_t(\theta)$ and $\widetilde{L}_n(\theta)$, respectively. Using (2.2.2), (2.2.5) and the inequality $\log(1+x) \leq x$, we have

$$\left|\log \widetilde{\lambda}_t(\theta) - \log \lambda_t(\theta)\right| = \left|\log\left(1 + \frac{\widetilde{\lambda}_t(\theta) - \lambda_t(\theta)}{\lambda_t(\theta)}\right)\right| \leq \frac{a_t}{\underline{\omega}}.$$

By (2.2.4), as $n \to \infty$

$$\sup_{\theta\in\Theta}\left|\widetilde{L}_n(\theta) - L_n(\theta)\right| \leq \frac{1}{n}\sum_{t=s+1}^{n}\left(a_t + X_t\frac{a_t}{\underline{\omega}}\right) \to 0 \qquad \text{a.s.} \tag{2.6.1}$$

Now note that, using (2.2.1), and again $\log(x) \leq x - 1$ for $x > 0$,

$$\begin{aligned} E\left\{\ell_1(\theta) - \ell_1(\theta_0)\right\} &= E\left\{\lambda_1(\theta_0)\log\frac{\lambda_1(\theta)}{\lambda_1(\theta_0)} - \lambda_1(\theta) + \lambda_1(\theta_0)\right\} \\ &\leq E\left\{\lambda_1(\theta_0)\left(\frac{\lambda_1(\theta)}{\lambda_1(\theta_0)} - 1\right) - \lambda_1(\theta) + \lambda_1(\theta_0)\right\} = 0 \end{aligned}$$

with equality iff $\theta = \theta_0$ by (2.2.8).

From (2.2.2) and (2.2.3), it can be seen that $|\log \lambda_1(\theta_0)|$ admits moments of any order. Hölder's inequality and (2.2.3) then entail that

$$E|X_1 \log \lambda_0(\theta_0)| \leq \|X_1\|_{1+\varepsilon}\|\log \lambda_1(\theta_0)\|_{1+1/\varepsilon} < \infty.$$

We thus have $E|\ell_1(\theta_0)| < \infty$. Therefore $E\left\{\ell_1(\theta) - \ell_1(\theta_0)\right\}$ belongs *a priori* to $[-\infty, 0]$, and one can deduce

$$E\ell_1(\theta) < E\ell_1(\theta_0), \quad \forall\theta \neq \theta_0. \tag{2.6.2}$$

For $k \in \mathbb{N}^*$ and $\theta_1 \in \Theta$, let $V_k(\theta_1)$ be the open ball of center θ_1 and radius $1/k$. Note that $\left\{\sup_{\theta\in V_k(\theta_1)\cap\Theta}\ell_t(\theta)\right\}_t$ is an ergodic stationary sequence, as a measurable function of the ergodic stationary process (X_t). Note also that $E\sup_{\theta\in V_k(\theta_1)\cap\Theta}\ell_t(\theta)$ belongs to $\mathbb{R}\cup\{-\infty\}$. In view of (2.6.1) and the ergodic theorem (see Billingsley, 2008, pp. 284 and 495) we thus obtain

$$\limsup_{n\to\infty}\sup_{\theta\in V_k(\theta_1)\cap\Theta}\widetilde{L}_n(\theta) = \limsup_{n\to\infty}\sup_{\theta\in V_k(\theta_1)\cap\Theta}L_n(\theta) \leq E\sup_{\theta\in V_k(\theta_1)\cap\Theta}\ell_1(\theta).$$

By Beppo Levi's theorem, $E\sup_{\theta\in V_k(\theta_1)\cap\Theta}\ell_1(\theta)$ decreases to $E\ell_1(\theta_1)$ as $k\to\infty$. In view of (2.6.2), we have shown that for all $\theta_1\neq\theta_0$ there exists a neighborhood $V(\theta_1)$ of θ_1 such that

$$\limsup_{n\to\infty}\sup_{\theta\in V(\theta_1)\cap\Theta}\widetilde{L}_n(\theta)<\limsup_{n\to\infty}\widetilde{L}_n(\theta_0)=E\ell_1(\theta_0). \tag{2.6.3}$$

The conclusion follows form a standard argument, using the compactness of Θ. $\square$

Proof of Theorem 2.2.2. First consider the impact of the initial values. We have

$$\begin{aligned}\sqrt{n}\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}\widetilde{L}_n(\theta)-\frac{\partial}{\partial\theta}L_n(\theta)\right\| &\le \frac{1}{\sqrt{n}}\sum_{t=s+1}^{n}\left\{b_t+X_t\left(\frac{a_t}{\underline{\omega}}d_t+\frac{b_t}{\underline{\omega}}\right)\right\}\\ &= o(1) \qquad (2.6.4)\end{aligned}$$

almost surely, by (2.2.15). For n large enough, $\widehat{\theta}_n$ does not lie at the boundary of Θ and thus we have

$$0=\sqrt{n}\frac{\partial}{\partial\theta}\widetilde{L}_n(\widehat{\theta}_n)\overset{o(1)}{=}\sqrt{n}\frac{\partial}{\partial\theta}L_n(\widehat{\theta}_n)=\sqrt{n}\frac{\partial}{\partial\theta}L_n(\theta_0)-J_n^*\sqrt{n}(\widehat{\theta}_n-\theta_0), \tag{2.6.5}$$

where $a\overset{c}{=}b$ stands for $a=b+c$, and J_n^* is a matrix whose generic term is of the form $-\partial^2L_n(\theta^*_{ij})/\partial\theta_i\partial\theta_j$, for some θ^*_{ij} between $\widehat{\theta}_n$ and θ_0. Note that

$$\sqrt{n}\frac{\partial}{\partial\theta}L_n(\theta_0)=\frac{1}{\sqrt{n}}\sum_{t=s+1}^{n}U_t,\quad U_t=\left(\frac{X_t}{\lambda_t(\theta_0)}-1\right)\frac{\partial\lambda_t(\theta_0)}{\partial\theta}, \tag{2.6.6}$$

where $\{U_t,\mathcal{F}_t\}$ is a stationary martingale difference, $\mathcal{F}_t$ denoting the σ-field generated by $\{X_u,u\le t\}$. In view of (2.2.11) and (2.2.12) we have $EU_tU_t'=I$. The central limit theorem of Billingsley (1961) for square-integrable stationary martingale difference then entails that

$$\sqrt{n}\frac{\partial}{\partial\theta}L_n(\theta_0)\xrightarrow{d}\mathcal{N}(0,I)\qquad\text{as } n\to\infty. \tag{2.6.7}$$

Let $V_m(\theta_0)$ be the ball of center θ_0 and radius $1/m$. Assume that m is large enough so that $V_m(\theta_0)$ is included in the neighborhood $V(\theta_0)$ defined in (2.2.14). Suppose that n is sufficiently large, so that $\theta^*_{ij}\in V_m(\theta_0)$. With probability one,

$$\begin{aligned}|J_n^*(i,j)-J(i,j)| &\le \frac{1}{n}\sum_{t=s+1}^{n}\sup_{\theta\in V_m(\theta_0)}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta)-E\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta_0)\right|\\ &\to E\sup_{\theta\in V_m(\theta_0)}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta)-E\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta_0)\right| \qquad (2.6.8)\end{aligned}$$

as $n \to \infty$. Under (2.2.14), the Lebesgue dominated convergence theorem entails that

$$\lim_{m\to\infty} E \sup_{\theta\in V_m(\theta_0)} \left| \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta) - E\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta_0) \right| = 0. \tag{2.6.9}$$

It follows that

$$J_n^* \to J \quad a.s.$$

The conclusion follows from (2.6.5), (2.6.7) and (2.2.13). □

Proof of Theorem 2.2.3. For all $\theta \in \Theta$, a second order Taylor expansion of $\widetilde{L}_n(\theta)$ at θ_0 yields

$$\widetilde{L}_n(\theta) - \widetilde{L}_n(\theta_0) = \frac{\partial L_n(\theta_0)}{\partial\theta'}(\theta-\theta_0) - \frac{1}{2}(\theta-\theta_0)'J(\theta-\theta_0) + R_n(\theta),$$

where

$$R_n(\theta) = \left\{ \frac{\partial\widetilde{L}_n(\theta_0)}{\partial\theta'} - \frac{\partial L_n(\theta_0)}{\partial\theta'} \right\}(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)'\left(\frac{\partial^2\widetilde{L}_n(\theta^*)}{\partial\theta\partial\theta'} + J\right)(\theta-\theta_0),$$

and θ^* is between θ and θ_0. Note that $\partial L_n(\theta_0)/\partial\theta'$ has to be understood as a vector of right derivatives. Even if θ_0 contains null components, this vector of right-derivatives is well defined, and is equal to $n^{-1}\sum_{t=s+1}^{n} U_t$, as in (2.6.6). Introducing the vector

$$Z_n = J^{-1}\sqrt{n}\frac{\partial L_n(\theta_0)}{\partial\theta},$$

we can write

$$\widetilde{L}_n(\theta) - \widetilde{L}_n(\theta_0) = \frac{1}{2n}\|Z_n\|_J^2 - \frac{1}{2n}\left\|Z_n - \sqrt{n}(\theta-\theta_0)\right\|_J^2 + R_n(\theta).$$

Let the projection of Z_n on $\mathcal{C}$

$$Z_n^{\mathcal{C}} = \arg\inf_{C\in\mathcal{C}} \|C - Z_n\|_J.$$

Define also

$$\theta_{Z_n} = \arg\inf_{\theta\in\Theta} \left\|\sqrt{n}(\theta-\theta_0) - Z_n\right\|_J.$$

In view of (2.2.22), we have

$$\sqrt{n}(\theta_{Z_n} - \theta_0) = Z_n^{\mathcal{C}} \text{ for } n \text{ large enough.} \tag{2.6.10}$$

By definition of θ_{Z_n} and $\widehat{\theta}_n$, and Lemma 2.6.1 below, we have

$$\begin{aligned} 0 &\leq \|\sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n\|_J^2 - \|\sqrt{n}(\theta_{Z_n} - \theta_0) - Z_n\|_J^2 \\ &= 2n\left\{\widetilde{L}_n(\theta_{Z_n}) - \widetilde{L}_n(\widehat{\theta}_n)\right\} + 2n\left\{R_n(\widehat{\theta}_n) - R_n(\theta_{Z_n})\right\} \\ &\leq 2n\left\{R_n(\widehat{\theta}_n) - R_n(\theta_{Z_n})\right\} = o_P(1). \end{aligned}$$

By (2.6.10) it follows that

$$\|\sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n\|_J^2 - \|Z_n^{\mathcal{C}} - Z_n\|_J^2 = o_P(1).$$

In view of (2.2.24) we have

$$\begin{aligned} \|\sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n\|_J^2 &= \|\sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n^{\mathcal{C}}\|_J^2 + \|Z_n^{\mathcal{C}} - Z_n\|_J^2 \\ &\quad + 2\left\langle \sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n^{\mathcal{C}}, Z_n^{\mathcal{C}} - Z_n \right\rangle_J \\ &\geq \|\sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n^{\mathcal{C}}\|_J^2 + \|Z_n^{\mathcal{C}} - Z_n\|_J^2. \end{aligned}$$

We thus obtain

$$\|\sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n^{\mathcal{C}}\|_J^2 \leq \|\sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n\|_J^2 - \|Z_n^{\mathcal{C}} - Z_n\|_J^2 = o_P(1).$$

Noting that, by central limit theorem of Billinsgley (1961),

$$Z_n \xrightarrow{d} Z \sim \mathcal{N}\left(0, J^{-1}IJ^{-1}\right) \qquad \text{as } n \to \infty, \tag{2.6.11}$$

we have $Z_n^{\mathcal{C}} \xrightarrow{d} Z^{\mathcal{C}}$ and the conclusion follows. □

Lemma 2.6.1. *Under the assumptions of Theorem 2.2.3, $R_n(\theta_{Z_n}) = o_P(n^{-1})$ and $R_n(\widehat{\theta}_n) = o_P(n^{-1})$ as $n \to \infty$.*

Proof. First consider the impact of the initial values on the second-order derivatives of the objective function. We have

$$\begin{aligned} &\sup_{\theta\in\Theta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}\widetilde{L}_n(\theta) - \frac{\partial^2}{\partial\theta\partial\theta'}L_n(\theta)\right\| \\ &\leq \frac{1}{n}\sum_{t=s+1}^{n}\left\{c_t + X_t\left(\frac{a_t}{\underline{\omega}}e_t + \frac{c_t}{\underline{\omega}} + \frac{a_t}{\underline{\omega}}d_t^2 + \frac{b_t}{\underline{\omega}}d_t\right)\right\} = o(1) \end{aligned} \tag{2.6.12}$$

almost surely, by (2.2.23). In view of (2.6.4), (2.6.12) and (2.6.8)-(2.6.9), as $n \to \infty$ we have

$$nR_n(\theta_n) = o_P\left\{\sqrt{n}(\theta_n - \theta_0)\right\} + o_P\left\{n\|\theta_n - \theta_0\|^2\right\} \tag{2.6.13}$$

when $\theta_n - \theta_0 = o_P(1)$. Therefore

$$nR_n(\theta_n) = o_P(1) \quad \text{when} \quad \sqrt{n}(\theta_n - \theta_0) = O_P(1). \tag{2.6.14}$$

By definition of θ_{Z_n}, and since $\theta_0 \in \Theta$, we also have

$$\left\|\sqrt{n}(\theta_{Z_n} - \theta_0) - Z_n\right\|_J \leq \|Z_n\|_J .$$

The Minskowski inequality then entails that

$$\left\|\sqrt{n}(\theta_{Z_n} - \theta_0)\right\|_J \leq \left\|\sqrt{n}(\theta_{Z_n} - \theta_0) - Z_n\right\|_J + \|Z_n\|_J \leq 2\|Z_n\|_J .$$

By (2.6.11), we have $\|Z_n\|_J = O_P(1)$, and thus $\sqrt{n}(\theta_{Z_n} - \theta_0) = O_P(1)$. In view of (2.6.14), this entails $nR_n(\theta_{Z_n}) = o_P(1)$.

It remains to show the second convergence. By definition of $\widehat{\theta}_n$, we have

$$0 \leq 2n\widetilde{L}_n(\widehat{\theta}_n) - 2n\widetilde{L}_n(\theta_0) = \|Z_n\|_J^2 - \left\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta_0)\right\|_J^2 + 2nR_n(\widehat{\theta}_n).$$

It follows that

$$\begin{aligned}
\left\|\sqrt{n}(\widehat{\theta}_n - \theta_0)\right\|_J^2 &\leq 2\left(\left\|\sqrt{n}(\widehat{\theta}_n - \theta_0) - Z_n\right\|_J^2 + \|Z_n\|_J^2\right) \\
&\leq 4\|Z_n\|_J^2 + 4nR_n(\widehat{\theta}_n).
\end{aligned}$$

The consistency of $\widehat{\theta}_n$ and (2.6.13) entail that $nR_n(\widehat{\theta}_n) = o_P\left(\left\|\sqrt{n}(\widehat{\theta}_n - \theta_0)\right\|_J^2\right)$. It follows that $\sqrt{n}(\widehat{\theta}_n - \theta_0) = O_P(1)$, and the conclusion comes from (2.6.14). □

Proof of Corollary 2.2.1. Note that, because $\lambda_t(\theta_0)$ is fixed, we have $\mu_2 = E(X_t^2 \mid \mathcal{F}_{t-1})$. For notational convenience, write the information matrices in the case $q = 3$. We have

$$J = \frac{1}{\omega_0}\begin{pmatrix} 1 & \omega_0 & \omega_0 & \omega_0 \\ \omega_0 & \mu_2 & \omega_0^2 & \omega_0^2 \\ \omega_0 & \omega_0^2 & \mu_2 & \omega_0^2 \\ \omega_0 & \omega_0^2 & \omega_0^2 & \mu_2 \end{pmatrix} \qquad \text{and} \qquad I = \frac{\mu_2 - \omega_0^2}{\omega_0} J.$$

We thus obtain

$$\Sigma = \begin{pmatrix} (q-1)\omega_0^2 + \mu_2 & -\omega_0 & \cdots & -\omega_0 \\ -\omega_0 & & & \\ \vdots & & I_q & \\ -\omega_0 & & & \end{pmatrix}.$$

By the arguments given in Section 8.3.4 of FZ, the conclusion then follows from Theorem 2.2.3 and Remarks 2.2.3 and 2.2.4. □

Sketch of proof of (2.3.3)**.** The detailed proof, which is long and tedious, is available from the authors. We only give here the main computations showing that the condition (2.3.3) is necessary for the existence of EX_t^4. Let $\epsilon_t = X_t - \lambda_t$. Note that for all $r_0 \geq 2$, $EX_t^{r_0} < \infty$ iff $E|\epsilon_t|^{r_0} < \infty$. We have $E(\epsilon_t^2 \mid \mathcal{F}_{t-1}) = \lambda_t + \lambda_t^2/r$. Therefore $EX_t^2 < \infty$ iff $E\lambda_t^2 < \infty$. Writing $\lambda_t = \omega_0 + \alpha_0\epsilon_{t-1} + (\alpha_0 + \beta_0)\lambda_{t-1}$, and assuming that λ_t is stationary with second-order moments, we obtain

$$\begin{aligned} E\lambda_t^2 &= \omega_0^2 + \alpha_0^2 E\epsilon_{t-1}^2 + (\alpha_0+\beta_0)^2 E\lambda_{t-1}^2 + 2\omega_0(\alpha_0+\beta_0)E\lambda_{t-1} \\ &= \left\{ \frac{\alpha_0^2}{r} + (\alpha_0+\beta_0)^2 \right\} E\lambda_t^2 + K, \end{aligned}$$

where, here and in the sequel, K denotes a generic positive constant whose value is unimportant. Therefore $EX_t^2 < \infty$ entails (2.3.2), which was already known from Christou & Fokianos (2014). Now, from the expression of the centred third and fourth order moments of the negative binomial distribution, we have

$$E\left(\epsilon_t^3 \mid \mathcal{F}_{t-1}\right) = 2\frac{\lambda_t^3}{r^2} + R_t^{(2)}, \qquad E\left(\epsilon_t^4 \mid \mathcal{F}_{t-1}\right) = (6+3r)\frac{\lambda_t^4}{r^3} + R_t^{(3)},$$

where, for $i = 2, 3$, $R_t^{(i)}$ is a polynomial in λ_t of degree i with positive coefficients. Therefore $EX_t^4 < \infty$ iff $E\lambda_t^4 < \infty$ and, after some tedious computations,

$$E\lambda_t^4 = \alpha_0^4 E\epsilon_{t-1}^4 + 8\alpha_0^3(\alpha_0+\beta_0)\frac{E\lambda_{t-1}^4}{r^2} + 6\alpha_0^2(\alpha_0+\beta_0)^2\frac{E\lambda_{t-1}^4}{r} + (\alpha_0+\beta_0)^4 E\lambda_{t-1}^4 + K,$$

and the conclusion follows. □

2.7 Conclusion

The PQMLE provides a general approach for estimating the conditional mean parameters of time series of counts. If the conditional mean is correctly specified, under some regularity conditions, the PQMLE is CAN, even if the conditional distribution is not Poisson. For the asymptotic variance, it is however important to employ the robust expression $\Sigma = J^{-1}IJ^{-1}$ instead of the expression $\Sigma = J^{-1}$ which may be invalid when the conditional distribution is not Poisson. When the parameter belongs to the boundary of the parameter space, the asymptotic distribution of the PQMLE is no more Gaussian. This is a usual framework which appears, for instance, when we have an over-identified INGARCH(p, q) model (*i.e.* p or q is larger than necessary). When assessing the significance of the estimated parameters, we thus have to take into account the fact that the PQMLE has a special non Gaussian asymptotic distribution under the null. This leads to adequacy tests with chi-bar-square distributions instead of usual chi-square distributions. Note that these particular distributions of the estimator and its related tests also hold for the MLE (*i.e.* when the conditional distribution is Poisson).

In view of Fokianos (2012), the INGARCH model admits a weak ARMA representation. The result still holds true when the conditional distribution is not Poisson. In principle, we could thus use general diagnostic checking tools of weak ARMA models (as in Francq et al. 2005) for identifying the orders p and q of a conditional mean of the INGARCH(p, q) form (2.2.9). The problem deserves however more thought, and is left for future work. Other possible extensions of the present work include models for time series valued in $\mathbb{Z}$ (see *e.g.* Kachour & Truquet, 2011 and Andersson & Karlis, 2014) which appear naturally, in particular when a count time series is differenced, and for which a QMLE could be searched.

References

Andersson, J. & Karlis, D. (2014). A parametric time series model with covariates for integers in Z. *Statistical Modelling*, 14, 135–156.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67, 1341–1383.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12, 788–792.

Billingsley, P. (2008). *Probability and measure.* New York : John Wiley & Sons.

Christou, V. (2014). *Statistical theory for mixed poisson time series models.* PhD thesis, University of Cyprus.

Christou, V. & Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35, 55–78.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., & Lauritzen, S. L. (1981). Statistical analysis of time series : some recent developments. *Scandinavian Journal of Statistics*, 8, 93–115.

Davis, R. A. & Liu, H. (2016). Theory and inference for a class of observation-driven models with application to time series of counts. *Statistica Sinica, 26, 1673-1707.*

Douc, R., Doukhan, P., & Moulines, E. (2013). Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Processes and their Applications*, 123, 2620–2647.

Drost, F. C., Akker, R. v. d., & Werker, B. J. (2009). Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 71, 467–485.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81, 709–721.

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27, 923–942.

Fokianos, K. (2012). Count time series models. *Handbook in statistics. Time series analysis : Methods and Applications*, 30, 315–348.

Fokianos, K., Rahbek, A., & Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104, 1430–1439.

Fokianos, K. & Tjøstheim, D. (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102, 563–578.

Fokianos, K. & Tjøstheim, D. (2012). Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64, 1205–1225.

Francq, C., Roy, R., & Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association*, 100, 532–544.

Francq, C., Wintenberger, O., & Zakoïan, J.-M. (2013). GARCH models without positivity constraints : Exponential or log garch ?. *Journal of Econometrics*, 177, 34–46.

Francq, C. & Zakoïan, J.-M. (2009a). Bartlett's formula for a general class of nonlinear processes. *Journal of Time Series Analysis*, 30, 449–465.

Francq, C. & Zakoïan, J.-M. (2009b). Testing the nullity of GARCH coefficients : correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association*, 104, 313–324.

Francq, C. & Zakoian, J.-M. (2010). *GARCH models : structure, statistical inference and financial applications*. Chichester : John Wiley & Sons.

Gonçalves, E., Mendes-Lopes, N., & Silva, F. (2015). Infinitely divisible distributions in integer-valued Garch models. *Journal of Time Series Analysis*, 36, 503–527.

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods : Theory. *Econometrica*, 52, 681–700.

Grunwald, G. K., Hyndman, R. J., Tedesco, L., & Tweedie, R. L. (2000). Theory & methods : Non-gaussian conditional linear AR(1) models. *Australian & New Zealand Journal of Statistics*, 42, 479–495.

Heinen, A. (2003). Modelling time series count data : an autoregressive conditional poisson model. *Discussion Paper, vol. 2003/62. Center for operations research and econometrics (CORE), Catholic University of Louvain, Belgium.*

Kachour, M. & Truquet, L. (2011). A p-order signed integer-valued autoregressive (SINAR(p)) model. *Journal of Time Series Analysis*, 32, 223–236.

Kedem, B. & Fokianos, K. (2002). *Regression models for time series analysis*. Hoboken, New Jersey : John Wiley & Sons.

Lee, S., Lee, Y., & Chen, C. W. (2015). Parameter change test for zero-inflated generalized poisson autoregressive models. *Statistics : A Journal of Theoretical and Applied Statistics*, 100, 540–557.

Liu, H. (2012). *Some models for time series of counts*. PhD thesis, University of Colombia.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, 11, 59–67.

Neumann, M. H. (2011). Absolute regularity and ergodicity of poisson count processes. *Bernoulli*, 17, 1268–1284.

Pavlopoulos, H. & Karlis, D. (2008). INAR(1) modeling of overdispersed count series with an environmental application. *Environmetrics*, 19, 369–393.

Pedeli, X., Davison, A. C., & Fokianos, K. (2014). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association*, (published online DOI :10.1080/01621459.2014.983230).

Tjøstheim, D. (2012). Some recent theory for autoregressive count time series. *Test*, 21, 413–438.

Tjøstheim, D. (2016). *Count time series with observation driven autoregressive parameter dynamics.* In Davis, R., Holan, S., Lund, R., Ravishanker, N. (eds). Handbook of Discrete-Valued Time Series, pp.77-100. London : Chapman & Hall.

Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, 439–447.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, 75, 621–629.

Zeger, S. L. & Qaqish, B. (1988). Markov regression models for time series : a quasi-likelihood approach. *Biometrics*, 44, 1019–1031.

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, 32, 54–67.

Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications*, 389, 58–71.

Zhu, F. (2012b). Modeling time series of counts with COM-poisson INGARCH models. *Mathematical and Computer Modelling*, 56, 191–203.

Zhu, F. (2012c). Zero-inflated poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference*, 142, 826–839.

Appendix

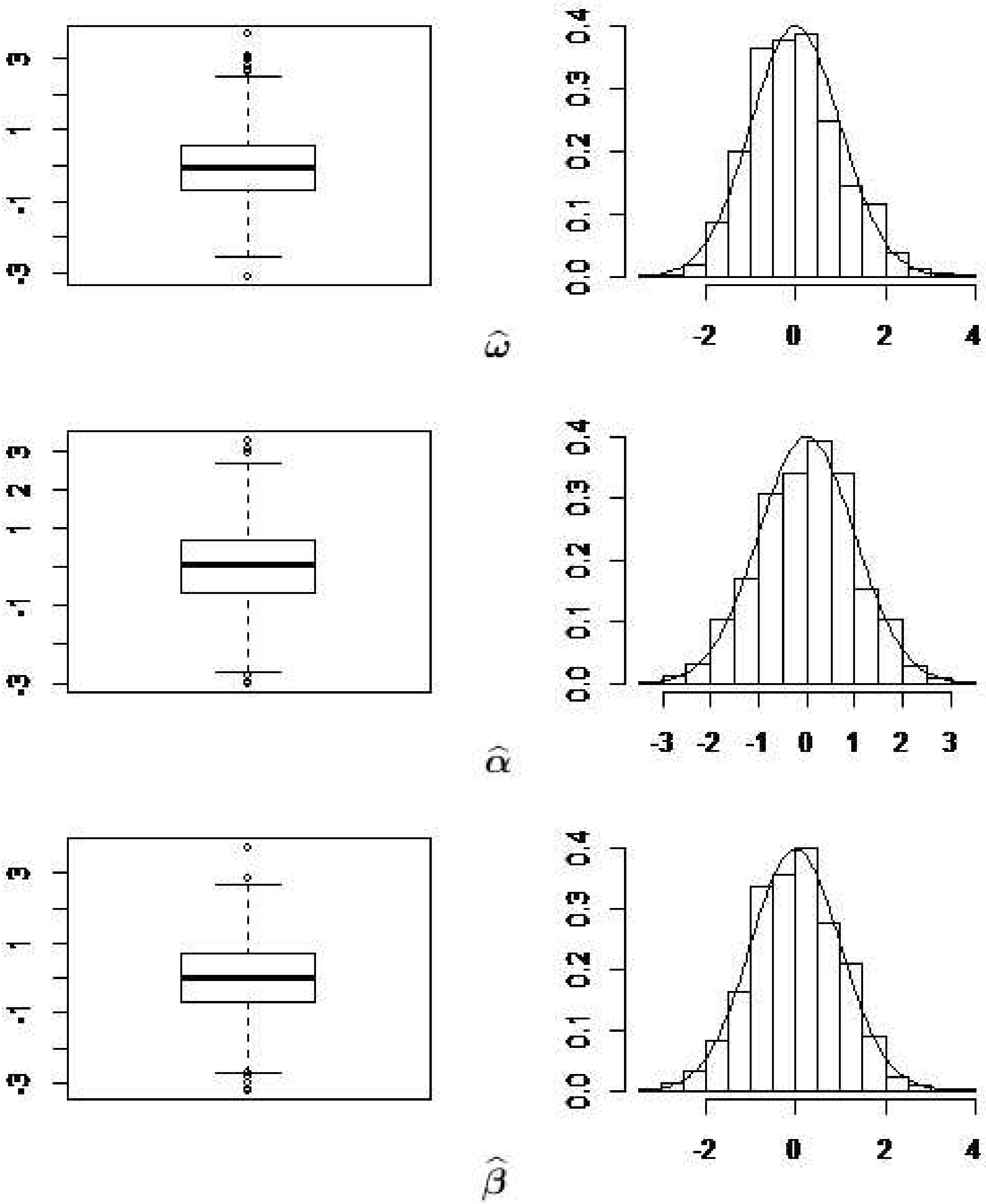

FIGURE 2.4 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}$, $\widehat{\alpha}$, $\widehat{\beta})$ when $\theta_0 = (2, 0.3, 0.6)$ for an INGARCH(1,1) model with Poisson conditional distribution $\mathcal{P}(\lambda_t)$. The number of simulations is N=1000 and the sample size is n=3000. Superimposed is the standard normal density function.

