



HAL
open science

Modèles de séries temporelles multivariées non-linéaires avec régresseurs exogènes

Zinsou Max Debaly

► To cite this version:

Zinsou Max Debaly. Modèles de séries temporelles multivariées non-linéaires avec régresseurs exogènes. Statistiques [math.ST]. Ecole Nationale de la Statistique et de l'Analyse de l'Information [Bruz], 2022. Français. <NNT : 2022NSAIM002>. <tel-03888646>

HAL Id: tel-03888646

<https://theses.hal.science/tel-03888646v1>

Submitted on 7 Dec 2022

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

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



HAL Authorization

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE DE LA STATISTIQUE
ET DE L'ANALYSE DE L'INFORMATION

ÉCOLE DOCTORALE N° 601

Mathématiques et Sciences et Technologies

de l'Information et de la Communication

Spécialité : *Mathématiques et leurs interactions*

Par

Zinsou Max DEBALY

**Modèles de séries temporelles multivariées non-linéaires avec ré-
gresseurs exogènes**

Thèse présentée et soutenue à ENSAI, le 11/10/2022

Unité de recherche : UMR CNRS 9194

Rapporteurs avant soutenance :

Andréas HEINEN Professeur des universités, Université de Cergy Pontoise (78)

Michael H. NEUMANN Professeur, Friedrich Schiller Universität (Iéna - Allemagne)

Composition du Jury :

Président : Paul DOUKHAN Professeur des universités, Université de Cergy Pontoise (78)

Examineurs : Paul DOUKHAN Professeur des universités, Université de Cergy Pontoise (78)

Marie-Pierre ETIENNE Maître de conférences, Agrocampus Ouest (35)

Andréas HEINEN Professeur des universités, Université de Cergy Pontoise (78)

Michael H. NEUMANN Professeur, Friedrich Schiller Universität (Iéna - Allemagne)

Laurence REBOUL Maître de conférences, Faculté de Luminy (13)

Dir. de thèse : Lionel Truquet Enseignant-Chercheur au CREST-ENSAI

REMERCIEMENTS

Je tiens à remercier en premier lieu mon directeur de thèse Lionel Truquet. Merci pour la grande gentillesse que tu as toujours eu à mon égard, pour ta patience, ta générosité et ta grande disponibilité. Merci d'avoir relu à de nombreuses reprises mes différentes productions. Merci à Adrien Saumard pour cet échange ce jeudi après midi durant le trajet Beaulieu-République qui a débouché sur ma rencontre avec Lionel Truquet. Je remercie également Myriam Vimond pour son accompagnement pour mon projet méthodologique en Master 2. Merci aussi à Valentin Patiléa pour tous ses conseils et ses mots encourageants.

Je remercie également mes rapporteurs de thèse Andréas HEINEN et Michael H. NEUMANN et les autres membres de mon jury Marie-Pierre ETIENNE, Laurence REBOUL et Paul DOUKHAN.

C'est avec un réel plaisir que j'ai effectué ma thèse à l'ENSAI, où il règne une atmosphère conviviale. Mes remerciements vont aussi à l'ensemble des enseignants-chercheurs qui font vivre le laboratoire pour leur accueil et leur bienveillance ; au personnel administratif, notamment à Cécile Terrien pour son soutien dans les démarches administratives liées au doctorat et à Corinne Barzic et Aurélie Duchesne pour l'organisation de mes heures de TD. Je tiens également à remercier Basile de Loynes, Matthieu Marbac Lourdelle, Francois Coquet et Salima EL Kolei de m'avoir fait confiance pour encadrer des TD.

Ce projet doctorale n'aurait pas été possible sans tous ces formateurs que j'ai eu la chance de rencontrer depuis le collège et qui ont su me transmettre leur amour des mathématiques. Un grand merci à tout le corps professoral intervenu dans le cycle de formation ITS à l'ENSAE-Sénégal entre 2013 et 2017 en particulier au responsable des enseignements Dr. Souleymane Fofana.

Merci aux autres doctorants de l'ENSAI : Edouard, Steven, Amandine, Camille 1, Camille 2, Elie, Sunny, Daphné, Guillaume 1, Guillaume 2, Omar et Assan. Ces trois années de thèse n'auraient pas été les mêmes sans vous. Merci à ma famille et à mes amis pour leur soutien durant ces trois années, en particulier Yves, Judicaël et Caleb.

TABLE OF CONTENTS

Synthèse des travaux	9
Introduction	11
1 A note on the stability of multivariate non-linear time series with an application to time series of counts	19
1.1 Introduction	19
1.2 Main result	20
1.3 Multivariate count autoregressions	21
1.3.1 Poisson Linear models	21
1.3.2 Poisson log-linear models	23
1.3.3 Multivariate GINAR(q) process	24
1.4 Proof of the main results	25
1.4.1 Proof of Theorem 1	25
1.4.2 Proof of Theorem 2	27
1.4.3 Proof of Theorem 3	31
1.4.4 Proof of Theorem 4	33
1.5 Appendix	34
2 Iterations of dependent random maps and exogeneity in nonlinear dynamics	36
2.1 Introduction	36
2.2 General results	41
2.2.1 Conditional contraction on average	41
2.2.2 Comments on Assumption A2	43
2.2.3 An additional result for strictly exogenous regressors	45
2.2.4 Example	46
2.2.5 A result for higher-order autoregressions	48
2.3 Functional dependence measure	49
2.3.1 Dependence coefficients for general iterated random functions	50

TABLE OF CONTENTS

2.3.2	Dependence coefficients for higher-order autoregressions with exogenous covariates	53
2.3.3	A central limit theorem	53
2.4	Examples	54
2.4.1	CHARN models	54
2.4.2	GARCH processes	57
2.4.3	Poisson autoregressions	58
2.4.4	Dynamic binary choice model	59
2.4.5	Categorical time series with covariates	61
2.4.6	Categorical time series and coalescence of the paths	62
2.5	Conclusion	63
2.6	Proofs of the results	64
2.6.1	Proof of Theorem 6	64
2.6.2	Proof of Theorem 7	66
2.6.3	Proof of Proposition 1	68
2.6.4	Proof of Theorem 8	69
2.6.5	Proof of Proposition 2	70
2.6.6	Proof of Theorem 9	71
2.6.7	Proof of Proposition 3	71
2.6.8	Proof of Proposition 4	73
2.6.9	Proof of Proposition 5	73
2.6.10	Proof of Proposition 6	75
2.6.11	Proof of Proposition 7	76
3	Multivariate time series models for mixed data	78
3.1	Introduction	78
3.1.1	Univariate observation-driven models	80
3.1.2	Extension to multivariate mixed time series models	81
3.2	Stability properties	83
3.2.1	Existence of stationary solutions	83
3.2.2	Specific results for linear type dynamics	85
3.2.3	Examples of linear dynamics	85
3.2.4	Two specific examples	89
3.3	Statistical inference	91

3.3.1	Estimation of dynamic parameters	92
3.3.2	Estimation of copula parameters	94
3.3.3	Asymptotic results for inference of autoregressive parameters	95
3.3.4	Sufficient conditions for A7 and A12	97
3.3.5	Examples	99
3.3.6	Guidelines for more general models	101
3.3.7	Asymptotic results for inference of copula parameters	103
3.4	Numerical experiments and real data applications	107
3.4.1	Numerical experiments	109
3.4.2	An application to sleep data	109
3.4.3	An application to high-frequency transactions in finance	110
3.5	Conclusion and perspectives	111
3.6	Supplementary material	113
3.6.1	Proof of Theorem 10	113
3.6.2	Proof of Proposition 9	113
3.6.3	Proof of Proposition 8	113
3.6.4	Proof of Proposition 10	114
3.6.5	Proof of Proposition 11	114
3.6.6	Proof of Lemma 4	115
3.6.7	Proof of Lemma 5	116
3.6.8	Proof of Lemma 6	116
3.6.9	Proof of Proposition 14	117
3.6.10	Proof of Proposition 12	119
3.6.11	Proof of Proposition 13	120
3.6.12	Proof of Proposition 15	120
3.6.13	Proof of Theorem 13	121
3.6.14	Proof of Corollary 2	127
3.6.15	Proof of Corollary 3	128
3.6.16	Checking Assumptions A14-A15 for Clayton copula	128
3.6.17	A multivariate binary/logINGARCH/logGARCH model	130
3.6.18	Combining a binary-logistic and an ARMA(1, 1) dynamic	132
3.6.19	Two useful lemmas	133
3.6.20	Approximation results for linear latent processes	135
3.6.21	Numerical experiments	138

TABLE OF CONTENTS

4 Autoregressive models for time series of random sums of positive variables	143
4.1 Introduction	144
4.2 Models and stability results	146
4.3 Estimation and asymptotics properties	150
4.4 Application	152
4.4.1 Simulation	152
4.4.2 Application to the white spruce growth series	153
4.5 Proofs for the main results	155
Bibliography	169

SYNTHÈSE DES TRAVAUX

Cette thèse de doctorat est constituée de quatre parties. Dans la première partie, nous considérons les processus autorégressifs multivariés et nous montrons moyennant une condition de contraction l'existence d'une unique solution stationnaire et ergodique pour les équations. Dans le chapitre 2, un résultat similaire est présenté pour des dynamiques non linéaires avec covariables. Le chapitre 3 est consacré aux modèles multivariés avec des composantes discrètes et/ou continues. Le chapitre 4 contient une application de nos différentes méthodes en écologie forestière.

Le chapitre 1 est une version étendue de l'article [30] paru dans *Statistics and Probability Letters*. Nous y considérons les processus multivariés autorégressifs $(Y_t)_{t \in \mathbb{Z}}$ où $Y_t = F(Y_{t-1}, \epsilon_t)$ avec $(\epsilon_t)_{t \in \mathbb{Z}}$ une suite de variables aléatoires *i.i.d.* Nous introduisons une condition de contraction équation par équation sur F qui permet de conclure à l'existence d'une unique distribution stationnaire à partir des résultats d'itération d'applications aléatoires de Wu and Shao [134]. Ce résultat est appliqué aux modèles multivariés de séries temporelles de comptage introduits par [57] et ont permis d'améliorer les résultats des auteurs. Les résultats établis dans cette partie ont permis d'établir l'existence de moments exponentiels de ces modèles.

Le chapitre 2 se base sur l'article [32] publié dans *Econometric Theory*. Dans cette partie, nous considérons des modèles de séries temporelles non linéaires avec des covariables. Le but de cette partie est de trouver des conditions suffisantes pour l'existence de solution stationnaire de l'équation $Y_t = F(Y_{t-1}, X_{t-1}, \epsilon_t)$. Dans ce cas, nous faisons une hypothèse d'exogénéité séquentielle sur la suite $(X_t)_{t \in \mathbb{Z}}$ i.e que la variable ϵ_t est indépendante de la filtration $\mathcal{F}_{t-1} = \sigma(X_s, Y_s, \epsilon_s; s \leq t-1)$. L'existence de solution de cette équation est établie à partir d'une condition de contraction en espérance conditionnelle uniformément sur les valeurs des covariables. Nous étudions également le coefficient de dépendance fonctionnelle de Wu (2005) de la solution à partir de la dépendance fonctionnelle de la suite de covariables.

Le chapitre 3 est basé sur l'article [33] publié dans *Bernoulli*. Il présente une approche générale pour construire des séries temporelles multivariées dont les coordonnées peuvent être à la fois discrètes et/ou continues. Pour y arriver, nous tirons profit de la théorie des

copules en appliquant des fonctions de répartitions inverses qui dépendent d'un unique paramètre réel à une suite $(U_t)_{t \in \mathbb{Z}}$ de variables aléatoires i.i.d dont la distribution est donnée par une copule. Nous donnons un résultat pour l'existence d'une solution stationnaire de l'équation qui définit le modèle multivarié mixte. De plus, nous proposons un estimateur de maximum de pseudo-vraisemblance pour les paramètres du modèle. Nous montrons que les paramètres d'autorégression sont consistents et suivent asymptotiquement une distribution gaussienne. De même, nous établissons un résultat de consistance sur la distribution conditionnelle qui correspond à la valeur estimée du paramètre de copule puisque ce paramètre n'est en général pas identifiable. Cependant, nous présentons un résultat d'identifiabilité pour la copule gaussienne.

Le chapitre 4 est une version étendue de l'article [34] paru dans *Ecological Modelling*. Nous y présentons un modèle semi-paramétrique autorégressif pour des séries temporelles à valeurs positives. Nous étudions dans ce chapitre les propriétés probabilistes de stabilité comme la stationnarité et l'ergodicité de la solution du modèle proposé. Nous avons également proposé un estimateur de quasi-maximum de vraisemblance exponentiel pour estimer les paramètres du modèle. Nous montrons que cet estimateur est consistant et asymptotiquement normal. Nous présentons également un test statistique pour l'adéquation du modèle.

INTRODUCTION

Les modèles de séries temporelles non linéaires ont fait l'objet de beaucoup de recherches ces dernières années [40, 131, 27] car ils trouvent des applications dans plusieurs domaines scientifiques. En santé publique par exemple, le nombre de nouveaux cas d'une épidémie est modélisé par des séries temporelles de comptage [79, 55]. Les modèles *Garch* [44, 58] sont utilisés en finance pour modéliser la volatilité des actifs. En économie, les modèles de séries temporelles binaires permettent d'expliquer les pics des prix de l'énergie sur les places boursières [95, 107]. Cependant, la majorité des travaux sur les séries temporelles non linéaires se sont concentrés sur les modèles univariés. Deux exceptions concernent toutefois les modèles *GARCH* et *VARMA* [94, 60].

Une suite de variables aléatoires $(Y_t)_{t \in \mathbb{Z}}$ à valeurs dans $\mathcal{Y} \subseteq \mathbb{R}^d$ (pour tout entier naturel $d > 1$) est un processus VAR (vectoriel autorégressif) si $(Y_t)_{t \in \mathbb{Z}}$ est solution de l'équation

$$Y_t = AY_{t-1} + \epsilon_t \quad (1)$$

avec A une matrice carré de d lignes et $(\epsilon_t)_{t \in \mathbb{Z}}$ une suite de variables aléatoires de moyennes nulles identiquement distribuées à valeurs dans $E \subseteq \mathbb{R}^d$ telle que la variable ϵ_t est indépendante de Y_t . La suite $(\epsilon_t)_{t \in \mathbb{Z}}$ est considérée comme le bruit du modèle. Lorsque $(\epsilon_t)_{t \in \mathbb{Z}}$ est une suite de variables indépendantes, on dit que (1) est un modèle VAR(1) pour spécifier que l'ordre maximal de retard dans (1) est 1. Dans le cas où la suite $(\epsilon_t)_{t \in \mathbb{Z}}$ est une combinaison linéaire des termes d'une suite *i.i.d.*, ie $\epsilon_t = B\eta_{t-1} + \eta_t$, B une matrice carré de d lignes et $(\eta_t)_{t \in \mathbb{Z}}$ une suite de variables aléatoires indépendantes, le modèle (1) est désigné par le terme *Vectorial Autoregressive Moving Average* d'ordre (1,1) noté VARMA(1,1). Il est également possible de définir le modèle VARMA pour des ordres de retards p et q quelconques. Cependant pour illustrer notre propos dans cette section, nous allons seulement considérer le modèle VAR(1).

Intéressons nous maintenant à un exemple de modèle multivarié non linéaire : le modèle *MINGARCH* [57]. Il est employé pour modéliser les vecteurs de comptage. En effet, les données de comptage constituent une autre classe importante de séries temporelles non linéaires. Elles ont récemment fait l'objet de beaucoup de recherches [23, 24, 86, 57]. Plusieurs auteurs ont défini des modèles de comptage qui imitent la structure linéaire de

(1). Ainsi, [57] en se basant sur les résultats des processus de comptage multivarié [80] définit le modèle *MINGARCH* comme suit.

$$Y_t = N_{\lambda_t}^{(t)}, \quad \lambda_t = \omega + \sum_{i=1}^q A_i \lambda_{t-i} + \sum_{i=1}^q B_i Y_{t-i}, \quad (2)$$

avec A_i et B_i des matrices $d \times d$ de termes positifs, ω est un vecteur de \mathbb{R}_+^d . In (2), $(N_{\cdot}^{(t)})_{t \in \mathbb{Z}}$ *i.i.d* de processus de comptage dans \mathbb{R}^d . Pour $t \in \mathbb{Z}$, $N^{(t)} = (N_1^{(t)}, \dots, N_d^{(t)})$ et pour $i = 1, \dots, p$, les processus de comptage marginaux $N_i^{(t)} = (N_{i,s}^{(t)})_{s \geq 0}$ sont des processus de Poisson d'intensité 1.

Nous pouvons remarquer qu'il est possible de définir les modèles (1) et (2) de manière générale par

$$Y_t = F(\lambda_t, \epsilon_t); \quad \lambda_t = \omega + B\lambda_{t-1} + AY_{t-1}. \quad (3)$$

Dans l'exemple (1), la matrice B et le vecteur ω sont nuls et dans (2), le bruit $(\epsilon_t)_{t \in \mathbb{Z}}$ du modèle est représenté par la suite $(N_{\cdot}^{(t)})_{t \in \mathbb{Z}}$. Selon la terminologie introduit par [9] et repris par plusieurs auteurs [47, 57, 54], lorsque la matrice B est nulle, (3) est dit de type *ARCH* et dans le cas contraire, on dit que (3) est de type *GARCH*. Mais il est tout a fait envisageable de spécifier une expression récursive générale pour le processus $(\lambda_t)_{t \in \mathbb{Z}}$ de (3) ie : $\lambda_t = g(\lambda_{t-1}, Y_{t-1}), t \in \mathbb{Z}$.

L'inclusion de covariables est également une question importante pour les modèles multivariés et surtout dans le cas des dynamiques non-linéaires. En effet, pour la grande majorité des études, le processus d'intérêt $(Y_t)_{t \in \mathbb{Z}}$ est observé conjointement avec d'autres processus $(X_t)_{t \in \mathbb{Z}}$ qui sont susceptibles de l'impacter. Pour prendre en compte cette relation, on modifie la seconde équation dans (3) en y incluant X_t comme par exemple :

$$\lambda_t = \omega + B\lambda_{t-1} + AY_{t-1} + \Gamma X_{t-1}$$

ou plus généralement

$$\lambda_t = g(\lambda_{t-1}, Y_{t-1}, X_{t-1}), t \in \mathbb{Z}$$

[68, 69, 3, 54, 107]. La difficulté lorsqu'on est amené à considérer les dynamiques avec des covariables réside dans le fait qu'il faut préciser la structure de dépendance entre les suites $(X_t)_{t \in \mathbb{Z}}$ et $(\epsilon_t)_{t \in \mathbb{Z}}$. Ceci permet de définir différentes notions d'exogénéité. Par exemple, lorsque les suites $(X_t)_{t \in \mathbb{Z}}$ et $(\epsilon_t)_{t \in \mathbb{Z}}$ sont indépendantes, le processus $(X_t)_{t \in \mathbb{Z}}$ est dit strictement exogène [119, 19]. Une deuxième notion plus faible d'exogénéité consiste à supposer que ϵ_t est indépendant de $X_s, s < t$; on dit dans ce cas que les régresseurs sont

prédéterminés.

Objectifs de la thèse

Dans cette thèse, nous considérons **des modèles de séries temporelles avec covariables exogènes, en particulier des séries temporelles multivariées présentant des composantes discrètes**. Ils peuvent par exemple servir sur le marché financier à modéliser conjointement la volatilité de prix sur un actif et le nombre de transactions réalisées sur ce dernier. Dans ce cas, il s'agit d'un modèle bivarié dont une coordonnée peut être un processus *GARCH* [44] et l'autre un processus de *INGARCH* [47]. Un autre exemple en écologie consiste à modéliser simultanément la densité d'un ravageur de culture (mesurée comme le nombre moyen de ravageurs par plant) et la perte de production enregistrée. Nous pouvons alors envisager un modèle avec une marginale de loi Beta [115, 66] et l'autre de loi de Poisson [47].

L'inclusion de covariables exogènes sera un point important de notre recherche. En effet, si en pratique, les auteurs incluent systématiquement des covariables dans les autoregressions [107, 115], il n'y a pas pour autant des travaux qui donnent des garanties théoriques sur les propriétés de stabilité des modèles. A titre d'exception nous pouvons citer Francq and Thieu [59] qui étudie les conditions de stationnarité des modèles *GARCH* lorsque le processus de bruit et celui de covariables forment un processus stationnaire. Un autre exemple porte sur les processus de Poisson autoregressif considéré par Agosto et al. [2] quand les régresseurs exogènes sont une chaîne de Markov. Nous introduisons dans ce travail un résultat général qui permet de fournir des garanties théoriques sur un grand nombre de modèles autoregressifs avec des régresseurs exogènes. Nous abordons également la question de l'estimation statistique. **Nous proposons dans ce travail une procédure d'estimation valide basée sur les estimateurs de minimum de contraste [141, 88, 124] permettant à la fois d'estimer les paramètres de l'autoregression et ceux de dépendance jointe.**

Nos différentes contributions

Ce travail est divisé en quatre parties. Dans la première partie, nous étudions plus en détail le cas particulier des modèles d'autoregression multivariés pour les séries temporelles de comptage. La deuxième partie étend le cadre de la première aux modèles avec covariables exogènes. Elle est consacrée à l'étude des propriétés de stabilité pour les pro-

cessus autoregressifs de ceux-ci. La troisième partie présente un cadre unifié pour définir des autoregressions dont les coordonnées sont de différents types. On y aborde longuement la question de l'inférence statistique. Dans la quatrième partie, nous abordons un cas d'application en écologie des modèles de séries temporelles non linéaires. Ce travail a été réalisé dans le cadre d'une collaboration avec deux chercheurs en sciences forestières de l'université de Québec au Canada. Le modèle développé a été fortement motivé par le cas d'application.

Première partie : Séries temporelles de comptage multivariées

Ce travail a été publié dans statistics and probability letters [30].

Nous donnons des garanties théoriques pour l'existence et l'unicité de solution pour des équations récursives stochastiques. Il s'agit de suites de variables aléatoires $(Y_t)_{t \in \mathbb{Z}}$ définies par des récursions de la forme

$$Y_t = F(Y_{t-1}, \varepsilon_t), \quad t \in \mathbb{Z}, \quad (4)$$

avec $F : E \times G \rightarrow E$ une fonction mesurable, $(\varepsilon_t)_{t \in \mathbb{Z}}$ une suite *i.i.d* à valeurs dans un espace mesurable G et E un ensemble de \mathbb{R}^k pour un entier naturel strictement positif k . Nous avons appliqué nos résultats à plusieurs modèles multivariés de Poisson. Nos travaux se basent sur ceux de Wu and Shao [135] combinant des techniques de contraction et des résultats sur les chaînes de Markov. La nouveauté dans ce travail consiste en l'introduction d'une technique de *pseudo*-contraction basée sur l'ordre lexicographique. Cela revient à introduire des hypothèses de contraction équation par équation sur les coordonnées du modèle. Nous retrouvons dans ce travail les mêmes conditions d'existence et d'unicité de solution pour les modèles $GINAR(p)$ que celles de [86]. Mais nos résultats ont également permis de déduire des conditions pour obtenir l'existence de certains moments de la solution. Nous arrivons à trouver des conditions nécessaires et suffisantes d'existence de solution intégrable pour les modèles $INGARCH$ introduits par [57] ainsi que des résultats sur les moments exponentiels. Nous complétons les résultats de [57] sur les modèles $log-INGARCH$ en trouvant de nouvelles conditions suffisantes pour l'existence et d'unicité de solution stationnaire ainsi que l'existence de moments exponentiels.

Deuxième partie : Modèles de séries temporelles non linéaires avec covariables exogènes

*Ce travail a été publié dans *Econometric Theory* [32].*

Ce travail se focalise sur l'inclusion de covariables dans les dynamiques non linéaires. Les équations récursives considérées ici sont de la forme :

$$Y_t = F(Y_{t-1}, X_{t-1}, \varepsilon_t), \quad t \in \mathbb{Z}, \quad (5)$$

avec $(\varepsilon_t)_{t \in \mathbb{Z}}$ un processus de bruit et $(X_t)_{t \in \mathbb{Z}}$ le processus de covariables. Dans cette partie, nous discutons les différentes notions d'exogénéité en économétrie [94] et considérons tout au long de nos travaux celle de la pré-détermination des régresseurs. L'objectif est d'obtenir des conditions suffisantes d'existence d'une solution stationnaire pour (5). Avec l'inclusion des régresseurs $(X_t)_{t \in \mathbb{Z}}$, la suite $(Y_t)_{t \in \mathbb{Z}}$ n'est plus une chaîne de Markov. Notre approche est basée sur l'étude des itérations d'applications aléatoires dans un cadre dépendant. Nous donnons un résultat général d'existence d'une solution stationnaire à partir d'une hypothèse de contraction en espérance conditionnelle. Avec un processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire et moyennant quelques conditions de régularité sur F et de moment sur ε_0 , on arrive à définir des solutions stationnaires pour (5) qui de plus présentent de bonnes propriétés probabilistes utiles à l'estimation statistique. Nos résultats ont été appliqués à plusieurs exemples dont les modèles *INGARCH* [2], *CHARN* [71] ou encore *GARCH-X* [110] et ont permis d'améliorer et de déduire de nombreux nouveaux résultats. Nous donnons également un résultat sur la manière dont les propriétés de dépendance [138] des régresseurs se répercutent sur celles de la solution de (5).