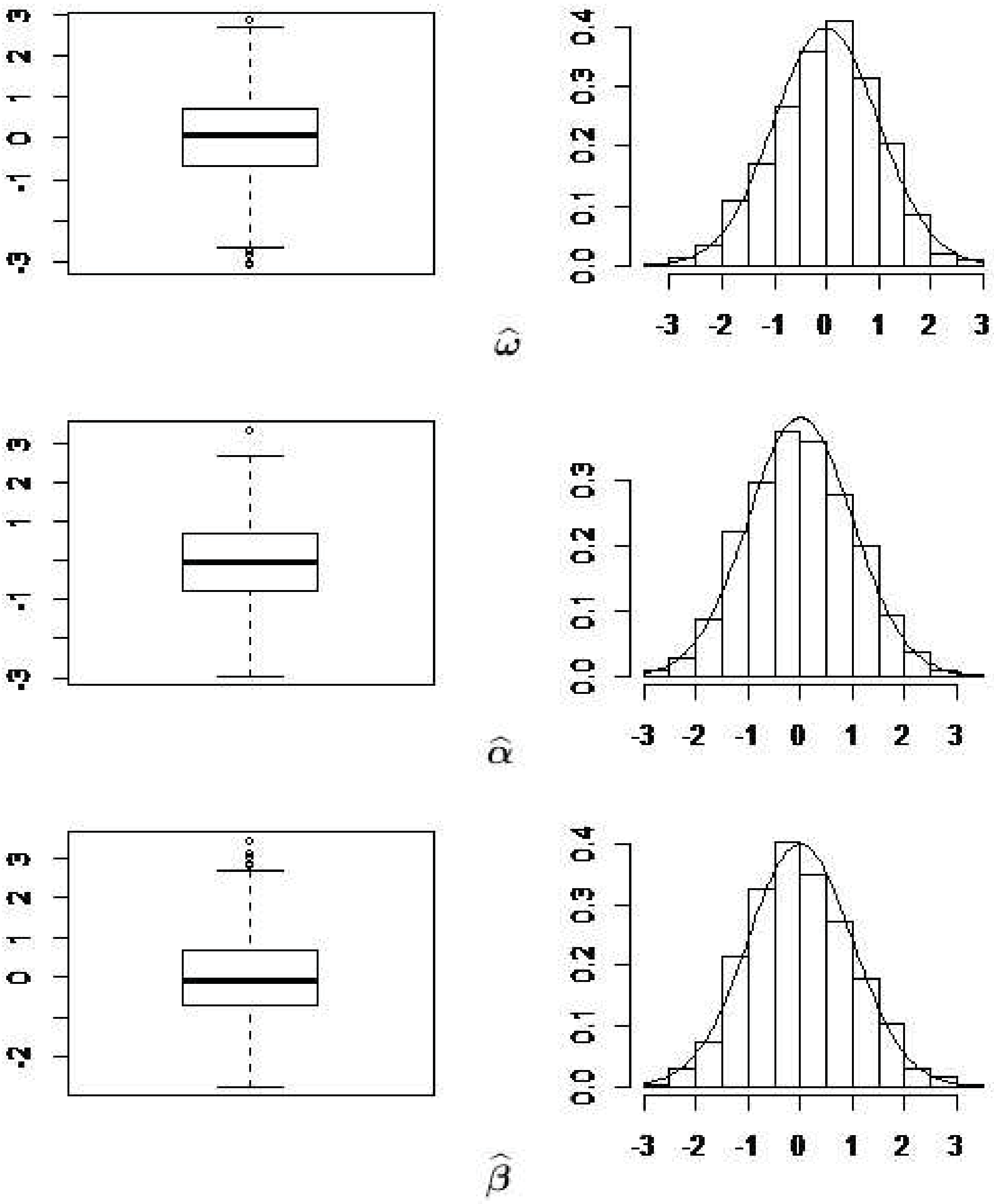

Figure 2.5 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}$, $\widehat{\alpha}$, $\widehat{\beta})$ when θ_0 = (2, 0.3, -0.6) for a log-linear(1,1) model with Poisson conditional distribution $\mathcal{P}(\lambda_t)$. The number of simulations is N=1000 and the sample size is n=3000. Superimposed is the standard normal density function.

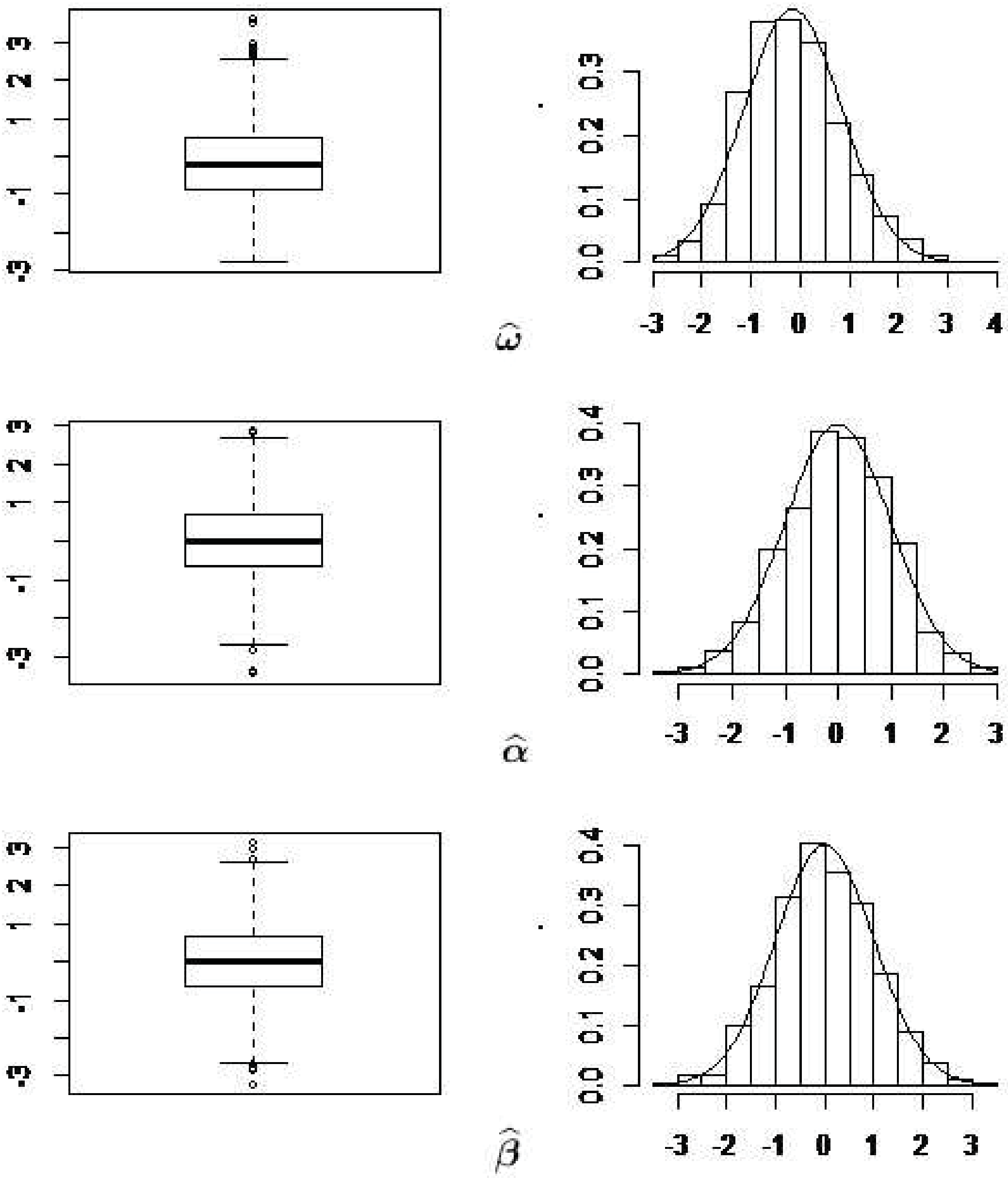

Figure 2.6 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}$, $\widehat{\alpha}$, $\widehat{\beta})$ when $\theta_0 = (2,\ 0.3,\ 0.6)$ for an INGARCH(1,1) model with double Poisson conditional distribution $\mathcal{DP}(\lambda_t, 2)$. The number of simulations is N=1000 and the sample size is n=3000. Superimposed is the standard normal density function.

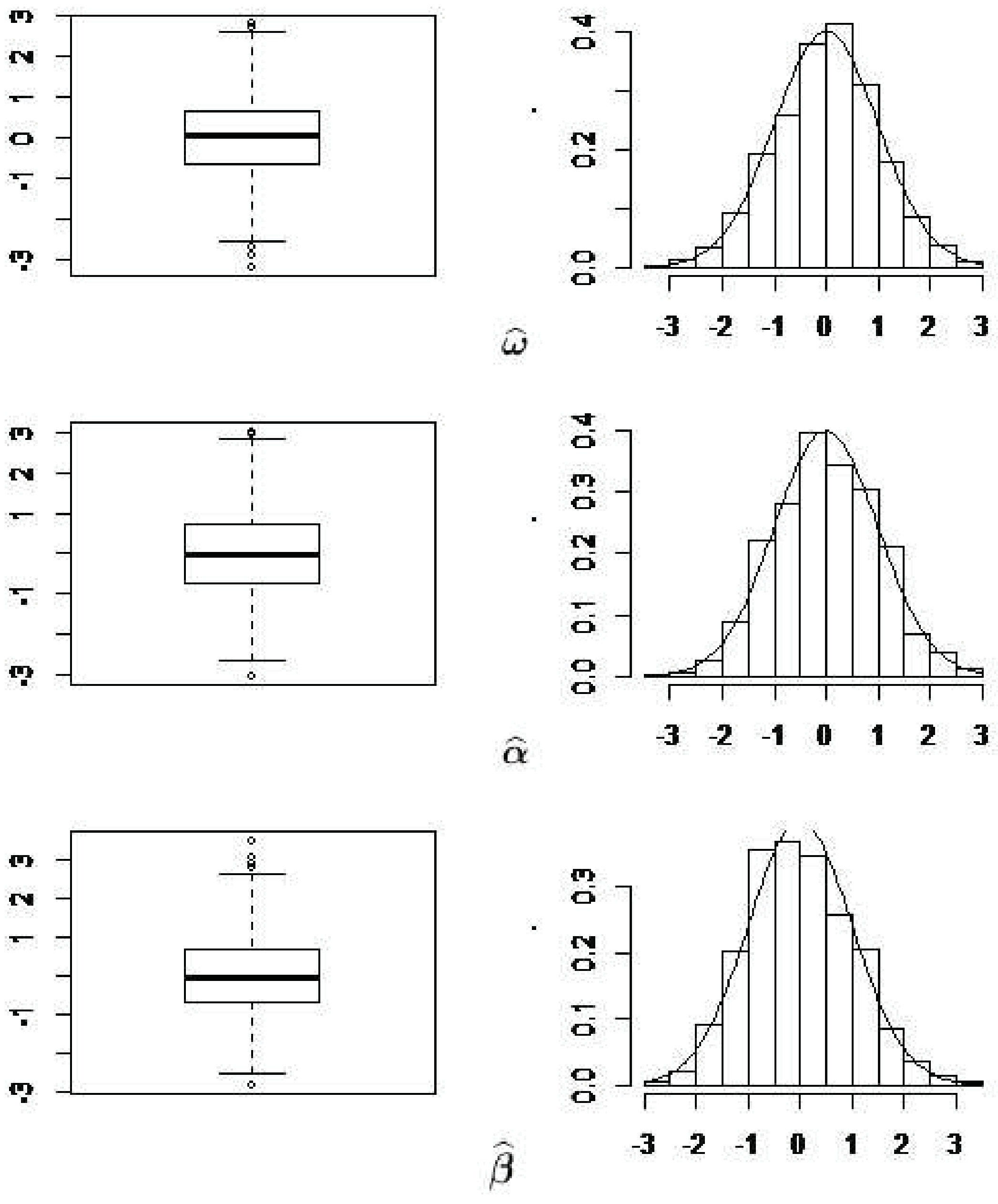

FIGURE 2.7 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}$, $\widehat{\alpha}$, $\widehat{\beta})$ when θ_0 = (2, 0.3, -0.6) for a log-linear(1,1) model with double Poisson conditional distribution $\mathcal{DP}(\lambda_t, 2)$. The number of simulations is N=1000 and the sample size is n=3000. Superimposed is the standard normal density function.

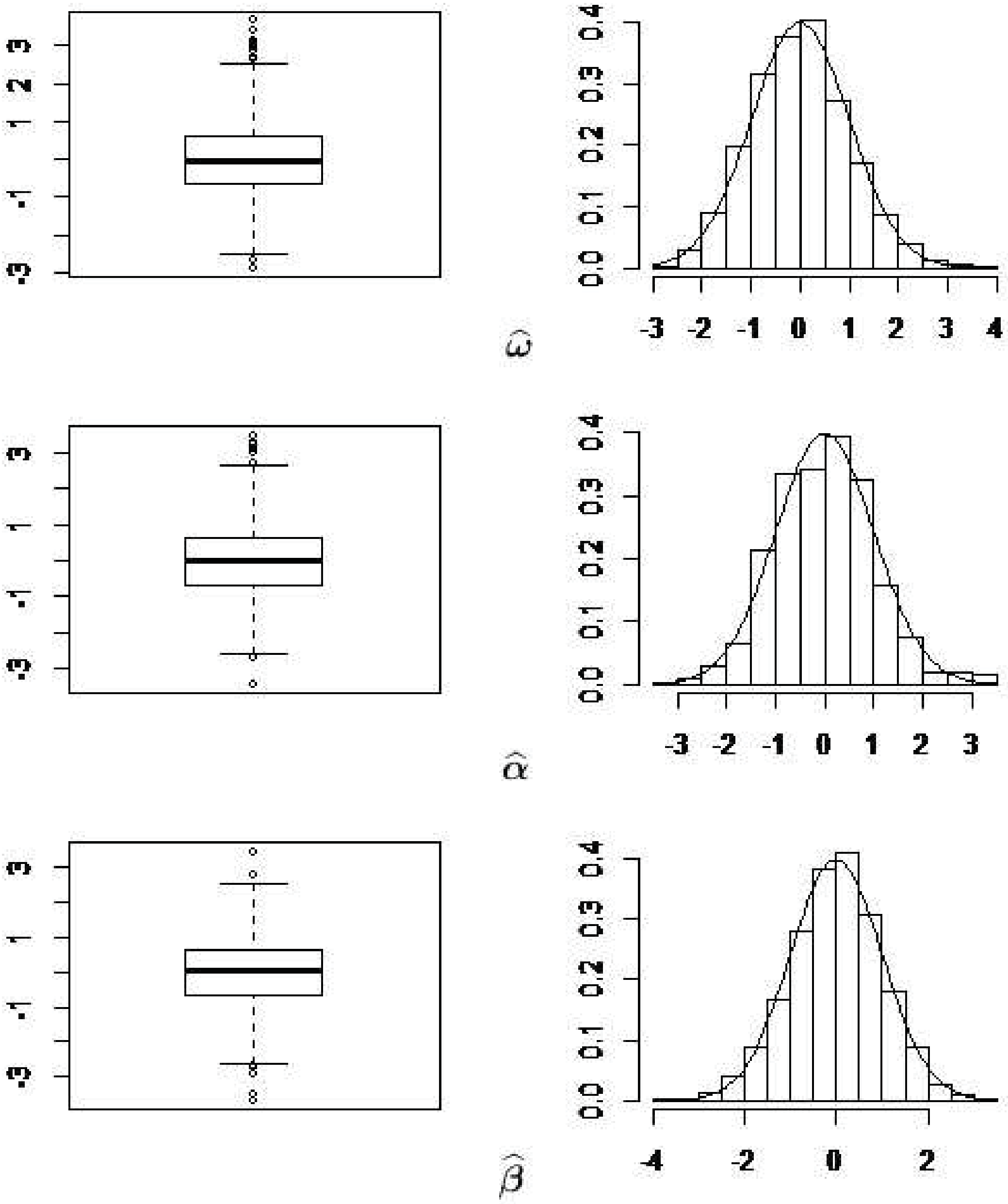

Figure 2.8 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}$, $\widehat{\alpha}$, $\widehat{\beta})$ when θ_0 =(2, 0.3, 0.6) for an INGARCH(1,1) model with negative binomial conditional distribution $\mathcal{NB}(3, p_t)$. The number of simulations is N=1000 and the sample size is n=3000. Superimposed is the standard normal density function.

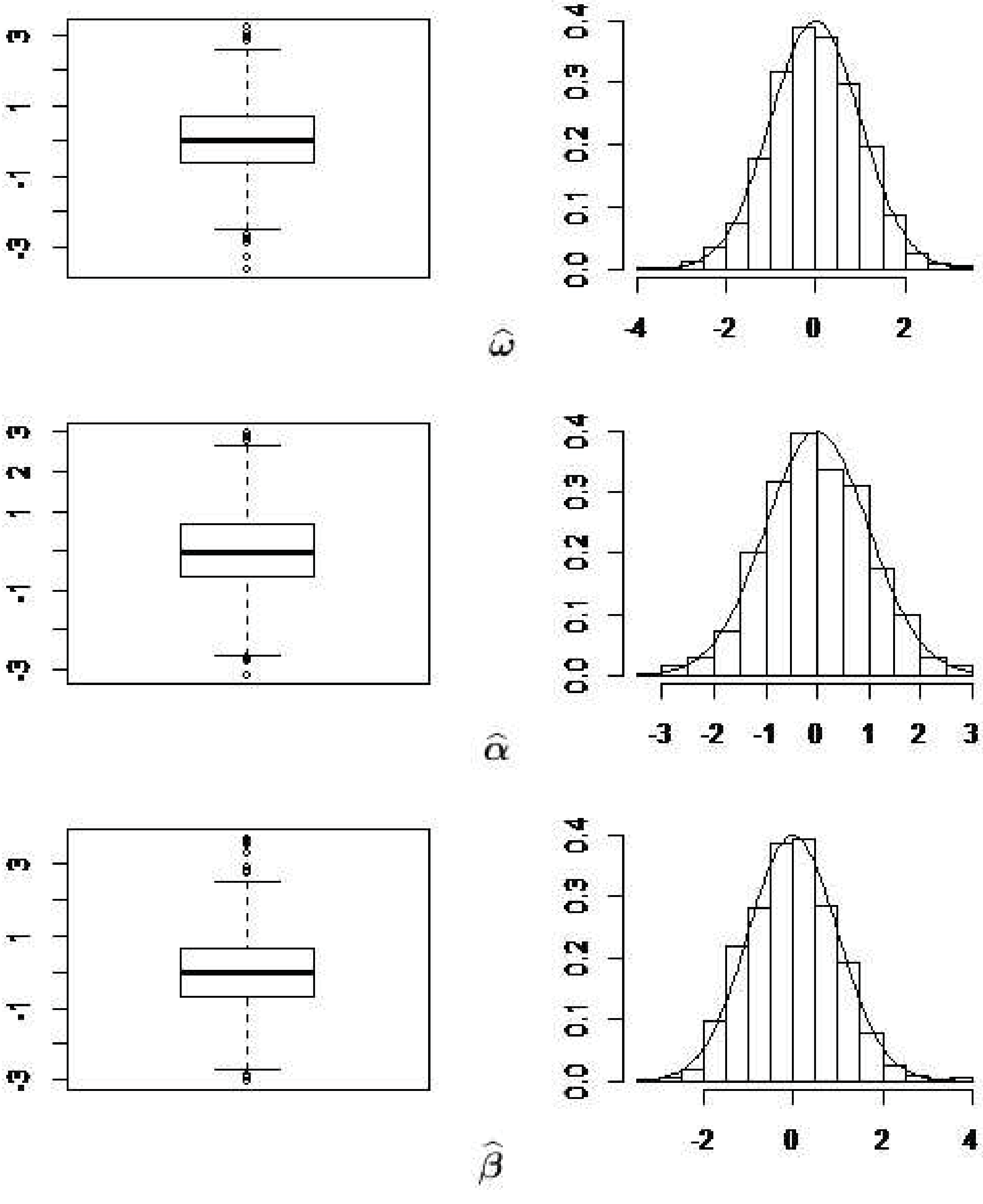

FIGURE 2.9 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}$, $\widehat{\alpha}$, $\widehat{\beta})$ when θ_0 =(2, 0.3, -0.6) for a log-linear(1,1) model with negative binomial conditional distribution $\mathcal{NB}(3, p_t)$. The number of simulations is N=1000 and the sample size is n=3000. Superimposed is the standard normal density function.

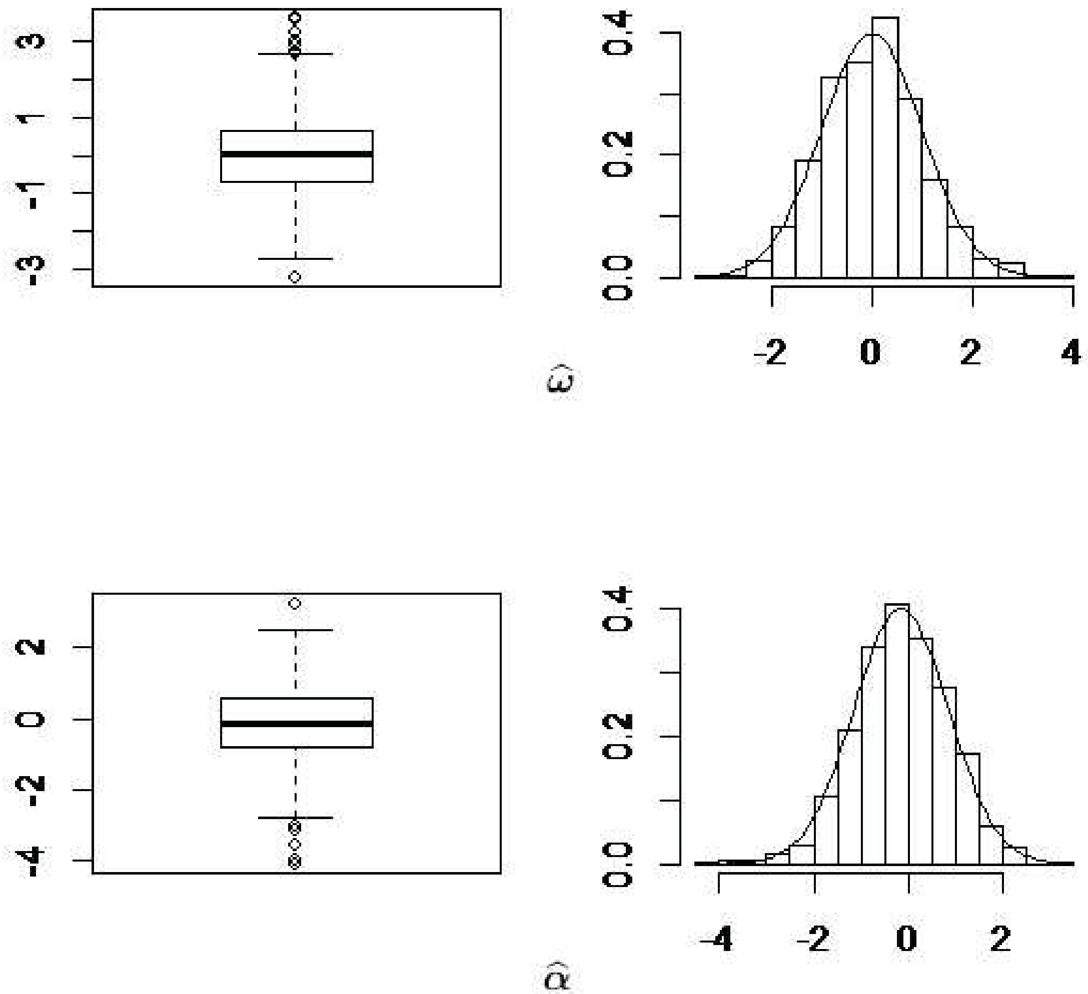

Figure 2.10 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}\ ,\ \widehat{\alpha})$ for an INAR(1) model when the distribution of the innovation is geometric $\mathcal{G}(0.6)$ and α_0=0.9. The number of simulations is N=1000 and the sample size is n=3000. Superimposed is the standard normal density function.

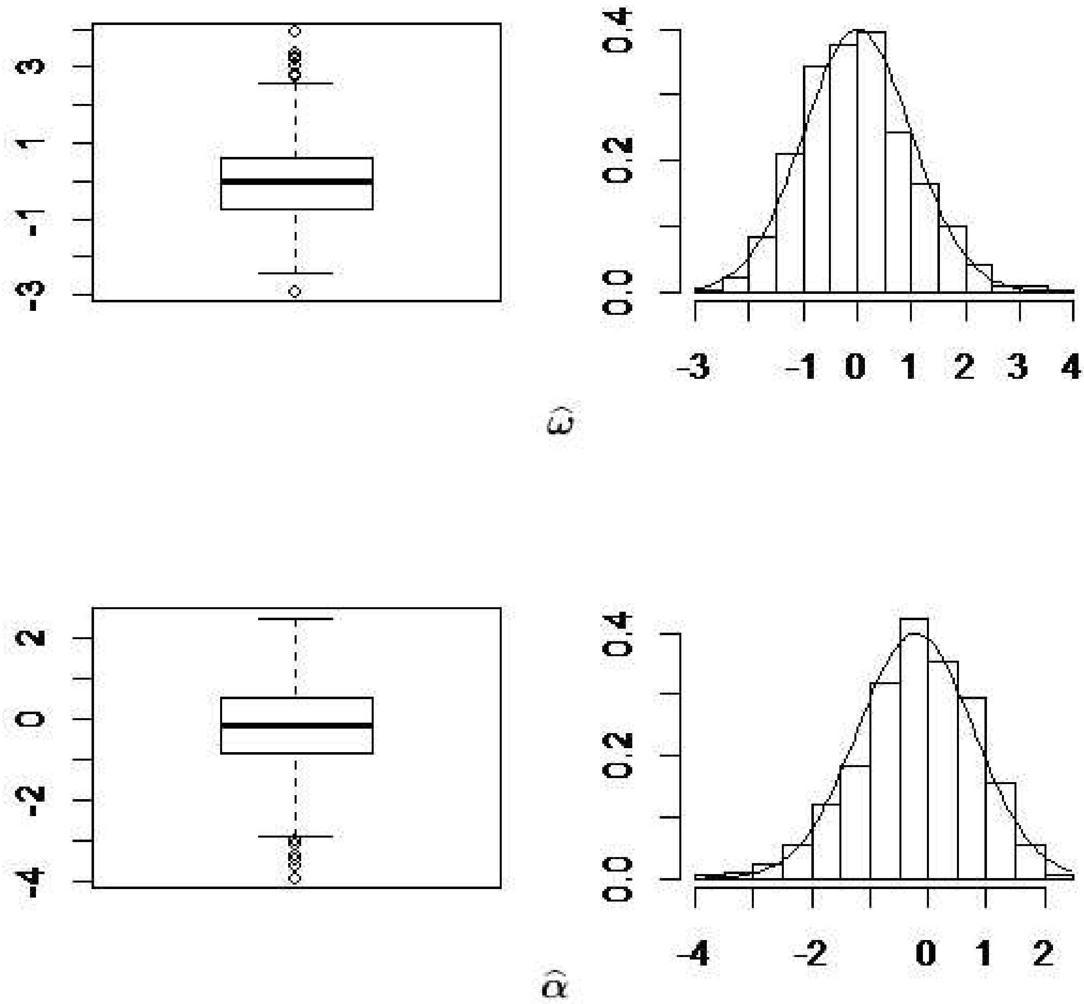

Figure 2.11 – Boxplot and histogram of the standardized distribution of $\widehat{\theta} = (\widehat{\omega}, \widehat{\alpha})$ for an INAR(1) model when the distribution of the innovation is Poisson $\mathcal{P}(3)$ and α_0=0.9. The number of simulations is N=1000 and the sample size is n=3000. Superimposed is the standard normal density function.