Troisième partie : Modèles de séries temporelles multivariées pour des données mixtes

*Ce travail est à paraître dans *Bernoulli* [33].*

Nous proposons ici une méthode générale pour construire des modèles multivariés dont certaines coordonnées sont de nature discrète. Cette nouvelle construction exploite les propriétés des copules [63] et généralise les travaux de [120] sur les données individuelles avec les copules gaussiennes. En effet, toutes les suites $(Y_t)_{t \in \mathbb{Z}}$ dont les familles de distributions marginales conditionnelles sont connues à l'avance $F_{1,\lambda_{1,t}}, \dots, F_{d,\lambda_{d,t}}$ et ne dépendent que

des paramètres $\lambda_{1,t}, \dots, \lambda_{d,t}$, peuvent s'écrire sous la forme :

$$Y_{i,t} = F_{i,\lambda_{i,t}}^{-1}(\varepsilon_{i,t}), 1 \leq i \leq d,$$

avec F_{i,s_i}^{-1} les fonctions de répartitions inverses des distributions marginales et $(\varepsilon_{i,t}, 1 \leq i \leq d)$ une suite *i.i.d* dont la distribution est donnée par une copule C_R qui dépend d'un paramètre R . Les paramètres s_i représentent les moyennes de distributions ou des fonctions connues de celles-ci. Dans le cadre de ce travail, nous avons supposé que le vecteur $\lambda_t = (\lambda_{1,t}, \dots, \lambda_{d,t})$ vérifie l'équation

$$\lambda_t = g_\theta(\lambda_{t-1}, Y_{t-1}, X_{t-1})$$

avec g_θ une fonction mesurable qui dépend d'un paramètre θ et $(X_t)_{t \in \mathbb{Z}}$ le processus de covariables. Le paramètre R permet de modéliser la dépendance instantanée jointe de Y_t et θ est le paramètre d'autorégression. En appliquant les résultats des deux premières parties, nous arrivons à trouver des conditions de stabilité des modèles ainsi définis. Nous proposons une procédure d'estimation en deux étapes. Premièrement, nous construisons une fonction de coût pour θ à partir d'une large classe de fonctions de coût sur les coordonnées de $(Y_t)_{t \in \mathbb{Z}}$ incluant les moindres carrées ou encore les fonctions log-vraisemblance. Celle-ci se base sur le principe de la pseudo-vraisemblance en supposant que conditionnellement au passé, les coordonnées de Y_t sont indépendantes entre elles. Sous certaines hypothèses de régularité, on montre que les paramètres ainsi estimés ont de bonnes propriétés statistiques telles que la consistance et la normalité asymptotique. Le paramètre R peut alors être estimé quand on remplace dans la vraisemblance du modèle les valeurs estimées de θ . On montre que les paramètres estimés pour la copule gaussienne ou de Clayton sont également consistants. Nous considérons pour illustrer nos résultats les modèles bivariés *GARCH/INGARCH* et *Logistic/Poisson*.

Application des modèles de séries temporelles pour des données de croissance de cernes d'arbres en périodes épidémiques

Ce travail a été publié dans Ecological modelling [34].

Ce travail a été réalisé dans le cadre d'une mobilité internationale avec des chercheurs de l'institut de recherche sur les forêts du Canada. A partir des différents résultats obtenus dans les parties précédentes, nous développons ici un nouveau modèle pour évaluer l'effet combiné du climat et de l'épidémie de la tordeuse du bourgeon d'épinette sur la croissance

de cernes de l'épinette noire au Canada. Ce modèle est fortement motivé par la nature des données et la question scientifique formulée par les forestiers. En effet, les données en notre disposition sont constituées à chaque instant de n_t surfaces de croissance radiale d'arbres *i.e* le nombre d'observations varie dans le temps. Nous avons admis ici que le nombre de sites K observés est fixe. En plus de ces données de cerne de croissance, nous disposons de données climatiques telles que les températures maximales au cours de l'été et du printemps et des données d'éco-épidémiologie telles que le niveau de défoliation des conifères dû à l'épidémie de la tordeuse de bourgeon d'épinette. Le modèle proposé s'écrit

$$Y_{k,t} = \sum_{l=1}^{n_{k,t}} \zeta_{l,k,t}, \quad 1 \leq k \leq K, \quad (6)$$

où conditionnellement à $n_{k,t}, X_{k,t}, n_{k,t}^- = (n_{k,t-s}, s \geq 1)$ et $Y_{k,t}^- = (Y_{k,t-s}, s \geq 1)$, les variables $\zeta_{l,k,t}, 1 \leq l \leq n_{k,t}$, qui représente la croissance radiale des arbres échantillonnés, sont et de même loi que $\zeta_{k,t}$ dont la moyenne est $\lambda_{k,t}$. Le processus de moyenne est donné par

$$\varphi_\delta(\lambda_{k,t}) =: \eta_{k,t} = \omega_k + \sum_{j=1}^p \alpha_j \frac{Y_{k,t-j}}{n_{k,t-j}} + \beta^\top X_{k,t}, \quad k = 1, \dots, K \text{ and } t = 1, \dots, T, \quad (7)$$

$\omega_k \in \mathbb{R}, \alpha_j \in \mathbb{R}, \beta = (\beta_1, \dots, \beta_m) \in \mathbb{R}^m$ et φ_δ est une fonction définie sur \mathbb{R}_+ qui peut dépendre d'un paramètre δ . Nous établissons dans cette partie les propriétés de stabilité de (6)-(7). Une procédure d'estimation statistique avec des bonnes propriétés asymptotiques est également fournie. Le modèle appliqué aux données de cernes de croissance a permis de tirer de nouvelles conclusions en écologie forestière utiles à l'aménagement territoriale.

A NOTE ON THE STABILITY OF MULTIVARIATE NON-LINEAR TIME SERIES WITH AN APPLICATION TO TIME SERIES OF COUNTS

Abstract. We introduce a simple criterion for studying stationarity and moments properties of some multivariate Markovian autoregressive processes, under a contracting mapping assumption. We apply our results to various multivariate models for count data included the Poisson INGARCH model and to one of its multivariate extension recently introduced in the literature. In particular, we obtain optimal stationarity conditions and existence of some exponential moments.

Long version of [31]

1.1 Introduction

The aim of this note is to give a new criterion for the existence of stationary solutions for q -order autoregressive processes of the form

$$X_t = F(X_{t-1}, \dots, X_{t-q}, \varepsilon_t), \quad t \in \mathbb{Z}, \quad (1.1)$$

where $F : E^q \times G \rightarrow E$ is a measurable function, $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of *i.i.d.* random variables taking values in a measurable space G and E is a subset of \mathbb{R}^k for some positive integer k . In particular, we provide a set of sufficient conditions for applying a result given in Wu and Shao [135] for the convergence of iterated random maps. The originality of our approach is to consider Lipschitz type conditions obtained equation by equation. Our result, which has an independent interest, will be particularly useful for studying the

classical Poisson INGARCH model and a multivariate extension of this model introduced recently in Fokianos et al. [57]. In Section 1.2, we state and prove our general result. In Section 1.3, we give an application to the aforementioned models and sharpen existing results by providing optimal stationarity conditions and existence of some exponential moments. The proofs of our main results are postponed to Section 1.4. Finally, two technical lemmas are provided in an appendix section.

1.2 Main result

In what follows, for any positive integer n , we denote by $|\cdot|_1$ the ℓ_1 -norm on \mathbb{R}^n , i.e. $|x|_1 = \sum_{i=1}^n |x_i|$ for $x \in \mathbb{R}^n$. We also denote by \mathcal{M}_n the set of square matrices with real coefficients and n rows and if $A \in \mathcal{M}_n$, $\rho(A)$ the spectral radius of the matrix A . An element x of \mathbb{R}^n will be identified with the column vector of its coordinates in the canonical basis and a cartesian product of type $(\mathbb{R}^p)^n$ will be identified with \mathbb{R}^{pn} . Moreover, for $x \in \mathbb{R}^n$ and $r \in \mathbb{R}_+$, the vector $(|x_1|^r, \dots, |x_n|^r)$ of \mathbb{R}^n will be denoted by $|x|_{vec}^r$. Finally, we introduce the usual partial ordering on \mathbb{R}^n such that $x \preceq x'$ means $x_i \leq x'_i$ for $i = 1, \dots, n$.

To state our main result, the following assumptions will be needed.

A1 For any $y \in E^q$, $\mathbb{E}[|F(y, \varepsilon_0)|_1] < \infty$.

A2 There exist some matrices $A_1, \dots, A_q \in \mathcal{M}_k$ with nonnegative elements, satisfying $\rho(A_1 + \dots + A_q) < 1$ and such that for $y, y' \in E^q$,

$$\mathbb{E}[|F(y, \varepsilon_1) - F(y', \varepsilon_1)|_{vec}] \preceq \sum_{i=1}^q A_i |y_i - y'_i|_{vec}.$$

A3 For an integer $\bar{k} > 1$, there exists a vector $\phi := (\phi_1, \dots, \phi_{\bar{k}})$ of continuous functions from $E \rightarrow \mathbb{R}_+$, a real number $r \geq 1$ and some matrices $D_1, \dots, D_q \in \mathcal{M}_{\bar{k}}$ with nonnegative elements such that $\rho(D_1 + \dots + D_q) < 1$ and $c \in \mathbb{R}_+^{\bar{k}}$ such that for $y \in E^q$,

$$\|\phi(F(y, \varepsilon_1))\|_{r,vec} \preceq c + \sum_{i=1}^q D_i \phi(y_i),$$

where for a random vector $Z = (Z_1, \dots, Z_{\bar{k}})$, $\|Z\|_{r,vec} := (\mathbb{E}^{1/r}[|Z_1|^r], \dots, \mathbb{E}^{1/r}[|Z_{\bar{k}}|^r])$.

Theorem 1 *Let Assumptions A1-A2 hold true.*

1. There then exists a unique stationary and non-anticipative process $(X_t)_{t \in \mathbb{Z}}$ solution of (1.1) such that $\mathbb{E}[|X_0|_1] < \infty$.
2. If in addition, **A3** holds true, then $\mathbb{E}[|\phi(X_0)|_1^r] < \infty$.

By non-anticipative, we mean that X_t is measurable with respect to $\sigma(\varepsilon_s : s \leq t)$.

1.3 Multivariate count autoregressions

In what follows, we denote by $|\cdot|_2$ the Euclidean norm on \mathbb{R}^n and by $|\cdot|_\infty$ the infinity norm. The corresponding operator norms (the matrix norms) will be denoted in the same way.

1.3.1 Poisson Linear models

We consider the multivariate count autoregression introduced recently by Fokianos et al. [57].

It is defined by

$$Y_t = N_{\lambda_t}^{(t)}, \quad \lambda_t = d + \sum_{i=1}^q A_i \lambda_{t-i} + \sum_{i=1}^q B_i Y_{t-i}, \quad (1.2)$$

where the A_i 's and the B_i 's are $p \times p$ matrices of nonnegative elements, d is a vector of \mathbb{R}_+^p . In (1.2), $(N^{(t)})_{t \in \mathbb{Z}}$ is a sequence of independent and identically distributed p -dimensional count processes. In particular, for any $t \in \mathbb{Z}$, we have $N^{(t)} = (N_1^{(t)}, \dots, N_p^{(t)})$ and for $i = 1, \dots, p$, we assume that the univariate count process $N_i^{(t)} = (N_{i,s}^{(t)})_{s \geq 0}$ is a Poisson process with intensity 1. A general construction of point processes of this type is given in Fokianos et al. [57]. Going back to (1.2) and setting $Y_t = (Y_{1,t}, \dots, Y_{p,t})$, we then have $Y_{i,t} = N_{i,\lambda_{i,t}}^{(t)}$.

Let us note that when $p = 1$, the process coincides with the INGARCH model developed by [47] and Fokianos et al. [52] and for which the conditional distribution of Y_t given past values is a Poisson distribution with random intensity λ_t . For $p > 1$, a particular case of (1.2) is obtained when $N^{(t)}$ is a vector of independent Poisson processes. Fokianos et al. [57] provided a more general approach using copula. While all the coordinates of $N^{(t)}$ are still Poisson processes with intensity 1, they can have a quite general dependence structure. However, all the results given in Fokianos et al. [57] about existence of stationary solutions and their marginal moments are independent from this dependence structure.

This will be also the case for the results given in the present paper. Finally note that if the orders of the lag values λ_{t-i} and Y_{t-i} are different, for example ℓ and k , the order q in (1.2) will be set to $q = \max(\ell, k)$, setting $A_i = 0$ or $B_j = 0$ if $i > \ell$ or $j > k$.

The following result provides a necessary and sufficient condition for existence of a stationary solution for (1.2).

Theorem 2 1. Assume that $\rho(\sum_{i=1}^q(A_i + B_i)) < 1$.

(a) There then exists a unique non-anticipative, stationary and integrable solution $(Y_t)_{t \in \mathbb{Z}}$ for (1.2).

(b) For any $r > 1$, we have $\mathbb{E}[|Y_t|^r] < \infty$.

(c) In contrast, assume that $\sum_{i=1}^q(|A_i|_1 + |B_i|_1) < 1$ or $|\sum_{i=1}^q(A_i + B_i)|_\infty < 1$. There then exists $\delta > 0$ such that $\mathbb{E}[\exp(\delta|Y_0|_1)] < \infty$ and $\mathbb{E}[\exp(\delta|\lambda_0|_1)] < \infty$.

2. Conversely, if (1.2) admits a stationary integrable solution and all the components of d are positive, then $\rho(\sum_{i=1}^q(A_i + B_i)) < 1$.

Notes

- When $q = 1$, Fokianos et al. [57] provided various sufficient conditions for existence of a stationary solution for (1.2). In contrast, Theorem 2 provides an optimal condition for stationarity. In particular, the condition on the spectral radius $\rho(\sum_{i=1}^q(A_i + B_i)) < 1$ is implied by any contraction condition of the form $|\sum_{i=1}^q(A_i + B_i)| < 1$, where $|\cdot|$ is a matrix norm. We recall that for any operator norm $|\cdot|$ and any matrix A of size $p \times p$, we have $|A| \geq \rho(A)$. Contraction conditions derived from the matrix norms $|\cdot|_1$ or $|\cdot|_2$ are used in Fokianos et al. [57]. To enlighten the difference, consider the case $q = 1$, $A_1 = 0$ and $B_1 = \begin{pmatrix} \alpha & \beta \\ 0 & \alpha \end{pmatrix}$. Conditions $|B_1|_1 < 1$ or $|B_1|_\infty < 1$ means $\alpha + \beta < 1$, while $|B_1|_2 \geq \sqrt{\alpha^2 + \beta^2}$. In contrast $\rho(B_1) = \alpha$ and the condition $\alpha < 1$ is a substantial improvement of the restrictions obtained from contractions with respect to the previous norms.
- The last point of Theorem 2 provides conditions for existence of some exponential moments. Assume that $p = 1$. In the univariate case, the various contraction conditions are equivalent to $\sum_{i=1}^q(A_i + B_i) < 1$ which is an optimal condition for existence of a stationary and integrable solution. Under this assumption, existence of polynomial moments have been widely discussed in the literature. See for instance Ferland et al. [47] or Fokianos et al. [52]. Theorem 2 provides a stronger

result by showing that this stationarity condition is sufficient for existence of some exponential moments.

3. It is also possible to show that Theorem 2 is still valid if we work with the alternative autoregressive process

$$Y_{i,t} = F_{i,\lambda_{i,t}}^{-1}(U_{i,t}), \quad 1 \leq i \leq p; \quad \lambda_t := (\lambda_{1,t}, \lambda_{2,t})' = \omega + AY_{t-1} + B\lambda_{t-1} \quad (1.3)$$

where $F_{i,\lambda}^{-1}$ stands for the inverse of the cumulative probability function of Poisson distribution of mean λ and $(U_t = (U_{i,t})_{1 \leq i \leq p})_{t \in \mathbb{Z}}$ is a sequence of independent random vectors following an arbitrary copula distribution.

4. Other types of multivariate count autoregressive models can be found in the literature. See for instance Cui and Zhu [23] and [24]. In both papers, the authors consider some specific versions of bivariate count distributions to construct their models. Applying our results in this case requires to find an autoregressive structure of the form (1.2) or (1.3), that is a bivariate point process or a copula density in accordance with these bivariate distributions. We did not find a satisfying answer for this. But note that the two constructions are of different nature. While the bivariate point process or the copula approach can be used to generate models with Poisson marginals and general simultaneous dependence properties, there are no guarantees that the family of conditional distributions obtained in this way contain some existing versions of bivariate count distributions.

1.3.2 Poisson log-linear models

We now consider a second model called log-linear in the literature. See in particular Fokianos and Tjøstheim [54] for the univariate case and Fokianos et al. [57] for the multivariate case. In the multivariate case, the model is defined similarly to (1.2) except that

$$\lambda_t = \exp(\mu_t), \quad \mu_t = d + \sum_{j=1}^q A_j \mu_{t-j} + \sum_{j=1}^q B_j \log(\mathbf{1} + Y_{t-j}). \quad (1.4)$$

Here, the functions \exp and \log are applied component-wise and the matrices A_j, B_j can now have negative elements. As before, we adopt the convention of column vectors.

Theorem 3 *Consider the log linear model (1.4).*

1. Assume that $\rho(\sum_{i=1}^q (|A_i|_{vec} + |B_i|_{vec})) < 1$. Then there exists a unique non anticipative, stationary and integrable process $(Y_t)_{t \in \mathbb{Z}}$ with $\mathbb{E}(|Y_t|_1) < \infty$.
2. Assume that $|\sum_{i=1}^q (|A_i|_{vec} + |B_i|_{vec})|_\infty < 1$. There then exists $\delta > 0$ such that $\mathbb{E}(\exp(\delta|Y_0|_1)) < \infty$ and $\mathbb{E}(\exp(\delta|\lambda_0|_1)) < \infty$.

Notes

1. To compare our results with that of Fokianos et al. [57], we assume $q = 1$. Using contraction properties of autoregressive processes, Fokianos et al. [57] used the condition $|A_1|_1 + |B_1|_1 < 1$ for studying existence of a stationary solution for (1.4). Our condition in point 1 of Theorem 3 is weaker. On the other hand, using perturbation methods, Fokianos et al. [57] showed that one can approximate the stationary solution from an ergodic Markov chain when $|A_1|_2 + |B_1|_2 < 1$. When $p = 1$, the latter condition coincides with ours but when $p > 1$ they cannot be compared. Our condition is only guaranteed to be weaker when the matrices A_1 and B_1 have nonnegative coefficients. We then provide a different result which complements the existing ones.
2. We obtain directly existence of exponential moments for the solution using a contraction condition for the norm $|\cdot|_\infty$. We are not aware of such result even in the univariate case. Fokianos et al. [57] only studied existence of polynomial moments but once again their assumptions are based on the norm $|\cdot|_1$ and $|\cdot|_2$ and cannot be compared directly with ours, even for nonnegative matrices. Nevertheless, for $p = 1$, all the conditions are equivalent to $|A_1| + |B_1| < 1$, and we improve existing results by showing existence of exponential moments.

1.3.3 Multivariate GINAR(q) process

This model has been studied by Latour [86] and using Theorem 1, we will recover many results but also get additional moment properties. The model writes

$$X_t = \sum_{j=1}^q A_{t,j} \circ X_{t-j} + U_t, \quad t \in \mathbb{Z}, \quad (1.5)$$

where for $x \in \mathbb{N}^p$,

$$A_{t,j} \circ x = \left(\sum_{\ell=1}^p A_{t,j}(i, \ell) \circ x_\ell \right)_{1 \leq i \leq p}$$

and for $y \in \mathbb{N}$,

$$A_{t,j}(i, \ell) \circ y = \sum_{s=1}^y Y_s^{t,j,i,\ell}.$$

The latter operator \circ is called the thinning operator. We assume that $(U_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. integrable random vectors in \mathbb{N}^p and independent from the family

$$\left\{ Y_s^{t,j,i,\ell} : (s, t, j, i, \ell) \in \mathbb{Z}^2 \times \{1, \dots, q\} \times \{1, \dots, p\}^2 \right\}$$

which is itself composed of independent integrable and integer-valued random variables and such that for $(s, t, s', t', j, i, \ell) \in \mathbb{Z}^4 \times \{1, \dots, q\} \times \{1, \dots, p\}^2$, $Y_s^{t,j,i,\ell}$ and $Y_{s'}^{t',j,i,\ell}$ have the same distribution with mean $A_j(i, \ell)$. When $q = 1$, this process coincides with a Galton-Watson process with immigration. Note that for a process $(X_t)_{t \in \mathbb{Z}}$ defined by (1.5), we have

$$\mathbb{E}[X_t | X_{t-1}, \dots, X_{t-q}] = \sum_{j=1}^q A_j X_{t-j} + \mathbb{E}[U_0].$$

We will use the two following assumptions.

G1 The spectral radius of the matrix $A_1 + \dots + A_q$ is less than 1.

G2 There exists $r > 1$ such that for all $(j, i, \ell) \in \{1, \dots, q\} \times \{1, \dots, p\}^2$, $Y_0^{0,j,i,\ell}$ and U_0 have a moment of order r .

Theorem 4 *Assume that Assumption G1 holds true. There then exists a unique stationary, non-anticipative and integrable solution to the recursions (1.5). If in addition, Assumption G2 is valid, we have $\mathbb{E}|X_0|_1^r < \infty$.*

Note. We obtain the same result as Latour [86] for the existence of a square integrable stationary solution for the recursions (1.5). However, using our formalism, we avoid lengthy computations to check such results. We also provide conditions for existence of a moment of arbitrary order $r > 1$, a problem not investigated in Latour [86].

1.4 Proof of the main results

1.4.1 Proof of Theorem 1

1. Define the following random map $f_t(u_1, \dots, u_q) = (F(u_1, \dots, u_q, \varepsilon_t), u_1, \dots, u_{q-1})$ and the sigma fields $\mathcal{F}_t = \sigma(\varepsilon_s : s \leq t)$, $t \in \mathbb{Z}$. We first note that a process $(X_t)_{t \in \mathbb{Z}}$

satisfies (1.1) if and only if the process $(U_t)_{t \in \mathbb{Z}}$ defined by $U_t = (X_t, \dots, X_{t-q+1})$ satisfies the recursions $U_t = f_t(U_{t-1})$, $t \in \mathbb{Z}$. It then only remains to study existence of stationary solutions for the recursions defined by the random functions f_t , $t \in \mathbb{Z}$. We set $u = (x_1, \dots, x_q) \in E^q$ and for $1 \leq t \leq q$, $X_t(u) = x_{q-t+1}$. Next for $t \geq q+1$, we define $X_t(u)$ recursively by

$$X_t(u) = F(X_{t-1}(u), \dots, X_{t-q}(u), \varepsilon_t).$$

We then have for $t \geq q+1$, $(X_t(u), \dots, X_{t-q+1}(u)) = f_{q+1}^t(u) := f_t \circ \dots \circ f_{q+1}(u)$. Using our assumptions, we have for $t \geq q+1$,

$$\mathbb{E}[|X_t(u) - X_t(u')|_{vec} | \mathcal{F}_{t-1}] \preceq \sum_{i=1}^q A_i |X_{t-i}(u) - X_{t-i}(u')|_{vec}.$$

Setting $w_t = \mathbb{E}[|X_t(u) - X_t(u')|_{vec}]$ for $t \geq 1$, we have $w_t \preceq \sum_{i=1}^q A_i w_{t-i}$ for $t \geq q+1$. From Lemma 1 given in the Appendix, there exist constants $C > 0$ and $\rho \in (0, 1)$ such that $|w_t|_1 \leq C\rho^t |u - u'|_1$. We then get

$$\mathbb{E}\left[|f_{q+1}^t(u) - f_{q+1}^t(u')|_1\right] = \sum_{j=0}^{q-1} |w_{t-j}|_1 \leq qC\rho^{t-q+1} |u - u'|_1.$$

Using **A1**, we then deduce that the random maps satisfies the geometric moment contraction condition of Theorem 2 in Wu and Shao [135] and the existence of an integrable and stationary solution for the recursions $U_t = f_t(U_{t-1})$, $t \in \mathbb{Z}$, follows. Note that we also have $U_t = (X_t, X_{t-1}, \dots, X_{t-q+1}) = \lim_{n \rightarrow \infty} f_{t-n}^t(u)$ a.s. and the solution $(U_t)_{t \in \mathbb{Z}}$ and then the solution $(X_t)_{t \in \mathbb{Z}}$ of (1.1) writes as a Bernoulli shift, i.e. $X_t = \lim_{n \rightarrow \infty} H(\varepsilon_t, \varepsilon_{t-1}, \dots)$ so it is non-anticipative.

Next, let $(Y_t)_{t \in \mathbb{Z}}$ be a another integrable non-anticipative stationary solution of (1.1). We have for any $t \in \mathbb{Z}$,

$$\mathbb{E}[|X_t - Y_t|_{vec}] \preceq \sum_{j=1}^q A_j \mathbb{E}[|X_{t-j} - Y_{t-j}|_{vec}].$$

Using Lemma 1, we deduce that $\mathbb{E}[|X_t - Y_t|_{vec}] = 0$ and then $X_t = Y_t$ a.s. This shows the uniqueness.

2. Using the notations of the previous point, we have

$$\|\phi(X_t(u))\|_{r,vec} \preceq c + \sum_{i=1}^q D_i \|\phi(X_{t-i}(u))\|_{r,vec}.$$

Using Lemma 1 (2.) and the triangular inequality, we get

$$\mathbb{E}^{1/r} [|\phi(X_t(u))|_1^r] \leq C\bar{\rho}^t \sum_{i=1}^q |\phi(x_i)|_1 + D,$$

for some constants $C, D > 0$ and $\bar{\rho} \in (0, 1)$ and only depending on c, D_1, \dots, D_q . Remembering that $f_{q+1}^t(u) = (X_t(u), \dots, X_{t-q+1}(u)) := U_t(u)$ and setting $V(u) = (\sum_{i=1}^q |\phi(x_i)|_1)^r$, we have, using Fatou's lemma and the fact that $f_{q+1}^t(u)$ and $f_{-t+q+1}^0(u)$ have the same probability distribution,

$$\mathbb{E}(V(U_0)) \leq \liminf_{t \rightarrow \infty} \mathbb{E}(V(U_t(u))) \leq (qD)^r,$$

which leads to the result. \square

1.4.2 Proof of Theorem 2

1. We set $E = \mathbb{N}^p \times \mathbb{R}_+^p$.

(a) We first note that any solution $X_t = (Y_t, \lambda_t)$ of the problem satisfies the recursions

$$X_t = F(X_{t-1}, \dots, X_{t-q}, N^{(t)}),$$

where $\forall j = 1, \dots, q, x_j = (y_j, s_j) \in E$,

$$F(x_1, \dots, x_q, N^{(t)}) = (N_{f(x_1, \dots, x_q)}^{(t)}, f(x_1, \dots, x_q)),$$

$$f(x_1, \dots, x_q) = d + \sum_{j=1}^q A_j s_j + \sum_{j=1}^q B_j y_j.$$

Going back to the definition of the model (1.1), we can consider that G equals to H^p , where H is the subset of the cadlag functions $D(\mathbb{R}_+, \mathbb{R}_+)$ that are piecewise constant and take integer values. For $x \in E^q$, we have the equality

$$\mathbb{E} \left[|F(x, N^{(1)})|_1 \right] = 2\mathbf{1}^T \left(d + \sum_{j=1}^q A_j s_j + \sum_{j=1}^q B_j y_j \right) < \infty,$$

where $\mathbb{1}^\top$ is the transpose of the column vector for which all the coordinates are equal to 1. Moreover, for $x, x' \in E^q$ with $x = (x_1, \dots, x_q), x' = (x'_1, \dots, x'_q)$, where $x_j = (y_j, s_j)$ and $x'_j = (y'_j, s'_j)$ for $1 \leq j \leq q$, we have

$$\mathbb{E} \left[\left| F(x, N^{(1)}) - F(x', N^{(1)}) \right|_{vec} \right] \preceq \sum_{j=1}^q \begin{pmatrix} B_j & A_j \\ B_j & A_j \end{pmatrix} |x_j - x'_j|_{vec}.$$

One can notice that the matrices $\Gamma = \sum_{j=1}^q \begin{pmatrix} B_j & A_j \\ B_j & A_j \end{pmatrix}$ and $\sum_{j=1}^q (A_j + B_j)$ have the same spectral radius (see lemma 3). The result then follows from Theorem 1.

- (b) Let $\delta > 0$ such that $(1 + \delta) \sum_{j=1}^q (A_j + B_j)$ has a spectral radius less than one. If $\Gamma_\delta := (1 + \delta) \sum_{j=1}^q \begin{pmatrix} B_j & A_j \\ B_j & A_j \end{pmatrix}$, then we also have $\rho(\Gamma_\delta) < 1$. Next, from Lemma 2 given in the Appendix, there exists $b > 0$ such that

$$\|N_{f(x_1, \dots, x_q)}^{(t)}\|_{r, vec} \preceq (1 + \delta) |f(x_1, \dots, x_q)|_{vec} + b\mathbb{1},$$

where $\mathbb{1}$ denotes the vector of \mathbb{R}^p for which all the coordinates are equal to 1. Since

$$|f(x_1, \dots, x_q)|_{vec} \preceq d + \sum_{j=1}^q B_j |y_j|_{vec} + \sum_{j=1}^q A_j |s_j|_{vec} \preceq d + \sum_{j=1}^q (B_j \ A_j) |x_j|_{vec},$$

we then get

$$\begin{aligned} \|F(x_1, \dots, x_q, N^{(1)})\|_{r, vec} &\preceq \begin{pmatrix} (1 + \delta)d + b\mathbb{1} \\ d \end{pmatrix} + \sum_{j=1}^q \begin{pmatrix} (1 + \delta)B_j & (1 + \delta)A_j \\ B_j & A_j \end{pmatrix} |x_j|_{vec} \\ &\preceq \begin{pmatrix} (1 + \delta)d + b\mathbb{1} \\ d \end{pmatrix} + (1 + \delta) \sum_{j=1}^q \begin{pmatrix} B_j & A_j \\ B_j & A_j \end{pmatrix} |x_j|_{vec}. \end{aligned}$$

Since $\rho(\Gamma_\delta) < 1$, Theorem 1 (3.) applied when ϕ is the identity leads to the result.

- (c) Assume first that $\gamma := \sum_{i=1}^q (|A_i|_1 + |B_i|_1) < 1$. Let $\delta > 0$ to be chosen later

and $\phi(y, s) = (\exp(\delta|y|_1), \exp(\delta|s|_1))$ for $y, s \in \mathbb{R}^p$. From convexity of the exponential function and matrix norm inequalities, we can write if $t \geq q + 1$:

$$\begin{aligned} \mathbb{E} [\exp(\delta|f(x_1, \dots, x_q)|_1)] &\leq \exp \left(\delta|d|_1 + \sum_{j=1}^q [\delta|B_j|_1|y_j|_1 + \delta|A_j|_1|s_j|_1] \right) \\ &\leq c + \sum_{j=1}^q [|B_j|_1 \exp(\delta|y_j|_1) + |A_j|_1 \exp(\delta|s_j|_1)], \end{aligned}$$

with $c = (1 - \gamma) \exp\left(\frac{\delta|d|_1}{1-\gamma}\right)$. Furthermore, from Hölder inequality, we have setting $\bar{\lambda} = f(x_1, \dots, x_q)$,

$$\mathbb{E} \left[\exp(\delta|N_{\bar{\lambda}}^{(1)}|_1) \right] = \mathbb{E} \left[\prod_{j=1}^p \exp \left(\delta N_{j, \bar{\lambda}_j}^{(1)} \right) \right] \leq \prod_{j=1}^p \mathbb{E}^{1/p} \left[\exp \left(p \delta N_{j, \bar{\lambda}_j}^{(1)} \right) \right].$$

Note that

$$\prod_{j=1}^p \mathbb{E}^{1/p} \left[\exp \left(p \delta N_{j, \bar{\lambda}_j}^{(1)} \right) \right] = \prod_{j=1}^p \left(\exp \left(\bar{\lambda}_j [\exp^{p\delta} - 1] \right) \right)^{1/p} = \exp \left(|\bar{\lambda}|_1 \frac{\exp^{p\delta} - 1}{p} \right).$$

But $\frac{\exp^x - 1}{x} \downarrow 1$ as $x > 0$ tends to 0. If $\epsilon > 0$ is such that $(1 + \epsilon)\gamma < 1$, let us choose $\delta = \delta(\epsilon)$ such that $\exp^{p\delta} - 1 \leq (1 + \epsilon)\delta p$. For the couple $(\epsilon, \delta = \delta(\epsilon))$, one can write :

$$\begin{aligned} \mathbb{E} \left[\exp(\delta|N_{\bar{\lambda}}^{(1)}|_1) \right] &\leq \exp \left(|\bar{\lambda}|_1 (1 + \epsilon)\delta \right) \\ &\leq c' + \sum_{j=1}^q [(1 + \epsilon)|A_j|_1 \exp(\delta|s_j|_1) + (1 + \epsilon)|B_j|_1 \exp(\delta|y_j|_1)], \end{aligned}$$

with $c' = (1 - (1 + \epsilon)\gamma) \exp\left(\frac{\delta(1+\epsilon)|d|_1}{1-(1+\epsilon)\gamma}\right)$. The second inequality follows from the convexity of the exponential function, as previously. We then obtain

$$\begin{aligned} \mathbb{E} \left[\phi \left(F(x, N^{(1)}) \right) \right] &\preceq \begin{pmatrix} c' \\ c \end{pmatrix} + \sum_{j=1}^q \begin{pmatrix} (1 + \epsilon)|B_j|_1 & (1 + \epsilon)|A_j|_1 \\ |B_j|_1 & |A_j|_1 \end{pmatrix} \phi(x_j) \\ &\preceq \begin{pmatrix} c' \\ c \end{pmatrix} + (1 + \epsilon) \sum_{j=1}^q \begin{pmatrix} |B_j|_1 & |A_j|_1 \\ |B_j|_1 & |A_j|_1 \end{pmatrix} \phi(x_j). \end{aligned}$$

Since the eigenvalues of $\Gamma := \sum_{j=1}^q \begin{pmatrix} |B_j|_1 & |A_j|_1 \\ |B_j|_1 & |A_j|_1 \end{pmatrix}$ are γ and 0, the spectral radius of $(1 + \epsilon)\Gamma$ is less than 1. The result then follows from Theorem 1 (2.). Next, we assume that $\gamma := |\sum_{i=1}^q (A_i + B_i)|_\infty < 1$. Some arguments previously used yield $\forall k = 1, \dots, p$,

$$\mathbb{E} \left[\exp(\delta \bar{\lambda}_k) \right] \leq e_k + \sum_{j=1}^q \sum_{l=1}^p [B_j(k, l) \exp(\delta y_{\ell, j}) + A_j(k, l) \exp(\delta s_{\ell, j})],$$

$$\mathbb{E} \left[\exp(\delta N_{k, \bar{\lambda}_k}^{(1)}) \right] \leq \exp(\delta(1+\epsilon)\bar{\lambda}_{k,t}) \leq e'_k + (1+\epsilon) \sum_{j=1}^q \sum_{l=1}^p [A_j(k, l) \exp(\delta s_{\ell, j}) + B_j(k, l) \exp(\delta y_{\ell, j})],$$

with $e_k = (1 - \gamma_k) \exp\left(\frac{\delta d_k}{1 - \gamma_k}\right)$, $e'_k = (1 - (1 + \epsilon)\gamma_k) \exp\left(\frac{\delta(1+\epsilon)d_k}{1 - (1+\epsilon)\gamma_k}\right)$ and $\gamma_k = \sum_{j=1}^q \sum_{l=1}^p [A_j(k, l) + B_j(k, l)]$ where $(\epsilon, \delta = \delta(\epsilon))$ satisfy $(1 + \epsilon) \sup_k \gamma_k = (1 + \epsilon)\gamma < 1$ and $\exp^\delta - 1 \leq (1 + \epsilon)\delta$. Therefore,

$$\begin{pmatrix} \mathbb{E} \left[\exp(\delta N_{\bar{\lambda}}^{(1)}) \right]_{vec} \\ \mathbb{E} \left[\exp(\delta \bar{\lambda}) \right]_{vec} \end{pmatrix} \preceq \begin{pmatrix} e'_{vec} \\ e_{vec} \end{pmatrix} + (1 + \epsilon) \sum_{j=1}^q \begin{pmatrix} B_j & A_j \\ B_j & A_j \end{pmatrix} \begin{pmatrix} \exp(\delta y_j)_{vec} \\ \exp(\delta s_j)_{vec} \end{pmatrix}$$

With $M = \sum_{j=1}^q (A_j + B_j)$, condition $|M|_\infty = \max_{1 \leq j \leq q} \sum_{l=1}^q M(j, l) < 1$ ensures that the spectral radius of the matrix $\sum_{j=1}^q \begin{pmatrix} B_j & A_j \\ B_j & A_j \end{pmatrix}$ is less than 1. Without loss of generality, we may assume that ϵ also provides that

$$\Gamma'_\epsilon = (1 + \epsilon) \sum_{j=1}^q \begin{pmatrix} B_j & A_j \\ B_j & A_j \end{pmatrix}$$

is less than 1. The result then follows from Theorem 1 (2.), setting $r = 1$ and for $y, s \in \mathbb{R}^p$, $\phi((y, s)) = (\exp(\delta|y_1|), \dots, \exp(\delta|y_p|), \exp(\delta|s_1|), \dots, \exp(\delta|s_p|))$.

2. If $(Y_t)_{t \in \mathbb{Z}}$ is a stationary and integrable solution of (1.2), we have

$$\mathbb{E}(Y_t) = \mathbb{E}(\lambda_t) = d + \sum_{j=1}^q B_j \mathbb{E}(Y_{t-j}) + \sum_{j=1}^q A_j \mathbb{E}(\lambda_{t-j}).$$

Setting $m = \mathbb{E}(Y_t)$, we have $m = d + Em$ with $E = \sum_{j=1}^q (A_j + B_j)$. We then obtain $m = d + Ed + \dots + E^{n-1}d + E^n m$ for any integer $n \geq 1$. Since all the quantities

are non negative, the series $\sum_{i=0}^{\infty} E^i d$ is convergent line by line. This implies that $\lim_{n \rightarrow \infty} E^n d = 0$. If $d_- = \min_{1 \leq i \leq p} d_i > 0$, we deduce that $E^n \rightarrow 0$, element-wise. This entails $\rho(E) < 1$. \square

1.4.3 Proof of Theorem 3

1. For any solution, setting $X_t = (\log(\mathbb{1} + Y_t)', \log(\lambda_t)')'$. We can write : $X_t = F(X_{t-1}, \dots, X_{t-q}, N^{(t)})$ with

$$F(x_1, \dots, x_q, N^{(t)}) = \left(\log(\mathbb{1} + N_{f(x_1, \dots, x_q)}^{(t)})', \log(f(x_1, \dots, x_q))' \right)',$$

where $\forall j = 1, \dots, q, x_j = (u_j, v_j) \in (\mathbb{R}_+^p \times \mathbb{R}^p)$ and

$$\log(f(x_1, \dots, x_q)) = b_0 + \sum_{j=1}^q B_j u_j + \sum_{j=1}^q A_j v_j.$$

For $x \in (\mathbb{R}_+^p \times \mathbb{R}^p)^q$, Jensen's inequality leads to the bound

$$\mathbb{E} \left[\left| F(x_1, \dots, x_q, N^{(1)}) \right|_1 \right] \preceq \mathbb{1}' \left(\left| \log(f(x_1, \dots, x_q)) \right|_{vec} + \log(\mathbb{1} + f(x_1, \dots, x_q)) \right) < \infty.$$

Using Jensen's inequality to the function $x \mapsto \log(1 + x)$ and Poisson process properties, we obtain that, for a given Poisson process N (for more details, see Fokianos and Tjøstheim [54], proof of Lemma 2.1),

$$\mathbb{E} \left(\log \left(\frac{1 + N_t}{1 + N_s} \right) \right) \leq \log(t) - \log(s).$$

For $x = (x_1, \dots, x_q), x' = (x'_1, \dots, x'_q) \in E^q$,

$$\begin{aligned} \mathbb{E} \left[\left| F(x, N^{(1)}) - F(x', N^{(1)}) \right|_{vec} \right] &= \left(\mathbb{E} \left(\left| \log \left(\frac{1 + N_{f(x)}^{(1)}}{1 + N_{f(x')}^{(1)}} \right) \right|_{vec} \right) \right) \\ &\preceq \left(\left| \log(f(x)) - \log(f(x')) \right|_{vec} \right) \\ &\preceq \sum_{j=1}^q \begin{pmatrix} |B_j|_{vec} & |A_j|_{vec} \\ |B_j|_{vec} & |A_j|_{vec} \end{pmatrix} \left| x_j - x'_j \right|_{vec}, \end{aligned}$$

where for $x \in \mathbb{R}^d$ and $y \in \mathbb{R}_*^d$, $\frac{x}{y} = (\frac{x_1}{y_1}, \dots, \frac{x_d}{y_d})'$. Note that the matrices $\Gamma =$

$\sum_{j=1}^q \begin{pmatrix} |B_j|_{vec} & |A_j|_{vec} \\ |B_j|_{vec} & |A_j|_{vec} \end{pmatrix}$ and $\sum_{j=1}^q (|A_j|_{vec} + |B_j|_{vec})$ have the same spectral radius. The result then follows from theorem 1 (1.).

2. We will use Theorem 1 (3.) with $r = 1$ and for $(u, v) \in \mathbb{R}_+^p \times \mathbb{R}^p$,

$$\phi(u, v) = (\exp(\delta |\exp(u_1) - 1|), \dots, \exp(\delta |\exp(u_p) - 1|), \exp(\delta \exp(|v_1|)), \dots, \exp(\delta \exp(|v_p|))),$$

with $\delta > 0$ to be specified latter. Setting for $x_i = (u_i, v_i) \in \mathbb{R}_+^p \times \mathbb{R}^p$ for $1 \leq i \leq p$, $\bar{\lambda} = f(x_1, \dots, x_q)$ and $\bar{\mu} = \log(\bar{\lambda})$. We have for $1 \leq k \leq p$,

$$\exp(|\bar{\mu}_k|) \leq e_k + \sum_{j=1}^q \sum_{l=1}^p |B_j(k, l)| [\exp(u_{\ell, j}) - 1] + |A_j(k, l)| \exp(|v_{\ell, j}|),$$

and

$$\mathbb{E}[\exp(\delta \exp(|\bar{\mu}_k|))] \leq e'_k + \sum_{j=1}^q \sum_{l=1}^p |B_j(k, l)| \exp(\delta [\exp(u_{\ell, j}) - 1]) + |A_j(k, l)| \exp(\delta \exp(|v_{\ell, j}|)),$$

with

$$e_k = (1 - \gamma_k) \exp\left(\frac{d_k}{1 - \gamma_k}\right) + \sum_{j=1}^q \sum_{l=1}^p |B_j(k, l)|, \quad e'_k = (1 - \gamma_k) \exp\left(\frac{\delta e_k}{1 - \gamma_k}\right)$$

and $\gamma_k = \sum_{j=1}^q \sum_{l=1}^p |A_j(k, l)| + |B_j(k, l)|$. We also have

$$\begin{aligned} \mathbb{E}\left[\exp\left(\delta N_{k, \bar{\lambda}_k}^{(1)}\right)\right] &\leq \exp(\delta(1 + \epsilon)\bar{\lambda}_k) \\ &\leq \tilde{e}_k + (1 + \epsilon) \sum_{j=1}^q \sum_{l=1}^p |B_j(k, l)| \exp(\delta [\exp(u_{\ell, j}) - 1]) + |A_j(k, l)| \exp(\delta \exp(|v_{\ell, j}|)) \end{aligned}$$

with $\tilde{e}_k = (1 - (1 + \epsilon)\gamma_k) \exp\left(\frac{\delta(1 + \epsilon)e_k}{1 - (1 + \epsilon)\gamma_k}\right)$ where $(\epsilon, \delta = \delta(\epsilon))$ satisfy $(1 + \epsilon) \sup_k \gamma_k \leq 1$ and $\exp^\delta - 1 \leq (1 + \epsilon)\delta$. Therefore,

$$\mathbb{E}\left[\phi\left(F(x, N^{(1)})\right)\right] \preceq \begin{pmatrix} \tilde{e}_{vec} \\ e'_{vec} \end{pmatrix} + (1 + \epsilon) \sum_{j=1}^q \begin{pmatrix} |B_j|_{vec} & |A_j|_{vec} \\ |B_j|_{vec} & |A_j|_{vec} \end{pmatrix} \begin{pmatrix} \exp(\delta (\exp(u_j) - \mathbf{1})) \\ \exp(\delta \exp(|v_j|_{vec})) \end{pmatrix}$$

With $M = \sum_{j=1}^q (|B_j|_{vec} + |A_j|_{vec})$, condition $|M|_\infty < 1$ ensures that the spectral

radius of the matrix $\sum_{j=1}^q \begin{pmatrix} |B_j|_{vec} & |A_j|_{vec} \\ |B_j|_{vec} & |A_j|_{vec} \end{pmatrix}$ is less than 1. Then, one can find ϵ such that the spectral radius of

$$\Gamma_\epsilon = (1 + \epsilon) \sum_{j=1}^q \begin{pmatrix} |B_j|_{vec} & |A_j|_{vec} \\ |B_j|_{vec} & |A_j|_{vec} \end{pmatrix}$$

is less than 1. Theorem 1 (3.) then leads to the result. \square

1.4.4 Proof of Theorem 4

The noise at time t , denoted by ϵ_t , is a vector with components U_t and random sequences $Y^{t,j,i,\ell}$ for $1 \leq j \leq q$, $1 \leq i, \ell \leq p$. Define $F(x_1, \dots, x_q, \epsilon_t) = \sum_{j=1}^q A_j \circ x_j + \epsilon_t$. Due to the properties of the thinning operator, we have for $x_1, \dots, x_q, y_1, \dots, y_q \in \mathbb{N}^p$,

$$\mathbb{E} |F(x_1, \dots, x_q, \epsilon_0) - F(y_1, \dots, y_q, \epsilon_0)|_{vec} \preceq \sum_{j=1}^q A_j |x_j - y_j|_{vec}.$$

This shows Assumption **A2**. Assumption **A1** is automatically satisfied. The first part of the theorem follows from Theorem 1 (1.)

Finally, we check **A3** when ϕ is the identity function. We decompose

$$F(x_1, \dots, x_q, \epsilon_0) = \sum_{j=1}^q A_j x_j + S(x_1, \dots, x_q) + U_0,$$

where $S(x_1, \dots, x_q) = \sum_{j=1}^q A_j \circ x_j - \sum_{j=1}^q A_j x_j$ is a vector of sums of independent random variables. Using Burkholder's inequality we have for $y \in \mathbb{N}$, $\|A_0^{0,j,i,\ell} \circ y - A_j(i, \ell)y\|_r \leq C y^{1/\max(r,2)}$ where $C > 0$ depends on r and \mathbb{L}^r -norm of the counting sequences. One can then take the same constant C for all the counting sequences. We denote by H the matrix $p \times p$ with all components equal to 1. For any $\varepsilon > 0$, there exists $b_{\varepsilon,r} > 0$ only depending on ε, r and such that for $y \in \mathbb{N}$, $y^{1/\max(r,2)} \leq \varepsilon y + b_{\varepsilon,r}$. We choose $\varepsilon > 0$ such that $\rho(A_1 + \dots + A_q + C\varepsilon qH) < 1$. We then obtain

$$\begin{aligned} \|F(x_1, \dots, x_q, \epsilon_0)\|_{r,vec} &\preceq \sum_{j=1}^q A_j x_j + \|S(x_1, \dots, x_q)\|_{r,vec} + \|U_0\|_{r,vec} \\ &\preceq \sum_{j=1}^q (A_j + C\varepsilon qH) x_j + qpCb_{\varepsilon,r}\mathbf{1} + \|U_0\|_{r,vec}. \end{aligned}$$

Setting $D_j = A_j + C\varepsilon qH$ for $j = 1, \dots, q$, the result follows from Theorem 1 (2.) \square

1.5 Appendix

Lemma 1 *Let E_1, \dots, E_q be square matrices of size $e \times e$, with nonnegative elements and such that $\rho(E_1 + \dots + E_q) < 1$.*

1. *We have $\rho(F) < 1$ where F denotes the companion matrix associated to E_1, \dots, E_q , i.e.*

$$F = \begin{pmatrix} E_1 & E_2 & \cdots & E_q \\ & I_{(q-1)e} & & 0_{(q-1)e,e} \end{pmatrix}.$$

2. *Let also $(v_n)_{n \geq 1}$ be a sequence of vectors of \mathbb{R}_+^e and $b \in \mathbb{R}_+^e$ such that*

$$v_n \preceq \sum_{i=1}^q E_i v_{n-i} + b, \quad n \geq q + 1.$$

Let $|\cdot|_1$ be the ℓ_1 -norm on \mathbb{R}^e . There exists $C > 0$ and $\bar{\rho} \in (0, 1)$, not depending on $(v_n)_{n \geq 1}$, such that $|v_n|_1 \leq C\bar{\rho}^n \sum_{i=1}^q |v_i|_1 + \frac{C|b|_1}{1-\bar{\rho}}$.

Proof of Lemma 1

1. Let $E = E_1 + \dots + E_q$. Since $\rho(E) < 1$, we have $E^n \rightarrow 0$. Suppose that λ is an eigenvalue of F of modulus greater than 1. If $v = (v_1, \dots, v_q) \in \mathbb{R}^{qe} \setminus \{0\}$ is such that $Fv = \lambda v$, we have $\lambda v_1 = \sum_{i=1}^q E_i v_i = \sum_{i=1}^q \lambda^{1-i} E_i v_1$. This yields

$$|v_1|_{vec} \preceq \sum_{i=1}^q |\lambda|^{-i} E_i |v_1|_{vec} \preceq E |v_1|_{vec} \preceq E^n |v_1|_{vec}.$$

Letting $n \rightarrow \infty$, we get $v_1 = 0$ and then $v_2, \dots, v_q = 0$. This contradicts $v \neq 0$. Hence $|\lambda| < 1$ and then $\rho(F) < 1$.

2. For $n \geq q$, set $u_n = (v_n, \dots, v_{n-q+1})$ and $B = (b, 0_{1,(q-1)e})$. For $n \geq q + 1$, we have

$$u_n \preceq F u_{n-1} + B \preceq F^{n-q} u_q + \sum_{i=0}^{n-q-1} F^i B,$$

with F being the companion matrix associated to the matrices E_1, \dots, E_q and which is defined in the previous point. We still denote by $|\cdot|_1$ the ℓ_1 -norm on \mathbb{R}^{eq} . From

the previous point, we have $\rho(F) < 1$ and then if $\varepsilon > 0$ is such that $\bar{\rho} = \rho(F) + \varepsilon < 1$, we have $|F^n|_1 \leq C\bar{\rho}^n$ for $C > 0$ only depending on the matrix F . Then if $n \geq q + 1$,

$$|v_n|_1 \leq |u_n|_1 \leq C\bar{\rho}^{n-q}|u_q|_1 + C \sum_{i=0}^{n-q-1} \bar{\rho}^i |B|_1.$$

Since $|B|_1 = |b|_1$ and $|u_q|_1 \leq \sum_{i=1}^q |v_i|_1$, this leads to the result. \square

Lemma 2 *Let $\lambda > 0$ and X_λ be a Poisson variable with parameter λ . Then, $\forall r \geq 1$ and any $\delta \in (0, 1)$, there exists $b_{r,\delta}$, not depending on λ and such that $\|X_\lambda\|_r \leq (1 + \delta)\lambda + b_{r,\delta}$.*

Proof of Lemma 2 We have the equality $\mathbb{E}(X_\lambda^r) = \sum_{i=1}^r \lambda^i \left\{ \begin{matrix} r \\ i \end{matrix} \right\}$ with $\left\{ \begin{matrix} r \\ i \end{matrix} \right\}$ are the Sterling's numbers of second kind. See for instance Johnson et al. [76].