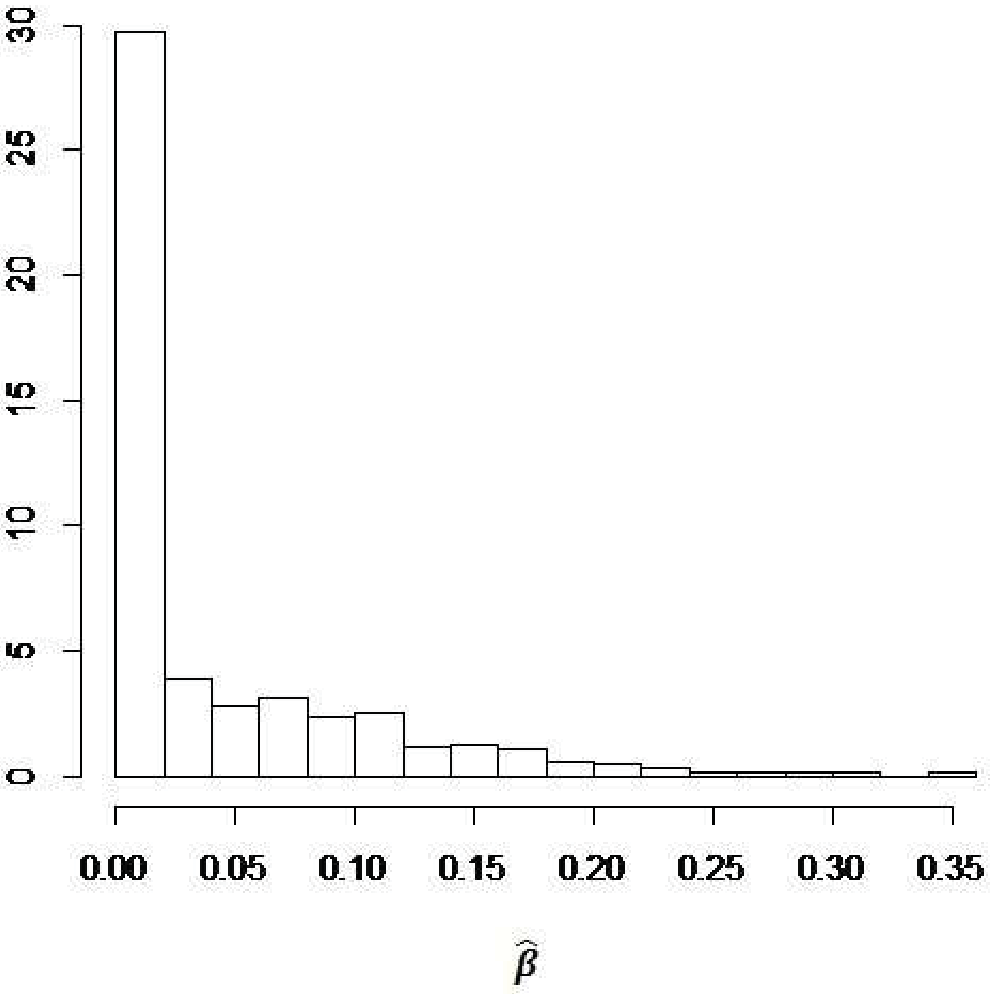

Figure 2.12 – Histogram of the distribution of $\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right)$ when θ_0= (ω_0 , α_0, β_0)= (2, 0.3, 0) for an INGARCH(1,1) model with negative binomial conditional distribution $\mathcal{NB}(p_t, 3)$. The number of simulations is N=1000 and the sample size is n=1000.

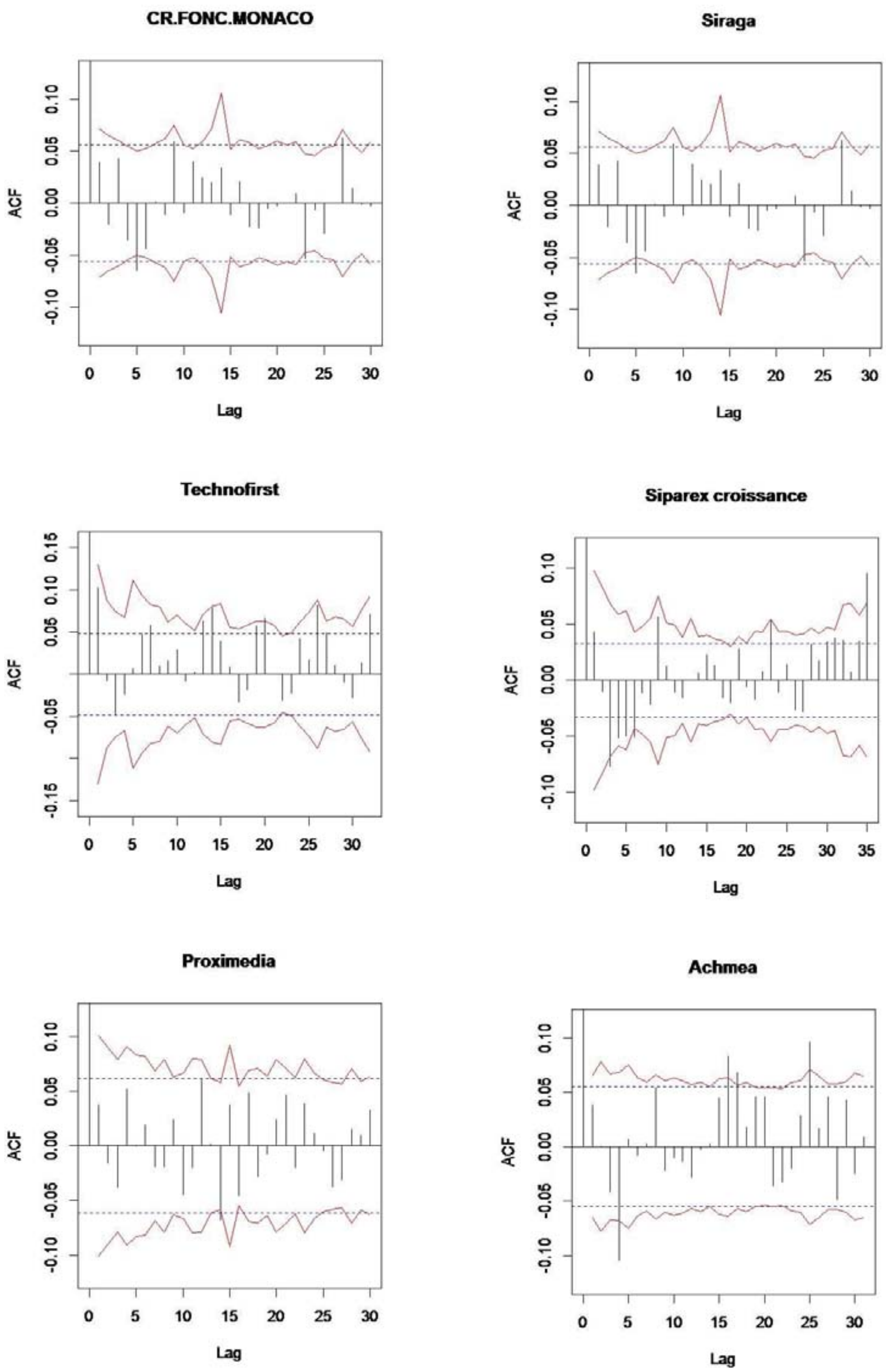

Figure 2.13 – Sample autocorrelation functions for the residuals of the series, together with the standard significance bands in dotted. The significance bands in full line are obtained from the generalized Bartlett's formula (see Francq & Zakoïan, 2009a).

Chapitre 3

Portmanteau test for count time series models

Abstract. We propose a robust and general goodness-of-fit test for the count time series models. The test represents a tool for checking the adequacy of fit for wide and important classes of integer-valued time series models, for example, the INGARCH(p, q), the log-linear$(1, 1)$, the non-linear$(1, 1)$ models with diverse conditional distributions and the INAR(p) models with different marginal distributions of the innovations. The asymptotic distribution of the statistic is derived and its finite sample properties are studied through Monte Carlo simulations. Moreover, the results are illustrated by a real financial data application.

Keywords. Portmanteau, Goodness-of-fit, INGARCH models, INAR models, Log-linear models, Non-linear models, Poisson quasi-maximum likelihood estimator, Time series of counts.

3.1 Introduction

The portmanteau test is considered as one of the most important tools for evaluating the goodness-of-fit in the context of the time series analysis. Firstly, this test based on the sum of squared autocorrelation functions of residuals was limited to be used when the residuals are independent and identically distributed (iid) (see Box & Pierce, 1970 and Ljung & Box, 1978 for more details). Lots of modifications have been proposed in the literature to make this test able to treat more general residual cases (see *e.g* Francq et al. (2005) and the references therein). For the time series of counts which received, in the recent years, a remarkable interest, the goodness-of-fit tests are less developed. However, several important breakthroughs were achieved in this field. For instance, Zhu & Wang (2010) introduced a collection of tests for the goodness-of-fit of the Poisson INARCH(p) model, but some of these tests do not take into account the influence of parameter estimation errors which affect the robustness of the test, especially when the parameters are relatively large, and other tests depend on arbitrary parameters. Neumann (2011) proposed to use the conditional equi-dispersion assumption of the Poisson INGARCH(p, q) model for testing the adequacy of the specification of the intensity process. In addition, Fokianos & Neumann (2013) studied a non parametric goodness-of-fit test for the Poisson INGARCH(p, q) models. Recently, Meintanis & Karlis (2014) propsed a goodness-of-fit test for the innovation distribution of the Poisson INAR(1) model. Hudecová et al. (2015) introduced a goodness-of-fit test based of the probability generating function for both Poisson INAR(p) and Poisson INARCH(p) models. Schweer (2016) introduced a more general test in which the empirical joint probability generating function is considered for testing the adequacy of wide class of count time series models, but this test depends on arbitrary parameters. In addition, the generalization of the methodology of this test to higher-order models seems quite challenging. In this contribution, the adopted approach is simpler and more general than the previous mentioned tests. Under appropriate conditions, the test can be used as a diagnostic checking tool for INGARCH(p, q), log-linear and non-linear models with a large variety of exponential discrete conditional distributions having non negative integer-valued supports. Moreover, the test can also be used to evaluate the adequacy of fit for the INAR(p) model under several distributional assumptions on the innovations. The methodology of the test is similar, to some extent, to the one used to evaluate the adequacy of fit for the generalized autoregressive conditional heteroskedasticity models (GARCH) in Carbon & Francq (2010), where the quasi maximum likelihood (QMLE) is used to estimate the models and to derive the asymptotic distribution of the statistic of test. The test proposed in the present chapter is based on the residual autocovariances obtained after estimating the model using the Poisson quasi maximum likelihood estimator (PQMLE) studied by Ahmad & Francq (2016), which represents a general way for estimating the conditional mean count time series models. The chapter is organized as follows. In Section 3.2, we present a general definition of the model and show the properties of the residuals which will be used in the test, we present some examples of models which satisfy the assumptions required and we recall the Poisson QMLE which is used to estimate the models. Section 3.3 contains the main results concerning the proposed test, the asymptotic distribution of the squared residual autocovariances and its estimation. Sections 3.4 and 3.5 show, respectively, the Monte Carlo simulation results and an application on financial series. The proofs are collected in Section 3.6, and Section 3.7 concludes.

3.2 Model and assumptions

Assume that $\{X_t \in \mathbb{N}\}$ is a count time series, such that

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda_t(\theta_0) = \lambda(X_{t-1}, X_{t-2}, \ldots; \theta_0), \tag{3.2.1}$$

where $\mathcal{F}_{t-1}$ denotes the $\sigma-$field generated by $(X_u, u < t)$,

$$\lambda \text{ is a measurable function valued in } (\underline{\omega}, +\infty) \text{ for some } \underline{\omega} > 0 \tag{3.2.2}$$

and θ_0 is an unknown parameter belonging to some parameter space $\Theta \subset \mathbb{R}^d$. We also assume that

$$\theta \mapsto \lambda_t(\theta) \text{ is almost surely continuous, } \Theta \text{ is a compact set,} \tag{3.2.3}$$

and

$$\text{the support of the conditional distribution of } X_t \text{ given its past contains at least three different values.} \tag{3.2.4}$$

For all $\theta \in \Theta$, $x_0 \in \mathbb{R}^+$ and $t \geq 1$, let

$$\lambda_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \ldots; \theta) \text{ and } \widetilde{\lambda}_t(\theta) = \lambda(X_{t-1}, X_{t-2}, \ldots, X_1, x_0, x_0, \ldots; \theta).$$

Note that $\widetilde{\lambda}_t(\theta)$ will serve as a proxy for $\lambda_t(\theta)$. It is obtained by setting to some positive real value x_0 the unknown initial values $X_0, X_{-1}, \ldots$ involved in $\lambda_t(\theta)$. This value x_0 can either be a fixed integer, for instance $x_0 = 0$, or a value depending on θ, or a value depending on the observations. We assume that

$$\widetilde{\lambda}_t(\theta) \geq \underline{\omega} \text{ for some } \underline{\omega} > 0, \forall t > 1 \text{ and } \theta \in \Theta. \tag{3.2.5}$$

Moreover, we assume that the fourth-order moment of the marginal distribution of X_t exists

$$EX_t^4 < \infty. \tag{3.2.6}$$

We define the residuals as follows

$$\epsilon_t(\theta_0) = X_t - \lambda_t(\theta_0). \tag{3.2.7}$$

Note that the assumption (3.2.6) implies that $E(\epsilon_t(\theta_0)^4) < \infty$ as well as $E(\lambda_t(\theta_0)^4) < \infty$.

If the conditional mean is correctly specified, under Assumption **A2** (see Appendix), one can show that $\epsilon_t(\theta_0)$ is a white noise sequence, where

$$E(\epsilon_t(\theta_0)) = E\left(E\left(X_t - \lambda_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) = 0$$

and

$$\begin{aligned}
Var(\epsilon_t(\theta_0)) &= E\left(Var\left(\epsilon_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) + Var\left(E\left(\epsilon_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) \\
&= E\left(Var\left(\epsilon_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) \text{ [since } Var\left(E\left(\epsilon_t(\theta_0)|\mathcal{F}_{t-1}\right)\right) = 0] \\
&= E\left(E\left(\epsilon_t(\theta_0)^2|\mathcal{F}_{t-1}\right)\right) \\
&= E\left(E\left((X_t - \lambda_t(\theta_0))^2|\mathcal{F}_{t-1}\right)\right) \\
&= E\left(Var\left(X_t|\mathcal{F}_{t-1}\right)\right).
\end{aligned} \tag{3.2.8}$$

Moreover, for $h > 0$, we have

$$Cov(\epsilon_t(\theta_0), \epsilon_{t+h}(\theta_0)) = E\left(\epsilon_t(\theta_0) E\left(\epsilon_{t+h}(\theta_0) | \mathcal{F}_{t+h-1}\right)\right) = 0.$$

The above formulation can accommodate a large variety of count time series models, for example :

The INGARCH(p, q) models : The INGARCH(p, q) models are defined by assuming that the conditional mean takes the following general linear representation

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda_t(\theta_0) = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} X_{t-i} + \sum_{j=1}^{p} \beta_{0j} \lambda_{t-j}(\theta_0), \tag{3.2.9}$$

where $\omega_0 > 0$, $0 \leq \alpha_{0i} < 1$ $(i = 1, .., q)$ and $0 \leq \beta_{0j} < 1$ $(j = 1, .., p)$. When the conditional distribution of X_t given its past is Poisson $\mathcal{P}(\lambda_t)$, the model is called Poisson INGARCH(p, q) model (see Ferland et al., 2006). The negative binomial INGARCH(p, q) model is studied by Christou & Fokianos (2014) and is defined by assuming that the conditional distribution is negative binomial $\mathcal{NB}(p_t, \nu)$ with $\lambda_t = p_t/(1 - p_t)\nu$. The previously mentioned models can deal only with the over-dispersed series. To overcome this limitation, Heinen (2003) proposed to use the double-Poisson $\mathcal{DP}(\lambda_t, \gamma)$ conditional distribution of Efron (1986) for modelling both over and under dispersed series. Another alternative is the generalized-Poisson INGARCH(p, q) model of Zhu (2012) which is defined as in (3.2.9) by assuming that the conditional distribution of X_t given its past is generalized-Poisson $\mathcal{GP}(\lambda_t^*, \kappa)$, where $\lambda_t^* > 0$ and $max(-1, -\lambda_t^*/4) \leq \kappa < 1$. The conditional mean of X_t given its past is equal to $\lambda_t = \frac{\lambda_t^*}{1-\kappa}$. Under the following condition :

$$\sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j} < 1, \tag{3.2.10}$$

all the models presented above, except the double-Poisson INGARCH(p,q) model, are stationary and ergodic (see Gonçalves et al., 2015). For the later model, ergodicity is not established, but one can conjecture that it holds under the same condition. In the Poisson case, Ferland et al. (2006)) have shown that under the condition (3.2.10), and when $p = q = 1$, the model (3.2.9) has moment of all orders, while the existence of the fourth-order moments for the negative binomial INGARCH(p, q) model requires a more complicated condition (see Section 3.2 of Ahmad & Francq, 2016). Furthermore, to our knowledge, the high order characterizations of the double-Poisson INGARCH(p, q) and the generalized-Poisson INGARCH(p, q) models are not yet studied in the literature.

The log-linear model : The log-linear model was introduced by Fokianos & Tjøstheim (2011) and is defined by assuming that the conditional distribution of X_t given its past is Poisson with intensity parameter $\lambda_t = e^{\upsilon_t}$, where

$$\upsilon_t = \omega_0 + \alpha_0 \log(X_{t-1} + 1) + \beta_0 \upsilon_{t-1}. \tag{3.2.11}$$

The main feature of this model is its capacity to model the count time series with a negative autocorrelation. Under the condition

$$|\alpha_0 + \beta_0| \vee |\alpha_0| \vee |\beta_0| < 1, \tag{3.2.12}$$

Douc et al. (2013) showed that the Poisson log-linear model with the log-intensity (3.2.11) admits a stationary and ergodic solution. In Fokianos & Tjøstheim (2011), it is shown that in the Poisson case and under a condition similar to (3.2.12), the moment conditions (3.2.6) is verified.

The non-linear model : One of the commonly used non linear specification of the model (3.2.1) is the following (see Gao et al., 2009) :

$$\lambda_t(\theta_0) = \frac{\omega_0}{(1 + X_{t-1})^{c_0}} + \alpha_0 X_{t-1} + \beta_0 \lambda_{t-1}(\theta_0), \tag{3.2.13}$$

where the parameters ω_0, c_0, α_0 and β_0 are positive. When the conditional distribution of X_t given its past is Poisson, Fokianos & Tjøstheim (2012) and Christou (2014) have shown that if $max\{\alpha_0, \omega_0 c_0 - \alpha_0\} + \beta_0 < 1$, the model (3.2.13) is stationary, ergodic and it has moments of all orders.

The INAR(p) model : The INAR(p) model has been studied by Alzaid & Al-Osh (1990) and is defined as follows :

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + r_t, \tag{3.2.14}$$

where $\alpha_i \circ X_{t-i} = \sum_{j=1}^{X_{t-i}} y_j$, with the convolution $\alpha_i \circ X_{t-i} = 0$ if $X_{t-i} = 0$, y_j are independent Bernoulli random variables with probability of success α_i, and r_t has a non negative integer-valued support. The variable λ_t is obtained by taking the conditional expectation of X_t given its past, which gives an INARCH(p) model with $\omega = E(r_t)$. In Theorem 2.3 of Silva (2005), it is shown that, if r_t has a finite fourth-order moment, the process (3.2.14) has moments of all orders.

According to Ahmad & Francq (2016), the aforementioned models can be consistently estimated by PQMLE which is defined as any measurable solution of

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \widetilde{L}_n(\theta), \quad \widetilde{L}_n(\theta) = \frac{1}{n}\sum_{t=s+1}^{n} \widetilde{\ell}_t(\theta). \tag{3.2.15}$$

The value of s is asymptotically unimportant, but it can affect the finite sample behaviour of the PQMLE by reducing the impact of the initial value x_0. Under (3.2.1)-(3.2.6), (3.2.5), the assumptions **A1-A4** and the assumptions **A6-A9** (see Appendix), the PQMLE is consistent and $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ is asymptotically normal with mean 0 and covariance matrix $\Sigma := J^{-1}IJ^{-1}$, where

$$J = E\left(\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}\right), \quad I = E\left(\frac{Var(X_t|\mathcal{F}_{t-1})}{\lambda_t^2(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'}\right). \tag{3.2.16}$$

3.3 Portmanteau test

As mentioned in the introduction, the goodness-of-fit test is based on the residual autocovariances. Firstly, the residuals considered in the test are defined as follows

$$\widetilde{\epsilon}_t(\widehat{\theta}_n) = X_t - \widetilde{\lambda}_t(\widehat{\theta}_n).$$

We denote by $\widehat{\gamma}_m = (\widehat{\gamma}_{(1)}, .., \widehat{\gamma}_{(m)})'$ the vector of the residuals autocovariance functions, where, for $m < n$ and $h \in \{1, .., m\}$, its elements are given by

$$\widehat{\gamma}_{(h)} = \frac{1}{n}\sum_{t=h+1}^{n} \widetilde{\epsilon}_t(\widehat{\theta}_n)\widetilde{\epsilon}_{t-h}(\widehat{\theta}_n).$$

Define the $m \times m$ matrix $\widehat{\Sigma}_{\gamma_m}$, whose elements for $(h, l) \in \{1, .., m\}$ are given as follows

$$\widehat{\Sigma}_{\gamma_m}(h, l) = \frac{1}{n}\sum_{t=max(h,l)+1}^{n} \widetilde{\epsilon}_t^2(\widehat{\theta}_n)\widetilde{\epsilon}_{t-h}(\widehat{\theta}_n)\widetilde{\epsilon}_{t-l}(\widehat{\theta}_n).$$

Let $\widehat{\Sigma}$ a consistent estimator of the matrix Σ defined in Section 3.2

$$\widehat{\Sigma} = \widehat{J}^{-1}\widehat{I}\widehat{J}^{-1},$$

where

$$\widehat{J} = \frac{1}{n}\sum_{t=s+1}^{n}\frac{1}{\widetilde{\lambda}_t(\widehat{\theta}_n)}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta'} \quad \text{and} \quad \widehat{I} = \frac{1}{n}\sum_{t=s+1}^{n}\left(\frac{X_t}{\widetilde{\lambda}_t(\widehat{\theta}_n)}-1\right)^2\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta'}.$$

Let d denote the number of parameters of the model. We define the $m \times d$ matrix $\widehat{C}_m$ and the $d \times m$ matrix $\widehat{\Sigma}_{\theta,\gamma_m}$. The elements of these two matrices, for $1 \leq k \leq d$ and $1 \leq h \leq m$, are obtained respectively by

$$\widehat{C}_m(h,k) = -\frac{1}{n}\sum_{t=h+1}^{n}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta_k}\widetilde{\epsilon}_{t-h}(\widehat{\theta}_n)$$

and

$$\widehat{\Sigma}_{\theta,\gamma_m}(k,h) = \widehat{J}^{-1}\frac{1}{n}\sum_{t=h+1}^{n}\frac{\partial\widetilde{\lambda}_t(\widehat{\theta}_n)}{\partial\theta_k}\frac{\widetilde{\epsilon}_t^2(\widehat{\theta}_n)}{\widetilde{\lambda}_t(\widehat{\theta}_n)}\widetilde{\epsilon}_{t-h}(\widehat{\theta}_n).$$

Now, we define the matrix $\widehat{\Sigma}_{\widehat{\gamma}_m}$ as follows

$$\widehat{\Sigma}_{\widehat{\gamma}_m} = \widehat{\Sigma}_{\gamma_m} + \widehat{C}_m\widehat{\Sigma}\widehat{C}_m' + \widehat{C}_m\widehat{\Sigma}_{\theta,\gamma_m} + \widehat{\Sigma}_{\theta,\gamma_m}'\widehat{C}_m'$$

and we assume that

$$\text{the matrix } \widehat{\Sigma}_{\widehat{\gamma}_m} \text{ is invertible.} \tag{3.3.1}$$

This condition will be shown hold in particular models (see Section 3.6)

Theorem 3.3.1. *Under (3.2.1)-(3.2.6), (3.3.1) and the assumptions **A1-A10** (see Appendix), we have*

$$n\widehat{\gamma}_m'\widehat{\Sigma}_{\widehat{\gamma}_m}^{-1}\widehat{\gamma}_m \xrightarrow{\mathcal{L}} \chi_m^2.$$

The adequacy of model is rejected at the asymptotic level $\underline{\alpha}$ when

$$n\widehat{\gamma}_m'\widehat{\Sigma}_{\widehat{\gamma}_m}^{-1}\widehat{\gamma}_m > \chi_m^2(1-\underline{\alpha}).$$

3.4 Numerical illustrations

To examine the finite sample behaviour of the test defined in Theorem 3.3.1, we report a Monte Carlo simulation with 1000 independent replications. Four important classes of models are considered in this experiment, namely the INGARCH(p, q), the INAR(p), the log-linear$(1, 1)$ and the non-linear$(1, 1)$ models. For all phases of simulations, we simulate the INGARCH(p, q), the log-linear$(1, 1)$ and the non-linear$(1, 1)$ models with three conditional distributions : Poisson $\mathcal{P}(\lambda_t)$, negative binomial $\mathcal{NB}(p_t, \nu)$ and double-Poisson $\mathcal{DP}(\lambda_t, \gamma)$ distributions. We also consider three distributional cases of the innovation r_t for the INAR(p) models : Poisson $\mathcal{P}(\lambda)$, geometric $\mathcal{G}(p)$ and negative binomial $\mathcal{NB}(p, \nu)$ distributions. The models are estimated using PQMLE. To ensure the stationarity of the simulated series of size n we generated $n + 100$ values, discarding the first 100 and keeping the rest assuming implicitly that stationarity has been achieved. For the size of test, we simulate two versions for each of INGARCH(p, q) and INAR(p) models. For the INGARCH(p, q) models, we consider an INGARCH$(1, 1)$ and an INGARCH$(1, 2)$ models. For the INAR(p) models, we evaluate the size of test for INAR(1) and INAR(2) models. For the power, we fit an INGARCH$(1, 1)$ model, we consider two alternatives : INGARCH$(1, 2)$ and INGARCH$(1, 3)$. For the INAR(p) models, we fit an INAR(1) model against two alternatives : INAR(2) and INAR(3) models. For the log-linear$(1, 1)$ and the non-linear$(1, 1)$ models, similarly to the previous experiments, we evaluate the size of test for several conditional distributions. For the power, we fit log-linear$(1, 1)$ and non-linear$(1, 1)$ models and we consider the INGARCH$(1, 1)$ models as alternatives. For each replication, we carried out the portmanteau test for evaluating the adequacy of the fitted models at three asymptotic levels $\underline{\alpha} = \{1\%, 5\%, 10\%\}$. In view of Tables 3.1, 3.3, 3.5 and 3.7, we can note that the empirical sizes for all the models are satisfactorily close to their theoretical nominal levels, especially when the sample size is large. Moreover, the results summarized in Tables 3.2, 3.4, 3.6 and 3.8 show that the proposed test achieves good power even when the sample size is relatively small. It remains to point out that the sizes and the powers of the test appear to be influenced by the number of autocovariances used in the test. Especially, when the sample size is small (n=500), the sizes tend to be less than the asymptotic levels as

Table 3.1 – Size of test for the INGARCH model (in %)

	INGARCH(1,1) $\omega_0 = 2,\ \alpha_0 = 0.3,\ \beta_0 = 0.6$								
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$
	$\underline{\alpha} = 1\%$								
500	1.0	0.9	0.7	0.3	0.6	0.7	0.3	0.3	0.5
1000	0.6	1.0	0.7	0.3	0.7	0.5	0.6	0.4	0.6
4000	0.8	1.7	1.4	0.7	1.0	1.2	1.4	0.8	0.8
	$\underline{\alpha} = 5\%$								
500	5.1	5.5	4.9	3.2	3.7	4.3	3.1	3.2	3.9
1000	4.2	5.2	4.2	4.3	4.4	5.5	4.3	3.5	4.4
4000	4.7	6.2	4.9	5.1	5.7	5.4	5.0	4.5	3.9
	$\underline{\alpha} = 10\%$								
500	10.5	9.8	11.1	7.5	8.5	8.8	8.8	5.6	7.4
1000	9.0	9.7	9.5	10.3	10.3	11.2	8.5	8.4	9.3
4000	10.4	11.9	10.3	10.5	11.0	10.8	10.8	8.9	9.6
	INGARCH(1,2) $\omega_0 = 2,\ \alpha_{01} = 0.4,\ \alpha_{02} = 0.3,\ \beta_0 = 0.2$								
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$
	$\underline{\alpha} = 1\%$								
500	1.2	0.8	1.1	1.0	0.4	0.2	0.9	0.3	0.6
1000	0.9	1.2	1.4	0.7	0.6	1.0	1.2	0.3	0.3
4000	1.0	1.3	1.0	0.9	0.8	1.1	0.6	0.7	0.5
	$\underline{\alpha} = 5\%$								
500	4.5	4.3	5.8	5.7	3.3	3.8	4.8	2.8	3.4
1000	5.5	4.7	4.4	3.7	3.2	4.9	4.5	3.7	4.3
4000	4.7	5.7	4.5	4.0	3.9	4.7	3.9	4.6	4.1
	$\underline{\alpha} = 10\%$								
500	9.9	10.2	12.6	9.0	7.4	7.7	10.3	8.0	8.4
1000	9.5	9.0	9.5	8.8	6.5	9.5	10.8	7.9	9.6
4000	10.9	10.9	10.5	8.3	8.9	9.9	8.9	9.3	8.7

m increases. Similarly, the powers decrease as m increases.