Then

$$\mathbb{E}(X_\lambda^r) = \lambda^r + \sum_{i=1}^{r-1} \lambda^i \left\{ \begin{matrix} r \\ i \end{matrix} \right\} \leq \lambda^r + C_r(\lambda + \lambda^{r-1}),$$

where $C_r > 0$ only depends on r . But, we can notice that, for any $\delta > 0$, there exists $\tilde{b}_{\delta,r} > 0$ such that for all $x \geq 0$: $x + x^{r-1} \leq \delta'x^r + \tilde{b}_{\delta,r}$ with $\delta' = \frac{(1+\delta)^{r-1}}{C_r}$. Then $\mathbb{E}(X_\lambda^r) \leq (1 + C_r\delta')\lambda^r + C_r\tilde{b}_{\delta,r}$. Therefore $\|X_\lambda\|_r \leq (1 + C_r\delta')^{1/r}\lambda + C_r^{1/r}\tilde{b}_{\delta,r}^{1/r}$. Setting $b_{\delta,r} = C_r^{1/r}\tilde{b}_{\delta,r}^{1/r}$, we get the result. \square

Lemma 3 *Let E_1, E_2 be square matrices of size $e \times e$. The matrices $\begin{pmatrix} E_1 & E_2 \\ E_1 & E_2 \end{pmatrix}$ and $E_1 + E_2$ have the same nonzero eigenvalues.*

Prof of lemma 3 Let us denote $x = (x_1, x_2) \in \mathbb{R}^{2e}$ with $x_1, x_2 \in \mathbb{R}^e$ an eigenvector of the matrix $E = \begin{pmatrix} E_1 & E_2 \\ E_1 & E_2 \end{pmatrix}$ and λ a nonzero eigenvalue of E . Straightforwardly, $E_1x_1 + E_2x_2 = \lambda x_1$ and $E_1x_1 + E_2x_2 = \lambda x_2$. These equations entail $(E_1 + E_2)x_1 = \lambda x_1$ since λ is nonzero. \square

ITERATIONS OF DEPENDENT RANDOM MAPS AND EXOGENEITY IN NONLINEAR DYNAMICS

Abstract. We discuss the existence and uniqueness of stationary and ergodic nonlinear autoregressive processes when exogenous regressors are incorporated into the dynamic. To this end, we consider the convergence of the backward iterations of dependent random maps. In particular, we give a new result when the classical condition of contraction on average is replaced with a contraction in conditional expectation. Under some conditions, we also discuss the dependence properties of these processes using the functional dependence measure of Wu (2005) that delivers a central limit theorem giving a wide range of applications. Our results are illustrated with CHARN models, GARCH processes, count time series, binary choice models and categorical time series for which we provide many extensions of existing results.

Based on [32]

2.1 Introduction

Among the various contributions devoted to time series analysis, theoretical results justifying stationarity and ergodicity properties of some standard stochastic processes when exogenous covariates are incorporated in the dynamic are rather scarce. A notable exception concerns linear models, such as VARMA processes, for which such properties are a consequence of the linearity. See for instance [93], a standard reference for multivariate time series models. Moreover, linear models represent a very simple setup for discussing various exogeneity notions found in the literature. See for instance [46]. For nonlinear dynamics, a few contributions consider the problem of exogenous regressors. For general GARCH type processes, [59] recently studied stationarity conditions when the noise and

the covariate process form a stationary process. [2] considered a Poisson autoregressive process with exogenous regressors (PARX models), under a Markov chain assumption for the covariate process. [28] consider the case of dynamic binary choice models and provide results about stationarity and mixing properties of a 0/1-valued time series which is autoregressive and defined conditionally on some exogenous regressors. [51] studied stationarity and ergodicity of general categorical time series defined conditionally on a strictly exogenous covariate process.

In this paper, we give general results for getting stationarity, ergodicity and stochastic dependence properties for general nonlinear dynamics defined in terms of iterations of random maps. For simplicity, we explain our setup with the following example which represents the basis for studying other processes. Let us consider the following model

$$X_t = F(X_{t-1}, Z_{t-1}, \varepsilon_t), \quad t \in \mathbb{Z}, \quad (2.1)$$

where $(Z_t)_{t \in \mathbb{Z}}$ is a covariate process and $(\varepsilon_t)_{t \in \mathbb{Z}}$ a noise process. One can note that $X_t = f_t(X_{t-1})$ for the random function defined by $f_t(x) = F(x, Z_{t-1}, \varepsilon_t)$. The sequence $(f_t)_{t \in \mathbb{Z}}$ is a sequence of dependent random maps even if the ε_t 's are i.i.d. because typically the Z_t 's exhibit temporal dependence. A key point for getting existence of a stationary solution in (2.1) is to control the behavior of the backward iterations $\{f_t \circ f_{t-1} \circ \dots \circ f_{t-n}(x) : n \geq 1\}$. The convergence of such iterations of random maps has been extensively studied in the independent case. In this case, the process $(X_t)_{t \in \mathbb{Z}}$ is a Markov chain. We defer the reader to [89] and [36] for seminal papers on iterated independent random maps and to [135] for additional results useful in a time series context. The last contribution is particularly interesting for getting existence of some moments for the marginal X_t and also some dependence properties for the process $(X_t)_{t \in \mathbb{Z}}$ that are often needed for statistical applications. All these contributions use average contraction conditions and the interested reader is referred to the interesting survey of [123] for an overview of the available results. There also exist some contributions studying the more general case of iterated stationary random maps $(f_t)_{t \in \mathbb{Z}}$. For instance, [10] gives many results for studying what he calls stochastically recursive sequences, when the independence assumption is removed. See also [74] for a survey of some available results. The results obtained in the dependent case are based on Lyapunov type exponents and the convergence of the backward iterations is only studied almost surely. We recall the following result which can be found in [43] (see also [74], Theorem 6.2) and which generalizes a widely known result given in [14] or [11]

for iterations of affine random maps.

To do so, we introduce some notations and conditions. We assume that $f_t : E \rightarrow E$ are random Lipschitz functions where E denotes a locally compact Polish space endowed with a metric d . We define the Lipschitz constant of a measurable function $g : E \rightarrow E$ by

$$c(g) := \sup_{x \neq y \in E} \frac{d(g(x), g(y))}{d(x, y)}.$$

Moreover, for any integers $s < t$, we set $f_s^t = f_t \circ \dots \circ f_s$. For a positive real number x , we set $\log^+(x) = \log(x)$ if $x \geq 1$ and 0 otherwise.

Theorem 5 *Assume that the process $((Z_t, \varepsilon_t))_{t \in \mathbb{Z}}$ in (2.1) is stationary and ergodic. Assume further that $\mathbb{E}[\log^+ c(f_0)] < \infty$ and $\mathbb{E}[\log^+ d(x_0, f_0(x_0))] < \infty$ for some point $x_0 \in E$.*

1. *There exists a constant $\chi \in \mathbb{R} \cup \{-\infty\}$ called Lyapunov exponent and such that*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log c(f_1^n) = \chi \text{ a.s.}$$

Moreover

$$\chi = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\log c(f_1^n)] = \inf_{n \geq 1} \frac{1}{n} \mathbb{E}[\log c(f_1^n)].$$

2. *If the constant χ is negative, then the almost sure limit $f_{-\infty}^t = \lim_{k \rightarrow \infty} f_{t-k}^t(x)$ exists for any $x \in E$ and does not depend on x . Setting $X_t = f_{-\infty}^t$, the process $(X_t)_{t \in \mathbb{Z}}$ is stationary and ergodic and satisfies the recursions (2.1). Moreover, $(X_t)_{t \in \mathbb{Z}}$ is the unique stationary process satisfying (2.1).*

The affine random maps version of this result has been applied recently by [59] for studying stationarity of asymmetric power GARCH processes. For nonlinear random maps, Theorem 5 is less known in the time series literature. In this paper, we will make use of Theorem 5 for defining a general class of categorical time series with exogenous covariates. In particular, we will see in Section 2.4.4 how Theorem 5 can be applied to binary time series and lead to an improvement of a result of [28].

However, the result presented above has several limitations.

1. First, it requires the random maps f_t to be almost surely Lipschitz. Such a property is not always valid, for instance for the Poissonian autoregressions discussed in Section

2.4.3. When there are no exogenous covariates, [26] studied integer-valued time series by using a different contraction result, developed by [135].

2. Existence of some moments for the marginal distributions that are sometimes necessary for statistical applications cannot be obtained directly from this result.

3. For autoregressions with several lags, it is not straightforward to get an explicit condition on the parameters of the model to ensure that $\chi < 0$.

To overcome these drawbacks, we will adapt the approach used by [135] for independent random maps to the case of dependent random maps. Our main result, see Theorem 6 and its extension Theorem 8, is obtained by replacing the usual contraction on average condition by a contraction in conditional expectation. The assumptions that we use are very simple to check and the proof of our main result is straightforward but its merit is to provide an elegant way for presenting a general approach which encompasses most of the previous attempts to include exogenous regressors in nonlinear dynamics. For strictly exogenous regressors, i.e. the processes $(Z_t)_{t \in \mathbb{Z}}$ and $(\varepsilon_t)_{t \in \mathbb{Z}}$ are independent, we also provide an additional result, see Theorem 7, with weaker assumptions. In the context of Theorem 6 and Theorem 8, we will then discuss how to control the functional dependence measure of $(X_t)_{t \in \mathbb{Z}}$ introduced by [134], a dependence notion which is an alternative to the standard strong mixing condition and which can be more easily checked for iterations of contracting random maps. Let us mention that even in the independent case, mixing properties of the process $(X_t)_{t \in \mathbb{Z}}$ require restrictive assumptions on the noise distribution, otherwise such properties may fail. We refer the reader to the standard textbook of [41], section 2.4 for mixing properties of iterations of independent random functions and to [5] for a famous counterexample of a non strongly mixing sequence defined via iterations of random maps. In the dependent case, as in (2.1), getting usual strong mixing properties seems to be harder because the process $(X_t)_{t \in \mathbb{Z}}$ does not have a Markov structure in general and the criteria for getting mixing properties of Markov chains are useless.

This paper is mainly motivated by dynamics of type (2.1) with covariates that are not necessarily strictly exogenous, assuming that at any time t , the noise ε_t is independent from the past information $\sigma((Z_s, \varepsilon_s) : s \leq t - 1)$. The term predetermindness is sometimes used in the literature. This independence assumption is substantially weaker than the independence between the two processes $(\varepsilon_t)_{t \in \mathbb{Z}}$ and $(Z_t)_{t \in \mathbb{Z}}$. The latter independence condition implies strict exogeneity, a notion initially defined by [119] and extended to

general models by [19]. Strict exogeneity is useful for deriving the conditional likelihood of the X'_t s conditionally on the Z'_t s. However, strict exogeneity is a rather strong assumption. Under additional regularity conditions on the model, [19] has shown that this assumption is equivalent to the non Granger-causality, i.e. Z_t is independent of $(X_s)_{s \leq t}$ conditionally on $(Z_s)_{s \leq t-1}$. It roughly means that the covariate process $(Z_t)_{t \in \mathbb{Z}}$ evolves in a totally autonomous way. In contrast, our exogeneity condition allows general covariates of the form $Z_t = H(\eta_t, \eta_{t-1}, \dots)$ with H a measurable function and a sequence $((\eta_t, \varepsilon_t))_{t \in \mathbb{Z}}$ of i.i.d. random vectors, ε_t being possibly correlated with η_t . The error ε_t can then still have an influence on future values of the covariates. For linear models, the two technical independence conditions discussed above between the noise and the covariate processes are often used as a distinction between weak and strict exogeneity. See for instance [93], Section 10.2. Let us mention that there exist additional concepts of exogeneity that are introduced and discussed in [46], in particular a notion of weak exogeneity. However, this notion is related to the estimation of a specific parameter of the conditional distribution for the bivariate process (X_t, Z_t) and it is necessary to specify the joint dynamic of the process. Since we do not want to consider specific dynamics for the covariate process, we will not use it in this paper. Inclusion of exogenous regressors motivates our approach which is based on conditional average contraction conditions. But our results can be also applied without referring to these concepts of exogeneity, i.e. when $((Z_{t-1}, \varepsilon_t))_{t \in \mathbb{Z}}$ is a general stationary and ergodic process in (2.1). However, in the latter case, a closed form expression for the conditional distribution of X_t given \mathcal{F}_{t-1} cannot be obtained directly from the recursions (2.1). Our contribution is then the first one presenting a general framework for inclusion of covariates in nonlinear dynamics. Our results can be applied to any statistical procedure which require either ergodic properties or the use of some limit theorems developed from the notion of functional dependence introduced by [134]. Since the existing literature already contains many asymptotic results of this type, we do not discuss specific applications.

The rest of this paper is organized as follows. In Section 2.2, we give our main results for defining stationary and ergodic solutions for recursions of type (2.1). In Section 2.3, we study weak dependence properties of the process using the functional dependence measure of [134]. Many examples of nonlinear time series models satisfying our assumptions are given in Section 2.4 and we revisit some nonlinear dynamics discussed recently in the literature but we also consider new ones. A conclusion is given in Section 2.5. The proofs of our results are postponed to the last section of the paper.

2.2 General results

In this section, we state several results for controlling the convergence of the backward iterations in some \mathbb{L}^p spaces. We recall that for a random variable X and a real number $p \geq 1$, the quantity $\|X\|_p = \mathbb{E}^{1/p}(|X|^p)$ is called the \mathbb{L}^p -norm of X . Now let $(f_t)_{t \in \mathbb{Z}}$ be a sequence of random maps defined on a Polish space (E, d) and taking values in the same space. We assume for convenience that $f_t = F(\cdot, \zeta_t)$ where $(\zeta_t)_{t \in \mathbb{Z}}$ is a stochastic process taking values in another Polish space E' and $F : E \times E' \rightarrow E$ is a measurable map. In connection with our initial example (2.1), we have $\zeta_t = (Z_{t-1}, \varepsilon_t)$. In the latter case, we will assume throughout the paper that $E' = E'_1 \times E'_2$ where E'_1 is a Borel subset of \mathbb{R}^e and E'_2 is another Polish space.

For $s < t$, we set $f_s^t = f_t \circ f_s^{t-1}$ with the convention $f_t^t = f_t$ and $f_t^{t-1}(x) = x$. Moreover, we consider a filtration $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ for which $(\zeta_t)_{t \in \mathbb{Z}}$ is adapted.

2.2.1 Conditional contraction on average

We first give a general and useful result for getting a convergence in some \mathbb{L}^p spaces. For some real numbers $p \geq 1$, $L > 0$, $\kappa \in (0, 1)$ and an integer $m \geq 1$, we consider the two following assumptions.

A1 There exists $x_0 \in E$ such that $\sup_{t \in \mathbb{Z}} \mathbb{E}[d^p(f_t(x_0), x_0)] < \infty$.

A2 For every $t \in \mathbb{Z}$, almost surely, the following inequalities hold for every $(x, y) \in E^2$.

$$\mathbb{E}[d^p(f_t(x), f_t(y)) | \mathcal{F}_{t-1}] \leq L^p d^p(x, y) \text{ and } \mathbb{E}[d^p(f_t^{t+m-1}(x), f_t^{t+m-1}(y)) | \mathcal{F}_{t-1}] \leq \kappa^p d^p(x, y).$$

Theorem 6 *Suppose that Assumptions A1-A2 hold.*

1. For every $(x, t) \in E \times \mathbb{Z}$, there exists a E -valued random variable $X_t(x)$ such that

$$\sup_{t \in \mathbb{Z}} \|d(X_t(x), x_0)\|_p < \infty, \quad \sup_{t \in \mathbb{Z}} \|d(f_{t-s}^t(x), X_t(x))\|_p = O(\kappa^{s/m}).$$

Moreover the sequence $(f_{t-s}^t(x))_{s \geq 0}$ converges almost surely to $X_t(x)$.

2. For $x \neq y$, we have $\mathbb{P}(X_t(x) \neq X_t(y)) = 0$. We then set $X_t = X_t(x)$.
3. The process $((X_t, \zeta_t))_{t \in \mathbb{Z}}$ is stationary and also ergodic if the process $(\zeta_t)_{t \in \mathbb{Z}}$ is itself stationary and ergodic.

4. If $(Y_t)_{t \in \mathbb{Z}}$ is a non-anticipative process (i.e. $Y_t \in \mathcal{F}_t$) such that $Y_t = f_t(Y_{t-1})$ for $t \in \mathbb{Z}$ and $\sup_{j \in \mathbb{Z}} \mathbb{E}[d^p(Y_j, x_0)] < \infty$, then $Y_t = X_t$ a.s.

Notes

1. The bounds given in Assumption **A2** are required to hold for all $(x, y) \in E^2$ at the same time. Since a conditional expectation is only unique up to a set with measure 0 for the probability measure \mathbb{P} , the bound given in Assumption **A2** has to be understood in term of regular conditional distribution, i.e. there exists a regular version of the conditional distribution of $(\zeta_t, \dots, \zeta_{t+m-1})$ given \mathcal{F}_{t-1} . On Polish spaces, a regular version always exists. See [77], Chapter 5.
2. When $\zeta_t = (Z_{t-1}, \varepsilon_t)$ forms a stationary process and $\mathcal{F}_t = \sigma((Z_j, \varepsilon_j) : j \leq t)$, Theorem 6 guarantees existence and uniqueness of a stationary process possessing a moment of order p and solution of (2.1). However, stationarity of the covariate/error process is not required for applying this result. In particular, when $(\zeta_t)_{t \in \mathbb{Z}}$ is non stationary, one can still define solutions of (2.1) provided that **A1-A2** are satisfied. In general, these solutions will be non stationary and the classical law of large numbers is not valid. In this case, studying the asymptotic properties of some classical inferential procedures such as conditional likelihood estimation requires a specific analysis.
3. Setting $p = 1$ and $d(x, y) = |x - y|^o$ for $x, y \in E = \mathbb{R}$ and some $o \in (0, 1)$, one can consider stochastic recursions (2.1) with heavy-tailed covariate processes.
4. When the process $(\zeta_t)_{t \in \mathbb{Z}}$ is stationary and the recursions are initialized at time $t = 0$ with a given state $x \in E$, the probability distribution of the forward iterations $f_1^t(x)$ coincides with the probability distribution of the backward iterations $f_{-t+1}^0(x)$. Since $\lim_{t \rightarrow \infty} f_{-t+1}^0(x) = X_0$ a.s., $f_1^t(x)$ converges in distribution to X_0 . The same property holds true when the iterations are initialized with a random variable \bar{X}_0 independent from $(\zeta_t)_{t \geq 1}$. The main interest of the convergence of the backward iterations is to define the good random initialization $X_0 = f_{-\infty}^0(x)$ in order to get a stationary process $(X_t)_{t \geq 0}$.
5. Assumption **A2** is a conditional contraction property in \mathbb{L}^p which is crucial for getting the convergence of the backward iterations in \mathbb{L}^p norms. Relaxing this assumption by introducing a random coefficient κ_{t-1} instead of κ can be problematic for getting a similar result. We discuss this point below.

2.2.2 Comments on Assumption A2

Let us consider the dynamic (2.1), set $f_t(x) = F(x, Z_{t-1}, \varepsilon_t)$, with $E = \mathbb{R}$, $p = 1$ and assume that for every t , ε_t is independent from $\mathcal{F}_{t-1} = \sigma((Z_j, \varepsilon_j) : j \leq t-1)$. Then Assumption **A2** is satisfied for $m = 1$ if and only if there exists $\kappa \in (0, 1)$ such that

$$\sup_{z \in E_1^t} \mathbb{E} |F(x, z, \varepsilon_0) - F(y, z, \varepsilon_0)| \leq \kappa |x - y|.$$

At a first sight, the latter condition is quite strong and it is natural to wonder if the following weaker assumption can be used, i.e. there exists a measurable function $\kappa : G \rightarrow (0, \infty)$ such that for every $z \in G$,

$$\mathbb{E} |F(x, z, \varepsilon_0) - F(y, z, \varepsilon_0)| \leq \kappa(z) |x - y|.$$

Of course, the challenging question concerns the convergence of the backward iterations when the function κ may take values larger than 1. However, a problem occurs for applying the successive contraction properties to the iterated random maps. Consider the iterations $f_t \circ f_{t-1}$. We have for $(x, y) \in E^2$,

$$\begin{aligned} \mathbb{E} |f_t \circ f_{t-1}(x) - f_t \circ f_{t-1}(y)| &= \mathbb{E} [\mathbb{E} [|f_t \circ f_{t-1}(x) - f_t \circ f_{t-1}(y)| \mathcal{F}_{t-1}]] \\ &\leq \mathbb{E} [\kappa(Z_{t-1}) |f_{t-1}(x) - f_{t-1}(y)|]. \end{aligned}$$

If the random variable $\kappa(Z_{t-1})$ depends on past values of the error ε_{t-j} , $j \geq 1$, it is stochastically dependent on the random map f_{t-1} and also on \mathcal{F}_{t-2} . It is then not possible to use the contraction property of f_{t-1} unless the function κ can be bounded by a constant. To show that the successive iterations lose the memory with respect to initialization, this constant has to be smaller than 1. Of course, this does not prove that the convergence of the iterations in \mathbb{L}^1 is not possible.

To show that the convergence of the backward iterations in \mathbb{L}^1 is problematic, we now consider a map f_t linear in x , a case for which an explicit solution is available. Accordingly, we assume that

$$f_t(x) = \kappa(Z_{t-1})x + \varepsilon_t,$$

where the function κ is bounded but not necessarily by 1 and an integrable noise ε_0 . The dynamic is then given by an AR process with a random lag coefficient and it is widely

known that the unique solution can be written as

$$X_t = \sum_{j \geq 1} \prod_{i=1}^j \kappa(Z_{t-i}) \varepsilon_{t-j} + \varepsilon_t, \quad (2.2)$$

provided that $\mathbb{E} \log \kappa(Z_0) < 0$. The series (2.2) converges almost surely. The \mathbb{L}^1 -convergence is guaranteed from Theorem 6, as soon as $\kappa := \|\kappa(Z_0)\|_\infty < 1$, where for a random variable X , $\|X\|_\infty$ denotes its supremum norm. If $\kappa \geq 1$, convergence in \mathbb{L}^1 of the series 2.2 is much more difficult to get because of the possible stochastic dependence between the coordinates of the process $(Z_t)_{t \in \mathbb{Z}}$. Let us first note that such a problem occurs in the non ergodic case, when $Z_t = Z_0$ a.s. In this case, we have

$$\mathbb{E} \left(\prod_{i=1}^j \kappa(Z_{t-i}) |\varepsilon_{t-j}| \right) = \mathbb{E}(|\varepsilon_0|) \cdot \mathbb{E}(\kappa(Z_0)^j)$$

and since $\mathbb{E}(\kappa(Z_0)^j) \geq \mathbb{P}(\kappa(Z_0) \geq 1)$, one cannot get convergence of the series (2.2) if $\mathbb{P}(\kappa(Z_0) \geq 1) > 0$. In what follows, we also stress that a similar problem of convergence also occurs in the ergodic case. To this end, set $\phi(p) = \|\kappa(Z_0)\|_p$ for $p \geq 1$. The function ϕ is non decreasing and $\phi(\infty) = \|\kappa(Z_0)\|_\infty$. Assumption **A2** is satisfied as soon as $\|\kappa(Z_0)\|_\infty < 1$. It is then tempting to study the convergence of the solution only assuming that $\|\kappa(Z_0)\|_p < 1$ but $\|\kappa(Z_0)\|_q \geq 1$ for some $1 \leq p < q$. However for any value of the pair (p, q) , there always exists an example of a process $(\kappa(Z_t))_{t \in \mathbb{Z}}$ such that the series (2.2) is not converging in \mathbb{L}^1 . To this end, assuming without loss of generality that q is an integer, we define $\kappa(z) = z$ and $Z_{t-1} = a_{t-1} \cdots a_{t-q}$ where $(a_t)_{t \in \mathbb{Z}}$ is a process of i.i.d. nonnegative random variables, independent of $(\varepsilon_t)_{t \in \mathbb{Z}}$, and such that $\|a_0\|_p < 1$ and $\|a_0\|_q \geq 1$. Since for $j \geq q$,

$$\prod_{i=1}^j \kappa(Z_{t-i}) = \prod_{i=1}^{q-1} a_{t-i}^i \cdot \prod_{i=q}^{j+1} a_{t-i}^q \prod_{i=0}^{q-2} a_{t-j-k+i}^{i+1},$$

we find

$$\mathbb{E} \left(\prod_{i=1}^j \kappa(Z_{t-i}) \right) = \prod_{i=1}^{q-1} \mathbb{E}^2(a_0^i) \cdot \mathbb{E}^{j-q+2}(a_0^q).$$

Hence the previous expectation does not converge to 0 when $j \rightarrow \infty$ and the series (2.2) cannot converge in \mathbb{L}^1 . The analysis of this linear case enlightens that a tail condition on $\kappa(Z_t)$ is not sufficient for getting this kind of convergence. In particular, the dependence structure of the process $(\kappa(Z_t))_{t \in \mathbb{Z}}$ is also of major importance. This contrasts with AR

processes with i.i.d. random coefficients, since in this case the condition $\mathbb{E}\kappa(Z_0) < 1$ is necessary and sufficient for the convergence of the series (2.2) in \mathbb{L}^1 . However, imposing an independence assumption on the covariate process is not reasonable.

In the next section, we show that one can investigate a different mode of convergence for the iterations in model (2.1) and which allow to relax Assumption **A2**. However, it is necessary to impose a strict exogeneity assumption on the covariate process.

2.2.3 An additional result for strictly exogenous regressors

In this subsection, we consider specifically equation (2.1) when the covariate process is independent of the error process. In this case, conditionally on Z , the process is a time-inhomogeneous Markov chain. The terminology Markov chain in random environments is often used in the literature. See for instance [122]. The following result will not be central in the rest of the paper because substantial efforts could be needed to derive moment and weak dependence properties for the corresponding solution and it could be also difficult to obtain explicit conditions for dealing with higher-order autoregressive processes. This is why we only provide a result when the f_t 's satisfied a one-step contraction (i.e. $m = 1$ in **A2**). We assume that there exist a real number $p \geq 1$ and a state $x_0 \in E$ such that the three following conditions are fulfilled.

A0 The process $Z := (Z_t)_{t \in \mathbb{Z}}$ is stationary and ergodic, $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a process of i.i.d. random variables taking values in E'_2 and is independent of Z .

A1' For every $z \in E'_1$, we have $\mathbb{E}[d^p(F(x_0, z, \varepsilon_1), x_0)] < \infty$.

A2' There exists a measurable function $\kappa : E' \rightarrow (0, \infty)$ satisfying $\mathbb{E}(\log^+ \kappa(Z_0)) < \infty$, $\mathbb{E} \log \kappa(Z_0) < 0$ and such that for every $(x, y) \in E^2$,

$$\mathbb{E}[d^p(F(x, z, \varepsilon_0), F(y, z, \varepsilon_0))] \leq \kappa^p(z)d^p(x, y).$$

Moreover,

$$\mathbb{E}\left[\log^+ \int d^p(x_0, F(x_0, Z_0, u)) d\mathbb{P}_{\varepsilon_1}(u)\right] < \infty.$$

We remind the notation $f_t(x) = F(x, Z_{t-1}, \varepsilon_t)$. Here, we set $\mathcal{F}_t = \sigma((Z_j, \varepsilon_j) : j \leq t)$ and $\mathbb{E}[X|Z]$ will denote the expectation of a random variable X conditionally on the covariate process $(Z_t)_{t \in \mathbb{Z}}$.

Theorem 7 *Suppose that Assumptions **A0**, **A1'**-**A2'** hold.*

1. *For every $(x, t) \in E \times \mathbb{Z}$, there exists a random variable $X_t(x)$ such that $\mathbb{E}[d^p(X_t(x), x_0) | Z] < \infty$ a.s. and $\lim_{s \rightarrow \infty} \mathbb{E}[d^p(f_{t-s}^t(x), X_t(x)) | Z] = 0$ a.s. The sequence $(f_{t-s}^t(x))_{s \geq 0}$ also converges almost surely to $X_t(x)$.*
2. *For $x \neq y$, we have $\mathbb{P}(X_t(x) \neq X_t(y)) = 0$. We then set $X_t = X_t(x)$.*
3. *The process $((X_t, Z_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.*
4. *If $(Y_t)_{t \in \mathbb{Z}}$ is a non-anticipative process (i.e. $Y_t \in \mathcal{F}_t$) such that $((Y_t, Z_t))_{t \in \mathbb{Z}}$ is stationary, for every $t \in \mathbb{Z}$, $Y_t = f_t(Y_{t-1})$ and $\mathbb{E}[d^p(Y_0, x_0) | Z_0, Z_{-1}, \dots] < \infty$ a.s., then $Y_t = X_t$ a.s.*

Notes

1. The contraction inequality in Assumption **A2'** can be restated as

$$\mathbb{E}[d^p(f_t(x), f_t(y)) | Z] \leq \kappa^p(Z_{t-1})d^p(x, y) \text{ a.s.}$$

It is then another example of contraction in conditional average.

2. In our context, our result can be seen as an improvement of Theorem 1 given in [122] for Markov chains in random environments. In particular, we do not assume a uniform contraction with respect to the environment which is given by the exogenous process Z in our random maps f_t .

2.2.4 Example

We compare the contraction conditions necessary to apply Theorem 5, Theorem 6 or Theorem 7 on a specific example. Let $((\varepsilon_t, Z_t))_{t \in \mathbb{Z}}$ be a stationary sequence of pair of random variables and for $t \in \mathbb{Z}$, set $\mathcal{F}_t = \sigma((Z_j, \varepsilon_j) : j \leq t)$. Assume that $\mathbb{E}(\varepsilon_1^2 | \mathcal{F}_0) = 1$ and $\mathbb{E}(\varepsilon_1 | \mathcal{F}_0) = 0$. For $i = 1, 2$, let $a_i : E'_1 \rightarrow \mathbb{R}$ and $b_i : E'_1 \rightarrow \mathbb{R}_+$ be some measurable maps such that $\mathbb{E} \log^+ a_i(Z_0) < \infty$ and $\mathbb{E} \log^+ b_i(Z_0) < \infty$. We consider the following AR-ARCH model with functional coefficients

$$X_t = a_0(Z_{t-1}) + a_1(Z_{t-1})X_{t-1} + \varepsilon_t \sqrt{b_0(Z_{t-1}) + b_1(Z_{t-1})X_{t-1}^2}.$$

Here we set $E = \mathbb{R}$ and $d(x, y) = |x - y|$ and $p = 2$. Setting $\sigma(z, x) = \sqrt{b_0(z) + b_1(z)x^2}$, one can note that $|\sigma(z, x) - \sigma(z, y)| \leq \sqrt{b_1(z)}|x - y|$.

1. To apply Theorem 5, we compute the Lipschitz constant $c(f_1)$ of the random map f_1 . We have

$$c(f_1) = \sup_{v \in \mathbb{R}} |f_1'(v)| = \sup_{v \in \mathbb{R}} \left| a_1(Z_0) + \frac{\varepsilon_1 b_1(Z_0)v}{\sqrt{b_0(Z_0) + b_1(Z_0)v^2}} \right|.$$

Making the change of variable $\bar{v} = \text{sign}(\varepsilon_1)\text{sign}(a_1(Z_0))v$, we have

$$c(f_1) = \sup_{\bar{v} \in \mathbb{R}} \left[|a_1(Z_0)| + \frac{|\varepsilon_1| b_1(Z_0)\bar{v}}{\sqrt{b_0(Z_0) + b_1(Z_0)\bar{v}^2}} \right].$$

We then obtain $c(f_1) = |a_1(Z_0)| + \sqrt{b_1(Z_0)}|\varepsilon_1|$ and Theorem 5 applies as soon as

$$\mathbb{E} \log c(f_1) < 0. \tag{2.3}$$

2. To apply Theorem 6, note that

$$\begin{aligned} \mathbb{E} \left[|f_t(x) - f_t(y)|^2 \mid \mathcal{F}_{t-1} \right] &= a_1(Z_{t-1})^2(x - y)^2 + (\sigma(Z_{t-1}, x) - \sigma(Z_{t-1}, y))^2 \\ &\leq (a_1(Z_{t-1})^2 + b_1(Z_{t-1})) \cdot |x - y|^2. \end{aligned}$$

One can then show that Theorem 6 applies as soon as

$$\kappa^2 := \sup_z (a_1(z)^2 + b_1(z)) < 1. \tag{2.4}$$

3. If the two processes $(\varepsilon_t)_{t \in \mathbb{Z}}$ and $(Z_t)_{t \in \mathbb{Z}}$ are independent and the ε_t 's are i.i.d., Theorem 7 applies as soon as

$$\mathbb{E} \log (a_1(Z_0)^2 + b_1(Z_0)) < 0. \tag{2.5}$$

Under the strict exogeneity assumption, we note that (2.5) is weaker than (2.4). However, (2.4) ensures the existence of a second order moment for the solution whereas (2.5) only guarantees that $\mathbb{E}(X_t^2|Z) < \infty$ a.s. On the other hand, (2.3), which only ensures existence of a stationary solution, is not necessarily weaker than (2.4) or (2.5). For instance, if the noise process has a Rademacher distribution, $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = 1/2$ and the functional coefficients are constant, (2.3) writes $|a_1| + \sqrt{b_1} < 1$ which is more restrictive than (2.4) or (2.5). But if a_1 is identically equal to 0, (2.3) writes as $\frac{1}{2}\mathbb{E} \log b_1(Z_0) +$

$\mathbb{E} \log |\varepsilon_0| < 0$ which is weaker than (2.5), since from Jensen's inequality, we have

$$\mathbb{E} \log |\varepsilon_0| \leq \log \mathbb{E} |\varepsilon_0| \leq \log \mathbb{E}^{1/2} (\varepsilon_0^2) = 0.$$

2.2.5 A result for higher-order autoregressions

In this subsection, we extend Theorem 6 to higher-order autoregressive processes. We only consider stationary processes in this part. The main result, Theorem 8, is particularly interesting for multivariate autoregressions for which Lipschitz type properties can be obtained equation by equation. See Section 2.4 for an application of Theorem 8 to various examples.

For a given real number $0 < o \leq 1$, we define the distance Δ on \mathbb{R} , by $\Delta(u, v) = |u - v|^o$ for $u, v \in \mathbb{R}$. Let E be a subset of \mathbb{R}^k and $\|\cdot\|$ an arbitrary norm on \mathbb{R}^k . Our aim is to study existence of solutions for the following recursive equations :

$$X_t = F(X_{t-1}, \dots, X_{t-q}, \zeta_t), \quad t \in \mathbb{Z}, \quad (2.6)$$

where $F : E^q \times E' \rightarrow E$ is a measurable function. Note that one can always associate a random map f_t on E^q to the dynamic (2.6). To this end, for $t \in \mathbb{Z}$ and $x_1, \dots, x_q \in E$, we set

$$f_t(x_1, \dots, x_q) = (F(x_1, \dots, x_q, \zeta_t), x_1, \dots, x_{q-1}).$$

We first introduce additional notations. We denote by \mathcal{M}_k the set of square matrices with real coefficients and k rows and if $A \in \mathcal{M}_k$, $\rho(A)$ the spectral radius of the matrix A . Moreover, for $x, y \in \mathbb{R}^k$ and $p \geq 1$, the vector $(\Delta^p(x_1, y_1), \dots, \Delta^p(x_k, y_k))'$ will be denoted by $\Delta_{vec}^p(x, y)$. Finally, we introduce a partial order relation \preceq on \mathbb{R}^k and such that $x \preceq y$ means $x_i \leq y_i$ for $i = 1, \dots, k$.

The following assumptions will be needed.

- B1** The process $(\zeta_t)_{t \in \mathbb{Z}}$ is stationary and ergodic adapted to a filtration $(\mathcal{F}_t)_{t \in \mathbb{Z}}$.
- B2** For any $y \in E^q$, $\mathbb{E} [\|F(y, \zeta_1)\|^{op}] < \infty$.
- B3** There exist some matrices $A_1, \dots, A_q \in \mathcal{M}_k$ with nonnegative elements, satisfying $\rho(A_1 + \dots + A_q) < 1$ and such that for $y, y' \in E^q$,

$$\mathbb{E} [\Delta_{vec}^p(F(y, \zeta_t), F(y', \zeta_t)) | \mathcal{F}_{t-1}] \preceq \sum_{i=1}^q A_i \Delta_{vec}^p(y_i, y'_i).$$

Though the following result is stated for an arbitrary pair $o \in (0, 1), p \geq 1$, the two interesting cases are $o \in (0, 1), p = 1$ and $o = 1, p > 1$.

Theorem 8 *Suppose that Assumptions **B1-B3** hold. There then exists a unique stationary and non-anticipative process $(X_t)_{t \in \mathbb{Z}}$ solution of (2.6) and such that $\mathbb{E} [\|X_t\|^{op}] < \infty$. Moreover, this process is ergodic.*

Notes

1. If $\zeta_t = (Z_{t-1}, \varepsilon_t)$ with ε_t independent of $\mathcal{F}_{t-1} = \sigma((Z_j, \varepsilon_j) : j \leq t)$ and $k = o = 1$, Assumption **B3** writes

$$\sup_z \mathbb{E} \left[\left| F(y_1, \dots, y_q, z, \varepsilon_1) - F(y'_1, \dots, y'_q, z, \varepsilon_1) \right|^p \right] \leq \sum_{i=1}^q A_i |y_i - y'_i|^p,$$

with $\rho(\sum_{i=1}^q A_i) = \sum_{i=1}^q A_i < 1$. This provides a quite simple criterion for application to autoregressive processes.

2. In the spirit of Section 2.2.4, the previous criterion can be checked for models with varying parameters, directly constructed from smooth parametric autoregressive processes. Consider the model

$$Y_t = \bar{F}_\theta(Y_{t-1}, \dots, Y_{t-q}, \varepsilon_t), \quad t \in \mathbb{Z}, \theta \in \Theta \subset \mathbb{R}^e.$$

If

$$\mathbb{E} \left[\left| \bar{F}_\theta(y_1, \dots, y_q, \varepsilon_1) - \bar{F}_\theta(y'_1, \dots, y'_q, \varepsilon_1) \right|^p \right] \leq \sum_{i=1}^q \bar{A}_i(\theta) |y_i - y'_i|^p,$$

the model

$$Y_t = \bar{F}_{\theta(Z_{t-1})}(Y_{t-1}, \dots, Y_{t-q}, \varepsilon_t)$$

satisfies **B3** as soon as $\sum_{i=1}^q \|\bar{A}_i\|_\infty < 1$. We then obtain a model with exogenous covariates by replacing parameter θ with a varying parameter $\theta(Z_{t-1})$ where $\theta : \mathbb{R}^e \rightarrow \Theta$ is a measurable map. See also the note after Proposition 3 for a discussion.

2.3 Functional dependence measure

The functional dependence measure has been introduced by [134] and is particularly interesting for autoregressive processes which are not necessarily strong mixing or for

which getting strong mixing conditions requires additional regularity conditions on the noise distribution. The single requirement is to get a Bernoulli shift representation of the stochastic process of interest, i.e. $X_t = H(\xi_t, \xi_{t-1}, \dots)$ where $(\xi_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. random variables taking values in a measurable space (G, \mathcal{G}) . The functional dependence measure is expressed in terms of some coefficients which evaluate for $t \geq 0$ the \mathbb{L}^p -distance between X_t and a copy \bar{X}_t , obtained by replacing ξ_0 with ξ'_0 , ξ'_0 following the same distribution as ξ_0 and being independent from the sequence $(\xi_t)_{t \in \mathbb{Z}}$. See below for the definition of these coefficients. Most of the limit theorems and deviation inequalities have been derived under such dependence measures. See for instance [138] and [137]. Such asymptotic results have been applied to various statistical problems. See for instance [136] for kernel estimation for time series, [139] for covariance estimation or [90] for spectral density estimation. The notion of functional dependence is then an attractive alternative to the usual strong mixing when the process is defined by stochastic recursions. Our aim in this section is to show that under the assumptions of Theorem 6 or Theorem 8, when the process $(\zeta_t)_{t \in \mathbb{Z}}$ or more specifically the covariate process $(Z_t)_{t \in \mathbb{Z}}$ satisfies this kind of dependence, the functional dependence measure of the solution $(X_t)_{t \in \mathbb{Z}}$ can be controlled. We will then provide a new wide class of examples for which the aforementioned references provide an important number of statistical applications. In this section, we assume that the state space E is a subspace of \mathbb{R}^k and the distance d is given by $d(x, y) = |x - y|^o$ where $|\cdot|$ is an arbitrary norm on \mathbb{R}^k and $0 < o \leq 1$.

2.3.1 Dependence coefficients for general iterated random functions

Assume that the process ζ has a Bernoulli shift representation, i.e. $\zeta_t = H(\xi_t, \xi_{t-1}, \dots)$ for some measurable map $H : G^{\mathbb{N}} \rightarrow E'$ and $(\xi_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d., G -valued random variables. We then have for every $(x, t) \in E \times \mathbb{Z}$,

$$f_t(x) = F(x, H(\xi_t, \xi_{t-1}, \dots))$$

and the map f_t has itself a Bernoulli shift representation. To define the functional measure coefficients, we then define a new sequence $(\bar{\xi}_t)_{t \in \mathbb{Z}}$ such that $\bar{\xi}_0 = \xi'_0$ and $\bar{\xi}_t = \xi_t$ for $t \neq 0$. Here ξ'_0 is a copy of ξ_0 which is assumed to be independent from $(\xi_t)_{t \in \mathbb{Z}}$. Moreover, for $t > 0$, let

$$\bar{f}_t(x) = F(x, H(\bar{\xi}_t, \bar{\xi}_{t-1}, \dots)).$$

we define for $t \geq 0$ and $p \geq 1$,

$$\theta_{p,t} = \mathbb{E}^{1/p} \left[d \left(f_{-\infty}^t, \bar{f}_{-\infty}^t \right)^p \right].$$

Moreover, for $h \in \mathbb{N}$, let $\Theta_{p,h} = \sum_{t \geq h} \theta_{p,t}$. Two cases of interest are $p = 1$ and $o = 1, p > 1$.

Our aim is to get an upper bound for the functional dependence coefficients $\Theta_{p,h}$. To this end, we add other assumptions. Here we set for $t \in \mathbb{Z}$, $\mathcal{F}_t = \sigma(\xi_{t-j} : j \geq 0)$.

A3 There exists a measurable function $S : E \rightarrow \mathbb{R}_+$ and $r, s \geq p$ such that $r^{-1} + s^{-1} = p^{-1}$ and $\mathbb{E} \left[S(f_{-\infty}^0)^s \right] < \infty$ and for all $x \in E$ and $t \geq 1$,

$$\mathbb{E}^{1/p} \left[d^p \left(\bar{f}_t(x), f_t(x) \right)^p \mid \sigma(\xi'_0) \vee \mathcal{F}_{t-1} \right] \leq S(x) H_{t-1},$$

where H_{t-1} is a random variable measurable with respect to $\sigma(\xi'_0) \vee \mathcal{F}_{t-1}$ and such that $\mathbb{E} |H_{t-1}|^r < \infty$.

An immediate consequence of Assumption **A3** is that for any random variable V_{t-1} , measurable with respect to $\sigma(\xi'_0) \vee \mathcal{F}_{t-1}$, we have $\mathbb{E} \left[d^p \left(\bar{f}_t(V_{t-1}), f_t(V_{t-1}) \right) \right] \leq \mathbb{E} \left[S(V_{t-1})^p H_{t-1}^p \right]$ and from Hölder's inequality, we get

$$\|d \left(\bar{f}_t(V_{t-1}), f_t(V_{t-1}) \right)\|_p \leq \|S(V_{t-1})\|_s \|H_{t-1}\|_r.$$

When $\zeta_t = (Z_{t-1}, \varepsilon_t)$, with Z_t taking values in a Borel subset E'_1 of \mathbb{R}^e and ε_t taking values in a Polish space E'_2 , we will still denote by $|\cdot|$ an arbitrary norm on \mathbb{R}^e and we also set $d(z, z') = |z - z'|^o$ for $z, z' \in \mathbb{R}^e$ to avoid additional notations. We will use two specific assumptions.

A3' There exists a measurable function $S : E \rightarrow \mathbb{R}_+$ and $r, s \geq p$ such that $r^{-1} + s^{-1} = p^{-1}$ and $\mathbb{E} \left[S(f_{-\infty}^0)^s \right] < \infty$, $\mathbb{E} |Z_0|^{ro} < \infty$ and for all $x \in E$ and $t \geq 1$,

$$\mathbb{E}^{1/p} \left[d^p \left(F(x, z, \varepsilon_0), F(x, z', \varepsilon_0) \right) \right] \leq S(x) d(z, z'),$$

A4 Let $(\eta_t)_{t \in \mathbb{Z}}$ be a sequence of random variables taking values in a measurable space (G_1, \mathcal{G}_1) and such that $Z_t = H'(\eta_t, \eta_{t-1}, \dots)$ for a measurable function H' . Moreover, setting $\xi_t = (\varepsilon_t, \eta_t)$, we assume that $(\xi_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. of random variables taking values in $G = E'_2 \times G_1$.

From **A4**, we have the specific representation

$$\zeta_t = H(\xi_t, \xi_{t-1}, \dots) := (H'(\eta_{t-1}, \eta_{t-2}, \dots), \varepsilon_t), \quad t \in \mathbb{Z}.$$

The map H takes values in $E' = E'_1 \times E'_2$. Note that our formulation allows the covariate process to have a general form, including a VARMA or GARCH process among others.

Proposition 1 1. *Suppose that Assumptions **A1-A4** hold. For any $h \geq 2$, there then exists $C_1 > 0$ not depending on h such that*

$$\Theta_{p,h} \leq C_1 \left[\kappa^{h/m} + \sum_{i=0}^{h-1} \kappa^{i/m} \eta_{r,h-i} + \sum_{i \geq h} \kappa^{i/m} \eta_{r,1} \right], \quad (2.7)$$

with

$$\eta_{r,j} = \sum_{t \geq j} \|H_{t-1}\|_r, \quad j \geq 1.$$

In particular, if $\eta_{r,1} < \infty$, there exists $C > 0$, not depending on h , such that

$$\Theta_{p,h} \leq C \left[\kappa^{h/m} + \sum_{i=0}^{h-1} \kappa^{i/m} \eta_{r,h-i} \right]. \quad (2.8)$$

2. *Suppose that Assumptions **A1-A2-A4** and **A3'** hold. Then we get the bound (2.8) with $\eta_{r,j} = \Theta_{r,j-1}(Z)$.*

Notes

1. Let us comment on Assumption **A4**. Under this assumption, the η'_t s are i.i.d. as well as the ε'_t s and for any $t \in \mathbb{Z}$, ε_t is independent from $\mathcal{F}_{t-1} = \sigma(\xi_s : s \leq t-1)$. Note that we allow simultaneous dependence between ε_t and η_t . For instance, we can set $\eta_t = K(\varepsilon_t, U_t)$ where K is a measurable function and U is a sequence of i.i.d. random variables, independent from the sequence ε . This assumption is then more flexible than the complete independence between the two error processes ε and η , which implies strict exogeneity.
2. It can happen that our assumptions are satisfied with some p , leading to an upper bound for the functional dependence coefficients $\theta_{p,t}$, while it is required a condition on $\theta_{q,t}$ or $\Theta_{q,h}$ for $q > p$ for applying some limit theorems or statistical results. This is still possible if one can prove finiteness of higher-order moments for the solution,

e.g. if $\mathbb{E} \left[|f_{-\infty}^0|^{q'o} \right] < \infty$ for some $q' > q$. Indeed, from Hölder's inequality, we have

$$\theta_{q,t} \leq \theta_{p,t}^{\frac{p(q'-q)}{q(q'-p)}} \theta_{q',t}^{\frac{q'(q-p)}{q(q'-p)}}.$$

Moreover, $\theta_{q',t} \leq 2 \| |f_{-\infty}^0|^o \|_{q'}$.

2.3.2 Dependence coefficients for higher-order autoregressions with exogenous covariates

Here, we revert to higher-order autoregressions considered in Section 2.2.5. We consider directly the case $\zeta_t = (Z_{t-1}, \varepsilon_t)$ with Assumption **A4** being satisfied. Additionally to Assumptions **B1-B2-B3** and **A4**, the following assumption, which is the analogue of **A3'**, will be needed.

B4 If $r > 0$ and $s \in \mathbb{R}_+ \cup \{\infty\}$ are such that $r^{-1} + s^{-1} = p^{-1}$, there exists a measurable function $S : E^q \rightarrow \mathbb{R}$ such that $\mathbb{E} [S(X_q, \dots, X_1)^s] < \infty$, $\mathbb{E} |Z_0|^{or} < \infty$ and for all $z, z' \in E_1^r$ and $x_1, \dots, x_q \in E^q$,

$$\mathbb{E}^{1/p} [d^p (F(x_1, \dots, x_q, z, \varepsilon_0), F(x_1, \dots, x_q, z', \varepsilon_0))] \leq S(x_1, \dots, x_q) d(z, z').$$

The result is the following.

Proposition 2 *Suppose that Assumptions **B2 – B4** and **A4** hold. There then exists $C > 0$ and $\rho \in (0, 1)$, such that for all $h \geq 1$, $\Theta_{p,h}(X) \leq C \left[\rho^h + \sum_{i=1}^h \rho^i \Theta_{r,h-i}(Z) \right]$.*

Note. From the upper bound given in Proposition 2, we note that the decay of $\Theta_{p,h}(X)$ is polynomial (respectively geometric) in h if the decay of $\Theta_{r,h}$ is polynomial (respectively geometric) in h .

2.3.3 A central limit theorem

To illustrate the usefulness of our results, we give below a central limit theorem for partial sums

$$S_n = \sum_{i=1}^n f(X_t, X_{t-1}, \dots, X_{t-k}),$$

where f is some real-valued measurable function and $(X_t)_{t \in \mathbb{Z}}$ is a stochastic process solution of $X_t = F(X_{t-1}, \dots, X_{t-q}, \zeta_t)$ and the assumptions of either Proposition 1 or

Proposition 2 are satisfied. The following result, which is a straightforward corollary of the invariance principle given in [134], is not the most general as possible. In particular, when a moment of order greater than p is available for the stationary solution, different assumptions on the function f could be used.

Theorem 9 *Suppose that either Assumptions **A1-A4** or Assumptions **B2-B3-B4-A4** hold true for some $p > 2$ and $\Theta_{p,0}(X) < \infty$. If there exists $C > 0$ and $0 \leq \ell \leq \frac{p-2}{2}$ such that for $x_i, x'_i \in E$, $0 \leq i \leq k$,*

$$|f(x_0, \dots, x_k) - f(x'_0, \dots, x'_k)| \leq C \left[1 + \sum_{i=0}^k (|x_i|^{o\ell} + |x'_i|^{o\ell}) \right] \cdot \sum_{i=0}^k d(x_i, x'_i).$$

Then we have the weak convergence

$$\frac{1}{\sqrt{n}} (S_n - \mathbb{E}S_n) \Rightarrow \mathcal{N}(0, \sigma^2),$$

with $\sigma^2 = \sum_{j \in \mathbb{Z}} \text{Cov}(Y_j, Y_0)$, $Y_t = f(X_t, \dots, X_{t-k})$.

2.4 Examples

2.4.1 CHARN models

In this section, we consider conditional heteroscedastic autoregressive nonlinear (CHARN) models such as in [71] or [72] but that can encompass exogenous regressors. More precisely, we consider the dynamic

$$Y_t = f_1(Y_{t-1}, \dots, Y_{t-q}, Z_{t-1}) + \varepsilon_t f_2(Y_{t-1}, \dots, Y_{t-q}, Z_{t-1}), \quad (2.9)$$

where q is a positive integer and $f_1, f_2 : \mathbb{R}^q \times E'_1 \rightarrow \mathbb{R}$ are measurable functions. In order to study stationary solutions of the recursions (2.9), the following assumptions will be needed.

CH1 The process $((Z_t, \varepsilon_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.