3.5 Real data application

We report an application on real transactions data. The data consist of the number of transactions per minute for the stock Ericsson B, for one day period. The length of the data set is 460 observations. The original series, the sample autocorrelation function and the histogram of the series are plotted in Figure 3.1. The mean of the series is 9.909 while the sample variance is 32.836, that is the data are over-dispersed. The estimation results for the different models are summarized in Tables 3.9, 3.10, 3.11 and 3.12. The estimated parameters are shown in the rows '$\widehat{\theta}$'

Table 3.2 – Power of test for the INGARCH model (in %)

INGARCH$(1,2)$ vs INGARCH$(1,1)$									
$\omega_0 = 2,\ \alpha_{01} = 0.4,\ \alpha_{02} = 0.3,\ \beta_0 = 0.2$									
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$
$\underline{\alpha} = 1\%$									
500	45.0	13.5	54.0	23.1	5.0	24.5	13.4	2.4	13.0
1000	88.8	26.7	89.8	67.2	13.8	70.7	53.4	8.5	56.8
4000	100.0	81.4	100.0	100.0	71.1	100.0	100.0	63.0	100.0
$\underline{\alpha} = 5\%$									
500	69.3	31.6	74.3	48.4	18.3	50.6	33.9	12.1	37.3
1000	96.4	50.6	96.9	87.0	34.7	88.3	78.8	24.4	80.5
4000	100.0	90.9	100.0	100.0	85.0	100.0	100.0	80.4	100.0
$\underline{\alpha} = 10\%$									
500	79.9	44.7	83.6	62.5	29.6	63.2	48.2	23.2	50.2
1000	98.6	62.2	98.9	94.1	49.6	93.6	86.5	39.6	78.4
4000	100.0	94.1	100.0	100.0	90.8	100.0	100	87.5	100.0
INGARCH$(1,3)$ vs INGARCH$(1,1)$									
$\omega_0 = 2,\ \alpha_{01} = 0.2,\ \alpha_{02} = 0.4,\ \alpha_{03} = 0.2,\ \beta_0 = 0.1$									
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t, 6)$	$\mathcal{DP}(\lambda_t, 2)$
$\underline{\alpha} = 1\%$									
500	96.8	51.8	97.5	84.5	28.6	83.2	64.4	15.6	70.9
1000	100.0	82.7	100.0	99.8	66.5	100.0	99.3	51.8	99.5
4000	100.0	96.6	100.0	100.0	97.0	100.0	100.0	95.8	100.0
$\underline{\alpha} = 5\%$									
500	99.4	76.1	99.4	96.2	56.4	95.2	86.3	39.1	90.3
1000	100.0	93.5	100.0	100.0	82.9	100.0	99.9	73.4	100.0
4000	100.0	98.5	100.0	100.0	98.5	100.0	100.0	97.9	100.0
$\underline{\alpha} = 10\%$									
500	100.0	83.5	99.8	98.4	70.9	97.8	93.6	54.3	95.9
1000	100.0	97.0	100.0	100.0	89.5	100.0	100.0	83.1	100.0
4000	100.0	99.2	100.0	100.0	98.6	100.0	100.0	98.5	100.0

Table 3.3 – Size of test for INAR model (in %)

n	m=4 $\mathcal{P}(2)$	m=4 $\mathcal{G}(0.5)$	m=4 $\mathcal{NB}(2,0.5)$	m=12 $\mathcal{P}(2)$	m=12 $\mathcal{G}(0.5)$	m=12 $\mathcal{NB}(2,0.5)$	m=20 $\mathcal{P}(2)$	m=20 $\mathcal{G}(0.5)$	m=20 $\mathcal{NB}(2,0.5)$
INAR(1) $\alpha_0 = 0.9$									
$\underline{\alpha} = 1\%$									
500	0.8	0.5	1.0	0.4	0.4	0.5	0.3	0.3	0.6
1000	1.0	0.9	0.7	0.9	1.5	0.7	1.1	0.5	0.6
4000	1.1	0.9	1.0	1.0	1.3	0.9	1.1	0.6	1.2
$\underline{\alpha} = 5\%$									
500	4.5	4.1	4.2	3.5	4.2	3.3	3.3	2.5	3.8
1000	5.3	5.3	4.9	5.1	3.3	3.6	4.2	4.2	4.0
5000	4.1	5.6	5.0	3.8	6.0	5.0	4.1	4.4	4.2
$\underline{\alpha} = 10\%$									
500	10.4	8.3	9.2	8.8	9.3	6.6	7.6	6.6	8.5
1000	10.9	9.6	9.7	8.5	8.1	7.7	8.6	9.3	8.5
4000	9.0	10.2	10.3	9.6	10.3	9.7	8.3	9.8	9.0
INAR(2) $\alpha_{01} = 0.4,\ \alpha_{02} = 0.2$									
$\underline{\alpha} = 1\%$									
500	0.7	0.6	0.8	0.7	0.8	0.4	0.4	0.3	0.4
1000	1.5	0.8	0.9	1.7	1.0	0.5	0.7	0.9	0.6
4000	0.8	1.1	1.1	1.3	1.1	0.9	0.9	0.8	1.3
$\underline{\alpha} = 5\%$									
500	3.6	4.7	5.2	4.2	3.5	3.5	3.2	3.6	3.8
1000	5.4	4.1	4.6	3.8	4.7	4.4	4.6	4.4	4.0
4000	4.6	4.7	6.1	4.7	4.1	4.6	4.7	5.2	5.0
$\underline{\alpha} = 10\%$									
500	8.8	9.8	10.1	7.5	8.7	8.0	6.9	7.8	7.6
1000	10.6	9.8	10.4	9.3	11.1	9.4	9.0	8.8	9.6
4000	9.0	10.2	10.0	9.4	8.9	9.2	9.6	10.8	10.7

Table 3.4 – Power of test for INAR model (in %)

	INAR(2) vs INAR(1)								
	$\alpha_{01}=0.4,\ \alpha_{02}=0.2$								
	m=4			m=12			m=20		
n	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$
	$\underline{\alpha}=1\%$								
500	79.3	73.6	76.4	51.7	44.3	50.9	32.7	28.6	33.7
1000	99.9	98.6	99.8	97.3	93.2	95.6	91.3	85.2	88.9
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\underline{\alpha}=5\%$								
500	93.2	90.7	92.2	78.4	71.6	76.4	60.3	56.9	62.1
1000	100.0	99.8	99.9	99.5	97.8	97.4	98.5	95.3	96.0
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\underline{\alpha}=10\%$								
500	96.2	95.0	96.4	86.0	82.8	86.2	74.1	72.2	74.8
1000	100.0	100.0	100.0	99.6	99.5	99.5	99.3	98.0	99.1
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	INAR(3) vs INAR(1)								
	$\alpha_{01}=0.4,\ \alpha_{02}=0.2,\ \alpha_{03}=0.1$								
	m=4			m=12			m=20		
n	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$	$\mathcal{P}(2)$	$\mathcal{G}(0.5)$	$\mathcal{NB}(2,0.5)$
	$\underline{\alpha}=1\%$								
500	97.7	97.0	96.2	88.2	85.2	84.1	75.5	67.1	70.2
1000	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.7	100.0
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\underline{\alpha}=5\%$								
500	99.7	99.2	98.8	97.3	95.5	96.2	92.4	89.3	88.9
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\underline{\alpha}=10\%$								
500	99.9	99.7	99.2	98.5	97.6	97.4	96.0	94.4	93.8
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 3.5 – Size of test for the log-linear$(1,1)$ model (in %)

	log-linear$(1,1)$								
	$\omega_0=2,\ \alpha_0=0.3,\ \beta_0=-0.6$								
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$
	$\underline{\alpha}=1\%$								
500	1.6	0.9	0.9	1.6	1.4	1.2	1.4	0.7	0.4
1000	1.3	1.2	0.6	1.3	0.9	1.0	1.1	0.9	1.0
4000	0.9	1.2	1.1	1.1	1.2	1.2	1.0	1.2	0.9
	$\underline{\alpha}=5\%$								
500	7.1	4.4	6.2	5.9	5.3	6.2	6.8	6.1	3.8
1000	6.2	4.6	4.0	6.0	4.8	5.4	5.9	5.7	4.2
4000	4.4	5.4	4.8	4.6	5.1	5.1	5.2	4.9	4.5
	$\underline{\alpha}=10\%$								
500	13.1	9.8	11.2	12.2	11.1	12.0	12.8	8.9	8.4
1000	10.7	10.1	9.6	10.7	10.9	10.4	9.3	10.6	10.1
4000	9.4	9.9	11.1	9.2	10.1	10.0	10.2	10.4	10.6

Table 3.6 – Power of test for the log-linear$(1,1)$ model (in %)

	log-linear$(1,1)$ vs INGARCH$(1,1)$ $\omega_0 = 2,\ \alpha_0 = 0.3,\ \beta_0 = -0.6$								
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$
				$\underline{\alpha} = 1\%$					
500	99.7	97.5	98.8	98.7	97.1	96.9	99.4	95.5	94.2
1000	100.0	100.0	100.0	100.0	99.8	100.0	100.0	99.6	99.8
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
				$\underline{\alpha} = 5\%$					
500	99.8	98.6	99.4	99.1	98.9	98.9	99.7	97.6	97.6
1000	100.0	100.0	100.0	100.0	99.8	100.0	99.9	99.8	100.0
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
				$\underline{\alpha} = 10\%$					
500	99.8	98.9	98.6	99.5	99.6	99.1	99.7	98.0	97.5
1000	100.0	100.0	100.0	100.0	99.9	100.0	100.0	99.9	99.5
4000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 3.7 – Size of test for the non-linear$(1,1)$ model (in %)

	Non-linear$(1,1)$ $\omega_0 = 2,\ \alpha_0 = 0.3,\ \beta_0 = -0.6$								
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$
				$\underline{\alpha} = 1\%$					
500	1.2	0.5	0.7	0.7	0.5	0.5	0.7	0.4	0.4
1000	0.8	0.9	1.0	0.7	0.7	0.5	0.6	0.6	0.7
4000	1.0	0.9	1.1	0.8	0.9	0.9	0.8	0.9	0.8
				$\underline{\alpha} = 5\%$					
500	5.0	3.5	4.0	4.0	3.1	4.1	2.5	2.5	2.2
1000	4.7	4.8	4.6	4.2	3.9	4.4	4.1	3.6	3.9
4000	5.1	5.6	4.9	4.5	4.8	4.6	4.2	4.4	4.6
				$\underline{\alpha} = 10\%$					
500	8.4	8.5	8.3	8.2	7.3	8.8	6.8	6.8	6.2
1000	6.9	8.9	8.9	8.7	7.8	9.1	9.6	8.0	7.4
4000	10.4	9.2	9.4	8.9	9.3	9.4	9.9	9.2	9.1

TABLE 3.8 – Power of test for the non-linear$(1,1)$ model (in %)

	Non-linear$(1,1)$ vs INGARCH$(1,1)$ $\omega_0 = 2,\ \alpha_0 = 0.3,\ \beta_0 = -0.6$								
	m=4			m=12			m=20		
n	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$	$\mathcal{P}(\lambda_t)$	$\mathcal{NB}(p_t,6)$	$\mathcal{DP}(\lambda_t,2)$
				$\underline{\alpha} = 1\%$					
500	53.1	10.6	80.9	30.0	3.0	55.3	17.9	2.3	32.1
1000	92.1	24.4	99.7	78.8	14.0	98.0	66.2	7.9	80.2
4000	100.0	93.9	100.0	100.0	82.2	100.0	100.0	68.5	92.1
				$\underline{\alpha} = 5\%$					
500	75.9	29.3	93.1	55.4	13.1	81.4	43.7	9.4	60.8
1000	96.7	53.5	100.0	92.5	34.7	98.0	85.7	24.1	99.3
4000	100.0	98.6	100.0	100.0	93.9	100.0	100.0	89.6	100.0
				$\underline{\alpha} = 10\%$					
500	83.8	43.3	96.5	71.0	23.4	90.7	57.8	17.6	74.9
1000	98.4	67.8	100.0	96.9	48.1	99.7	92.7	37.0	94.0
4000	100.0	99.6	100.0	100.0	96.5	100.0	100.0	88.2	100.0

and their corresponding asymptotic standard errors computed by using the robust sandwich matrix $\widehat{\Sigma}$ are given in the rows 'ASE'. Moreover, the rows 'T.Wald' present the p-values of the test of the nullity of the corresponding coefficient proposed in Remark 2.4 of Ahmad & Francq (2016) : the testing problem is

$$H_0 : \theta_{0i} = 0 \quad \text{against} \quad H_1 : \theta_{0i} > 0.$$

and its rejection region is given by

$$\frac{n\widehat{\theta}_{ni}^2}{\widehat{\Sigma}(i,i)} \geq \chi_1^2(1-2\underline{\alpha}), \tag{3.5.1}$$

where θ_{0i} is the ith component of θ_0 and $\underline{\alpha}$ is the asymptotic level. We used the function nlminb() of R to minimize the opposite of the quasi log-likelihood function. The values of minimized function are provided in the rows '-Log-Like'. The rows 'MSE' give the values of the mean square error (MSE) of the Pearson residuals defined by

$$MSE = \sum_{t=1}^{n} \frac{\tilde{e}_t^2(\widehat{\theta}_n)}{n-d}, \quad \text{where } \tilde{e}_t(\widehat{\theta}_n) = \frac{X_t - \tilde{\lambda}_t(\widehat{\theta}_n)}{\sqrt{\tilde{\lambda}_t(\widehat{\theta}_n)}},$$

and d is the number of parameters. The last rows 'T.Portmanteau' contain the p-values of the goodness-fit-test proposed in Theorem 3.3.1. The number of autocovariance functions taken into account in the test is $m = \{4, 12, 20\}$. In view of Table 3.9, we note that the estimated parameters $\widehat{\alpha}_2$ and $\widehat{\alpha}_3$ are not statistically significant. In addition, the p-values of the portmanteau test for different values of m, indicate that there are significant residual autocorrelations. Thus, the INGARCH$(1,3)$ is not adequate for modelling the series. We suppressed the two non significant parameters and repeated the estimation for fitting an INGARCH$(1,1)$ model. Table 3.10 shows that all the parameters are significant and the p-values of the portmanteau test refer to absence of residual autocorrelation. Table 3.11 shows the results for the log-linear$(1,1)$ model. We can also note that all the parameters are significant and the p-values of the portmanteau test indicate that there are no significant residual autocorrelations. The results of the non-linear$(1,1)$ model which are presented in Table 3.12 show that the parameter $\widehat{c}$ is not significant. Thus, the non linearity is rejected for this series. Note that the test (3.5.1) can be used for the parameter c_0 in the model (3.2.13) because it is also positively constrained. By comparison between the INGARCH(1,1) and the log-linear$(1,1)$ models, we find that the value of "-Log-Like" in log-linear(1,1) is smaller than that of INGARCH(1,1) model, but on the other hand, the value of MSE of INGARCH$(1,1)$ model is smaller than that of log-linear$(1,1)$ model. Therefore, we can consider that the two models are adequate for modelling the series. In fact, when the data are positively correlated (see Figure 3.1), then models INGARCH(1,1) and log-linear$(1,1)$ will yield similar conclusions. It is anticipated that the log-linear model provides a better fit when either there exists negative correlation among the data or when covariates need to be taken into account for the data analysis. Finally, it is interesting to note that for both INGARCH(1,1) and log-linear(1,1) models, the sum of the estimated values of the α and β coefficients is close to 0.9, which indicates a strong persistence in the dynamics. This is in accordance with the clusters of high values that are observed on the series plotted in Figure 3.1.

3.6 Proofs

Proof of Theorem 3.3.1

We denote by $\gamma_m(\theta_0) = (\gamma_{(1)}, .., \gamma_{(m)})'$ the vector of autocovariances, where

$$\gamma_{(h)} = \frac{1}{n} \sum_{t=h+1}^{n} \epsilon_t(\theta_0)\epsilon_{t-h}(\theta_0) \quad \text{for} \quad h \in \{1, .., m\}.$$

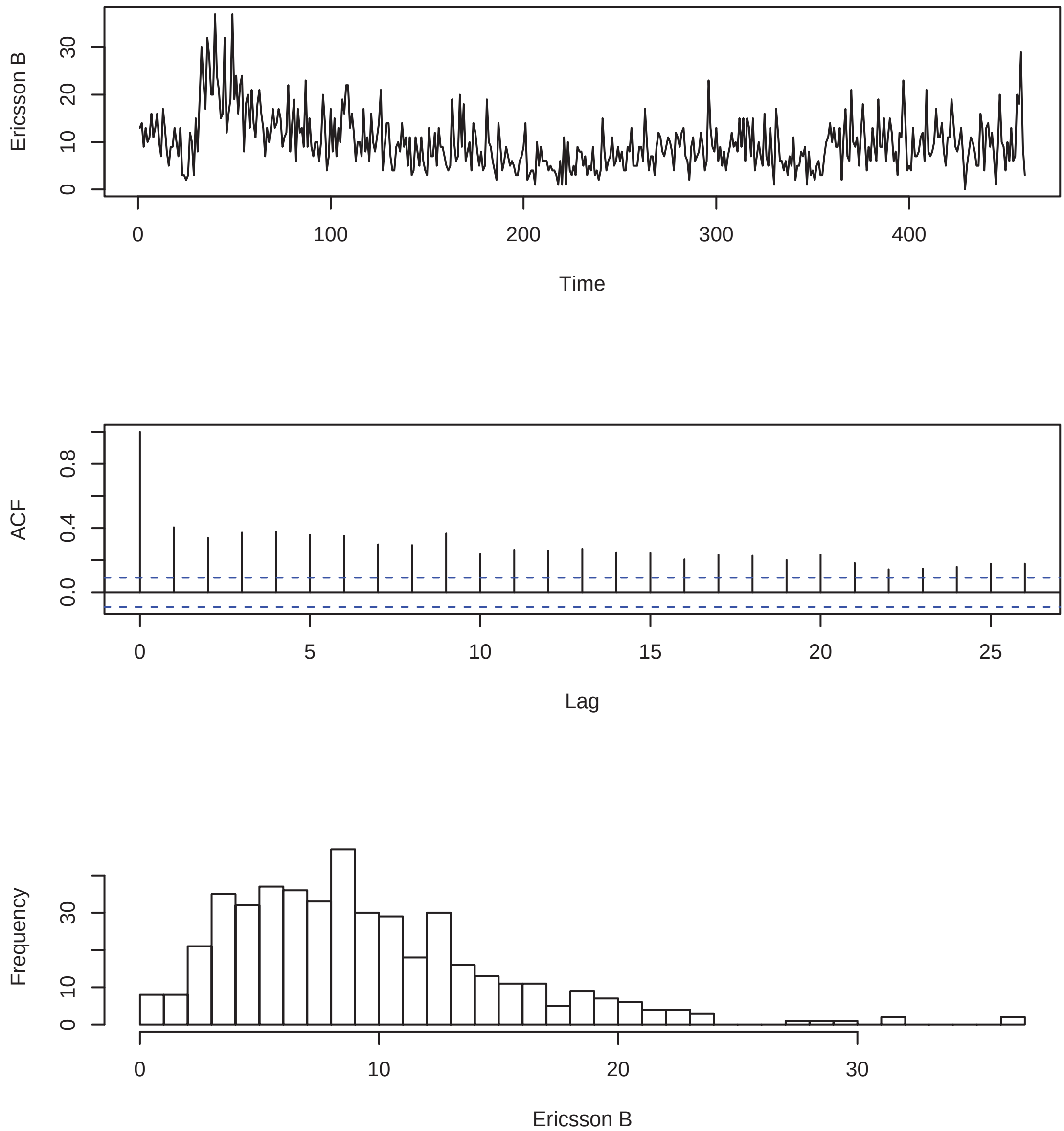

Figure 3.1 – The time series plot, the sample autocorrelation function and the histogram of the data

Table 3.9 – INGARCH$(1,3)$

$\widehat{\theta}$	$\widehat{\omega}$	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$	$\widehat{\beta}_1$
	0.344	0.129	0.000	0.028	0.809
ASE	0.170	0.051	0.064	0.040	0.052
T.Wald		0.011	1	0.484	0.000
$-loglike$	-13.108				
MSE	2.342				
m	4	12	20		
T.Portmanteau	0.000	0.000	0.002		

Table 3.10 – INGARCH$(1,1)$

$\widehat{\theta}$	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$
	0.291	0.139	0.832
ASE	0.140	0.034	0.041
T.Wald		0.000	0.000
$-loglike$	-13.107		
MSE	2.336		
m	4	12	20
T.Portmanteau	0.426	0.165	0.444

Table 3.11 – Log-linear$(1,1)$

$\widehat{\theta}$	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$
	0.105	0.208	0.746
ASE	0.051	0.040	0.050
T.Wald		0.000	0.000
$-loglike$	-13.170		
MSE	2.388		
m	4	12	20
T.Portmanteau	0.598	0.374	0.510

Table 3.12 – Non-linear$(1,1)$

$\widehat{\theta}$	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{c}$
	1.350	0.154	0.830	1.036
ASE	1.709	0.038	0.041	0.904
T.Wald		0.000	0.000	0.252
$-loglike$	-13.229			
MSE	2.332			
m	4	12	20	
T.Portmanteau	0.482	0.199	0.474	

Let $\widetilde{\gamma}_m(\theta_0)$ another vector defined as $\gamma_m(\theta_0)$ except that $\epsilon_t(\theta_0)$ is replaced by $\widetilde{\epsilon}_t(\theta_0) = X_t - \widetilde{\lambda}_t(\theta_0)$.