CH2 For $j = 1, 2$, there exist measurable functions $a_{i,j} : \mathbb{R}^d \times \mathbb{R}_+, 1 \leq i \leq q$ such that

$$\left| f_j(y_1, \dots, y_q, z) - f_j(y'_1, \dots, y'_q, z) \right| \leq \sum_{i=1}^q a_{i,j}(z) |y_i - y'_i|.$$

CH3 There exist a real number $p \geq 1$ such that $\|\varepsilon_1\|_p < \infty$ and $r, s \geq p$ such that $r^{-1} + s^{-1} = p^{-1}$, s can be infinite, $\mathbb{E}|Z_0|^r < \infty$ and two functions L_1, L_2 defined on \mathbb{R}^q and such that for $j = 1, 2$, $y_1, \dots, y_q \in \mathbb{R}$ and $z, z' \in E'_1$,

$$|f_j(y_1, \dots, y_q, z) - f_j(y_1, \dots, y_q, z')| \leq L_j(y_1, \dots, y_q)|z - z'|.$$

For $t \in \mathbb{Z}$ and $i = 1, \dots, q$, we set $c_{i,t} = a_{i,1}(Z_{t-1}) + a_{i,2}(Z_{t-1})|\varepsilon_t|$, We then define a sequence of random matrices $\mathbf{A} = (A_t)_{t \in \mathbb{Z}}$ by

$$A_t = \begin{pmatrix} c_{1,t} & c_{2,t} & \cdots & c_{q,t} \\ & & & 0 \\ & I_{q-1} & & \vdots \\ & & & 0 \end{pmatrix}.$$

Finally, we denote by $\chi(\mathbf{A})$ the Lyapunov exponent of the sequence \mathbf{A} , i.e.

$$\chi(\mathbf{A}) = \lim_{n \rightarrow \infty} \frac{\mathbb{E} [\log \|A_n \cdots A_1\|]}{n},$$

where $\|\cdot\|$ is an arbitrary norm on the space of square matrices of size $q \times q$.

Proposition 3 *Suppose that Assumptions CH1-CH2 hold.*

1. *Suppose that $\chi(\mathbf{A}) < 0$. There then exists a unique stationary process $(Y_t)_{t \in \mathbb{Z}}$ solution of (2.9) which is also ergodic.*
2. *Assume additionally that for every $t \in \mathbb{Z}$, ε_t is independent from \mathcal{F}_{t-1} . If there exist $\bar{x} \in \mathbb{R}^q$, $p \geq 1$ such that $f_1(\bar{x}, Z_0) + \varepsilon_1 f_2(\bar{x}, Z_0) \in \mathbb{L}^p$ and*

$$\sum_{i=1}^q \sup_z \|a_{i,1}(z) + a_{i,2}(z)|\varepsilon_1|\|_p < 1, \quad (2.10)$$

there then exists a unique stationary and non-anticipative process solution of (2.9) which is also ergodic and such that $\mathbb{E}|Y_1|^p < \infty$.

3. *Assume furthermore that Assumptions (CH3) and (A4) hold true with $\mathbb{E}L_j^s(Y_q, \dots, Y_1) < \infty$ for $j = 1, 2$. There then exists $C > 0$ and $\rho \in (0, 1)$, such that for all $h \geq 1$,*

$$\Theta_{p,h}(X) \leq C \left[\rho^h + \sum_{i=1}^h \rho^i \Theta_{r,h-i}(Z) \right].$$

Notes

1. Our results can be useful for dealing with models with functional coefficients in the spirit of the example given in Section 2.2.4. See also the notes after the statement of Theorem 8. For $j = 1, 2$, let $m_\theta^{(j)} : \mathbb{R}^q \rightarrow \mathbb{R}$ be some functions depending on some parameters $\theta \in \mathbb{R}^e$ and such that

$$\left| m_\theta^{(j)}(y_1, \dots, y_q) - m_\theta^{(j)}(y'_1, \dots, y'_q) \right| \leq \sum_{i=1}^q d_{i,j}(\theta) |y_i - y'_i|,$$

for some nonnegative real numbers $d_{i,j}(\theta)$, $1 \leq i \leq q$. If θ is replaced by a function $\theta(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^e$ and $f_j(y_1, \dots, y_q, z) = m_{\theta(z)}^{(j)}(y_1, \dots, y_q)$ for $j = 1, 2$, one can then consider some standard autoregressive processes and obtain a version with functional parameters depending on exogenous covariates. For instance, threshold autoregressions or power-ARCH volatility,

$$m_\theta^{(1)}(y_1, \dots, y_q) = \theta_0 + \sum_{i=1}^q (\theta_i y_i^+ + \theta_{i+q} y_i^-), \quad m_\theta^{(2)}(y_1, \dots, y_q) = \left(\theta_0 + \sum_{i=1}^q \theta_i |y_i|^\delta \right)^{1/\delta},$$

where x^+ and x^- denotes respectively the positive part and the negative part of a real number x and $\delta \geq 1$.

2. When ε_t is not necessarily independent from \mathcal{F}_{t-1} , (2.10) can be replaced with the following more abstract condition. There exists $\eta \in (0, 1)$ such that

$$\sum_{i=1}^q \mathbb{E}^{1/p} [(a_{i,1}(Z_0) + a_{i,2}(Z_0)|\varepsilon_1|)^p | \mathcal{F}_0] \leq 1 - \eta \text{ a.s.} \quad (2.11)$$

For $q = 1$, let us compare (2.11) with the condition $\chi(\mathbf{A}) < 0$, which reduces to

$$\mathbb{E} [\log (a_{1,1}(Z_0) + a_{1,2}(Z_0)|\varepsilon_1|)] < 0.$$

This latter condition is much weaker than (2.11). Indeed, (2.11) entails that

$$\mathbb{E} [a_{1,1}(Z_0) + a_{1,2}(Z_0)|\varepsilon_1|] \leq \|a_{1,1}(Z_0) + a_{1,2}(Z_0)|\varepsilon_1|\|_p < 1$$

and from Jensen's inequality, $\log \mathbb{E} [a_{1,1}(Z_0) + a_{1,2}(Z_0)|\varepsilon_1|] \leq \chi(\mathbf{A})$.

For $q \geq 2$, it is more difficult to obtain explicit conditions which guaranty that $\chi(\mathbf{A}) < 0$.

3. Using the results of [91], a nonparametric kernel estimation of the functions f and

g is possible. Proposition 3 gives precise assumptions under which it is possible to control the functional dependence measure of some CHARN models when the regressors include lag values of the response as well as exogenous covariates. We then obtain additional examples of time series models for which standard nonparametric estimators of the regression function are still consistent.

2.4.2 GARCH processes

GARCH processes with exogenous regressors have been considered recently by [110] or [59]. We consider here the asymmetric power GARCH studied by [59]. The model is defined as follows.

$$Y_t = \varepsilon_t h_t^{1/\delta}, \quad h_t = \pi' Z_{t-1} + \sum_{i=1}^q \left\{ \beta_i h_{t-i} + \alpha_{i+} (Y_{t-i}^+)^{\delta} + \alpha_{i-} (Y_{t-i}^-)^{\delta} \right\}, \quad (2.12)$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ and $(Z_t)_{t \in \mathbb{Z}}$ are two sequences of random variables taking values in \mathbb{R} and \mathbb{R}_+^d respectively, $\delta > 0$, $\pi \in \mathbb{R}_+^d$ and the β_i 's, α_{i+} 's and α_{i-} 's are nonnegative real numbers. Optimal stationarity properties of time series models defined by (2.12) have been obtained by [59], using a version of Theorem 5 for affine random maps. In contrast, we use our results to get existence of a moment of order δ for the unique stationary solution. The following assumptions will be needed.

G1 The process $((Z_t, \varepsilon_t))_{t \in \mathbb{Z}}$ is stationary and ergodic and $\mathbb{E}|Z_0| < \infty$.

G2 There exist s_-, s_+ such that $\mathbb{E}[(\varepsilon_t^+)^{\delta} | \mathcal{F}_{t-1}] \leq s_+$ and $\mathbb{E}[(\varepsilon_t^-)^{\delta} | \mathcal{F}_{t-1}] \leq s_-$ a.s. and $\gamma := \sum_{i=1}^p (\beta_i + s_+ \alpha_{i+} + s_- \alpha_{i-}) < 1$.

Proposition 4 *Suppose that Assumptions G1-G2 hold.*

1. *There then exists a unique stationary and non-anticipative solution $(Y_t)_{t \in \mathbb{Z}}$ of (2.12). This solution is ergodic and satisfies $\mathbb{E}|Y_0|^{\delta} < \infty$.*
2. *Additionally, assume that Assumption A4 holds true. Let $H_t = ((Y_t^+)^{\delta}, (Y_t^-)^{\delta}, h_t)$. There then exists $C > 0$ and $\rho \in (0, 1)$, such that for all $h \geq 1$, $\Theta_{1,h}(H) \leq C [\rho^h + \sum_{i=1}^h \rho^i \Theta_{1,h-i}(Z)]$. Moreover, if $\delta \geq 1$, we have the bound*

$$\theta_{\delta,t}(Y) \leq \theta_{1,t}^{1/\delta}(Y^+) + \theta_{1,t}^{1/\delta}(Y^-), \quad t \in \mathbb{N}.$$

Note. Let us consider the example of a GARCH process. We then set $\delta = 2$, $\alpha_{j+} = \alpha_{j-} = \alpha_j$ and we assume that $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a martingale difference, adapted to the filtration

$(\mathcal{F}_t)_{t \in \mathbb{Z}}$ with $\mathcal{F}_t = \sigma((\varepsilon_s, Z_s) : s \leq t)$. Set $v_{t-1} = \mathbb{E}[\varepsilon_t^2 | \mathcal{F}_{t-1}]$. If there exists a positive real number v_+ such that $v_{t-1} \leq v_+$ a.s., the contraction condition in **G2** reduces to $v_+ \sum_{j=1}^q (\alpha_j + \beta_j) < 1$. For GARCH processes with i.i.d. innovations ε_t , we recover a standard condition ensuring the existence of a solution with a finite second moment.

2.4.3 Poisson autoregressions

We consider the PARX model introduced in [2]. The idea is to model the conditional distribution of Y_t given \mathcal{F}_{t-1} by a Poisson distribution with a random intensity λ_t depending on past values and a covariate process. More precisely, we assume that

$$Y_t = N_{\lambda_t}^{(t)}, \quad \lambda_t = \beta_0 + \sum_{j=1}^q \beta_j \lambda_{t-j} + \sum_{j=1}^q \alpha_j Y_{t-j} + \pi' Z_{t-1}, \quad (2.13)$$

where $(N^{(t)})_{t \in \mathbb{Z}}$ is a sequence of i.i.d. Poisson processes with intensity 1, $\beta_0, \dots, \beta_q, \alpha_1, \dots, \alpha_q$ are nonnegative real numbers and π is a vector of \mathbb{R}^d with nonnegative coordinates.

PA1 We have $\gamma := \sum_{j=1}^q \alpha_j + \sum_{j=1}^q \beta_j < 1$.

PA2 The process $((Z_t, N^{(t)}))_{t \in \mathbb{Z}}$ is stationary, ergodic and adapted to a filtration $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ such that for all $t \in \mathbb{Z}$, $N^{(t)}$ is independent from \mathcal{F}_{t-1} . Moreover, $\mathbb{E}|Z_1| < \infty$.

Proposition 5 1. *Suppose that Assumptions **PA1-PA2** hold. There then exists a unique non-anticipative, stationary and ergodic process $(Y_t)_{t \in \mathbb{Z}}$ solution of (2.13).*

2. *Additionally, if Assumption **A4** is also satisfied with $\varepsilon_t = N^{(t)}$, there then exists $C > 0$ and $\rho \in (0, 1)$, such that for all $h \geq 1$, $\Theta_{1,h}((Y_t, \lambda_t)_t) \leq C [\rho^h + \sum_{i=1}^h \rho^i \Theta_{1,h-i}(Z)]$.*

Note. Our result extends substantially that of [2]. First, we prove ergodicity properties in PARX models without assuming that the covariate process $(Z_t)_{t \in \mathbb{Z}}$ is a Markov chain defined by a random map contracting in average. Secondly, for the stochastic dependence properties, we control the coefficient of functional dependence measure only assuming a general Bernoulli shift representation for $(Z_t)_{t \in \mathbb{Z}}$. For instance, $(Z_t)_{t \in \mathbb{Z}}$ can be defined by an infinite moving average process and is not necessarily Markovian.

2.4.4 Dynamic binary choice model

We consider the dynamic

$$Y_t = \mathbb{1}_{g(Y_{t-1}, \dots, Y_{t-q}, \zeta_t) > 0}, \quad (2.14)$$

where $(\zeta_t)_{t \in \mathbb{Z}}$ is a stationary process taking values in a measurable space E' and $g : \{0, 1\}^q \times E' \rightarrow \mathbb{R}$ is a measurable function. This kind of binary model is popular in econometrics for studying the dynamics of recessions. See [28] who studied the case g linear and [78] for a study of US recessions.

Proposition 6 *1. Assume that $(\zeta_t)_{t \in \mathbb{Z}}$ is a stationary and ergodic process such that*

$$\mathbb{P} \left(\min_{y \in \{0, 1\}^q, 1 \leq t \leq q} g(y, \zeta_t) > 0 \right) + \mathbb{P} \left(\max_{y \in \{0, 1\}^q, 1 \leq t \leq q} g(y, \zeta_t) \leq 0 \right) > 0. \quad (2.15)$$

There then exists a unique stationary and ergodic solution $(Y_t)_{t \in \mathbb{Z}}$ for the recursions (2.14).

*2. Assume that for some real numbers a_1, \dots, a_q and $\pi \in \mathbb{R}^e$, $g(y, \zeta_t) = \sum_{i=1}^q a_i y_i + \pi' Z_{t-1} + \varepsilon_t$, with $\zeta_t = (Z_{t-1}, \varepsilon_t)$ satisfying **A4** and the c.d.f. F_ε of ε_t being Lipschitz and taking values in $(0, 1)$. Moreover, setting $v_t = \pi' Z_{t-1} + \varepsilon_t$, we assume that there exists $\delta > 0$ and a positive integer K such that*

$$\mathbb{P} \left(\phi_- + \min_{1 \leq t \leq q} v_t > 0 \mid \mathcal{F}_{-K} \right) + \mathbb{P} \left(\phi_+ + \max_{1 \leq t \leq q} v_t \leq 0 \mid \mathcal{F}_{-K} \right) \geq \delta \text{ a.s.}, \quad (2.16)$$

where

$$\phi_+ = \max \left\{ \sum_{i=1}^q a_i y_i : (y_1, \dots, y_n) \in \{0, 1\}^n \right\}, \quad \phi_- = \min \left\{ \sum_{i=1}^q a_i y_i : (y_1, \dots, y_n) \in \{0, 1\}^n \right\}.$$

There then exists $C > 0$ and $\rho \in (0, 1)$, such that for all $h \geq 1$,

$$\Theta_{1,h}(Y) \leq C \left[\rho^h + \sum_{i=1}^h \rho^i \Theta_{1,h-i}(Z) \right].$$

Notes

1. Consider the case of g linear as in the second point of Proposition 6. In this case, [28] derived existence of a unique stationary and ergodic solution for when Condition

(2.16) holds true. As shown in [28], Condition (2.16) holds in particular when the process $(v_t)_{t \in \mathbb{Z}}$ is m -dependent or for some infinite moving averages. Condition (2.15) is much weaker since it holds as soon as

$$\mathbb{P}\left(\phi_+ + \max_{1 \leq t \leq q} v_t \leq 0\right) + \mathbb{P}\left(\phi_- + \min_{1 \leq t \leq q} v_t > 0\right) > 0. \quad (2.17)$$

Condition (2.17) holds true as soon as the random vector (v_1, \dots, v_q) has full support. Another sufficient condition for (2.17) is the following. If $(v_t)_{t \in \mathbb{Z}}$ is adapted to a filtration $(\mathcal{F}_t)_{t \in \mathbb{Z}}$, we assume that for any $x \in \mathbb{R}$ and $t \in \mathbb{Z}$, $\mathbb{P}(v_t \leq x | \mathcal{F}_{t-1}) > 0$ a.s. or for any $t \in \mathbb{Z}$ and $x \in \mathbb{R}$, $\mathbb{P}(v_t > x | \mathcal{F}_{t-1}) > 0$ a.s. Recall that $v_t = \pi' Z_{t-1} + \varepsilon_t$. The latter condition is valid in particular when ε_t has full support and is independent from $\mathcal{F}_{t-1} = \sigma((\varepsilon_s, Z_s) : s \leq t-1)$.

2. As the proof of Proposition 6 will show, the condition (2.16) implies Assumption **A2**. Condition (2.15) is only used for applying Theorem 5. However, (2.15) does not entail mixing properties. In contrast, Condition (2.16) does. See [28], Theorem 2. Our results (see point 2. of Proposition 6) give a complement when the covariate process is not necessarily strongly mixing and has a Bernoulli shift representation.
3. When $\zeta_t = (Z_{t-1}, \varepsilon_t) \in \mathbb{R}^{d+1}$ in (2.14), one can allow interactions between lag values of the response and the covariates. For example,

$$g(y, \zeta_t) = \sum_{i=1}^d c_i y_i + \sum_{i=1}^q \sum_{j=1}^d [a_{i,j} y_i + b_{i,j} (1 - y_i)] Z_{j,t-i} + \varepsilon_t.$$

When ε_t is independent of $\mathcal{F}_{t-1} = \sigma((\varepsilon_{t-j}, Z_{t-j}) : j \geq 1)$, one can show, using the same arguments as in the previous point, that condition (2.15) is satisfied as soon as the distribution of ε_t has support equal to the whole real line. We will not give a control of the functional dependence measure for this model because we were not able to check **A2** when the covariate process $(Z_t)_{t \in \mathbb{Z}}$ is not bounded. However when the cdf of ε_t is known (e.g. for the logistic or the probit model), it is widely known that ergodicity of the process is sufficient for showing consistency and asymptotic normality of conditional pseudo likelihood estimators of the parameters.

2.4.5 Categorical time series with covariates

We consider a finite set $E = \{1, 2, \dots, N\}$, an integer $q \geq 1$, a process $(Z_t)_{t \in \mathbb{Z}}$ taking values in $\mathcal{Z} \subset \mathbb{R}^d$ and a family $\{K_z(\cdot|\cdot) : z \in \mathcal{Z}\}$ of probability kernels from E^q to E . Our aim is to construct a process $(Y_t)_{t \in \mathbb{Z}}$, taking values in E and such that

$$\mathbb{P}(Y_t = i | Y_{t-1}^-, Z_{t-1}^-) = K_{Z_{t-1}^-}(i | Y_{t-1}^-, \dots, Y_{t-q}^-).$$

A particular example is given by the multinomial autoregression, i.e.

$$K_z(i, y_1, \dots, y_q) = \frac{\exp\left(\sum_{j=1}^q a_{i,j} y_j + \gamma'_i z\right)}{\sum_{k=1}^N \exp\left(\sum_{j=1}^q a_{k,j} y_j + \gamma'_k z\right)}$$

and is a classical model for categorical time series. See [53]. In econometrics, [116] studied the dynamic of price changes using this kind of model but with a more general observation-driven form such as in GARCH models and that will not fall into our framework.

For applying our results, we now define some random maps. For $t \in \mathbb{Z}$, let ε_t be a random variable uniformly distributed over $[0, 1]$. For $u \in [0, 1]$, $z \in E'_1$, $y \in E^q$ and $u \in [0, 1]$, we set

$$K_z^-(u|y) = \inf \left\{ i = 1, \dots, N : \sum_{j=1}^i K_z(j|y) \geq u \right\}$$

and

$$f_t(y_1, \dots, y_q) = \left(K_{Z_{t-1}^-}(\varepsilon_t | y_1, \dots, y_q), y_1, \dots, y_{q-1} \right)'$$

We introduce the following assumptions.

- C1** The probability kernels K_z are lower bounded by a positive constant, i.e. for any $z \in E'$, $\eta(z) := \min_{(i,y) \in E^{q+1}} K_z(i|y) > 0$.
- C2** The process $((Z_t, \varepsilon_t))_{t \in \mathbb{Z}}$ is stationary and ergodic. Moreover, for $t \in \mathbb{Z}$, ε_t is independent from $\mathcal{F}_{t-1} = \sigma((Z_s, \varepsilon_s) : s \leq t-1)$.
- C3** There exists a constant $C > 0$ such that for all $y_1, \dots, y_q \in E$,

$$\sum_{i=1}^N |K_z(i|y_1, \dots, y_q) - K_{\bar{z}}(i|y_1, \dots, y_q)| \leq C|z - \bar{z}|.$$

Proposition 7 *Suppose that Assumptions C1-C2 hold.*

1. There exists a unique stationary process satisfying the recursions

$$Y_t = K_{Z_{t-1}}^-(\varepsilon_t | Y_{t-1}, \dots, Y_{t-q}), \quad t \in \mathbb{Z}. \quad (2.18)$$

Moreover, the process $((Y_t, Z_t))_{t \in \mathbb{Z}}$ is ergodic.

2. Additionally, assume that Assumption **A4** and **C3** hold true and that $\eta_- = \inf_{z \in E'_1} \eta(z) > 0$ in **C1**. There then exist $C > 0$ and $\rho \in (0, 1)$, such that for all $h \geq 1$, $\Theta_{1,h}(Y) \leq C \left[\rho^h + \sum_{i=1}^h \rho^i \Theta_{1,h-i}(Z) \right]$.

Note. A proof of the first point of Proposition 7 is based on Theorem 5 and provides a general result for existence of stationary categorical time series with covariates. In particular, probit, logistic and multinomial autoregressions can be considered without restriction for the covariate process $(Z_t)_{t \in \mathbb{Z}}$. However, the derivation of the dependence properties in the second point imposes a more restrictive assumption on the transition kernel K because it is necessary to check Assumption **A2**. Recently, [51] studied categorical time series under the strict exogeneity assumption for the covariate process. For the recursions (2.18), strict exogeneity holds true as soon as the two processes $(Z_t)_{t \in \mathbb{Z}}$ and $(\varepsilon_t)_{t \in \mathbb{Z}}$ are independent. Assumption **C2** is weaker than strict exogeneity in general.

2.4.6 Categorical time series and coalescence of the paths

In this section, we give another interpretation of the convergence of the backward iterations for categorical time series. This interpretation has a link with some perfect simulation schemes that are widely known for Markov chains. See [111]. Since the state space is discrete, the iterations should be automatically constant after some steps. Figure 2.1 illustrates the convergence when $q = 1$ and $N = 3$. In this case, $f_t(j) = K_{Z_{t-1}}^-(\varepsilon_t | j)$ for $j = 1, 2, 3$. Setting

$$T = \inf \left\{ k \geq 1 : \varepsilon_{t-k} \leq \min_{i,j} K_{Z_{t-k-1}}(i | j) \right\},$$

we know that from the ergodicity assumption **C2** and the positivity assumption **C1**, T is finite almost surely. In this case, $f_{t-T}(j) = 1$ a.s., $f_{-\infty}^t := \lim_{n \rightarrow \infty} f_{t-n}^t(j) = f_{t-T}^t(j)$ a.s. and the limit does not depend on the state j . All the paths corresponding to $f_{t-n}^t(j)$ for $n \geq T$ coalesce through State 1.

When $q \geq 2$, our assumptions guarantee that it is possible to get q times successively

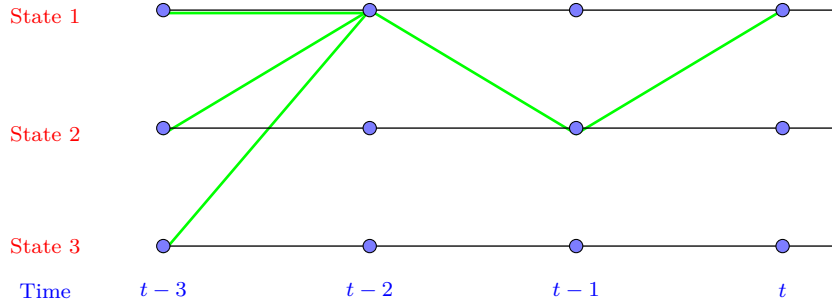


FIGURE 2.1 – Illustration of the convergence for $N = 3$ modalities and $q = 1$ lag

the value 1 for the time series whatever the previous values. A coalescence property for the paths will then also occur in this case.

This interpretation is also relevant for getting an interpretation of Condition (2.15) for binary choice models. When (2.15) is satisfied, it is possible to get, starting at any time t , either q times the value 1 or q -times the value 0 whatever the previous values of the binary time series. Running the backward iterations, we have coalescence of the paths at the first (random) time $t - T$ such that such an event occurs.

2.5 Conclusion

A general theoretical analysis of nonlinear autoregressive time series models with exogenous covariates is absent from the present time series literature and only a few references consider such a problem, mainly for specific examples. The aim of this paper was to provide some results for a reasonable class of nonlinear time series models for which the required assumptions can be checked. In particular, we provide two results, Theorem 6 and Theorem 8, which justify existence and uniqueness of stationary and ergodic solutions possessing some moments. The crucial assumption to check, **A2** or **B3**, involves a uniform conditional contraction condition. Assumption **B3** is the main assumption to check for autoregressive models with several lags. For some nonlinear models already considered in the literature, such as GARCH or autoregressive Poisson processes, this contraction

condition is easily checked because the exogenous covariates have an additive contribution in the expression of the latent process and play the role of a random intercept which does not modify the usual stability conditions. However, our results can also be applied to autoregressive processes for which lag parameters depend on the covariates (see Section 2.2.4, the notes after Theorem 8, Proposition 3 and Proposition 6 for some examples). In this case, a uniform control of the random lag parameters is necessary to check our assumptions which shows the limit of our approach.

It may be possible to weaken our uniform contraction condition, as shown in Theorem 7, at least under a strict exogeneity assumption. However, getting additional general results to ensure existence of some unconditional moments, to control dependence coefficients and to consider higher-order autoregressive processes would require substantial effort. The proposed framework is also useful for deriving weak dependence properties of the solution, leading to the possibility to apply many existing statistical inference procedures, the central limit theorem of Section 2.3.3 providing an illustration. To this end, the functional dependence measure discussed in Section 2.3 is of primary importance. Note that a general result for getting weak dependence properties of autoregressive processes with exogenous covariates is also new and it is another contribution of this paper. Finally, we also derived results for categorical time series in Sections 2.4.4-2.4.5. Apart from the weak dependence properties which can be derived from our general results, we also obtained stationarity conditions with weaker assumptions, applying Theorem 5. Note that whatever the results used in the paper (Theorems 5, 6 or 7), the convergence of the backward iterations of random maps appears to be a central point of view for considering many different types of autoregressive systems with exogenous regressors.

2.6 Proofs of the results

2.6.1 Proof of Theorem 6

We use the convention $f_t^{t-1}(x) = x$ for $(x, t) \in E \times \mathbb{Z}$. From Assumption **A2**, if $(t, s, s') \in \mathbb{Z}^3$ are such that $s' \leq s \leq t$ and $x, y \in E$,

$$\mathbb{E} \left[d^p \left(f_s^t(x), f_{s'}^t(y) \right) \mid \mathcal{F}_{t-1} \right] \leq L^p d^p \left(f_s^{t-1}(x), f_{s'}^{t-1}(y) \right),$$

and then

$$\|d \left(f_s^t(x), f_{s'}^t(y) \right)\|_p \leq L \|d \left(f_s^{t-1}(x), f_{s'}^{t-1}(y) \right)\|_p. \quad (2.19)$$

Applying (2.19) with $y = x_0$, $s = s' = t$, we get $\sup_{t \in \mathbb{Z}} \|d(f_t(x), f_t(x_0))\|_p < \infty$. Next, using **A1** and the triangular inequality, we get

$$\sup_{t \in \mathbb{Z}} \|d(f_t(x), y)\|_p < \infty \text{ for every } (x, y) \in E^2. \quad (2.20)$$

With the same kind of arguments, we get for $s \leq t - m$,

$$\|d(f_s^t(x), f_{s'}^t(y))\|_p \leq \kappa \|d(f_s^{t-m}(x), f_{s'}^{t-m}(y))\|_p. \quad (2.21)$$

1. We denote by $[z]$ the integer part of a real number z . To apply recursively the previous bounds, we note that for any integer i , $i+1 = r_1 m + r_2$ with $r_1 = [(i+1)/m]$ and $r_2 = i+1 - r_1 m$. We then get from (2.19) and (2.21),

$$\begin{aligned} \sum_{i \geq 0} \|d(f_{t-i}^t(x), f_{t-i-1}^t(x))\|_p &\leq \sum_{i \geq 0} \kappa^{[(i+1)/m]} L^{i+1-m[(i+1)/m]} \|d(x, f_{t-i-1}(x))\|_p \\ &\leq \frac{(L+1)^m \kappa^{(1-m)/m}}{1 - \kappa^{1/m}} \sup_{j \in \mathbb{Z}} \|d(x, f_j(x))\|_p. \end{aligned}$$

This latter bound entails that the series $\sum_{i \geq 0} d(f_{t-i}^t(x), f_{t-i-1}^t(x))$ is almost surely finite. By the Cauchy criterion, there exists a random variable $X_t(x)$ such that $\lim_{i \rightarrow \infty} d(f_{t-i}^t(x), X_t(x)) = 0$ a.s. Moreover, from the previous bound, (2.20) and the triangular inequality, we deduce that

$$\sum_{s, t \in \mathbb{Z}, s \leq t} \|d(y, f_s^t(x))\|_p < \infty \text{ for every } (x, y) \in E^2. \quad (2.22)$$

Next we note that the convergence also holds in \mathbb{L}^p , since from Fatou's lemma,

$$\begin{aligned} \|d(X_t(x), f_{t-s}^t(x))\|_p &\leq \liminf_{j \rightarrow \infty} \|d(f_{t-j}^t(x), f_{t-s}^t(x))\|_p \\ &\leq \sum_{i \geq s} \|d(f_{t-i}^t(x), f_{t-i-1}^t(x))\|_p \\ &\leq \kappa^{s/m} \frac{(L+1)^m \kappa^{(1-m)/m}}{1 - \kappa^{1/m}} \sup_{j \in \mathbb{Z}} \|d(f_j(x), x)\|_p \rightarrow 0 \text{ as } s \rightarrow \infty. \end{aligned}$$

Finally, using (2.22) and the triangular inequality, we get the last assertion $\sup_{t \in \mathbb{Z}} \|d(X_t(x), x_0)\|_p < \infty$.

2. If $x \neq y$, we have from the almost sure convergence and Fatou's lemma

$$\begin{aligned} \|d(X_t(x), X_t(y))\|_p &\leq \liminf_{s \rightarrow \infty} \|d(f_{t-s}^t(x), f_{t-s}^t(y))\|_p \\ &\leq \liminf_{s \rightarrow \infty} \kappa^{[(s+1)/m]} L^{s+1-[(s+1)/m]m} d(x, y) = 0. \end{aligned}$$

This shows the second point.

3. For the third point, we observe that for any $j \geq 1$, there exists a measurable function $H_j^{(x)} : E'^{j+1} \rightarrow E$ such that $f_{t-j}^t(x) = H_j^{(x)}(\zeta_t, \dots, \zeta_{t-j})$. Since $\lim_{j \rightarrow \infty} H_j^{(x)}$ exists $\mathbb{P}_{(\zeta_{t-j})_{j \geq 0}}$ a.s., it is then possible to define a measurable function $H : (E')^{\mathbb{N}} \rightarrow E$ such that $X_t = H((\zeta_{t-j})_{j \geq 0})$ a.s. The process $(X_t)_{t \in \mathbb{Z}}$ has a Bernoulli shift structure with dependent entries and is then stationary and ergodic provided that the process $(\zeta_t)_{t \in \mathbb{Z}}$ satisfies the same properties.

4. The last property follows from the following bounds which hold for any $j \geq 1$:

$$\begin{aligned} \|d(X_t, Y_t)\|_p &= \|d(f_{t-mj+1}^t(X_{t-mj}), f_{t-mj+1}^t(Y_{t-mj}))\|_p \\ &\leq \kappa^j \left[\sup_{j \in \mathbb{Z}} \|d(x_0, X_j)\|_p + \sup_{j \in \mathbb{Z}} \|d(x_0, Y_j)\|_p \right]. \quad \square \end{aligned}$$

2.6.2 Proof of Theorem 7

1. From Assumption **A0**, $f_{t-s}^t(x)$ is, conditionally on Z , an iteration of $s + 1$ independent random maps. Using Assumption **A2'**, we get

$$\sum_{s \geq 0} \mathbb{E}^{1/p} \left[d^p \left(f_{t-s}^t(x), f_{t-s-1}^t(x) \right) \mid Z \right] \leq \sum_{s \geq 0} \prod_{i=1}^{s+1} \kappa(Z_{t-i}) b_{t-s-2}(x),$$

with

$$b_t^p(x) = \int d^p(x, F(x, Z_t, u)) d\mathbb{P}_{\varepsilon_0}(u).$$

From Assumption **A2'** and the triangular inequality, we have $\mathbb{E} \log^+ b_t(x) < \infty$ for any $x \in E$. We are going to show that

$$\sum_{s \geq 0} \mathbb{E}^{1/p} \left[d^p \left(f_{t-s}^t(x), f_{t-s-1}^t(x) \right) \mid Z \right] < \infty \text{ a.s.} \quad (2.23)$$

This follows from the assumptions on the logarithmic moments. Indeed, $((\kappa(Z_t), b_t(x)))_{t \in \mathbb{Z}}$ is a stationary process and it is widely known that the stochastic recursions

$$Y_t = \kappa(Z_{t-1})Y_{t-1} + b_{t-1}(x)$$

have a unique stationary solution given by

$$Y_t = b_{t-1}(x) + \sum_{s \geq 0} \prod_{i=1}^{s+1} \kappa(Z_{t-i}) b_{t-s-2}(x),$$

the latter series being convergent almost surely. See for instance [14], Theorem 1. This shows (2.23). Using Minkowski's inequality and Fatou's lemma for conditional expectations, see for instance [77], Chapter 5, we then deduce that

$$\mathbb{E}^{1/p} [S(x)^p | Z] < \infty \text{ a.s. } S(x) := \sum_{s \geq 0} d(f_{t-s}^t(x), f_{t-s-1}^t(x)).$$

As a consequence, we have $\mathbb{P}(S(x) < \infty | Z) = 1$ a.s. and then $\mathbb{P}(S(x) < \infty) = 1$. From the Cauchy criterion, we then conclude the existence of a random variable $X_t(x)$ such that $\lim_{s \rightarrow \infty} f_{t-s}^t(x) = X_t(x)$ a.s. Note that from Assumption **A1'** and (2.23), we have $\mathbb{E} [d^p(f_{t-s}^t(x), x_0) | Z] < \infty$ a.s. for every positive integer s . The convergence $\lim_{s \rightarrow \infty} \mathbb{E} [d^p(f_{t-s}^t(x), X_t(x)) | Z] = 0$ a.s. and $\mathbb{E} [d^p(X_t(x), x_0) | Z] < \infty$ a.s. follow as in the proof of Theorem 6, using Fatou's lemma for conditional expectation.

2. For a positive integer s , we have from **A2**,

$$\mathbb{E}^{1/p} [d^p(f_{t-s}^t(x), f_{t-s}^t(y)) | Z] \leq \prod_{i=1}^s \kappa(Z_{t-i}) d(x, y) \rightarrow 0 \text{ a.s.}$$

Letting $s \rightarrow \infty$, we have $\mathbb{E} [d^p(X_t(x), X_t(y)) | Z] = 0$ a.s. and then $\mathbb{P}(X_t(x) \neq X_t(y) | Z) = 0$ a.s. Taking the expectation, we conclude that $\mathbb{P}(X_t(x) \neq X_t(y)) = 0$.

3. From the almost sure convergence of the sequence $(f_{t-s}^t(x))_{s \geq 0}$, stationarity and ergodicity of the process $((X_t, Z_t))_{t \in \mathbb{Z}}$ follows exactly as in the proof of point 3 of Theorem 6.
4. Let $(Y_t)_{t \in \mathbb{Z}}$ be a stochastic process satisfying the proposed conditions. If the process is non-anticipative, we have from **A2'**,

$$\begin{aligned} \mathbb{E} [d^p(X_t, Y_t) | Z] &= \mathbb{E} [d^p(f_{t-s}^t(X_{t-s-1}), f_{t-s}^t(Y_{t-s-1})) | Z] \\ &\leq \prod_{i=0}^s \kappa^p(Z_{t-i-1}) \cdot \mathbb{E} [d^p(X_{t-s-1}, Y_{t-s-1}) | Z]. \end{aligned}$$

Note that, from **A2'**, $\prod_{i=0}^s \kappa^p(Z_{t-i-1}) = o_{\mathbb{P}}(1)$. Moreover

$$\mathbb{E}^{1/p} [d^p (X_{t-s-1}, Y_{t-s-1}) | Z] \leq \mathbb{E}^{1/p} [d^p (X_{t-s-1}, x_0) | Z] + \mathbb{E}^{1/p} [d^p (x_0, Y_{t-s-1}) | Z].$$

Note that

$$\mathbb{E}^{1/p} [d^p (x_0, Y_{t-s-1}) | Z] = \mathbb{E}^{1/p} [d^p (x_0, Y_{t-s-1}) | Z_{t-s-1}, Z_{t-s-2}, \dots]$$

and if the process $((Z_t, Y_t))_{t \in \mathbb{Z}}$ is stationary, then the process $(V_t)_{t \in \mathbb{Z}}$ defined by

$$V_t = \mathbb{E}^{1/p} [d^p (x_0, Y_t) | Z_t, Z_{t-1}, \dots]$$

is also stationary and takes finite values from our assumptions. Then $V_{t-s-1} = O_{\mathbb{P}}(1)$ where s is the index of the sequence. The same property holds if Y_t is replaced with X_t . As a consequence

$$\mathbb{E} [d^p (X_{t-s-1}, Y_{t-s-1}) | Z] = O_{\mathbb{P}}(1).$$

We then conclude that $\mathbb{E} [d^p (X_t, Y_t) | Z] = 0$ a.s. Then $\mathbb{P} [X_t \neq Y_t | Z] = 0$ a.s. and by integration, we get the conclusion.

2.6.3 Proof of Proposition 1

1. We use the decomposition

$$\begin{aligned} \bar{f}_{-\infty}^t - f_{-\infty}^t &= \sum_{i=0}^{t-1} [f_{t-i}^t \circ \bar{f}_{-\infty}^{t-i-1} - f_{t-i-1}^t \circ \bar{f}_{-\infty}^{t-i-2}] \\ &+ \bar{f}_t \circ \bar{f}_{-\infty}^{t-1} - f_t \circ \bar{f}_{-\infty}^{t-1}. \end{aligned}$$

From Assumption **A2** and Assumption **A4**, we have, for $i = 0, \dots, t-2$,

$$\begin{aligned} \|d(f_{t-i}^t \circ \bar{f}_{-\infty}^{t-i-1}, f_{t-i-1}^t \circ \bar{f}_{-\infty}^{t-i-2})\|_p &\leq \kappa^{\frac{i+1}{m}-1} L^m \|d(\bar{f}_{t-i-1} \circ \bar{f}_{-\infty}^{t-i-2}, f_{t-i-1} \circ \bar{f}_{-\infty}^{t-i-2})\|_p \\ &\leq \kappa^{\frac{i+1}{m}-1} L^m \|S(\bar{f}_{-\infty}^{t-i-2}) H_{t-i-2}\|_p \\ &\leq \kappa^{\frac{i+1}{m}-1} L^m \|S(f_{-\infty}^0)\|_s \|H_{t-i-2}\|_r. \end{aligned}$$

If $i = t - 1$, we have $\|d(f_{t-i}^t \circ \bar{f}_{-\infty}^{t-i-1}, f_{t-i-1}^t \circ \bar{f}_{-\infty}^{t-i-2})\|_p \leq 2\|d(0, f_{-\infty}^0)\|_p \kappa^{t/m-1}$.
Using the triangular inequality, we get for $t \geq 2$,

$$\theta_{p,t} \leq \kappa^{-1} L^m \|S(f_{-\infty}^0)\|_s \sum_{i=0}^{t-2} \kappa^{(i+1)/m} \|H_{t-i-2}\|_r + \|S(f_{-\infty}^0)\|_s \|H_{t-1}\|_r + 2\kappa^{t/m-1} \|d(0, f_{-\infty}^0)\|_p.$$

The bound (2.7) is obtained by summation and entails the simpler bound (2.8).

2. From **A4'**, we have **A4** with $H_{t-1} = d(Z_{t-1}, \bar{Z}_{t-1})$ with

$$\bar{Z}_t = H'(\eta_t, \dots, \eta_1, \eta'_0, \eta_{-1}, \dots).$$

We then deduce the result from the previous point, noticing that $\eta_{r,j} = \Theta_{r,j-1}(Z)$.

2.6.4 Proof of Theorem 8

Define the following random map

$$f_t(u_1, \dots, u_q) = \left(F(u_1, \dots, u_q, \zeta_t)', u'_1, \dots, u'_{q-1} \right)'$$

We set $x = (u_1, \dots, u_q) \in E^q$ and for $1 \leq t \leq q$, $U_t(x) = u_{q-t+1}$. Next for $t \geq q + 1$, we define $U_t(x)$ recursively by

$$U_t(x) = F(U_{t-1}(x), \dots, U_{t-q}(x), \varepsilon_t).$$

We then have for $t \geq q + 1$,

$$(U_t(x), \dots, U_{t-q+1}(x)) = f_{q+1}^t(x).$$

Using our assumptions, we have for $t \geq q + 1$,

$$\mathbb{E} [\Delta_{vec}^p(U_t(x), U_t(x')) | \mathcal{F}_{t-1}] \leq \sum_{i=1}^q A_i \Delta_{vec}^p(U_{t-i}(x), U_{t-i}(x')).$$

We introduce the matrix

$$B = \begin{pmatrix} A_1 & \cdots & A_{q-1} & A_q \\ & I_{k(q-1)} & & 0_{k(q-1),1} \end{pmatrix}.$$

The condition $\rho(A_1 + \dots + A_q) < 1$ entails that $\rho(B) < 1$. Indeed, if $v = (v'_1, \dots, v'_q)' \in \mathbb{R}^{kq} \setminus \{0\}$ is such that $Bv = \lambda v$ for $|\lambda| \geq 1$, we get the equality $v_1 = [\lambda^{-1}A_1 + \dots + \lambda^{-q}A_q]v_1$. Since the coefficients of the A_i 's are nonnegative, we get

$$|v_1|_{vec} \preceq \sum_{i=1}^q |\lambda|^{-i} A_i |v_1|_{vec} \preceq \sum_{i=1}^q A_i |v_1|_{vec},$$

where $|v_1|_{vec}$ denotes the vector of the absolute values of the coordinates of v_1 . We then get $|v_1|_{vec} \preceq \left(\sum_{j=1}^q A_j\right)^k |v_1|_{vec}$ for any positive integer k . Letting $k \rightarrow \infty$, we obtain $v_1 = 0$. Since $v_i = \lambda v_{i+1}$ for $i = 1, \dots, q-1$, we get $v = 0$ which is a contradiction. Then $|\lambda| < 1$ and $\rho(B) < 1$. Next, we set

$$V_t(x) = (U_t(x)', \dots, U_{t-q+1}(x)'), \quad t \geq q+1.$$

Note that $V_t(x) = f_{q+1}^t(x)$. We then have

$$\mathbb{E}[\Delta_{vec}^p(V_t(x), V_t(x')) | \mathcal{F}_{t-1}] \preceq B \Delta_{vec}^p(V_{t-1}(x), V_{t-1}(x')) \preceq \dots \preceq B^{t-q} \Delta^p(x, x').$$

Setting for $v, v' \in \mathbb{R}^{kq}$, $\bar{d}(v, v') = \left(\sum_{i=1}^{kq} \Delta^p(v_i, v'_i)\right)^{1/p}$, we get

$$\mathbb{E}[\bar{d}^p(f_{q+1}^t(x), f_{q+1}^t(x')) | \mathcal{F}_q] \leq |\mathbb{1}' B^{t-q}|_\infty \bar{d}^p(x, x'),$$

where $\mathbb{1}$ denotes the vector of \mathbb{R}^{kq} having all its components equal to 1, and $|\cdot|_\infty$ the infinite norm in \mathbb{R}^{kq} . Since $\rho(B) < 1$, if t is large enough, we have $|\mathbb{1}' B^{t-q}|_\infty < 1$. We then conclude that **A2** is satisfied with $d = \bar{d}$, $m = \inf\{j \geq 1 : |\mathbb{1}' B^j|_\infty < 1\}$ and $\kappa^p = |\mathbb{1}' B^m|_\infty$. Moreover, **A1** is a direct consequence of **B2**. The result then follows from Theorem 6. \square

2.6.5 Proof of Proposition 2

Defining

$$f_t(x_1, \dots, x_q) = (F(x_1, \dots, x_q, Z_{t-1}, \varepsilon_t), x_1, \dots, x_{q-1}),$$

we set for $t \geq 0$, $\bar{Z}_t = H(\eta_t, \dots, \eta_1, \eta'_0, \eta_{-1}, \dots)$. Using Assumptions **A5-B5**, we have, for $t \geq 1$,

$$\mathbb{E}^{1/p} \left[d^p \left(f_t(x_1, \dots, x_q), \bar{f}_t(x_1, \dots, x_q) \right) | \mathcal{F}_{t-1} \vee \sigma(\xi'_0) \right] \leq S(x_1, \dots, x_q) d(Z_{t-1}, \bar{Z}_{t-1}).$$

We then check Assumption **A4**, setting $H_{t-1} = d(Z_{t-1}, \bar{Z}_{t-1})$. One can then apply Proposition 1 using for instance the ℓ_1 -distance on E^q defined by $\bar{d}(u, v) = \sum_{i=1}^q d(u_i, v_i)$ which is equivalent to the distance \bar{d} used in the proof of Theorem 8. Since $\mathbb{E}^{1/r} [d^r(Z_{t-1}, \bar{Z}_{t-1})] = \theta_{r,t-1}(Z)$, the proof of the proposition is now complete.

2.6.6 Proof of Theorem 9

The proof is a consequence of Theorem 3 in [134]. It is simply necessary to prove that

$$\Gamma(Y) = \sum_{t=0}^{\infty} \|Y_t - \bar{Y}_t\|_2 < \infty.$$

To this end, for $t \geq 0$, let \bar{X}_t be the random variables obtained by replacing ξ_0 by an independent copy ξ'_0 in its Bernoulli shift representation. We then have $\bar{Y}_t = f(\bar{X}_t, \dots, \bar{X}_{t-k})$ for $t \geq k$. From the assumption on the function f , stationarity and Hölder's inequality, we have, setting $\ell_1 = 2p/(p-2)$,

$$\theta_{2,t}(Y) \leq C \left(1 + 2k \| |X_0|^{o\ell} \|_{\ell_1}\right) \cdot \sum_{i=0}^k \theta_{p,t-i}(X).$$

The required condition easily follows by summation.

2.6.7 Proof of Proposition 3

1. Let d be the distance induced by the ℓ_1 -norm on \mathbb{R}^q , i.e. $d(x, y) = \sum_{i=1}^q |x_i - y_i|$. We use the notation $|x - y|$ instead of $d(x, y)$. For a square matrix A of size $q \times q$, we denote by $\|A\|$ the corresponding operator norm of A . We define the sequence of random maps as follows :

$$g_t(x) = (f_1(x_1, \dots, x_q, Z_{t-1}) + \varepsilon_t f_2(x_1, \dots, x_q, Z_{t-1}), x_1, \dots, x_{q-1})'$$

We then have

$$|g_t(x) - g_t(y)|_{vec} \preceq \left(\sum_{i=1}^q c_{i,t} |x_i - y_i|, |x_1 - y_1|, \dots, |x_{q-1} - y_{q-1}| \right)' = A_t \cdot |x - y|_{vec}.$$

Iterating the previous bound, we get for any positive integer t ,

$$\left|g_1^t(x) - g_1^t(y)\right|_{vec} \preceq A_t \cdots A_1 \cdot |x - y|_{vec}.$$

We then deduce that $c(g_1^t) \leq \|A_t \cdots A_1\|$. The result is then a consequence of Theorem 5, using the condition $\chi(\mathbf{A}) < 0$.

2. We check the assumptions of Theorem 8. First note that from Assumption **CH1**, the process $\zeta_t = (Z_{t-1}, \varepsilon_t)$ is ergodic. This entails **B1**. Next we set

$$F(x_1, \dots, x_q, \zeta_t) = f_1(x_1, \dots, x_q, Z_{t-1}) + \varepsilon_t f_2(x_1, \dots, x_q, Z_{t-1}).$$

Our assumptions guarantee that $F(\bar{x}, \zeta_1) \in \mathbb{L}^p$ and, using **CH2**, we deduce that $F(x, \zeta_1) \in \mathbb{L}^p$ for any $x \in \mathbb{R}^q$. This shows **B2**. Finally, we check **B3**. To this end, for $i = 1, \dots, q$, we set $\delta_i = \sup_z \|a_{i,1}(z) + a_{i,2}(z)|\varepsilon_1\|_p$.

Using Minkowski's inequality for conditional expectations (see for instance [38], Chapter XI, Section 3), we have

$$\begin{aligned} \mathbb{E}^{1/p} [|F(x_1, \dots, x_q, \zeta_t) - F(y_1, \dots, y_q, \zeta_t)|^p | \mathcal{F}_{t-1}] &\leq \mathbb{E}^{1/p} \left[\left(\sum_{i=1}^q c_{i,t} |x_i - y_i| \right)^p | \mathcal{F}_{t-1} \right] \\ &\leq \sum_{i=1}^q \mathbb{E}^{1/p} [c_{i,t}^p | \mathcal{F}_{t-1}] \cdot |x_i - y_i| \\ &\leq \sum_{i=1}^q \delta_i |x_i - y_i|. \end{aligned}$$

Next using convexity, we get

$$\mathbb{E} [|F(x_1, \dots, x_q, \zeta_t) - F(y_1, \dots, y_q, \zeta_t)|^p | \mathcal{F}_{t-1}] \leq \left(\sum_{i=1}^q \delta_i \right)^{p-1} \sum_{i=1}^q \delta_i |x_i - y_i|.$$

B3 is then a consequence of **CH3**.

3. We apply Proposition 2. From the previous points, it is only required to check **B4** which is a consequence of Assumption **CH3**.

2.6.8 Proof of Proposition 4

For the first part, we apply Theorem 8. To this end, we set $E = \mathbb{R}_+^3$, $F = (F_1, F_2, F_3)$, $F_2(y_1, \dots, y_q, \zeta_t) = (\varepsilon_t^+)^{\delta} y_{1,1}$, $F_3(y_1, \dots, y_q, \zeta_t) = (\varepsilon_t^-)^{\delta} y_{1,1}$ and

$$F_1(y_1, \dots, y_q, \zeta_t) = \pi' Z_{t-1} + \sum_{j=1}^q \beta_j y_{1,j} + \left(\alpha_{1+} (\varepsilon_t^+)^{\delta} + \alpha_{1-} (\varepsilon_t^-)^{\delta} \right) y_{1,1} + \sum_{j=2}^q (\alpha_{j+} y_{2,j} + \alpha_{j-} y_{3,j}).$$

We then deduce that Assumption **B3** holds true with

$$A_1 = \begin{pmatrix} \beta_1 + \alpha_{1+s_+} + \alpha_{1-s_-} & 0 & 0 \\ s_+ & 0 & 0 \\ s_- & 0 & 0 \end{pmatrix}, \quad A_j = \begin{pmatrix} \beta_j & \alpha_{j+} & \alpha_{j-} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad j \geq 2.$$

It is straightforward to show that the matrix $\Gamma := \sum_{j=1}^q A_j$ has eigenvalues 0 and $\frac{a \pm \sqrt{a^2 + 4(bd + ce)}}{2}$ with $a = \sum_{j=1}^q \beta_j + \alpha_{1+s_+} + \alpha_{1-s_-}$, $b = \sum_{j=2}^q \alpha_{j+}$, $c = \sum_{j=2}^q \alpha_{j-}$, $d = s_+$ and $e = s_-$. Condition $\rho(\Gamma) < 1$ is equivalent to $\gamma < 1$. It is then clear that **B1-B3** follow from **G1-G2**.