We now illustrate the asymptotic impact of the initial values on the statistic $\widehat{\gamma}_m$. We have $\epsilon_t(\theta_0)\epsilon_{t-h}(\theta_0) - \widetilde{\epsilon}_t(\theta_0)\widetilde{\epsilon}_{t-h}(\theta_0) = g_t + f_t$, where $g_t = (\widetilde{\epsilon}_t(\theta_0) - \epsilon_t(\theta_0))\epsilon_{t-h}(\theta_0)$ and $f_t = (\widetilde{\epsilon}_{t-h}(\theta_0) - \epsilon_{t-h}(\theta_0))\widetilde{\epsilon}_t(\theta_0)$. By using **A4** and **A5**, the sequences g_t and f_t tend to zero as $n \to \infty$. Note that for the linear specification (3.2.9), we can show that **A5** is verified by using the same arguments used to show (7.30) in Francq & Zakoian (2010) and the arguments of Remark 2.1 in Ahmad & Francq (2016), where we have $a_t \leq Kp^t$, for any $K > 0$ and $p \in (0,1)$. The conclusion then follows from the Borel-Cantelli lemma and the fact that $P(Kp^t\lambda_t(\theta_0) \geq r) \leq Kp^tE\lambda_t(\theta_0)/r$ and $P(Kp^t\widetilde{\lambda}_t(\theta_0) \geq r) \leq Kp^tE\widetilde{\lambda}_t(\theta_0)/r$, for some $r > 0$. Similarly, by using the assumption **A7**, we can show that the impact of the initial values on the first order derivative of $\widehat{\gamma}_m$ is asymptotically negligible. We thus have

$$\|\gamma_m(\theta_0) - \widetilde{\gamma}_m(\theta_0)\| = o_p(1) \quad \text{and} \quad \left\|\frac{\partial\gamma_m(\theta_0)}{\partial\theta'} - \frac{\partial\widetilde{\gamma}_m(\theta_0)}{\partial\theta'}\right\| = o_p(1). \tag{3.6.1}$$

Now, we derive the asymptotic distribution of $\widehat{\gamma}_m$. Let $\epsilon_{t-1:t-m}(\theta_0) = X_{t-1:t-m}(\theta_0) - \lambda_{t-1:t-m}(\theta_0) = (\epsilon_{t-1}(\theta_0), .., \epsilon_{t-m}(\theta_0))'$. The second order derivative of $\gamma_m(\theta_0)$ at θ, for any $\theta \in \Theta$ and $(i,j) \in \{1,..,d\}$, is given by :

$$\frac{\partial^2\gamma_m(\theta_0)}{\partial\theta_i\theta_j} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\lambda_{t-1:t-m}(\theta_0)}{\partial\theta_i}\frac{\partial\lambda_t(\theta_0)}{\partial\theta_j} - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\lambda(\theta_0)}{\partial\theta_i\theta_j}\epsilon_{t-1:t-m}(\theta_0). \tag{3.6.2}$$

There exists a neighborhood $V(\theta_0)$ of θ_0 such that

$$\lim_{n\to\infty} E\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2\gamma_m(\theta)}{\partial\theta_i\theta_j}\right\| \leq E\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial\lambda_{t-1:t-m}(\theta_0)}{\partial\theta_i}\frac{\partial\lambda_t(\theta_0)}{\partial\theta_j}\right\| + E\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2\lambda(\theta)}{\partial\theta_i\theta_j}\epsilon_{t-1:t-m}\right\|. \tag{3.6.3}$$

The right-hand side of the inequality (3.6.3) is bounded by using **A10**. For the linear model (3.2.9), **A10** can be shown using the arguments used to show (7.40) in Francq & Zakoian (2010). For θ^* between $\widehat{\theta}_n$ and θ_0, the Taylor expansion of $\widehat{\gamma}_m$ around θ_0 yields

$$\sqrt{n}\widehat{\gamma}_m = \sqrt{n}\widetilde{\gamma}_m(\theta_0) + \frac{\partial\widetilde{\gamma}_m(\theta^*)}{\partial\theta'}\sqrt{n}(\widehat{\theta}_n - \theta_0).$$

By using (3.6.1), we obtain

$$\sqrt{n}\widehat{\gamma}_m = \sqrt{n}\gamma_m(\theta_0) + \frac{\partial\gamma_m(\theta^*)}{\partial\theta'}\sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1) \tag{3.6.4}$$

By the ergodic theorem, the strong consistency of PQMLE and the fact that $E\left(\frac{\partial\lambda_{t-h}(\theta_0)}{\partial\theta}\epsilon_t(\theta_0)\right) = 0$, one can show that

$$\frac{\partial\gamma_m(\theta^*)}{\partial\theta'} = \frac{\partial\gamma_m(\theta_0)}{\partial\theta'} + o_p(1) \to C_m = (c_{(1)}, .., c_{(m)})' \text{ in probability as } n \to \infty,$$

where

$$c_{(h)} = -\frac{1}{n}\sum_{t=h+1}^{n}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\epsilon_{t-h}(\theta_0).$$

Thus, we can rewrite (3.6.4) as follows

$$\sqrt{n}\widehat{\gamma}_m = \sqrt{n}\gamma_m + C_m\sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1). \tag{3.6.5}$$

We also have

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = J^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\epsilon_t(\theta_0)}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta} + o_p(1) \tag{3.6.6}$$

Using (3.6.6), the central limit theorem of Billingsley (1961) applied to the martingale difference $\left\{\left(\frac{\partial\ell_t(\theta_0)}{\theta}, \epsilon_t(\theta_0)\epsilon_{t-1:m}(\theta_0)\right)', \sigma(\epsilon_u, u < t)\right\}$, shows that

$$\begin{aligned}\sqrt{n}\begin{pmatrix}\widehat{\theta}_n - \theta_0\\ \gamma_m(\theta_0)\end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\theta_0)\begin{pmatrix}J^{-1}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{1}{\lambda_t(\theta_0)}\\ \epsilon_{t-1:t-m}(\theta_0)\end{pmatrix}\\ &\xrightarrow{\mathcal{L}} N\left(0, \begin{pmatrix}\Sigma & \Sigma_{\theta,\gamma_m}\\ \Sigma'_{\theta,\gamma_m} & \Sigma_{\gamma_m}\end{pmatrix}\right),\end{aligned} \tag{3.6.7}$$

where Σ is given in Section 3.2, the matrix Σ_{γ_m} is given by

$$\Sigma_{\gamma_m} = Cov(\sqrt{n}\gamma_m(\theta_0), \sqrt{n}\gamma'_m(\theta_0)).$$

Under the moment assumption (3.2.6) and the fact that $\epsilon_t(\theta_0)$ is difference martingale sequence, $Cov(\sqrt{n}\gamma_m, \sqrt{n}\gamma'_m)$ is the $m \times m$ matrix whose elements, for $(h, l) \in \{1, .., m\}$, are given by

$$\Sigma_{\gamma_m}(h, l) = E\left(\epsilon_t^2(\theta_0)\epsilon_{t-h}(\theta_0)\epsilon_{t-l}(\theta_0)\right)$$

and $\Sigma_{\theta_n,\gamma_m}$ is the $d \times m$ matrix defined as follows

$$\begin{aligned}\Sigma_{\theta,\gamma_m} &= J^{-1}Cov\left(\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n}\frac{\epsilon_t(\theta_0)}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\right), \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\theta_0)\epsilon'_{t-1:t-m}(\theta_0)\right)\\ &= J^{-1}E\left(\frac{\epsilon_t^2(\theta_0)}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\epsilon'_{t-1:t-m}(\theta_0)\right),\end{aligned}$$

because, for $h \in \{1, .., m\}$, we have

$$Cov\left(\frac{\epsilon_t(\theta_0)}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}, \epsilon_s(\theta_0)\epsilon_{s-h}(\theta_0)\right) = \begin{cases} E\left(\frac{\epsilon_t^2(\theta_0)}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\epsilon_{t-h}(\theta_0)\right), & \text{when } s = t \\ 0, & \text{otherwise.} \end{cases}$$

Therefore

$$\begin{aligned} &Cov\left(\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n}\frac{\epsilon_t(\theta_0)}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\theta}\right), \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\theta_0)\epsilon_{t-h}(\theta_0)\right) \\ &= \frac{1}{n}\sum_{t,s}^{n} Cov\left(\frac{\epsilon_t(\theta_0)}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}, \epsilon_s(\theta_0)\epsilon_{s-h}(\theta_0)\right) \\ &= E\left(\frac{\epsilon_t^2(\theta_0)}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\epsilon_{t-h}(\theta_0)\right). \end{aligned}$$

Using (3.6.5), we have

$$\begin{aligned} Cov(\sqrt{n}\widehat{\gamma}_m, \sqrt{n}\widehat{\gamma}_m') = Cov(\sqrt{n}\gamma_m, \sqrt{n}\gamma_m') + C_m\Sigma C_m' + C_m Cov(\sqrt{n}(\widehat{\theta}_n - \theta_0), \sqrt{n}\gamma_m') \\ + Cov(\sqrt{n}\gamma_m, \sqrt{n}(\widehat{\theta}_n - \theta_0)')C_m' + o_p(1). \quad (3.6.8) \end{aligned}$$

Thus, the asymptotic distribution of $\sqrt{n}\widehat{\gamma}_m$ is given by

$$\sqrt{n}\widehat{\gamma}_m \to N(0, \Sigma_{\widehat{\gamma}_m}), \text{ where } \quad \Sigma_{\widehat{\gamma}_m} = \Sigma_{\gamma_m} + C_m\Sigma C_m' + C_m\Sigma_{\theta,\gamma_m} + \Sigma_{\theta,\gamma_m}' C_m'.$$

Finally, using **A4**, **A5**, **A7** and the fact that $\widehat{\theta}_n \to \theta_0$ almost surely as $n \to \infty$, by the ergodic theorem (see Billingsley, 2008), it can be shown that $\widehat{\Sigma}_{\widehat{\gamma}_m} \to \Sigma_{\widehat{\gamma}_m}$ almost surely as $n \to \infty$ and the conclusion follows from (3.3.1). □

Verification of the condition (3.3.1) for the linear and the log-linear models

For the linear models

To show the invertibility of the matrix $\Sigma_{\widehat{\gamma}_m}$ for the linear models (*i.e* the INGARCH(p, q) and INAR(p) models), we define the vector

$$V_t = \epsilon_t(\theta_0)\epsilon_{t-1:t-m}(\theta_0) + C_m J^{-1}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\epsilon_t(\theta_0)}{\lambda_t(\theta_0)}.$$

The matrix $\Sigma_{\widehat{\gamma}_m} = EV_tV_t'$ is singular if there exists a non null vector $\eta = \{\eta_1, ..., \eta_m\}'$, so that $\eta' V_t = 0$. We assume that

$$\eta' V_t = \epsilon_t(\theta_0)\left(\eta'\left(\epsilon_{t-1:t-m}(\theta_0) + C_m J^{-1}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{1}{\lambda_t(\theta_0)}\right)\right) = 0. \quad (3.6.9)$$

By squaring $\eta' V_t$ and taking the conditional expectation with respect to $\mathcal{F}_{t-1}$, we obtain

$$E((\eta' V_t)^2 \mid \mathcal{F}_{t-1}) = E(\epsilon_t(\theta_0)^2 \mid \mathcal{F}_{t-1}) \left(\eta' \left(\epsilon_{t-1:t-m}(\theta_0) + C_m J^{-1} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{1}{\lambda_t(\theta_0)} \right) \right)^2 = 0. \tag{3.6.10}$$

Note that, under the assumption **A8**, we have $E(\epsilon_t(\theta_0)^2 \mid \mathcal{F}_{t-1}) = Var(X_t \mid \mathcal{F}_{t-1}) > 0$. Thus, (3.6.10) implies that

$$\eta' \epsilon_{t-1:t-m}(\theta_0) + \zeta \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{1}{\lambda_t(\theta_0)} = 0, \tag{3.6.11}$$

with $\zeta = \eta' C_m J^{-1}$. Note that $\zeta = (\zeta_1, .., \zeta_{p+q+1}) \neq 0$. Otherwise, $\eta' \epsilon_{t-1:t-m}(\theta_0) = 0$ $a.s$, which implies that there exists $h \in \{1, .., m\}$ such that ϵ_{t-h} is measurable with respect to $\sigma\{t, t \neq h\}$. This is impossible because ϵ_t is a non degenerated difference martingale sequence. Let R_{t-2} any random variable measurable with respect to $\sigma\{\epsilon_u, u \leq t\}$. We have

$$\begin{aligned} (\alpha_1 X_{t-1} + R_{t-2})(\eta_1 \epsilon_{t-1} + R_{t-2}) + \zeta_2 X_{t-1} + R_{t-2} = 0, \\ \alpha_1 \eta_1 X_{t-1}^2 + X_{t-1} R_{t-2} + R_{t-2} = 0. \end{aligned} \tag{3.6.12}$$

By solving the quadratic equation (3.6.12), we have either $X_{t-1} = R_{t-2}$ which is impossible because, under the assumption (3.2.4), the conditional distribution of X_{t-1} given its past has at least three different values, or $\alpha_1 \eta_1 = 0$ which is also impossible because for the INGARCH(p,1) model we have $\alpha_1 > 0$. Let $\eta'_{2:m} = \{\eta_2, .., \eta_m\}$, then from (3.6.11), we have

$$(\alpha_1 X_{t-1} + R_{t-2})(\eta'_{2:m}) + \zeta_2 X_{t-1} + R_{t-2} = 0, \tag{3.6.13}$$

This equation entails that $X_{t-1} = R_{t-2}$. Therefore, the matrix $\Sigma_{\widehat{\gamma}_m}$ is not singular for the case INGARCH(p,1). For the general case, we show the invertibility of $\Sigma_{\widehat{\gamma}_m}$ by showing that (3.6.11) entails $\alpha_1 = .. = \alpha_q = 0$.

For the log-linear$(1, 1)$ models

For this model, the equation (3.6.11) becomes

$$\eta' \epsilon_{t-1:t-m}(\theta_0) + \zeta \frac{\partial v_t(\theta_0)}{\partial \theta} = 0. \tag{3.6.14}$$

The equation (3.6.14) entails that

$$R_{t-2} = \eta_1 X_{t-1} + \zeta_2 log(X_{t-1} + 1), \tag{3.6.15}$$

under (3.2.4), this equation is impossible by argements already given, except when $\eta_1 = \zeta_2 = 0$. Now, assume that $\eta_1 = \zeta_2 = 0$. Thus, we have

$$R_{t-3} = \eta_2 X_{t-2}, \tag{3.6.16}$$

which is also impossible under (3.2.4). Similarly, we can show that the other components of the vector η are null. The matrix $\Sigma_{\widehat{\gamma}_m}$ is thus invertible for the log-linear models. □

3.7 Conclusion

The test introduced in this chapter represents a general and reliable diagnostic tool for evaluating the goodness-of-fit for the conditional mean count time series models. This test helps to assess the adequacy of fit of a chosen model and to facilitate comparison between two (or more) model specifications. A similar methodology might be adopted for evaluating the adequacy of fit for multivariate count time series models (*e.g* Franke & Subba Rao, 1993, Heinen & Rengifo, 2007 and Pedeli & Karlis, 2011). However, this extension deserves more thought and is left for future work.

References

Ahmad, A. & Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis*, 37, 291–314.

Alzaid, A. & Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 27, 314–324.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12, 788–792.

Billingsley, P. (2008). *Probability and measure.* New York : John Wiley & Sons.

Box, G. E. & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65, 1509–1526.

Carbon, M. & Francq, C. (2010). Diagnostic checking in ARMA models with uncorrelated errors. *Austrian Journal of Statistics*, 40, 55–64.

Christou, V. (2014). *Statistical theory for mixed poisson time series models*. PhD thesis, University of Cyprus.

Christou, V. & Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35, 55–78.

Douc, R., Doukhan, P., & Moulines, E. (2013). Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Processes and their Applications*, 123, 2620–2647.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81, 709–721.

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27, 923–942.

Fokianos, K. & Neumann, M. H. (2013). A goodness-of-fit test for poisson count processes. *Electronic Journal of Statistics*, 7, 793–819.

Fokianos, K. & Tjøstheim, D. (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102, 563–578.

Fokianos, K. & Tjøstheim, D. (2012). Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64, 1205–1225.

Francq, C., Roy, R., & Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association*, 100, 532–544.

Francq, C. & Zakoian, J.-M. (2010). *GARCH models : structure, statistical inference and financial applications*. Chichester : John Wiley & Sons.

Franke, J. & Subba Rao, T. (1993). *Multivariate first order integer valued autoregressions*. Technical report, University of Kaiserslautern, Department of Mathematics.

Gao, J., King, M., Lu, Z., Tjøstheim, D., et al. (2009). Specification testing in nonlinear and nonstationary time series autoregression. *The Annals of Statistics*, 37, 3893–3928.

Gonçalves, E., Mendes-Lopes, N., & Silva, F. (2015). Infinitely divisible distributions in integer-valued Garch models. *Journal of Time Series Analysis*, 36, 503–527.

Heinen, A. (2003). Modelling time series count data : an autoregressive conditional poisson model. *Discussion Paper, vol. 2003/62. Center for operations research and econometrics (CORE), Catholic University of Louvain, Belgium.*

Heinen, A. & Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14, 564–583.

Hudecová, Š., Hušková, M., & Meintanis, S. G. (2015). Tests for time series of counts based on the probability-generating function. *Statistics*, 49, 316–337.

Ljung, G. M. & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303.

Meintanis, S. G. & Karlis, D. (2014). Validation tests for the innovation distribution in INAR time series models. *Computational Statistics*, 29, 1221–1241.

Neumann, M. H. (2011). Absolute regularity and ergodicity of poisson count processes. *Bernoulli*, 17, 1268–1284.

Pedeli, X. & Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical modelling*, 11, 325–349.

Schweer, S. (2016). A goodness-of-fit test for integer-valued autoregressive processes. *Journal of Time Series Analysis*, 37, 77–98.

Silva, I. M. M. (2005). *Contributions to the analysis of discrete-valued time series.* PhD thesis, University of Porto.

Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications*, 389, 58–71.

Zhu, F. & Wang, D. (2010). Diagnostic checking integer-valued ARCH(p) models using conditional residual autocorrelations. *Computational Statistics & Data Analysis*, 54, 496–508.

Appendix

For the consistency of PQMLE we assume that

A1 We have $\theta_0 \in \Theta$ where Θ is compact.

A2 The process X_t is stationary and ergodic.

A3 $\lambda_1(\theta) = \lambda_1(\theta_0)$ almost surely if and only if $\theta = \theta_0$

The next assumption is used to demonstrate that the initial values have no effect on the asymptotic properties of PQMLE.

A4 We have a.s. $\lim_{t\to\infty} a_t = 0$ and $\lim_{t\to\infty} X_t a_t = 0$, where $a_t = \sup_{\theta\in\Theta} \left|\widetilde{\lambda}_t(\theta) - \lambda_t(\theta)\right|$.

To show that the impact of the initial values on the statistic of $\widehat{\gamma}_m$ is asymptotically negligible, we need the next two additional convergences

A5 We have a.s. $\lim_{t\to\infty} \lambda_t(\theta_0)a_{t-h} = 0$ and $\lim_{t\to\infty} \widetilde{\lambda}_t(\theta_0)a_t = 0$.

For the asymptotic normality, we need the following assumptions

A6 $\theta_0 \in \mathring{\Theta}$, where $\mathring{\Theta}$ denotes the interior of Θ.

The next assumption is introduced to handle initial values.

A7 We have a.s. b_t, b_tX_t, $a_td_tX_t$ and $b_t\epsilon_{t-h}(\theta_0)$ are of order $O(t^{-\kappa})$ for some $\kappa > 1/2$, where

$$b_t = \sup_{\theta\in\Theta} \left\| \frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta} - \frac{\partial\lambda_t(\theta)}{\partial\theta} \right\|.$$

$$d_t = \sup_{\theta\in\Theta} \max\left\{ \left\| \frac{1}{\lambda_t(\theta)} \frac{\partial\lambda_t(\theta)}{\partial\theta} \right\|, \left\| \frac{1}{\widetilde{\lambda}_t(\theta)} \frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta} \right\| \right\}$$

Moreover, $\lim_{t\to\infty} \frac{\partial\widetilde{\lambda}_t(\theta_0)}{\partial\theta} a_{t-h} = 0$.

A8 The conditional variance of X_t given its past exists and

$$E \mid Var\left(X_t \mid \mathcal{F}_{t-1}\right) \mid^{1+\varepsilon} < \infty, \text{ for some } \varepsilon > 0.$$

A9 $\lambda_t(.)$, $\widetilde{\lambda}_t(.)$ admit continuous second-order derivatives, the matrices J and I exist and J is invertible.

A10 For any $\theta \in \Theta$, $(i, j) \in \{1, .., d\}$ and $h \in \{1, .., m\}$, there exists a neighborhood $V(\theta_0)$ of θ_0 such that

$$E \sup_{\theta\in V(\theta_0)} \left\| \frac{\partial\lambda_{t-h}(\theta_0)}{\partial\theta_i} \frac{\partial\lambda_t(\theta_0)}{\partial\theta_j} \right\| < \infty \text{ and } E \sup_{\theta\in V(\theta_0)} \left\| \frac{\partial^2\lambda(\theta)}{\partial\theta_i\theta_j} \epsilon_{t-h} \right\| < \infty.$$

Chapitre 4

Poisson QMLE of multivariate count time series models

Abstract. The Poisson quasi maximum likelihood estimator (PQMLE) is proposed to estimate the conditional mean parameters of a multivariate time series of counts equation-by-equation. The asymptotic behaviour of the estimator is studied and the main results are applied to particular multivariate count time series models, as the multivariate INAR and the multivariate INGARCH models. Numerical illustrations via Monte Carlo simulations and financial real data application are provided.

Keywords. Multivariate time series of counts, Poisson quasi maximum likelihood, Consistency and asymptotic normality, Copulas, Conditional mean.

4.1 Introduction

The multivariate time series of counts are found in a lot of applications in many scientific fields (see *e.g.* Boudreault & Charpentier, 2011 and Liu, 2012). Several models are proposed to deal with this kind of data, for example, the conditional mean models in which the means are assumed to follow, conditionally on past observations, a vector autoregressive model (see *e.g.* Heinen & Rengifo, 2007 and Liu, 2012) and the multivariate INAR models (see *e.g.* Franke & Subba Rao, 1993, Pedeli & Karlis, 2013 and Karlis & Pedeli, 2013). The study and the analysis of multivariate count time series pose several problems and questions. For instance, an important issue associated to the multivariate conditional mean models is to find an appropriate conditional distribution for modelling the negative contemporaneous correlation between the series. A similar difficulty is observed for the multivariate INAR models since the distribution of the innovations must accommodate negative correlation to produce negative correlation between the series. Indeed, the multivariate distributions with integer-valued supports, which are able to accommodate the negative contemporaneous correlation between the series, are not abundantly available in the literature. One of those distributions is the multivariate Poisson Log-Normal distribution of Aitchison & Ho (1989), but it is difficult to be used in the conditional mean models. This problem has been solved by using copulas (see *e.g.* Heinen & Rengifo, 2007 and Karlis & Pedeli, 2013). Nevertheless, copula theory still has limitations with discrete distributions and requires sometimes to transform the original count data into continuous using a continued extension. Another related difficulty is that the MLE of certain models, as the multivariate INAR models is often challenging. For high order multivariate INAR models, the MLE function becomes very complicated and difficult to manipulate. In this chapter, we propose a general and simple methodology to estimate such models. We extend the Poisson quasi maximum likelihood estimator (PQMLE) studied by Ahmad & Francq (2016) to estimate the conditional mean parameters of a multivariate time series of counts, whatever the contemporaneous correlation between the series (positive, negative or absence of correlation) and whatever the nature of dispersion of the series (under-dispersed, over-dispersed or equi-dispersed) [1]. In this estimator, we only need to specify the conditional means of the series. The multivariate models will be estimated equation-by-equation (EbE). We give general regularity conditions under which the EbE-PQMLE is consistent and asymptotically normal (CAN). The reader is referred to Francq & Zakoian (2016) for more details on EbE estimation of multivariate GARCH models using QMLE.

The chapter is organized as follows. Section 4.2 contains the general formulation of the model and the main results concerning the asymptotic behaviour of the EbE-PQMLE. Section 4.3 applies the general results to particular linear models. Section 4.4 studies the finite sample properties of the EbE-PQMLE via a set of Monte Carlo experiments. In Section 4.5, we use the EbE-PQMLE to fit multivariate INGARCH(1, 1) models for real series. The proofs are collected in Section 4.6, and Section 4.7 concludes.

4.2 Model and main results

Assume that X_t is a k-dimensional vector of time series of counts valued in $\mathbb{N}^k$. Assume also that the conditional mean of the dth component $(X_{d,t})$ of this vector given its past is defined as follows

$$E\left(X_{d,t} \mid \mathcal{F}_{t-1}\right) = \lambda_d\left(X_{t-1}, X_{t-2}, \ldots; \theta_{0,d}\right) = \lambda_{d,t}(\theta_{0,d}), \qquad (4.2.1)$$

where $\mathcal{F}_{t-1}$ denotes the $\sigma-$field generated by $(X_u, u < t)$, $\theta_{0,d}$ is an unknown parameter belonging to some parameter space Θ_d and, for all $(x_u,\ u < t) \in \mathbb{N}^k$ and all $\theta_d \in \Theta_d$,

$$\lambda_d(x_{t-1}, x_{t-2}, \ldots; \theta_d) \text{ is valued in } (\underline{\omega}, +\infty) \text{ for some } \underline{\omega} > 0. \qquad (4.2.2)$$

We assume also that the marginal distribution has a moment slightly greater than 1

$$EX_{d,t}^{1+\varepsilon} < \infty, \text{ for some } \varepsilon > 0, \text{ and for } \ d = 1, .., k, \qquad (4.2.3)$$

which entails the existence of the conditional mean. For all $\theta_d \in \Theta_d$ and $t \geq 1$, let

$$\widetilde{\lambda}_{d,t}(\theta_d) = \lambda_d(X_{t-1}, X_{t-1}, .., X_1, \widetilde{X}_0, \widetilde{X}_{-1}, ..; \theta_d) \text{ and } \lambda_{d,t}(\theta_d) = \lambda_d(X_{t-1}, X_{t-1}, ..; \theta_d),$$

1. The unconditional variance is smaller than, greater than or equal to the unconditional mean.

where $\widetilde{X}_i$, for $i \leq 0$, are arbitrary initial values. Note that $\widetilde{\lambda}_{d,t}(\theta_d)$ will serve as a proxy for $\lambda_{d,t}(\theta_d)$. It is obtained by setting to some integer values $\widetilde{X}_0,\ \widetilde{X}_{-1},..$ the unknown initial values $X_0,\ X_{-1},..$ involved in $\lambda_{d,t}(\theta_d)$. We assume that

$$\widetilde{\lambda}_{d,t}(\theta_d) > \underline{\omega},\ \forall t \geq 1 \text{ and } \theta_d \in \Theta. \tag{4.2.4}$$

We also assume that

$$\theta_d \mapsto \lambda_{d,t}(\theta_d) \text{ is almost surely continuous and } \Theta_d \text{ is a compact set.} \tag{4.2.5}$$

The EbE-PQMLE is defined as any solution of the following maximization problem

$$\widehat{\theta}_{n,d} = \arg\max_{\theta_d \in \Theta_d} \widetilde{L}_{n,d}(\theta_d), \quad \widetilde{L}_{n,d}(\theta_d) = \frac{1}{n}\sum_{t=s+1}^{n} \widetilde{\ell}_{d,t}(\theta_d), \tag{4.2.6}$$

where $\widetilde{\ell}_{d,t}(\theta_d) = -\widetilde{\lambda}_{d,t}(\theta_d) + X_{d,t}\log\widetilde{\lambda}_{d,t}(\theta_d)$ and s is asymptotically unimportant, but it can affect the finite sample behaviour of the PQMLE by reducing the impact of the initial values. Let also

$$L_{n,d}(\theta_d) = \frac{1}{n}\sum_{t=s+1}^{n} \ell_{d,t}(\theta_d), \text{ where } \ell_{d,t}(\theta_d) = -\lambda_{d,t}(\theta_d) + X_{d,t}\log\lambda_{d,t}(\theta_d).$$

4.2.1 Other technical assumptions

For the consistency of EbE-PQMLE, we assume that

A1 We have $\theta_{0,d} \in \Theta_d$ where Θ_d is compact.

A2 The process $(X_{d,t})$ is stationary and ergodic.

A3 $\lambda_{d,1}(\theta_d) = \lambda_{d,1}(\theta_{0,d})$ almost surely if and only if $\theta_d = \theta_{0,d}$

The next two assumptions are used to show that the initial values have no effect on the asymptotic properties of EbE-PQMLE.

A4 We have a.s. $\lim_{t\to\infty} a_{d,t} = 0$ and $\lim_{t\to\infty} X_{d,t}a_{d,t} = 0$,

where $a_{d,t} = \sup_{\theta_d \in \Theta_d}\left|\widetilde{\lambda}_{d,t}(\theta_d) - \lambda_{d.t}(\theta_d)\right|$.

For the asymptotic normality, we need the following assumptions

A5 $\theta_{0,d} \in \overset{\circ}{\Theta}_d$, where $\overset{\circ}{\Theta}_d$ denotes the interior of Θ_d.

The next assumption is introduced to handle the initial values.

A6 $b_{d,t}, b_{d,t}X_{d,t}$ and $a_{d,t}g_{d,t}X_{d,t}$ are of order $O(t^{-\kappa})$ in probabilty for some $\kappa > 1/2$, where

$$b_{d,t} = \sup_{\theta_d\in\Theta_d} \left\| \frac{\partial\widetilde{\lambda}_{d,t}(\theta_d)}{\partial\theta_d} - \frac{\partial\lambda_{d,t}(\theta_d)}{\partial\theta_d} \right\|,$$
$$g_{d,t} = \sup_{\theta_d\in\Theta_d} \max\left\{ \left\| \frac{1}{\lambda_{d,t}(\theta_d)} \frac{\partial\lambda_{d,t}(\theta_d)}{\partial\theta_d} \right\|, \left\| \frac{1}{\widetilde{\lambda}_{d,t}(\theta_d)} \frac{\partial\widetilde{\lambda}_{d,t}(\theta_d)}{\partial\theta_d} \right\| \right\}.$$

A7 The conditional variance of $X_{d,t}$ given its past exists and

$$E \mid Var\,(X_{d,t} \mid \mathcal{F}_{t-1}) \mid^{1+\varepsilon} < \infty, \text{ for some } \varepsilon > 0.$$

A8 $\lambda_{d,t}$, $\widetilde{\lambda}_{d,t}$ admit continuous second-order derivatives, the matrices

$$J_{dd} = E\left(\frac{1}{\lambda_{d,t}(\theta_{0,d})} \frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta_d} \frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta'_d} \right)$$

and

$$I_{dd} = E\left(\frac{Var\,(X_{d,t} \mid \mathcal{F}_{t-1})}{\lambda^2_{d,t}(\theta_{0,d})} \frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta_d} \frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta'_d} \right)$$

exist and J_{dd} is invertible.

Theorem 4.2.1. *Let $(X_{d,t})$ be a stationary and ergodic process satisfying (4.2.1)-(4.2.5) and the assumptions **A1-A4**. Let $\widehat{\theta}_{n,d}$ be the EbE-PQMLE, then*

$$\widehat{\theta}_{n,d} \to \theta_{0,d} \quad a.s. \qquad as\ n \to \infty.$$

Remark 4.2.1 (linear multivariate model)**.** *One of the most commonly used specification of the general model (4.2.1) is the linear or so-called multivariate INGARCH(p,q) model which satisfies the following vectorial representation*

$$\Lambda_t(\theta_0) = \Omega_0 + \sum_{i=1}^{q} A_{0,i}X_{t-i} + \sum_{j=1}^{p} B_{0,j}\Lambda_{t-j}(\theta_0), \tag{4.2.7}$$

where $\Lambda_t(\theta_0) = (\lambda_{1,t}(\theta_0), \ldots, \lambda_{k,t}(\theta_0))'$, $\Omega_0 = (\omega_{0,1}, \ldots, \omega_{0,k})'$ has strictly positive coefficients, and the matrices

$$A_{0,i} = \begin{pmatrix} \alpha_{0,11,i} & \alpha_{0,12,i} & \cdots & \alpha_{0,1k,i} \\ \alpha_{0,21,i} & \alpha_{0,22,i} & \cdots & \alpha_{0,2k,i} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{0,k1,i} & \alpha_{0,k2,i} & \cdots & \alpha_{0,kk,i} \end{pmatrix}, \qquad B_{0,j} = \begin{pmatrix} \beta_{0,11,j} & \beta_{0,12,j} & \cdots & \beta_{0,1k,j} \\ \beta_{0,21,j} & \beta_{0,22,j} & \cdots & \beta_{0,2k,j} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{0,k1,j} & \beta_{0,k2,j} & \cdots & \beta_{0,kk,j} \end{pmatrix}$$

are of size $k \times k$ with positive coefficients. To facilitate the procedures of estimation and to make the required assumptions more explicit, we assume that

$$\text{the matrices } B_{0,j}, \text{ for } j = 1, \ldots, p, \text{ are diagonal.} \tag{4.2.8}$$

Under (4.2.8), the dth equation of this model can be written as follows

$$\lambda_{d,t}(\theta_{0,d}) = \omega_d + \sum_{i=1}^{q} \alpha_{0,d1,i} X_{1,t-i} + \cdots + \sum_{i=1}^{q} \alpha_{0,dk,i} X_{k,t-i} + \sum_{j=1}^{p} \beta_{0,dd,j} \lambda_{d,t-j}. \tag{4.2.9}$$

*The identifiability assumption **A3** can be shown using **A2**, the assumptions (4.2.1)-(4.2.3) and the following conditions :*

(a) *For any $c \in \{1, \ldots, k\}$, there is no common root to $\mathcal{A}_\theta^{(dc)}(z)$ and $\mathcal{B}_\theta^{(d)}(z)$, where*

$$\mathcal{A}_\theta^{(dc)}(z) = \sum_{i=1}^{q} \alpha_{0,dc,i} z^i \quad \text{and} \quad \mathcal{B}_\theta^{(d)}(z) = 1 - \sum_{j=1}^{p} \beta_{0,dd,j} z^j.$$

(b) *If $p > 0$, $\sum_{j=1}^{p} \beta_{0,dd,j} < 1$. Thus $\mathcal{B}_\theta^{(d)}(z)$ is invertible.*

(c) *$Var(X_t)$ is a positive definite matrix.*

(d) *At least one $\alpha_{0,dc,i}$ is strictly positive, for $c = 1, \ldots, k$ and $i = 1, .., q$.*

Note that, under the condition (b), the equation (4.2.9) can be rewritten in vectorial autoregressive form, as

$$\underline{\lambda}_{d,t}(\theta_{0,d}) = \underline{c}_{d,t}(\theta_{0,d}) + B_d^* \underline{\lambda}_{d,t-1}(\theta_{0,d}), \tag{4.2.10}$$

where $\underline{\lambda}_{d,t}(\theta_{0,d}) = (\lambda_{d,t}, \ldots, \lambda_{d,t-p+1})'$,

$$\underline{c}_{d,t}(\theta_{0,d}) = \begin{pmatrix} \omega_d + \sum_{i=1}^{q} \alpha_{0,d1,i} X_{1,t-i} + \cdots + \sum_{i=1}^{q} \alpha_{0,dk,i} X_{k,t-i} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and

$$B_d^* = \begin{pmatrix} \beta_{0,dd,1} & \beta_{0,dd,2} & \cdots & \beta_{0,dd,p} \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}$$

*Thus, **A4** can be shown using similar arguments used in Ahmad & Francq (2016) and the arguments used to show (7.30) in Francq & Zakoian (2010).*

Theorem 4.2.2. *Assume that* $(X_{d,t})$ *satisfies the conditions of Theorem 4.2.1. Assume also* ***A5-A8****, then*

$$\sqrt{n}(\widehat{\theta}_{n,d} - \theta_{0,d}) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{dd} := J_{dd}^{-1} I_{dd} J_{dd}^{-1}\right) \qquad \text{as } n \to \infty.$$

It can be shown that, under the assumptions of Theorem 4.2.2, the matrix Σ_{dd} is consistently estimated by

$$\widehat{\Sigma}_{dd} = \widehat{J}_{dd}^{-1} \widehat{I}_{dd} \widehat{J}_{dd}^{-1} \tag{4.2.11}$$

with

$$\widehat{J}_{dd} = \frac{1}{n} \sum_{t=s+1}^{n} \frac{1}{\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})} \frac{\partial \widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial \theta_d} \frac{\partial \widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial \theta'_d}, \tag{4.2.12}$$

$$\widehat{I}_{dd} = \frac{1}{n} \sum_{t=s+1}^{n} \left(\frac{X_{d,t}}{\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})} - 1 \right)^2 \frac{\partial \widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial \theta_d} \frac{\partial \widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial \theta'_d}. \tag{4.2.13}$$

Remark 4.2.2 (linear multivariate model)**.** *Under the moment assumption* (4.2.3)*, one can show that* ***A6*** *and* ***A8*** *are satisfied by similar arguments used in Remark 2.3 of Ahmad & Francq (2016). The invertibility of matrix* J_{dd} *can be shown under* (4.2.3)*,* (4.2.2) *and the condition (c) of Remark 4.2.1.*

Denote by θ_0 the vector of parameters of all equations of the model. Even if the components of θ_0 are estimated equation-by-equation, the components of $\widehat{\theta}_n$ are not necessarily asymptotically independent in the presence of autocorrelation between the series. We assume that

- **A9** For (c,d) $\in \{1, \ldots, k\}$, $\lim_{n\to\infty} \sup_{\theta_d \in \Theta_d} \left\| \frac{\partial}{\partial \theta_d} \widetilde{L}_{n,d}(\theta_{0,d}) - \frac{\partial}{\partial \theta_d} L_{n,d}(\theta) \right\| h_{c,t} = 0, a.s,$

  where $h_{c,t} = \sup_{\theta_c \in \Theta_c} \max \left\{ \left\| \frac{\partial \lambda_{c,t}(\theta_c)}{\partial \theta_c} \right\|, \left\| \frac{\partial \widetilde{\lambda}_{c,t}(\theta_c)}{\partial \theta_c} \right\| \right\}$.

Theorem 4.2.3. *Assume that, for* $d \in \{1, \ldots, k\}$*,* $(X_{d,t})$ *satisfies the conditions of Theorems 4.2.1 and 4.2.2. Assume also* ***A9****, then*

$$\begin{pmatrix} \sqrt{n}(\widehat{\theta}_{n,1} - \theta_{0,1}) \\ \vdots \\ \sqrt{n}(\widehat{\theta}_{n,k} - \theta_{0,k}) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_\theta) \qquad \text{as } n \to \infty,$$

where

$$\Sigma_\theta = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} \\ \vdots & \ddots & \vdots \\ \Sigma'_{k1} & \cdots & \Sigma_{kk} \end{pmatrix}. \tag{4.2.14}$$

For (c,d) $\in \{1, \dots, k\}$, the sub-matrices of Σ_θ are given as follows

$$\Sigma_{dc} = J_{dd}^{-1} I_{dc} J_{cc}^{-1\prime}. \tag{4.2.15}$$

The matrix I_{dc} is obtained by

$$I_{dc} = \lim_{n\to\infty} \frac{1}{n} \sum_{t=s+1}^{n} \frac{\epsilon_{d,t}(\theta_{0,d})}{\lambda_{d,t}(\theta_{0,d})} \frac{\partial \lambda_{d,t}(\theta_{0,d})}{\partial \theta_d} \frac{\epsilon_{c,t}(\theta_{0,c})}{\lambda_{c,t}(\theta_{0,c})} \frac{\partial \lambda_{c,t}(\theta_{0,c})}{\partial \theta'_c}, \tag{4.2.16}$$

where $\epsilon_{d,t}(\theta_{0,d}) = X_{d,t} - \lambda_{d,t}(\theta_{0,d})$ and $\epsilon_{c,t}(\theta_{0,c}) = X_{c,t} - \lambda_{c,t}(\theta_{0,c})$, and it can be consistently estimated by

$$\widehat{I}_{dc} = \frac{1}{n} \sum_{t=s+1}^{n} \frac{\widetilde{\epsilon}_{d,t}(\widehat{\theta}_d)}{\widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})} \frac{\partial \widetilde{\lambda}_{d,t}(\widehat{\theta}_{n,d})}{\partial \widehat{\theta}_d} \frac{\widetilde{\epsilon}_{c,t}(\widehat{\theta}_{n,c})}{\widetilde{\lambda}_{c,t}(\widehat{\theta}_{n,c})} \frac{\partial \widetilde{\lambda}_{c,t}(\widehat{\theta}_{n,c})}{\partial \widehat{\theta}'_c}. \tag{4.2.17}$$

4.3 Applications to particular models

4.3.1 Multivariate INAR model

Among the most popular multivariate count time series models is the Multivariate INAR (M-INAR) model introduced by Franke & Subba Rao (1993). This model defines X_t as a k-dimensional vector valued in $\mathbb{N}^k$, such that

$$X_t = \sum_{i=1}^{q} A_{0,i} \circ X_{t-1} + E_t, \tag{4.3.1}$$

where $A_{0,i}$ is a $k \times k$ matrix with entries $\alpha_{0,dc,i} \in \{0, 1\}$ for $(d, c) \in \{1, ..., k\}$. The thinning operators' matrix $A_{0,i}\circ$ [2] acts as the usual matrix multiplication keeping in the same time the properties of the binomial thinning operator (see Steutel et al., 1979). The innovations' term $E_t = (\epsilon_{1,t}, \dots, \epsilon_{k,t})'$ is assumed to be an *iid* $\mathbb{N}^k$-valued random vector whose components have finite mean ω_d and finite variance $\sigma^2_{\epsilon_{d,t}}$. The M-INAR(1) process (*i.e.* $q = 1$) is stationary and ergodic if $\rho(A_0) < 1$, where $\rho(A_0)$

2. The thinning operator applied to an integer-valued variable X gives $\alpha \circ X = \sum_{i=1}^{X} y_i$ with the convention $\alpha \circ X = 0$ when $X = 0$, and where y_i's are an *iid* Bernoulli random variable independent of X and has probability of success α.

is the spectral radius of the matrix A_0, and if $\|E_t\| < \infty$. For more details on the properties of M-INAR(1) model, the reader is referred to Pedeli & Karlis (2013). One of the distributional choices for the innovations of the bivariate INAR(1) model (the case when $k = 2$) is the bivariate Poisson distribution which is denoted by $\mathcal{BP}(\lambda_{\epsilon_1}, \lambda_{\epsilon_2}, \phi)$ and has the following probability mass function

$$\begin{aligned} &P(\epsilon_{1,t} = r_1, \epsilon_{2,t} = r_2) \\ &= e^{-(\lambda_{\epsilon_1}+\lambda_{\epsilon_2}+\phi)} \frac{(\lambda_{\epsilon_1} - \phi)^{r_1}}{r_1!} \frac{(\lambda_{\epsilon_2} - \phi)^{r_2}}{r_2!} \sum_{i=0}^{min(r_1,r_2)} \binom{r_1}{i} \binom{r_2}{i} i! \left(\frac{\phi}{(\epsilon_1 - \phi)(\epsilon_2 - \phi)} \right)^i, \end{aligned} \tag{4.3.2}$$

where $(\lambda_{\epsilon_1}, \lambda_{\epsilon_2}) > 0$ and $\phi \in [0, min(\lambda_{\epsilon_1}, \lambda_{\epsilon_2}))$. The parameter ϕ is the covariance between the two random variables. Marginally, each random variable ϵ_1 and ϵ_2 follows a Poisson distribution with parameters λ_{ϵ_1} and λ_{ϵ_2} respectively. Therefore, the bivariate Poisson INAR(1) (BP INAR(1)) model is limited to deal only with the equi-dispersed time series. Another alternative model to deal with the over-dispersion, which is often observed in time series of counts, is the bivariate negative binomial INAR(1) (BNB INAR (1)) model. This model is given by assuming that the joint probability mass function of the two innovations' processes $(\epsilon_{1,t}, \epsilon_{2,t})$ follows the bivariate negative binomial distribution (see Marshall & Olkin, 1990, Boucher et al., 2008 and Cheon et al., 2009). We denote this distribution by $\mathcal{BNB}(\eta_{\epsilon_1}, \eta_{\epsilon_2}, \nu)$, where η_{ϵ_1}, η_{ϵ_2}, ν >0. This distribution can be obtained by several ways. For instance, we can generate a random variable g distributed as the gamma distribution with shape parameter ν and scale parameter $w = 1$. Then we generate independently two random variables following Poisson distribution with respectively intensity parameters $g\eta_1$ and $g\eta_2$. The mixture has thus negative binomial marginals and mass function

$$\begin{aligned} P(Y_1 = r_1, Y_2 = r_2) = &\frac{\Gamma(\nu + r_1 + r_2)}{\Gamma(\nu)\Gamma(r_1 + 1)\Gamma(r_2 + 2)} \\ &\left(\frac{\eta_1}{(\eta_1 + \eta_2 + 1)} \right)^{r_1} \left(\frac{\eta_2}{(\eta_1 + \eta_2 + 1)} \right)^{r_2} \left(\frac{1}{(\eta_1 + \eta_2 + 1)} \right)^{\nu}. \end{aligned} \tag{4.3.3}$$

The univariate marginal distributions are negative binomial with parameters ν and $p_d = \eta_d\nu/(\eta_d\nu + \nu)$. The two models presented above can not accommodate negative correlation between the series, which can be considered as a serious limitation. To overcome this problem, an alternative approach is proposed by Karlis & Pedeli (2013) for introducing negative cross-correlation between the series. This model is defined by assuming that the innovation of the bivariate INAR(1) model has Poisson $\mathcal{P}(\lambda)$ or negative binomial $\mathcal{NB}(p,\nu)$ marginals, connected through a Frank or Gaussian bivariate copulas. Copula theory goes back to the work of Sklar (1959), who showed that a joint distribution can be decomposed into its k marginal distributions and a copula, that describes the dependence between the variables. The bivariate Gaussian copula is defined as follows

$$C_{\sigma_C}(u,v) = \Phi_{\sigma_C}(\Phi^{-1}(u), \Phi^{-1}(v)), \tag{4.3.4}$$

where Φ is the $N(0,1)$ cdf, Φ^{-1} is the functional inverse of Φ and Φ_{σ_C} is the bivariate standard normal cdf with covariance σ_C. The bivariate Frank copula is given by

$$C_\delta(u,v) = -\frac{1}{\delta}\left(1 + \frac{(exp(-u\delta)-1)((exp(-v\delta)-1))}{(exp(-\delta)-1)}\right), \tag{4.3.5}$$

where $0 \leq u,v \geq 1$ and $\delta \in \mathbb{R}/0$. When δ is negative, the correlation between the variables will be also negative.

The autocovariance function of the M-INAR(1) model is given by :

$$\gamma(0) = A\gamma(0)A' + diag(Z\mu) + var(E_t), \tag{4.3.6}$$

where Z is $k \times k$ matrix and $[Z_{ij}] = \alpha_{ij}(1-\alpha_{ij})$, for $i,j = 1,...,k$. The vector μ has a dimension $k \times 1$ and contains the unconditional means of X_t. For the bivariate case, the elements of μ are given by :

$$\begin{aligned}\mu_1 &= \frac{(1-\alpha_{22})\omega_1 + \alpha_{12}\omega_2}{(1-\alpha_{11})(1-\alpha_{22}) - \alpha_{12}\alpha_{21}} \\ \mu_2 &= \frac{(1-\alpha_{11})\omega_2 + \alpha_{21}\omega_1}{(1-\alpha_{11})(1-\alpha_{22}) - \alpha_{12}\alpha_{21}}.\end{aligned}$$

Equation (4.3.6) can be simplified using the vectorization and Kronecker product, as

$$vec(\gamma(0)) = (I_{k^2} - A \otimes A)^{-1}\left(vec(diag(Z\mu)) + vec(var(E_t))\right). \tag{4.3.7}$$

One can note that by taking the conditional expectation of the M-INAR process, we will have a model similar to that defined in (4.2.7), where $E(X_t|\mathcal{F}_{t-1}) = \Lambda_t = \Omega + \sum_{i=1}^q A_{0,i}X_{t-i}$ and $E(E_t) = \Omega$.

The conditional variance of the dth series of this model is given as follows

$$var\left(X_{d,t} \mid \mathcal{F}_{t-1}\right) = \sum_{i=1}^{q} \alpha_{0,d1,i}(1-\alpha_{0,d1,i})X_{1.t-i} + \cdots + \sum_{i=1}^{q} \alpha_{0,dk,i}(1-\alpha_{0,dk,i})X_{k.t-i} + Var(\epsilon_{d.t}). \tag{4.3.8}$$

Thus, under (4.2.3) and that the innovation has finite variance. Assumption **A7** is satisfied. Since, $a_{d.t} = b_{d.t} = d_{d.t} = 0$ for $t \geq 2$. Assumptions **A4** and **A6** are thus fulfilled.

4.3.2 Bivariate Poisson INGARCH model

The bivariate Poisson INGARCH model is defined by assuming that the conditional distribution of X_t given its past is the bivariate Poisson distribution given in (4.3.2)

$$X_t|\mathcal{F}_{t-1} \rightsquigarrow \mathcal{BP}(\lambda_{1.t}, \lambda_{2.t}, \phi), \tag{4.3.9}$$

where $\Lambda_t = (\lambda_{1.t}, \lambda_{2.t})'$ satisfies (4.2.7). The conditional covariance between the series is ϕ. The limitation of this model is that it can only accommodate positive correlation between series. According to Proposition 4.3.1 of Liu (2012) the bivariate Poisson INGARCH model is stationary and ergodic if

$$2^{1-1/r} \sum_{i=1}^{q} \|A_{0,i}\|_r + \sum_{j=1}^{p} \|B_{o,j}\|_r < 1, \tag{4.3.10}$$

where, for $0 \leq r \leq \infty$, $\|g\|_r$ is the r-*norm* of a squared matrix g. Using the Minkowski's inequality and the fact that

$$\rho(g) \leq \|g\|_r,$$

we have

$$\begin{aligned} 1 > \|B_{o,1}\|_r + \cdots + \|B_{o,p}\|_r &\geq \|B_{o,1} + \cdots + B_{o,p}\|_r \\ &\geq \rho(B_{o,1} + \cdots + B_{o,p}). \end{aligned}$$

Under the assumption (4.2.8), we have $\rho(B_{o,1} + \cdots + B_{o,p}) = \|B_{o,1} + \cdots + B_{o,p}\|_\infty$, where $\|g\|_\infty$ is the maximum absolute row sum. Thus, the condition (4.3.10) implies that the condition (b) of Remark 4.2.1 is fulfilled. So, this model is within the scopes of Remark 4.2.1 and Remark 4.2.2. Hence, all the assumptions required for the Theorems 4.2.1 and 4.2.2 can be verified.

4.3.3 Copulas-based multivariate INGARCH models

An alternative approach to deal with both positive and negative cross correlation between the series is the copula-based multivariate INGARCH model studied by Heinen & Rengifo (2007). This model is defined as in (4.2.7) by assuming that the conditional distribution of $X_{d,t}$ given its past values is double Poisson $\mathcal{DP}(\lambda_{d,t}, \gamma_d)$ of Efron (1986) which allows both underdispersed and overdispersed count series. The components of the vector X_t are assumed to be correlated through Gaussian copula. To overcome the difficulties associated to using copulas with discrete distributions, Heinen & Rengifo (2007) proposed to use the continued extension of the count data (see Denuit & Lambert, 2005). In this situation, the joint density of X_t is :

$$h(X_{1.t}, .., X_{K.t}; \theta_1, .., \theta_K; \Sigma_C) = \prod_{d=1}^{K} \mathcal{DP}(X_{d.t}, \lambda_{d,t}, \gamma_d).C_{\Sigma_C}(q_t), \tag{4.3.11}$$

where $\mathcal{DP}(X_{d.t}, \lambda_{d.t}, \gamma_d)$ denotes the double Poisson density as a function of the observation $X_{d.t}$, the conditional mean $\lambda_{d,t}$ and the dispersion parameter γ_d, C_{Σ_C} is the multivariate Gaussian copula density with variance covariance matrix Σ_C, $q_t = (\Phi^{-1}(z_{1,t}), ..., \Phi^{-1}(z_{k,t}))$, Φ^{-1} is the inverse of the standard univariate normal distribution function and $z_{d,t}$ is the Probability Integral Transformation (PIT) of the continued extension of the count data, under the marginal densities :

$$z_{d,t} = F^*_{d,t}(X^*_{d,t}) = F_{d,t}(X_{d,t}) + f_{d,t}(X_{d,t})U_{d,t}$$

where $F^*_{d,t}$, $F_{d,t}$ and $f_{d,t}$ are the conditional cumulative distribution function (cdf) of the continued extension of the data, the conditional cdf and the conditional probability distribution function, respectively. The variable $X^*_{d,t}$ is the continued extension of the original count data $X_{d,t}$:

$$X^*_{d,t} = X_{d,t} + (U_{d,t} - 1).$$

The $U_{d,t}$ is uniform random variable on $[0, 1]$. The ergodicity of the model has not been established. According to Heinen & Rengifo (2007) the model has a stationary solution when the eigenvalues of $I_{k\times k} - \sum_{i=1}^{q} A_{0,i} - \sum_{j=1}^{p} B_{0,j}$ lie within the unit circle. Under the stationarity and ergodicity assumptions and the conditions of Remark 4.2.1, the assumptions required for Theorem 4.2.1 and 4.2.2 are thus explicit.

4.4 Numerical illustrations

The first part of this section examines the finite sample behaviour of the EbE-PQMLE. The second part presents a finite sample comparison between EbE-PQMLE and MLE for bivariate Poisson and bivariate negative Binomial INGARCH models. All the results of this section are based on $N = 1000$ independent replications of Monte Carlo simulations of different sample sizes n. For each simulation, the first 100 observations are omitted, so that the process approaches its stationary regime.

4.4.1 Finite sample behaviour of EbE-PQMLE

The EbE-PQMLE defined in (4.2.6) is optimized numerically using the PORT routine (implemented by the function nlminb() of R). Firstly, we simulate a bivariate INAR(1) model when the marginal distributions of the innovations $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are, respectively $\mathcal{P}(\lambda_1)$ and $\mathcal{P}(\lambda_2)$ and also when they follow marginally $\mathcal{NB}(\nu_1, p_1)$ and $\mathcal{NB}(\nu_2, p_2)$ distributions. In the two models, the marginal distributions of the innovations are assumed to be correlated through a bivariate Gaussian Cupola. In addition, we simulate a bivariate INAR(1) model when the innovations follow jointly $\mathcal{BP}(\lambda_1, \lambda_2, \phi)$ distribution. We also report a simulation for the copula based bivariate INGARCH (1,1) model when the conditional distributions are Poisson, negative binomial and double-Poisson. The results are summarized in Tables 4.1 and 4.2. The means of the estimated values of $\theta_{0,1}$ and $\theta_{0,2}$ are given in the rows "$\widehat{\theta}_1$" and "$\widehat{\theta}_2$". We present also three different estimators of the root-mean-square deviation $\sqrt{E\left(\widehat{\theta}_n - \theta_0\right)^2}$: the empirical standard errors (ESE), the estimated standard error based on the asymptotic theory (ASE) and the theoretical standard error based on the asymptotic theory (TSE) which is obtained by replacing $\widehat{\Sigma}_{dd}$ by Σ_{dd}. The means of the contemporaneous covariances between the series are given in the rows "$cov(X_{1t}, X_{2t})$".

Tables 4.1 and 4.2 show that, for all the models, the means of the estimated parameters are satisfactorily close to their theoretical values, especially for large sample sizes. Moreover, the three estimations of the standard deviations, the ESE, ASE and TSE, are very similar. The closeness between ESE and ASE means that the asymptotic theory provides a reliable view on the actual standard error of the EbE-

TABLE 4.1 – The finite sample behaviour of EbE-PQMLE for the bivariate INAR model

Bivariate INAR(1) model										
		Copula based bivariate INAR model, $\sigma_C = -0.5$						BP INAR		
$cov(X_{1t}, X_{2t})$		0.351			0.533			4.217		
n		$\epsilon_{1.t} \sim \mathcal{P}(2)$			$\epsilon_{1.t} \sim \mathcal{NB}(2, 0.5)$			$E_t \sim \mathcal{BP}(2, 3, 1)$		
		ω_{01}=2	α_{011}=0.2	α_{012}=0.4	ω_{01}=2	α_{011}=0.2	α_{012}=0.4	ω_{01}=3	α_{011}=0.2	α_{012}=0.4
500	$\widehat{\theta}_1$	2.034	0.195	0.399	2.023	0.198	0.399	3.065	0.194	0.399
	ESE	0.357	0.398	0.034	0.443	0.038	0.029	0.444	0.042	0.037
	ASE	0.365	0.039	0.035	0.434	0.037	0.030	0.434	0.040	0.037
	TSE	0.367	0.039	0.035	0.438	0.038	0.029	0.439	0.041	0.037
1000	$\widehat{\theta}_1$	2.016	0.198	0.399	2.004	0.190	0.400	3.028	0.198	0.399
	ESE	0.256	0.026	0.025	0.315	0.027	0.021	0.300	0.029	0.026
	ASE	0.257	0.028	0.025	0.307	0.026	0.021	0.307	0.029	0.026
	TSE	0.258	0.028	0.025	0.309	0.027	0.021	0.308	0.029	0.026
		$\epsilon_{2.t} \sim \mathcal{P}(3)$			$\epsilon_{2.t} \sim \mathcal{NB}(3, 0.6)$			$E_t \sim \mathcal{BP}(2, 3, 2)$		
		ω_{02}=3	α_{022}=0.3	α_{021}=0.4	ω_{02}=4.5	α_{022}=0.3	α_{021}=0.4	ω_{02}=4	α_{022}=0.3	α_{021}=0.4
500	$\widehat{\theta}_2$	3.039	0.297	0.397	4.569	0.295	0.398	4.056	0.296	0.398
	ESE	0.421	0.045	0.042	0.629	0.041	0.053	0.510	0.041	0.045
	ASE	0.412	0.039	0.044	0.618	0.041	0.053	0.478	0.040	0.044
	TSE	0.415	0.039	0.044	0.624	0.041	0.053	0.482	0.041	0.045
1000	$\widehat{\theta}_2$	3.016	0.298	0.399	4.541	0.297	0.400	4.008	0.299	0.399
	ESE	0.296	0.028	0.031	0.454	0.029	0.038	0.329	0.028	0.032
	ASE	0.291	0.028	0.031	0.437	0.029	0.037	0.337	0.029	0.032
	TSE	0.292	0.028	0.031	0.440	0.029	0.037	0.339	0.027	0.032

PQMLE. As expected, the standard errors decrease as the sample sizes increase. Figures 4.8 and 4.2 show boxplot and histogram of the $N = 1000$ values of the EbE-PQMLE (centred and reduced) for simulations of length $n = 3000$ of bivariate INAR(1) and bivariate INGARCH$(1, 1)$ models. In agreement with Theorem 4.2.2, the empirical distribution of the estimator resembles the standard Gaussian law. Table 4.3 gives the p-values of the Kolmogorov-Smirnov test of normality for the N values of the EbE-PQMLE, computed on simulations of size $n = 3000$ of each of the models considered in Tables 4.1 and 4.2. This table shows that the normality assumption is never rejected.

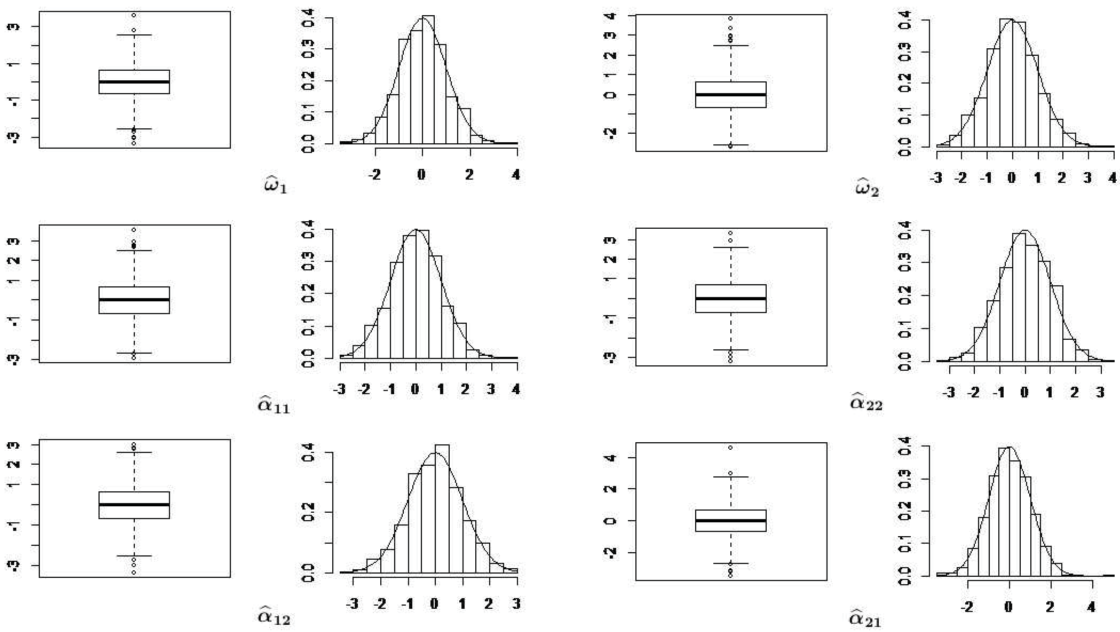

FIGURE 4.1 – Boxplot and histogram of the standardized distribution of EbE-PQMLE for bivariate Poisson INAR(1) model when $\theta_{01} = (\omega_{01}, \alpha_{11}, \alpha_{12})$= (3, 0.2, 0.4), $\theta_{02} = (\omega_{02}, \alpha_{022}, \alpha_{021})$= (4, 0.3, 0.4) and $(\lambda_{\epsilon_1}, \lambda_{\epsilon_2}, \phi)$= (2, 3, 1). The sample size n= 3000. Superimposed is the standard normal density function.

TABLE 4.2 – The finite sample behaviour of EbE-PQMLE for the copula based bivariate INGARCH(1,1) model, when $\theta_{01} = (\omega_{01},\ \alpha_{011},\ \beta_{011},\ \alpha_{012})$= (2, 0.2, 0.4, 0.3), $\theta_{02} = (\omega_{02},\ \alpha_{022},\ \beta_{022},\ \alpha_{021})$= (4, 0.3, 0.4, 0.2) and $\sigma_C = -0.5$.

Copula based bivariate INGARCH model													
$cov(X_{1t}, X_{2t})$		-3.692				-27.978				-5.271			
n		$\mathcal{P}(\lambda_{1.t})$				$\mathcal{NB}(3, p_{1.t})$				$\mathcal{DP}(\lambda_{1.t}, 0.5)$			
		$\widehat{\omega}_1$	$\widehat{\alpha}_{11}$	$\widehat{\beta}_1$	$\widehat{\alpha}_{12}$	$\widehat{\omega}_1$	$\widehat{\alpha}_{11}$	$\widehat{\beta}_1$	$\widehat{\alpha}_{12}$	$\widehat{\omega}_1$	$\widehat{\alpha}_{11}$	$\widehat{\beta}_1$	$\widehat{\alpha}_{12}$
500	$\widehat{\theta}_1$	2.137	0.193	0.401	0.301	2.245	0.190	0.399	0.300	2.596	0.194	0.391	0.297
	ESE	1.428	0.052	0.107	0.047	1.462	0.055	0.112	0.055	2.358	0.051	0.162	0.090
	ASE	1.385	0.047	0.098	0.046	1.467	0.049	0.103	0.053	2.219	0.047	0.147	0.086
	TSE	1.382	0.048	0.105	0.045	1.470	0.051	0.108	0.055	2.100	0.048	0.151	0.091
1000	$\widehat{\theta}_1$	2.004	0.197	0.403	0.300	2.094	0.196	0.400	0.300	2.233	0.198	0.391	0.303
	ESE	0.963	0.034	0.071	0.033	1.103	0.037	0.080	0.039	1.615	0.035	0.113	0.066
	ASE	0.947	0.033	0.069	0.032	1.026	0.035	0.074	0.038	1.502	0.034	0.105	0.062
	TSE	0.953	0.034	0.070	0.032	1.031	0.036	0.067	0.039	1.453	0.034	0.106	0.064
		$\mathcal{P}(\lambda_{2.t})$				$\mathcal{NB}(4, p_{2.t})$				$\mathcal{DP}(\lambda_{1.t}, 2)$			
		$\widehat{\omega}_2$	$\widehat{\alpha}_{22}$	$\widehat{\beta}_2$	$\widehat{\alpha}_{21}$	$\widehat{\omega}_2$	$\widehat{\alpha}_{22}$	$\widehat{\beta}_2$	$\widehat{\alpha}_{21}$	$\widehat{\omega}_2$	$\widehat{\alpha}_{22}$	$\widehat{\beta}_2$	$\widehat{\alpha}_{21}$
500	$\widehat{\theta}_2$	4.535	0.296	0.388	0.199	4.484	0.290	0.392	0.202	4.128	0.293	0.404	0.199
	ESE	1.911	0.051	0.114	0.051	1.869	0.056	0.111	0.051	1.308	0.049	0.083	0.026
	ASE	1.743	0.048	0.106	0.049	1.702	0.051	0.103	0.048	1.257	0.047	0.079	0.025
	TSE	1.666	0.048	0.105	0.051	1.657	0.053	0.104	0.049	1.256	0.048	0.079	0.025
1000	$\widehat{\theta}_2$	4.264	0.299	0.392	0.201	4.273	0.296	0.394	0.200	4.097	0.295	0.402	0.200
	ESE	1.209	0.035	0.075	0.036	1.228	0.036	0.074	0.034	0.917	0.035	0.058	0.018
	ASE	1.184	0.034	0.074	0.035	1.181	0.037	0.073	0.034	0.873	0.033	0.056	0.018
	TSE	1.154	0.034	0.074	0.036	1.161	0.037	0.074	0.035	0.871	0.034	0.056	0.018

4.4.2 Finite sample comparison between EbE-PQMLE and MLE

Table 4.4 gives results of a Monte Carlo experiment of estimation of bivariate Poisson INGARCH$(1,1)$ model. The model is fitted by two estimators : the first is EbE-PQMLE difined in (4.2.6), and the second is the true maximum likelihood estimator (MLE) obtained by maximizing the following function (see Liu (2012)) :

$$\ell_t^{BP}(\theta) = -\sum_{t=1}^{n}(\lambda_{1.t} + \lambda_{2.t} - \phi) + \sum_{t=1}^{n} X_{1.t} log(\lambda_{1.t} - \phi) + \sum_{t=1}^{n} X_{1.t} log(\lambda_{2.t} - \phi)$$
$$+ \sum_{t=1}^{n} log\left(\sum_{i=0}^{min(X_{1.t}, X_{2.t})} \binom{X_{1.t}}{i}\binom{X_{2.t}}{i} i! \left(\frac{\phi}{(\lambda_{1.t} - \phi)(\lambda_{2.t} - \phi)}\right)^i\right). \quad (4.4.1)$$

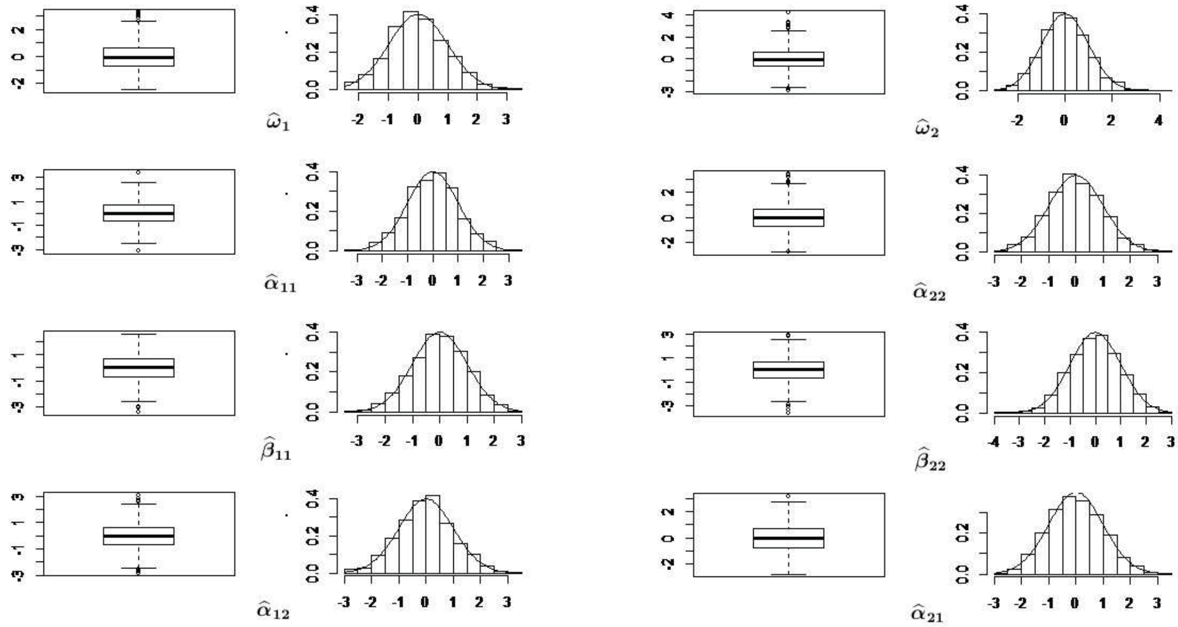

FIGURE 4.2 – Boxplot and histogram of the standardized distributions of EbE-PQMLE when $\theta_{01}=(\omega_{01},\alpha_{011},\beta_{011},\alpha_{012})$= (2, 0.2, 0.4, 0.3) and $\theta_{02}=(\omega_{02},\alpha_{022},\beta_{022},\alpha_{021})$= (4, 0.3, 0.4, 0.2) for a copula based bivariate INGARCH$(1,1)$ model with double Poisson conditional distributions when (γ_1,γ_2)= (2, 0.5) and σ_C= -0.5. The sample size n= 3000. Superimposed is the standard normal density function.

In addition, this table gives also a comparison between EbE-PQMLE and MLE for bivariate negative binomial INGARCH$(1,1)$ model. This model might be defined as in (4.2.7) by assuming that the conditional distribution is bivariate negative binomial defined in (4.3.3), such that

$$X_t|\mathcal{F}_{t-1}\rightsquigarrow \mathcal{BNB}(\eta_{1.t},\eta_{2.t},\nu). \tag{4.4.2}$$

With this parametrization, we have $\Lambda_t=(\lambda_{1.t}=\eta_{1.t}\nu,\lambda_{2.t}=\eta_{2.t}\nu)'$. To our knowledge, the bivariate INGARCH model with bivariate negative binomial conditional distribution has not been studied yet in the literature. Hence, the stationarity and ergodicity conditions are not established. However, we are mainly interested in providing an example showing the effect of equation-by-equation estimation using PQMLE on the efficiency. The model could be studied more deeply in a future work.

TABLE 4.3 – p-values of the Kolmogorov-Smirnov test of normality of the EbE-PQMLE

	Bivariate INGARCH			Bivariate INAR		
$\widehat{\theta}_1$	$\mathcal{P}(\lambda_{1.t})$	$\mathcal{NB}(3, P_{1.t})$	$\mathcal{DP}(\lambda_{1.t}, 0.5)$	$\epsilon_{1.t} \sim \mathcal{P}(2)$	$\epsilon_{1.t} \sim \mathcal{NB}(2, 0.5)$	$E_t \sim \mathcal{BP}(2, 3, 2)$
$\widehat{\omega}_1$	0.425	0.222	0.403	0.817	0.915	0.884
$\widehat{\alpha}_{11}$	0.737	0.884	0.979	0.912	0.732	0.541
$\widehat{\alpha}_{12}$	0.999	0.980	0.501	0.975	0.995	0.976
$\widehat{\beta}_{11}$	0.874	0.987	0.423			
$\widehat{\theta}_2$	$\mathcal{P}(\lambda_{2.t})$	$\mathcal{NB}(4, P_{2.t})$	$\mathcal{DP}(\lambda_{2.t}, 2)$	$\epsilon_{2.t} \sim \mathcal{P}(3)$	$\epsilon_{2.t} \sim \mathcal{NB}(3, 0.6)$	$E_t \sim \mathcal{BP}(2, 3, 2)$
$\widehat{\omega}_2$	0.248	0.225	0.851	0.507	0.909	0.836
$\widehat{\alpha}_{22}$	0.451	0.985	0.875	0.999	0.930	0.762
$\widehat{\alpha}_{21}$	0.989	0.791	0.489	0.869	0.645	0.997
$\widehat{\beta}_{22}$	0.756	0.264	0.977			

The MLE of this model is obtained by maximizing the following function

$$\ell_t^{BNB}(\theta) = \sum_{t=1}^{n} \left\{ log\left(\frac{\Gamma(\nu + X_{1.t} + X_{2.t})}{\Gamma(\nu)\Gamma(X_{1.t})\Gamma(X_{2.t})}\right) + X_{1.t}log\left(\frac{\lambda_{1.t}}{\nu + \lambda_{1.t} + \lambda_{2.t}}\right) \right.$$
$$\left. + X_{2.t}log\left(\frac{\lambda_{2.t}}{\nu + \lambda_{1.t} + \lambda_{2.t}}\right) + \nu log\left(\frac{\nu}{\nu + \lambda_{1.t} + \lambda_{2.t}}\right) \right\}. \qquad (4.4.3)$$

The means of estimated parameters are shown in the rows "$\widehat{\theta}_1$" and "$\widehat{\theta}_2$" and the related empirical standard errors are given in the rows "ESE". In view of Table 4.4, we can note that, for both estimators, the means of estimated parameters of the two models are satisfactorily close to their theoretical values and the standard errors decrease as the sample seizes increase. For the bivariate Poisson INGARCH model, when $n = 500$, the MLE is more efficient than EbE-PQMLE, but when $n = 1000$ the standard errors of both estimators become very similar. For the bivariate negative Binomial INGARCH model, MLE seems slightly more efficient than EbE-PQMLE even with increasing the sample size. This small loss of efficiency can be considered as a price to pay for not having to entirely specify the conditional distribution. However, in practice, we do not know the true conditional distribution of the data. Furthermore, the EbE-PQMLE is simpler and more flexible because it allows us to estimate the conditional mean parameters for large variety of count time series models regardless the nature of dispersion of the data and the sign of the autocorrelation between the series.

Table 4.4 – Comparison between MLE and EbE-PQMLE

n		MLE				EbE-PQMLE			
$(X_t \mid \mathcal{F}_{t-1}) \sim \mathcal{BP}(\lambda_{1.t}, \lambda_{2.t}, 1)$									
$\omega_{01} = 2$ $\alpha_{011} = 0.2$ $\beta_{011} = 0.4$ $\alpha_{012} = 0.3$									
500	$\widehat{\theta}_1$	1.750	0.181	0.407	0.305	1.667	0.189	0.401	0.302
	ESE	0.697	0.045	0.074	0.043	1.131	0.062	0.103	0.056
1000	$\widehat{\theta}_1$	1.915	0.205	0.382	0.297	1.758	0.196	0.398	0.302
	ESE	0.515	0.032	0.055	0.034	0.606	0.030	0.052	0.029
$\omega_{02} = 3$ $\alpha_{022} = 0.3$ $\beta_{022} = 0.4$ $\alpha_{021} = 0.2$									
500	$\widehat{\theta}_2$	2.523	0.296	0.411	0.196	2.726	0.293	0.396	0.202
	ESE	0.948	0.037	0.072	0.050	1.229	0.062	0.106	0.063
1000	$\widehat{\theta}_2$	2.633	0.291	0.408	0.199	2.823	0.297	0.397	0.200
	ESE	0.701	0.034	0.054	0.029	0.682	0.029	0.045	0.029
$(X_t \mid \mathcal{F}_{t-1}) \sim \mathcal{BNB}(\eta_{1.t}, \eta_{2.t}, 2)$									
$\omega_{01} = 2$ $\alpha_{011} = 0.2$ $\beta_{011} = 0.4$ $\alpha_{012} = 0.3$									
n		MLE				EbE-PQMLE			
500	$\widehat{\theta}_1$	2.131	0.196	0.397	0.299	2.321	0.193	0.396	0.292
	ESE	0.654	0.107	0.063	0.098	0.794	0.128	0.0722	0.120
1000	$\widehat{\theta}_1$	2.048	0.201	0.398	0.298	2.172	0.187	0.400	0.298
	ESE	0.280	0.053	0.029	0.049	0.376	0.066	0.036	0.061
$\omega_{02} = 3$ $\alpha_{022} = 0.3$ $\beta_{022} = 0.4$ $\alpha_{021} = 0.2$									
500	$\widehat{\theta}_2$	3.176	0.293	0.394	0.204	3.363	0.283	0.395	0.202
	ESE	0.794	0.107	0.067	0.113	0.923	0.128	0.075	0.135
1000	$\widehat{\theta}_2$	3.072	0.297	0.396	0.203	3.185	0.294	0.400	0.193
	ESE	0.355	0.053	0.031	0.056	0.445	0.067	0.0385	0.072

4.5 Real data application

We report a simple application of EbE-PQMLE to real intra-daily trading volume data. The estimation and prediction of the mean of intra-daily trading volume is helpful to improve the performance of trading algorithms and to minimize transaction costs by optimally placing orders (see Satish et al., 2014). The main idea is that the actual trading volume of a security may be not only influenced by its past values but also by the past values of the trading volumes of other securities. The data consists of intra-daily trading volume in interval of time of 5 minutes from 06 May 2015 to 06 June 2015 for the stocks of largest three banks in United States, namely JP Morgan Chase & Co (JPM), Bank of America (BAC) and Wells Fargo (WFC). The data is taken from $www.finam.ru$. Table 4.5 shows that all the series are over-dispersed and the contemporaneous cross correlations between the series, which are also visualized in Figure 4.4, are positive. The time series plots, the sample autocorrelation functions and the histograms are presented in Figure 4.3. We used the EbE-PQMLE defined in (4.2.6) to estimate the parameters of the conditional means of the multivariate model defined in (4.2.7) when $p = q = 1$. The estimated parameters and their standard errors are shown in (4.5.1).

Table 4.5 – Data summary

	mean	variance	cross correlation	
JPM	11965.52	96707437	JPM&BAC	0.381
BAC	44455.59	2550265224	JPM&WFC	0.565
WFC	11654.23	109833650	BAC&WFC	0.388

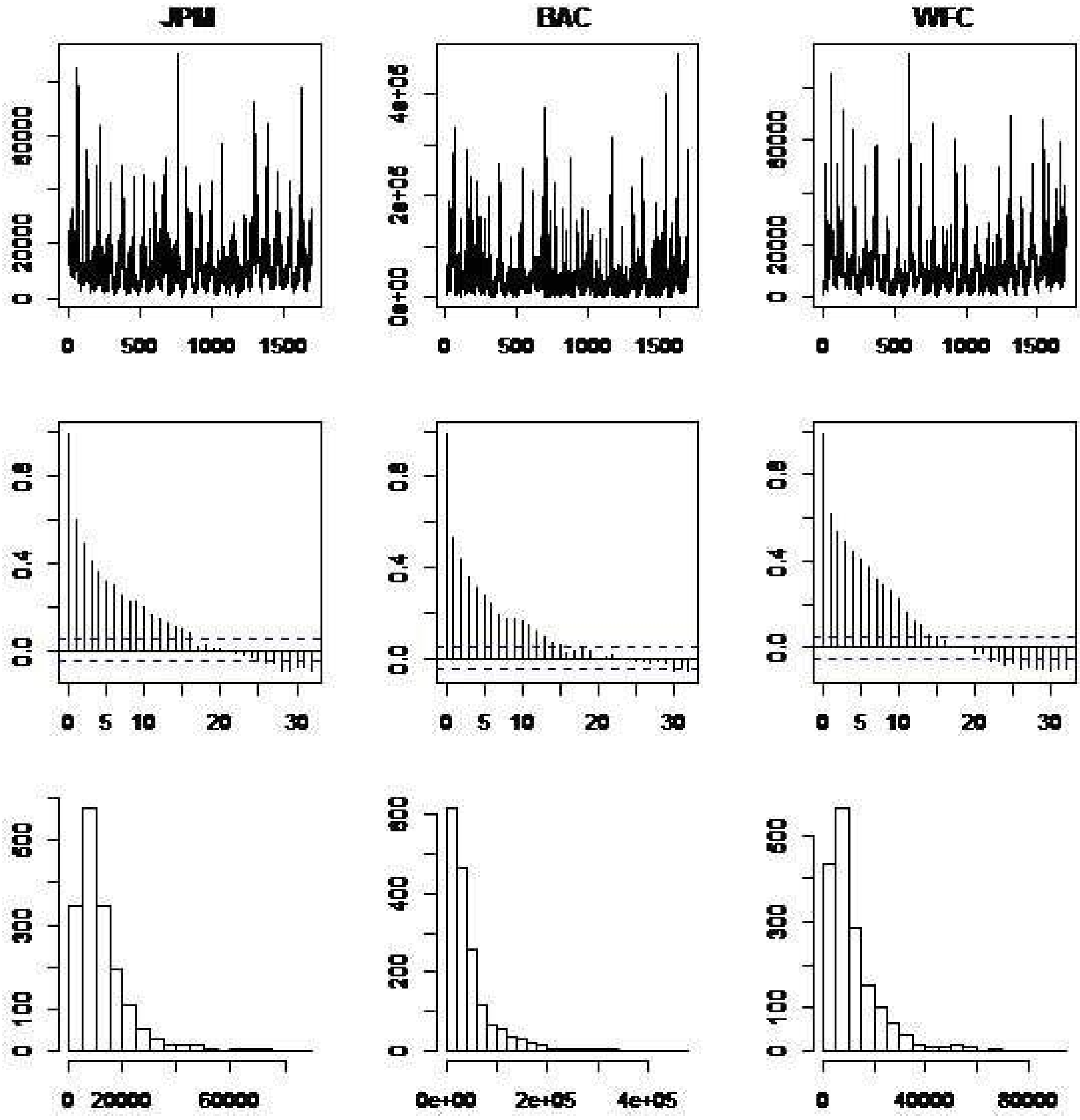

Figure 4.3 – The time series plots, the sample autocorrelation functions and the histograms of the data

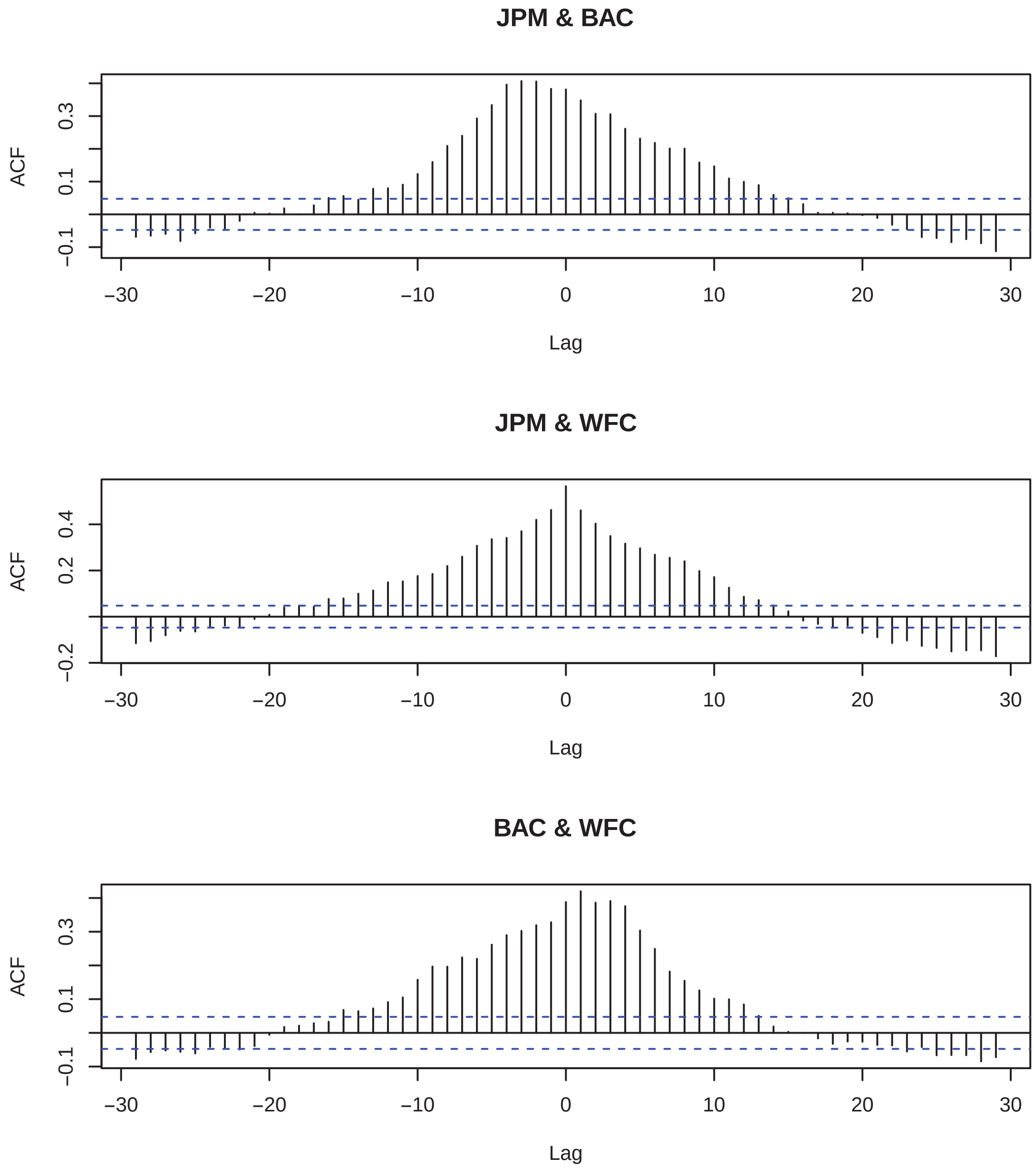

Figure 4.4 – The cross correlations between the series

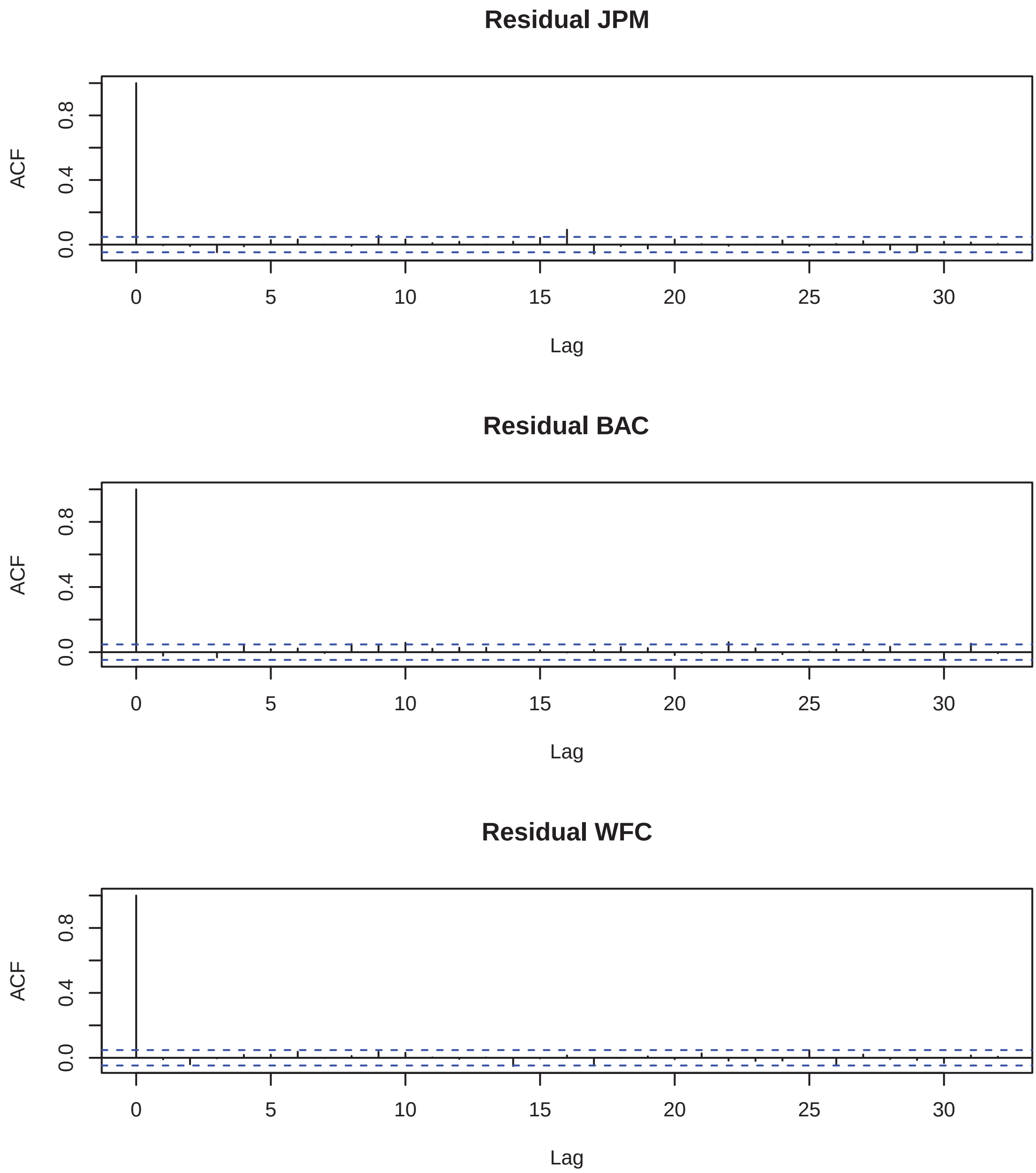

Figure 4.5 – The sample autocorrelations of the Pearson residuals of the series

$$\begin{pmatrix} \lambda_t^{JPM} \\ \lambda_t^{BAC} \\ \lambda_t^{WFC} \end{pmatrix} = \begin{pmatrix} \underset{(235.238)}{1482.691} \\ \underset{(1154.431)}{3290.295} \\ \underset{(159.936)}{727.968} \end{pmatrix} + \begin{pmatrix} \underset{(0.031)}{0.389} & \underset{(0.007)}{0.008} & \underset{(0.020)}{0.080} \\ \underset{(0.099)}{0.390} & \underset{(0.030)}{0.292} & \underset{(0.098)}{0.460} \\ \underset{(0.002)}{0.004} & \underset{(0.018)}{0.073} & \underset{(0.028)}{0.385} \end{pmatrix} \begin{pmatrix} JPM_{t-1} \\ BAC_{t-1} \\ WFC_{t-1} \end{pmatrix} \tag{4.5.1}$$
$$+ \begin{pmatrix} \underset{(0.045)}{0.380} & 0 & 0 \\ 0 & \underset{(0.052)}{0.410} & 0 \\ 0 & 0 & \underset{(0.034)}{0.464} \end{pmatrix} \begin{pmatrix} \lambda_{t-1}^{JPM} \\ \lambda_{t-1}^{BAC} \\ \lambda_{t-1}^{WFC} \end{pmatrix}$$

Figure 4.5 shows the sample autocorrelation functions of the Pearson residuals defined as follows

$$\widetilde{e}_{d,t}(\widehat{\theta}_d) = \frac{X_{d,t} - \widetilde{\lambda}_{d,t}(\widehat{\theta}_d)}{\sqrt{\widetilde{\lambda}_{d,t}(\widehat{\theta}_d)}}, \text{ for } t > 1 \text{ and } d = 1, 2, 3. \tag{4.5.2}$$

As expected, the residuals appear compatible with the white noise assumption.

4.6 Proofs

4.6.1 Proofs of Theorem 4.2.1 and Theorem 4.2.2

The detailed proofs are omitted because they are analogues to that of Theorems 2.1 and 2.2 in Ahmad & Francq (2016). We will only sketch the proofs. For showing the consistency of EbE-PQMLE, it suffices to show that

(i) $\lim_{n\to\infty} \sup_{\theta_d \in \Theta_d} \left| \widetilde{L}_{n,d}(\theta_d) - L_{n,d}(\theta_d) \right| = 0, a.s.$

(ii) $E\left|\ell_{d,1}(\theta_d)\right| < \infty$ and if $\theta_d \neq \theta_{0,d}$, $E\ell_{d,1}(\theta_d) < E\ell_{d,1}(\theta_{0,d})$

(iii) For any $\theta_{1,d} \neq \theta_{0,d}$, there exists a neighbourhood $V(\theta_{1,d})$ of $\theta_{1,d}$ such that

$$\limsup_{n\to\infty} \sup_{\theta_d \in V(\theta_{1,d})} \widetilde{L}_{n,d}(\theta_d) < E\ell_{d,1}(\theta_{0,d}).$$

We can easily show (i) by using (4.2.3), **A4** and the inequality $log(1 + x) \leq x$. In addition, (ii) can be shown using the assumptions (4.2.2), (4.2.3), **A3** and again the inequality $log(x) \leq x - 1$. The last point (iii) follows from the ergodic theorem and (ii).

For the asymptotic normality, we have to show

(iv) $\lim_{n\to\infty} \sup_{\theta_d \in \Theta_d} \left\| \frac{\partial}{\partial \theta_d} \widetilde{L}_{n,d}(\theta_{0,d}) - \frac{\partial}{\partial \theta_d} L_{n,d}(\theta_{0,d}) \right\| = 0, a.s.$

(v) $\sqrt{n} \frac{\partial}{\partial \theta_d} L_{n,d}(\theta_{0,d}) \xrightarrow{d} \mathcal{N}(0, I_{dd})$ as $n \to \infty$

(vi) $\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\theta_{d,i}\theta_{d,j}} \widetilde{\ell}_{d,t}(\theta_d^*) \to J_{dd}(i,j)$ a.s., where θ_d^* is between $\widehat{\theta}_{n,d}$ and $\theta_{0,d}$.

The first point concerning the impact of the initial values can be easily shown using **A4** and **A6**. Under the assumption **A5** ($\theta_{0,d}$ does not lie at the boundary of Θ_d) and using (iv), for θ_d^* is between $\widehat{\theta}_d$ and $\theta_{0,d}$, the Taylor expansion yields

$$\sqrt{n} \frac{\partial}{\partial \theta_d} L_{n,d}(\theta_{0,d}) = -\frac{\partial^2}{\partial \theta_{d,i} \partial \theta_{d,j}} L_{0,d}(\theta_d^*) \sqrt{n}(\widehat{\theta}_{n,d} - \theta_{0,d}). \tag{4.6.1}$$

The left-hand side of (4.6.1) is given by

$$\sqrt{n} \frac{\partial}{\partial \theta_d} L_{n,d}(\theta_{0,d}) = \frac{1}{\sqrt{n}} \sum_{t=s+1}^{n} U_{d,t}, \quad U_{d,t} = \left(\frac{X_{d,t}}{\lambda_{d,t}(\theta_{0,d})} - 1 \right) \frac{\partial \lambda_{d,t}(\theta_{0,d})}{\partial \theta_d}, \tag{4.6.2}$$

where $\{U_{d,t}, \mathcal{F}_t\}$ is a stationary martingale difference. In view of **A8** and under **A7**, we have $EU_{d,t}U'_{d,t} = I_{dd}$. The central limit theorem of Billingsley (1961) for square-integrable stationary martingale difference then entails that

$$\sqrt{n} \frac{\partial}{\partial \theta_d} L_{n,d}(\theta_{0,d}) \xrightarrow{d} \mathcal{N}(0, I_{dd}) \quad \text{as } n \to \infty. \tag{4.6.3}$$

It is easy to show that (vi) follows from **A8**, the ergodic theorem and the dominated convergence theorem. The conclusion follows from **A8**, (4.6.1) and (4.6.3). ☐

4.6.2 Proof of Theorem 4.2.3

Using (4.6.1) and (vi), one can show that

$$\begin{pmatrix} \sqrt{n}(\widehat{\theta}_{n,1} - \theta_{0,1}) \\ \vdots \\ \sqrt{n}(\widehat{\theta}_{n,k} - \theta_{0,k}) \end{pmatrix} = \begin{pmatrix} -J_{11}^{-1} & 0 & \cdots & 0 \\ 0 & -J_{22}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -J_{kk}^{-1} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=s+1}^{n} U_{1,t} \\ \vdots \\ \frac{1}{\sqrt{n}} \sum_{t=s+1}^{n} U_{k,t} \end{pmatrix} + o_p(1). \tag{4.6.4}$$

Using **A9** and (4.6.4), the cental limit theorem for the multivariate martingale differences shows that

$$\begin{pmatrix} \sqrt{n}(\widehat{\theta}_{n,1} - \theta_{0,1}) \\ \vdots \\ \sqrt{n}(\widehat{\theta}_{n,k} - \theta_{0,k}) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_\theta) \quad \text{as } n \to \infty, \tag{4.6.5}$$

where

$$\Sigma_\theta = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} \\ \vdots & \ddots & \vdots \\ \Sigma'_{k1} & \cdots & \Sigma_{kk} \end{pmatrix}. \tag{4.6.6}$$

For $(d, c) \in \{1, .., k\}$, the sub-matrices of (4.6.6) are given by

$$\Sigma_{dc} = J_{dd}^{-1} I_{dc} J_{cc}^{-1\prime},$$

where

$$\begin{aligned} I_{dc} &= \lim_{n\to\infty} E\left(U_{d,t}U'_{c,t}\right) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=s+1}^{n} \frac{\epsilon_{d,t}(\theta_{0,d})}{\lambda_{d,t}(\theta_{0,d})} \frac{\partial\lambda_{d,t}(\theta_{0,d})}{\partial\theta_d} \frac{\epsilon_{c,t}(\theta_{0,c})}{\lambda_{c,t}(\theta_{0,t})} \frac{\partial\lambda_{c,t}(\theta_{0,c})}{\partial\theta'_c}. \end{aligned} \tag{4.6.7}$$

The existence of I_{dc} is straightforward under **A8**. Indeed, the Cauchy-Schwarz inequality in (4.6.7) yields

$$\left\| \frac{\partial L_{n,d}(\theta_{0,d})}{\partial\theta_d} \frac{\partial L_{n,c}(\theta_{0,c})'}{\partial\theta_c} \right\|_1 \leq \left\| \frac{\partial L_{n,d}(\theta_{0,d})}{\partial\theta_d} \right\|_2 \left\| \frac{\partial L_{n,c}(\theta_{0,c})}{\partial\theta'_c} \right\|_2 < \infty.$$

□

4.7 Conclusion

The EbE-PQMLE provides a simple and general way for estimating the conditional mean parameters for a large variety of models of multivariate time series of counts. If the conditional mean is correctly specified, under some regularity assumptions, EbE-PQMLE is CAN, regardless the nature of dispersion and the sign of the cross correlation between the series. If the actual conditional distribution of the series is known, MLE is generally more efficient than EbE-PQMLE, but on the other hand, for some models such as the high order INAR models, MLE is very complicated and difficult to manipulate. The robust estimator of the asymptotic variance defined in (4.2.11) provides an estimation of the asymptotic variance using the estimated conditional mean parameters, without the need of additional parameters (*i.e.* the conditional distribution parameters). Several extensions of this work are possible, such as studying the asymptotic behaviour of EbE-PQMLE when $\theta_{0,d}$ lies at the boundary of parameter space Θ_d which can be considered for testing the significance of parameters (see Section 2.2.2. of Ahmad & Francq, 2016).

References

Ahmad, A. & Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis*, 37, 291–314.

Aitchison, J. & Ho, C. (1989). The multivariate poisson-log normal distribution. *Biometrika*, 76, 643–653.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12, 788–792.

Boucher, J.-P., Denuit, M., & Guillen, M. (2008). Models of insurance claim counts with time dependence based on generalisation of poisson and negative binomial distributions. *Variance*, 2, 135–162.

Boudreault, M. & Charpentier, A. (2011). Multivariate integer-valued autoregressive models applied to earthquake counts. *arXiv :1112.0929*.

Cheon, S., Song, S. H., & Jung, B. C. (2009). Tests for independence in a bivariate negative binomial model. *Journal of the Korean Statistical Society*, 38, 185–190.

Denuit, M. & Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93, 40–57.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81, 709–721.

Francq, C. & Zakoian, J.-M. (2010). *GARCH models : structure, statistical inference and financial applications*. Chichester : John Wiley & Sons.

Francq, C. & Zakoian, J.-M. (2016). Estimating multivariate GARCH and stochastic correlation models equation by equation. *Journal of Royal Statistics Society : Series B*, 67, 213–225.

Franke, J. & Subba Rao, T. (1993). *Multivariate first order integer valued autoregressions*. Technical report, University of Kaiserslautern, Department of Mathematics.

Heinen, A. & Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14, 564–583.

Karlis, D. & Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. *Communications in Statistics-Theory and Methods*, 42, 723–740.

Liu, H. (2012). *Some models for time series of counts.* PhD thesis, University of Colombia.

Marshall, A. W. & Olkin, I. (1990). Multivariate distributions generated from mixtures of convolution and product families. *Lecture Notes-Monograph Series*, 16, 371–393.

Nelsen, R. (2006). *An introduction to copulas, 2nd.* New York : Springer Science Business Media.

Pedeli, X. & Karlis, D. (2013). Some properties of multivariate INAR(1) processes. *Computational Statistics & Data Analysis*, 67, 213–225.

Satish, V., Saxena, A., et al. (2014). Predicting intraday trading volume and volume percentages. *The Journal of Trading*, 9, 15–25.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

Steutel, F., van Harn, K., et al. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7, 893–899.

Appendix

Bivariate copula

We quote the definition of the bivariate copula from Karlis & Pedeli (2013) : According to Nelsen (2006), a bivariate copula is a function from $[0,1]^2$ to $[0, 1]$ with the following properties.

- For every $\{u, v\} \in [0, 1]$

$$C(u, 0) = 0 = C(0, v) \text{ and } C(u, 1) = u, \quad C(1, v) = v.$$

- For every $\{u_1, u_2, v_1, v_2\} \in [0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, u_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

If $F(x)$ and $G(y)$ are the cumulative distribution functions (cdf's) of the univariate random variables X and Y, then $C(F(x), G(y))$ is a bivariate distribution for (X, Y) with marginal distribution F and G, respectively. Conversely, If H is a bivariate cdf with marginals cdf's F and G, then, according to Sklar (1959)'s theorem, there exists a bivariate copula C such that for all (X, Y), $H(x, y) = C(F(x), G(y))$. If F and G are continuous, then C is unique, otherwise, C is uniquely determined on *range* F $\times$ *range* G. This lack of uniqueness is not a problem in the practical applications as it implies that there may exist two copulas with identical properties.

Other simulation results

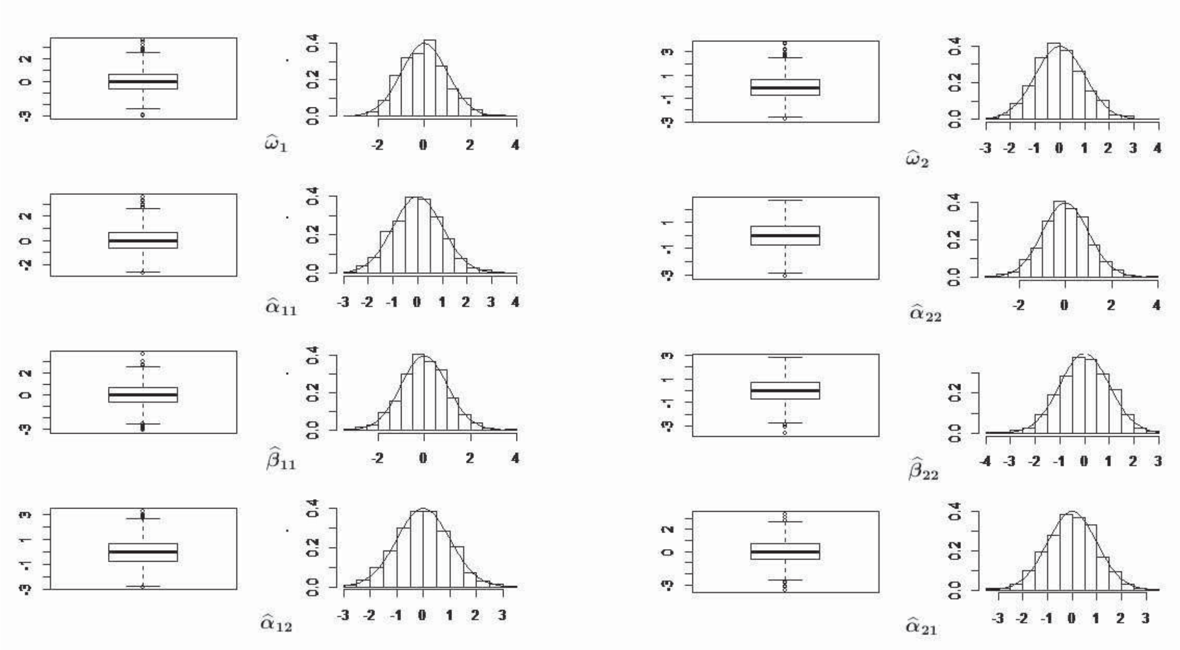

Figure 4.6 – Boxplot and histogram of the standardized distributions of EbE-PQMLE when $\theta_{01} = (\omega_{01}, \alpha_{011}, \beta_{011}, \alpha_{012})$= (2, 0.2, 0.4, 0.3) and $\theta_{02} = (\omega_{02}, \alpha_{022}, \beta_{022}, \alpha_{021})$= (4, 0.3, 0.4, 0.2) for a copula based bivariate INGARCH$(1, 1)$ model with Poisson conditional distributions and σ_C= -0.5. The sample size n= 3000. Superimposed is the standard normal density function.

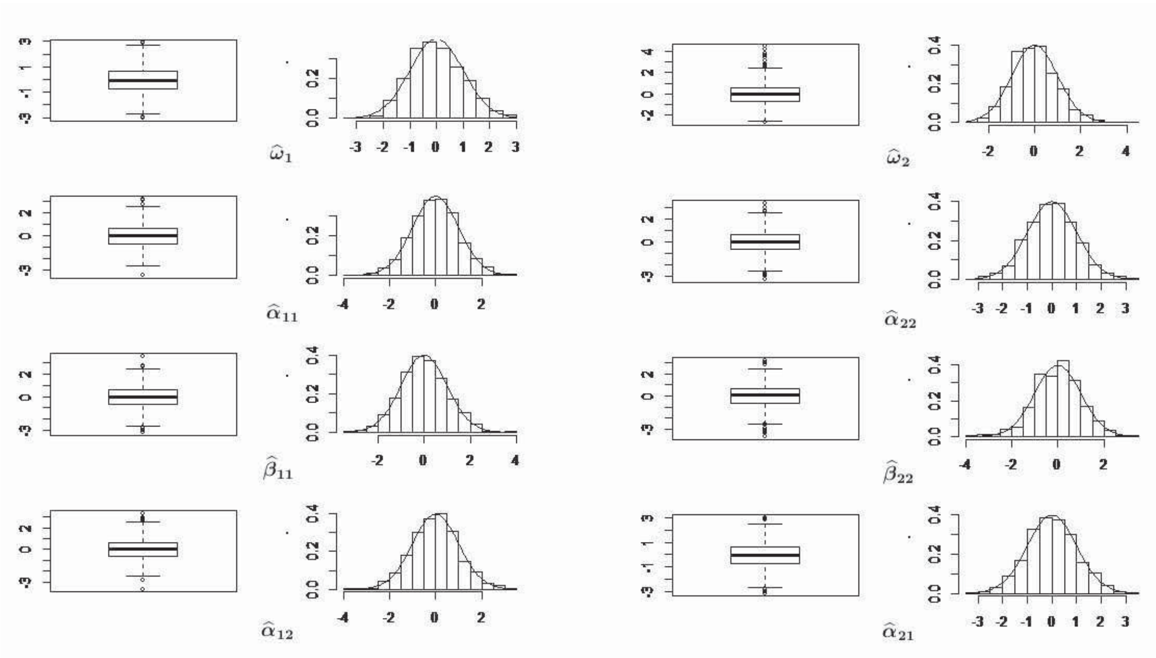

FIGURE 4.7 – Boxplot and histogram of the standardized distributions of EbE-PQMLE when $\theta_{01} = (\omega_{01}, \alpha_{011}, \beta_{011}, \alpha_{012})$= (2, 0.2, 0.4, 0.3) and $\theta_{02} = (\omega_{02}, \alpha_{022}, \beta_{022}, \alpha_{021})$= (4, 0.3, 0.4, 0.2) for a copula based bivariate INGARCH$(1, 1)$ model with negative binomial conditional distributions when (ν_1, ν_2)= (3, 4) and σ_C= -0.5. The sample size n= 3000. Superimposed is the standard normal density function.

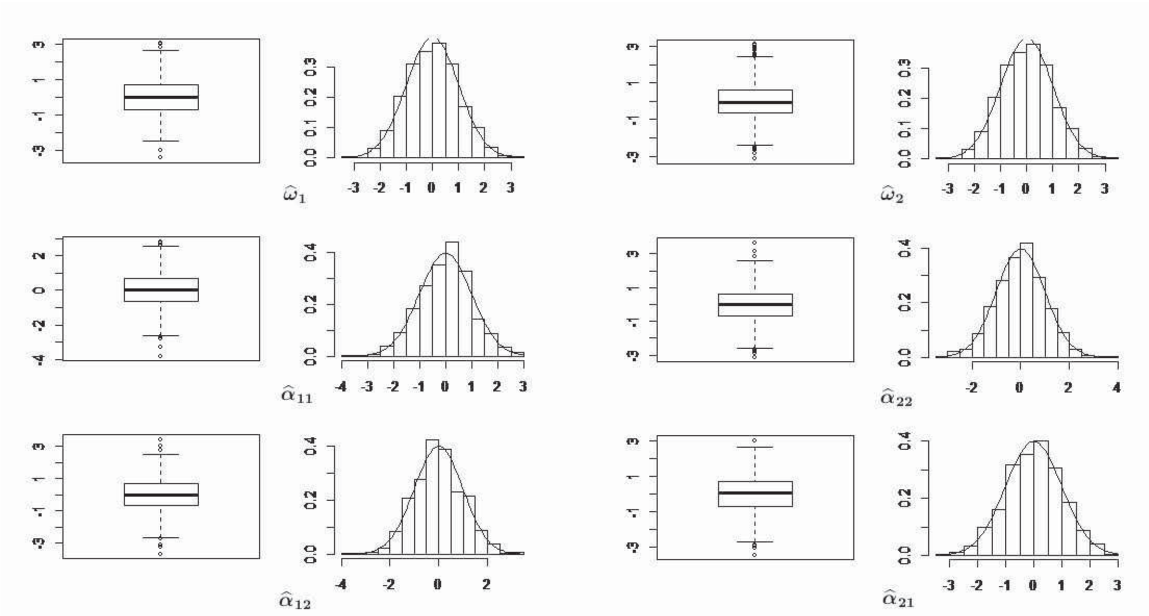

FIGURE 4.8 – Boxplot and histogram of the standardized distribution of EbE-PQMLE for the bivariate INAR(1) model when $\theta_{01} = (\omega_{01}, \alpha_{11}, \alpha_{12})$= (2, 0.2, 0.4) and $\theta_{02} = (\omega_{02}, \alpha_{022}, \alpha_{021})$= (4.5, 0.3, 0.4). The marginal distributions of the innovations are negative binomial with respectively parameters (ν_1, p_1)=(2,0.5) and (ν_2, p_2)=(3,0.6). The two marginals are connected thought a Gaussian copula with parameter σ_C= -0.5. The sample size n= 3000. Superimposed is the standard normal density function.

Chapitre 5

Conclusion générale

Dans cette thèse, nous avons étudié des modèles pour les moyennes conditionnelles des séries temporelles à valeurs entières. Sous des hypothèses générales de régularité, nous avons établi la consistance et la normalité asymptotique de l'EQMVP des paramètres de la moyenne conditionnelle pour des grandes classes de modèles. Par ailleurs, nous avons étudié la distribution asymptotique de l'EQMVP quand le paramètre est sur le bord de l'espace des paramètres et avons proposé des versions modifiées du test de Wald pour la significativité des paramètres et pour la moyenne conditionnelle constante.

Ensuite, nous avons abordé le problème de la validation des modèles des séries temporelles à valeurs entières. Nous avons proposé un test portmanteau qui a pour but de vérifier que les résidus des modèles estimés sont bien des estimations de bruits blancs. Dans un premier lieu, nous avons dérivé la distribution jointe de l'EQMVP et les fonctions des autocovariances empiriques des résidus. Ceci nous permet ensuite d'obtenir la distribution asymptotique des fonctions des autocovariances résiduelles estimées, et alors la statistique du test.

Enfin, nous avons proposé d'utiliser l'EQMVP pour estimer les paramètres de la moyenne conditionnelle des séries temporelles multivariées à valeurs entières équation-par-équation (EpE). Sous des conditions proches de celles requises pour l'EQMVP, nous avons établi la consistance et la normalité asymptotique de l'EQMVP-EpE et avons appliqué les résultats aux exemples des modèles multivariés linéaires.

Le travail réalisé dans cette thèse comporte de nombreuses extensions possibles. Par exemple, les séries temporelles à valeurs entières différenciées pourraient comporter des valeurs négatives. Autrement dit, elles seraient évaluées dans $\mathbb{Z}$. Cette situation ne se conforme pas aux hypothèses que nous avons données pour la consistance et la normalité asymptotique de l'EQMVP. Donc, il serait utile de rechercher un EQMV adapté à un tel cas.

Une autre voie de recherche est d'étudier dans quelle mesure le test de portmanteau, proposé dans le chapitre 3, pourrait être utilisé pour évaluer l'adéquation de l'ajustement des modèles des séries temporelles multivariées à valeurs entières. Surtout, la méthodologie d'estimation proposée dans chapitre 4 (EQMVP-EpE), offre un moyen général et simple pour estimer les paramètres de tels modèles.

Le chapitre 4 pourrait être complété en étudiant la distribution asymptotique de l'EQMVP-EpE quand le paramètre est sur le bord de l'espace des paramètres. De manière similaire au cas univarié, cette situation pourrait être prise en compte pour tester la nullité des paramètres des modèles multivariés.