For the second part, it is easily seen that **B4** is satisfied for a constant function S , $p = 1$ and $s = \infty$. This gives the bound for $\Theta_{1,h}(H)$. If $\delta \geq 1$, the last bound for $\theta_{\delta,t}(Y)$ can be obtained from the inequalities

$$|x - y|^{\delta} \leq |x^{\delta} - y^{\delta}|, \quad x, y \geq 0.$$

2.6.9 Proof of Proposition 5

1. To show the first point, we check the assumptions of Theorem 8. We set $\zeta_t = (N^{(t)}, Z_{t-1})$, $E = \mathbb{N} \times \mathbb{R}_+$ and the state space E is endowed with the ℓ_1 -norm. We first note that $(Y_t)_{t \in \mathbb{Z}}$ is a stationary solution of (2.13) if and only if $X_t = (Y_t, \lambda_t)'$ is the solution of $X_t = F(X_{t-1}, \dots, X_{t-q}, \zeta_t)$ with

$$F(x_1, \dots, x_q, \zeta_t) = \left(N_{f(x_1, \dots, x_q, Z_{t-1})}^{(t)}, f(x_1, \dots, x_q, Z_{t-1}) \right)'$$

and $f(x_1, \dots, x_q, Z_{t-1}) = \beta_0 + \sum_{j=1}^q \beta_j s_j + \sum_{j=1}^q \alpha_j y_j + \pi' Z_{t-1}$, $x_i = (y_i, s_i)$, $1 \leq i \leq q$.
 For $x \in (\mathbb{N} \times \mathbb{R}_+)^q$,

$$\mathbb{E}[|F(x, \zeta_1)|] = 2 \left(\beta_0 + \sum_{j=1}^q \beta_j s_j + \sum_{j=1}^q \alpha_j y_j + \pi' \mathbb{E}(Z_1) \right) < \infty$$

since $\mathbb{E}(|Z_1|) < \infty$.

We then have, for $(x, x') \in ((\mathbb{N} \times \mathbb{R}_+)^q)^2$ with $x = (x_1, \dots, x_q)$, $x' = (x'_1, \dots, x'_q) \forall j = 1, \dots, q$, $x_j = (y_j, s_j)$, $x'_j = (y'_j, s'_j)$,

$$\mathbb{E}[|F(x, \zeta_t) - F(x', \zeta_t)|_{vec} | \mathcal{F}_{t-1}] \preceq \sum_{j=1}^q \begin{pmatrix} \alpha_j & \beta_j \\ \alpha_j & \beta_j \end{pmatrix} |x_j - x'_j|_{vec}.$$

In the previous bounds, we have used the identity $\mathbb{E}[|N_{h_{t-1}}^{(t)} - N_{g_{t-1}}^{(t)}| | \mathcal{F}_{t-1}] = |h_{t-1} - g_{t-1}|$ which is valid for two nonnegative random variables h_{t-1}, g_{t-1} measurable with respect to \mathcal{F}_{t-1} . The previous equality follows from the properties of the Poisson process. Letting

$$\Gamma = \sum_{j=1}^q \begin{pmatrix} \alpha_j & \beta_j \\ \alpha_j & \beta_j \end{pmatrix},$$

the matrix Γ has two eigenvalues : 0 and γ . Assumption **PA1** then guarantees that $\rho(\Gamma) < 1$. Assumptions **B1-B3** of Theorem 8 are satisfied. Hence, according to Theorem 8, there exists a unique stationary and non-anticipative process $(X_t)_{t \in \mathbb{Z}}$ solution of (2.13) and such that $\mathbb{E}[|X_t|] < \infty$. This process is ergodic. This completes the proof of the first point.

2. For the second point, we use Proposition 2. To this end, it is only necessary to check **B4** for $p = r = 1$ and $s = \infty$. This is straightforward since we have the equality

$$\mathbb{E}[|F(x_1, \dots, x_q, z, N^{(t)}) - F(x_1, \dots, x_q, \bar{z}, N^{(t)})|] = 2|\pi'(z - \bar{z})|.$$

The proof of the second point is now complete.

2.6.10 Proof of Proposition 6

1. We apply Theorem 5. To this end, we define the random map from $E = \{0, 1\}^q$ to E by

$$f_t(x) = \left(\mathbb{1}_{g(x, \zeta_t) > 0}, x_1, \dots, x_{q-1} \right)'.$$

We set

$$\delta_t := \max_{y, y' \in \{0, 1\}^q} \left| \mathbb{1}_{g(y, \zeta_t) > 0} - \mathbb{1}_{g(y', \zeta_t) > 0} \right| \leq \mathbb{1}_{\max_{y \in \{0, 1\}^q} g(y, \zeta_t) > 0} - \mathbb{1}_{\min_{y \in \{0, 1\}^q} g(y, \zeta_t) > 0}.$$

Setting $(y_t(x), \dots, y_{t-q+1}(x))' = f_1^t(x)$ for $t \geq q$, we have

$$y_t(x) = 1 \text{ if and only if } g(y_{t-1}(x), \dots, y_{t-q}(x), \zeta_t) > 0.$$

We have, setting $c_t = |y_t(x) - y_t(x')|$,

$$c_t \leq \delta_t \max_{1 \leq j \leq q} c_{t-j}.$$

Using the fact that $\delta_t \leq 1$, a straightforward induction on $i = 0, \dots, q-1$ shows that

$$c_{t+i} \leq \delta_{t+i} \max_{1 \leq j \leq q} c_{t-j}.$$

Setting $d(y, y') = \max_{1 \leq i \leq q} |y_i - y'_i|$,

$$d(f_1^{t+q-1}(x), f_1^{t+q-1}(x')) \leq \max_{0 \leq i \leq q-1} \delta_{t+i} d(f_1^{t-1}(x), f_1^{t-1}(x')).$$

Setting $t = 1$, this shows in particular that

$$c(f_1^q) \leq \max_{1 \leq i \leq q} \delta_i \leq \mathbb{1}_{\max_{y, i} g(y, \zeta_i) > 0} - \mathbb{1}_{\min_{y, i} g(y, \zeta_i) > 0}. \quad (2.24)$$

From our assumptions, the last upper bound can vanish with positive probability, and then $\mathbb{E}[\log c(f_1^q)] = -\infty = \chi$. Theorem 5 leads to the result.

2. The result will follow from Proposition 1. To this end, we check Assumptions **A1-A3**. **A1** is automatic. We use the metric d on $\{0, 1\}^q$ which is bounded. We set $h_t = \max_{t-q+1 \leq i \leq t} \delta_i$. Note that in the linear case, we have

$$h_t \leq \mathbb{1}_{\phi_+ + \max_{t-q+1 \leq i \leq t} v_i > 0} - \mathbb{1}_{\phi_- + \min_{t-q+1 \leq i \leq t} v_i > 0}.$$

To check **A2**, we use the bound (2.24) and the inequality $h_s \leq 1$ for all $s \in \mathbb{Z}$ to get

$$c\left(f_{t-Jq+1}^t\right) \leq \prod_{j=0}^{J-1} h_{t-jq} \leq h_t.$$

Moreover using (2.16), we have

$$\begin{aligned} \mathbb{E}(h_t | \mathcal{F}_{t-Jq}) &\leq 1 - \mathbb{P}\left(\phi_+ + \max_{t-q+1 \leq i \leq t} v_i \leq 0 | \mathcal{F}_{t-Jq}\right) - \mathbb{P}\left(\phi_- + \min_{t-q+1 \leq i \leq t} v_i > 0 | \mathcal{F}_{t-Jq}\right) \\ &\leq 1 - \delta, \end{aligned}$$

provided that $Jq \geq q + K$. This guarantees **A2**, with $\kappa = 1 - \delta$ and $m = Jq$.

Finally, let us check **A3**. Setting $\bar{Z}_t = H(\eta_t, \dots, \eta_1, \eta'_0, \eta_{-1}, \dots)$. We have

$$\begin{aligned} \mathbb{E}\left[d\left(f_t(y), \bar{f}_t(y)\right) | \mathcal{F}_{t-1} \vee \sigma(\xi'_0)\right] &\leq \left|F_\varepsilon\left(-\sum_{i=1}^q a_i y_i - \pi' Z_{t-1}\right) - F_\varepsilon\left(-\sum_{i=1}^q a_i y_i - \pi' \bar{Z}_{t-1}\right)\right| \\ &\leq L_\varepsilon \cdot \max_{j=1}^d |\pi_j| \cdot \sum_{j=1}^d |Z_{j,t-1} - \bar{Z}_{j,t-1}|, \end{aligned}$$

where L_ε denotes the Lipschitz constant of F_ε .

2.6.11 Proof of Proposition 7

1. For the first point, we will apply Theorem 5 for the discrete metric $d(x, y) = \mathbf{1}_{x \neq y}$. Setting for $x \in E^q$ and $t \in \mathbb{Z}$, $f_{t-q+1}^t(x) = f_t \circ \dots \circ f_{t-q+1}(x)$, we have for $x, y \in E^q$,

$$\begin{aligned} \left\{f_{t-q+1}^t(x) = f_{t-q+1}^t(y)\right\} &\supset \left\{f_{t-q+1}^t(x) = f_{t-q+1}^t(y) = (1, \dots, 1)\right\} \\ &= \left\{\varepsilon_j \in [0, \eta(Z_{j-1})] : j = t - q + 1, \dots, t\right\}. \end{aligned}$$

We then obtain

$$d\left(f_{t-q+1}^t(x), f_{t-q+1}^t(y)\right) \leq \left(1 - \prod_{j=t-q+1}^t \mathbf{1}_{\varepsilon_j \in [0, \eta(Z_{j-1})]}\right) d(x, y).$$

Let us show that

$$p := \mathbb{P}\left(\varepsilon_j \in [0, \eta(Z_{j-1})] : j = t - q + 1, \dots, t\right) > 0. \quad (2.25)$$

Note that by stationarity, p does not depend on t . Assume that $p = 0$. From Assumption **C2** and the properties of the conditional expectations, we have

$$p = \mathbb{E} \left[\eta(Z_{t-1}) \prod_{j=t-q+1}^{t-1} \mathbb{1}_{\{\varepsilon_j \leq \eta(Z_{j-1})\}} \right] = 0.$$

Since η is positive, we get $\mathbb{P}(\varepsilon_j \in [0, \eta(Z_{j-1})] : j = t - q + 1, \dots, t - 1) = 0$. By finite induction, we deduce that $\mathbb{P}(\varepsilon_{t-q+1} \leq \eta(Z_{t-q})) = 0$. Since this latter probability equals to $\mathbb{E}(\eta(Z_{t-q})) > 0$, we obtain a contradiction. The property (2.25) is then valid and Theorem 5 applies, which leads to the conclusion.

2. For the second point, we will use Theorem 6 and Proposition 1. Since d is a bounded metric, **A1** is automatically satisfied. Next, observe that

$$\begin{aligned} \mathbb{P}(f_{t-q+1}^t(x) = f_{t-q+1}^t(y) | \mathcal{F}_{t-q}) &\geq \mathbb{P}(f_{t-q+1}^t(x) = f_{t-q+1}^t(y) = (1, \dots, 1) | \mathcal{F}_{t-q}) \\ &\geq \mathbb{P}(\varepsilon_t, \dots, \varepsilon_{t-q+1} \in [0, \eta_-] | \mathcal{F}_{t-q}) \\ &= \mathbb{P}(\varepsilon_t, \dots, \varepsilon_{t-q+1} \in [0, \eta_-]) \\ &\geq \eta_-^q. \end{aligned}$$

This yields to the bound $\mathbb{E} \left[d(f_{t-q+1}^t(x), f_{t-q+1}^t(y) | \mathcal{F}_{t-q}) \right] \leq 1 - \eta_-^q$, which shows the second part of **A2**. The first part is automatic.

It remains to check **A3'**. Note that for $i, j \in E$, we have $(N - 1)^{-1}|i - j| \leq \mathbb{1}_{i \neq j} \leq |i - j|$. Using the ℓ_1 -metric on E^q which is equivalent to the discrete metric, we have

$$\begin{aligned} \mathbb{E} \left[|K_z^-(\varepsilon_1 | y_1, \dots, y_q) - K_{z'}^-(\varepsilon_1 | y_1, \dots, y_q)| \right] &\leq \int_0^1 \left| K_z^-(u | y_1, \dots, y_q) - K_{z'}^-(u | y_1, \dots, y_q) \right| du \\ &\leq \sum_{j=1}^N \left| \sum_{i=1}^j K_z(i | y_1, \dots, y_q) - \sum_{i=1}^j K_{z'}(i | y_1, \dots, y_q) \right| \\ &\leq NC |z - z'|. \end{aligned}$$

Assumption **A3'** then follows with $s = \infty$ and $r = p = 1$. The bound for the functional dependence coefficients is then a direct consequence of Proposition 1. \square

MULTIVARIATE TIME SERIES MODELS FOR MIXED DATA

Abstract. We introduce a general approach which unifies some previous attempts for modeling the dynamic of multivariate time series or for regression analysis when the data are of mixed type (binary/count/continuous). Our approach is quite flexible since conditionally on past values, each coordinate at time t can have a distribution compatible with a standard univariate time series model such as GARCH, ARMA, INGARCH or logistic models whereas past values of the other coordinates play the role of exogenous covariates in the dynamic. The simultaneous dependence in the multivariate time series can be modeled with a copula. Additional exogenous covariates are also allowed in the dynamic. We first study some usual stability properties of these models and then show that autoregressive parameters can be consistently estimated equation-by-equation using a pseudo-maximum likelihood method, leading to a fast implementation even when the number of time series is large. Moreover, we prove consistency results when a parametric copula model is fitted to the time series and in the case of Gaussian copulas, we show that the likelihood estimator of the correlation matrix is strongly consistent. We carefully check all our assumptions for two prototypical examples : a GARCH/INGARCH model and logistic/log-linear INGARCH model. Our results are illustrated with numerical experiments as well as two real data sets.

Based on [33]

3.1 Introduction

Analyzing multivariate time series is now a common task in many fields. Many applications of multivariate time series historically come from econometrics or finance and many textbooks such as [94] or [129] now provide an overview of some interesting models in this context. But the development of multivariate time series analysis has been also

connected more recently to others important domains such as in biology [73], ecology [109] or industrial production [92] among others.

However, the literature of multivariate time series analysis is much less developed than in the univariate case. For univariate time series, there already exist many interesting autoregressive models for various types of data. Most of the existing works focus on continuous data with the development of ARMA models [15, 94] or GARCH models in financial econometrics (see [60] for an overview). But many time series are also related to count data [52] or categorical data [106]. Count time series data are systematically encountered when analyzing the dynamic of transaction numbers in finance or the number of disease cases in epidemiology [48] whereas categorical time series have to be analyzed when studying the dynamic of growth/recession period in economics [78], the dynamic of price changes in finance [116] or DNA sequence analysis [114], among others applications. In contrast, multivariate time series are mainly analyzed with continuous models such as vector ARMA models [94] or multivariate GARCH models [60]. It is then difficult to find a flexible approach for studying some multivariate time series when the corresponding univariate times series are of different nature (discrete or continuous). For instance, analyzing the number of transactions in finance or modeling the log-returns of the corresponding asset can only be done separately, though it is quite clear that a bivariate modeling could help to get a better understanding of the mutual interactions between these two quantities across the time. A few references are dedicated to multivariate time series for discrete data. For instance, [95] considered a multivariate binary time series models for analyzing the dynamic of electricity price spikes whereas [57] considered recently multivariate time series models for count data such as the numbers of transactions of several assets occurring across the time. Both contributions are based on some specific univariate dynamics (e.g. logistic or Poisson conditional distributions) and a copula model. In this paper, we will define a much more general framework, also providing additional results for inference in such discrete multivariate time series models.

Modeling mixed multivariate response is a challenging problem already occurring in the i.i.d. setting and finding suitable multivariate generalized linear models that extend the univariate ones is a non-trivial task, even in this context. One of the important difficulty is the lack of natural multivariate probability distributions for such data. For i.i.d. data, several approaches have been developed and [29] gives an interesting survey of some of them. However, the treatment of mixed data seems to be devoted to specific cases such as in [140] for the joint analysis of continuous/count data. A notable exception concerns

the general regression models considered in [120] with Gaussian copulas and our approach can be seen as a time series analogue of this modeling. Since our consistency results can be also applied to i.i.d. data, we then provide theoretical guarantees for inference in such regression models.

In the time series context, our approach consists in using (marginally) some standard univariate time series models called observation-driven. These univariate models are widely popular in the time series literature and provide a sufficiently rich class of dynamics from the GARCH models to Poisson autoregressive INGARCH models [47, 52] or logistic autoregressive processes [22, 106]. Let us also mention that our multivariate modeling is similar to [98], where a Gaussian copula is used to model panel data of mixed type with specified univariate dynamics. In contrast, our results apply to a larger class of dynamical models.

In what follows, we first recall the definition of univariate observation-driven models and we next discuss the multivariate extension we will consider throughout the paper.

3.1.1 Univariate observation-driven models

Let $P(\cdot|s)$ be a (non-degenerated) probability distribution on E (typically, $E = \{0, 1\}$, \mathbb{N} or \mathbb{R}) and depending on a real-valued parameter $s \in G$ (throughout the paper, $G = \mathbb{R}$ or $G = \mathbb{R}_+$), one can define a discrete-time stochastic process $(Y_t)_{t \in \mathbb{Z}}$ with one-point conditional distribution P in the following way. For a function $g : G \times E \times \mathbb{R}^m$, we assume that for $A \in \mathcal{B}(E)$,

$$\mathbb{P}(Y_t \in A | \mathcal{F}_{t-1}) = P(A | \lambda_t), \quad \lambda_t = g(\lambda_{t-1}, Y_{t-1}, X_{t-1}), \quad (3.1)$$

where $(X_t)_{t \in \mathbb{Z}}$ is a covariate process taking values in \mathbb{R}^m and for $t \in \mathbb{Z}$, $\mathcal{F}_t = \sigma((Y_s, X_s) : s \leq t)$.

Standard examples of popular time series models of this type are listed below.

- If $P(A|s) = \int_A f(y - s) dy$ with f a probability density on the real line and $g(s, y, x) = a + bs + cy + d'x$, we obtain the dynamic of the following ARMA(1, 1) process with exogenous regressors

$$Y_t = a + (b + c)Y_{t-1} + d'X_{t-1} + \varepsilon_t - b\varepsilon_{t-1}, \quad (3.2)$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. random variables with probability density f .

- If $P(A|\lambda) = \int_A \frac{1}{\lambda} f\left(\frac{y}{\lambda}\right) dy$, we obtain a GARCH type model with volatility process

$(\lambda_t)_{t \in \mathbb{Z}}$ and noise density f . Such dynamic is generally represented in more compact form in the literature, $Y_t = \varepsilon_t \lambda_t$ where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. random variables with probability density f .

- For count time series, a standard choice for $p(\cdot|\lambda)$ is the Poisson distribution with parameter λ . We obtain the well-known INGARCH processes.
- For binary time series, a classical approach consists in choosing a cdf F on the real line and to set $P(1|\lambda) = F(\lambda)$. When F is the cdf of the logistic distribution (resp. Gaussian distribution), we obtain respectively the logistic or probit autoregressive process.

Some parametric or semiparametric time series models satisfying (3.1) are obtained when $g = g_\theta$ depends on a finite-dimensional unknown vector of parameters θ . Most of the observation-driven models given above have been studied without exogenous covariates. But some recent theoretical guarantees for inclusion of exogenous covariates in non-linear time series models including those mentioned above have been obtained recently [50, 128, 42, 32]).

3.1.2 Extension to multivariate mixed time series models

Our aim is to consider multivariate time series models of type (3.1). Since there is no natural multivariate distribution P for considering mixed data, a possible approach is to consider multivariate distributions on some Cartesian products $E_1 \times \dots \times E_k$, $k \geq 1$, denoted by $P(\cdot|s)$, with parameter $s = (s_1, \dots, s_k) \in G_1 \times \dots \times G_k$ and with specific univariate marginal distributions $P_i(\cdot|s_i)$ for $1 \leq i \leq k$. A natural construction of this type can be obtained from a copula C , i.e. a probability distribution on $[0, 1]^k$ with uniform marginals. If for $1 \leq i \leq k$, F_{i,s_i} denotes the cdf of the probability distribution $P_i(\cdot|s_i)$ and $U = (U_1, \dots, U_k)$ follows the distribution C , then the random vector $(F_{1,s_1}^{-1}(U_1), \dots, F_{k,s_k}^{-1}(U_k))$ has marginal distributions $P_1(\cdot|s_1), \dots, P_k(\cdot|s_k)$. Here, for a cdf F , we denote by F^{-1} its quantile function, i.e. $F^{-1}(u) = \inf \{x \in \mathbb{R} : F(x) \geq u\}$ for $u \in (0, 1)$. Since the distribution C can have a general form, we then hope that the multivariate distributions $P(\cdot|\lambda)$ obtained in this way to be quite general. As pointed out in [63], copula for discrete data are not unique and lead to interpretation problems and identification issues. However, copula modeling is still a general and valid approach for modeling many stochastic dependence properties between the coordinates, even if some components are allowed to be discrete.

We now define multivariate time series models with a conditional distribution $P(\cdot|\cdot)$. To this end, we consider a sequence of i.i.d. random vectors $(U_t)_{t \in \mathbb{Z}}$ such that $U_1 = (U_{1,1}, \dots, U_{k,1})$ has a probability distribution denoted by C . We then set

$$Y_t = (Y_{1,t}, \dots, Y_{k,t}) = \left(F_{1,\lambda_{1,t}}^{-1}(U_{1,t}), \dots, F_{k,\lambda_{k,t}}^{-1}(U_{k,t}) \right)$$

and impose a recursive dynamic on the latent process $(\lambda_t)_{t \in \mathbb{Z}}$ as in (3.1).

Finally, we define a semiparametric model

$$Y_{i,t} = F_{i,\lambda_{i,t}}^{-1}(U_{i,t}), \quad 1 \leq i \leq k, \quad \lambda_t = g_{\theta_0}(\lambda_{t-1}, Y_{t-1}, X_{t-1}), \quad (3.3)$$

with $\theta_0 \in \Theta \subset \mathbb{R}^Q$. For conciseness, (3.3) will be written $Y_t = F_{\lambda_t}^{-1}(U_t)$.

Let us comment the structure of model (3.3). From the standard properties of quantile functions, the distribution of $Y_{i,t}$ conditionally on past observations is given by $P_i(\cdot|\lambda_{i,t})$ while the joint cdf of Y_t is given by

$$y \mapsto C\left(F_{1,\lambda_{1,t}}(y_1), \dots, F_{k,\lambda_{k,t}}(y_k)\right).$$

Conditionally on past values, the copula is used to model the mutual dependence between the coordinates and the specified marginal distributions are then compatible with the univariate observation-driven models defined in (3.1). Let us give an example for $k = 2$. If P_{1,s_1} is a Bernoulli distribution with parameter $(1 + \exp(-s_1))^{-1}$ and P_{2,s_2} is the Poisson distribution with parameter $\exp(s_2)$, we obtain a bivariate dynamic for which, setting $\mathcal{P}_{t-1} = ((Y_{t-j}, X_{t-j})_{j \geq 1})$,

$$\mathbb{P}(Y_{1,t} = 1 | \mathcal{P}_{t-1}) = (1 + \exp(-\lambda_{1,t}))^{-1}, \quad \mathbb{P}(Y_{2,t} = j | \mathcal{P}_{t-1}) = \exp(-\exp(\lambda_{2,t})) \frac{\exp(j\lambda_{2,t})}{j!},$$

for any non-negative integer j . In the regression case (i.e. g_{θ_0} only depends on X_{t-1}), we get a logistic regression for the first coordinate and a Poisson regression for the second one. This specific example will be studied in details throughout the paper.

The paper is organized as follows. In Section 8, we study stationarity properties of model (3.3) but we will mainly focus on linear type mappings g in the rest of the paper. In Section 3.3, we first study inference of the autoregressive parameter θ_0 . Here, we will use pseudo-maximum likelihood estimators and no specific assumptions will be required for the copula C . Moreover, some of the marginal cdf will be not necessarily specified,

as in the case of ARMA or GARCH components. We next consider a parametric model, with a copula C depending on a finite number of parameters. Here, all the marginal cdf will be specified. In this setting we assume that

$$C(du_1, \dots, du_k) = C_{R_0}(du_1, \dots, du_k) := c_{R_0}(u_1, \dots, u_k) du_1 \cdots du_k, \quad R_0 \in \Gamma \subset \mathbb{R}^S. \quad (3.4)$$

For the model defined by (3.3) and (3.4), the parameter of interest is the pair $(\theta'_0, R'_0)'$. Numerically, a joint estimation of this pair is a non trivial task. This is why we will plug a pseudo-likelihood estimator of parameter θ_0 in the expression of the true likelihood function and we then optimize it with respect to R_0 . We show that this two-step approach leads to a consistent estimation of the conditional distribution $P(\cdot|s) = P_{R_0}(\cdot|s)$, even if parameter R_0 is not identifiable. From this result, confidence regions for the prediction can be obtain from a Monte-Carlo procedure. For a Gaussian copula model, a sharper result is obtained. Since identification of the correlation matrix R_0 is automatic, the likelihood estimator of R_0 is shown to be strongly consistent. Throughout the paper, we illustrate our results with a bivariate GARCH/INGARCH model for continuous/count time series data and a bivariate logistic/INGARCH model for binary/count time series data. Numerical experiments and an application of our results to two real data sets are given in Section 3.4. A conclusion and some perspectives for our work are given in Section 3.5. Finally a supplementary material section 3.6 is provided at the end.

3.2 Stability properties

3.2.1 Existence of stationary solutions

We provide below a set of sufficient conditions ensuring existence and uniqueness of a stationary and ergodic solution for the recursions (3.3). In time series analysis, finding such conditions is often a first important step needed for studying consistency of inference procedures. Although the result given below can be applied to some non-linear functions g in (3.3), we will only consider linear type dynamics in the rest of the paper. See Subsection 3.2.2 below for a more specific result. In what follows, for any positive integer j , we denote by \preceq the classical ordering relation on \mathbb{R}^j , i.e. $x \preceq x'$ if and only if $x_i \leq x'_i$ for $i = 1, \dots, j$. Moreover $|\cdot|_1$ denotes the ℓ_1 norm on \mathbb{R}^j , i.e. $|x|_1 = \sum_{i=1}^j |x_i|$ for $x \in \mathbb{R}^j$. For any matrix C , we also denote by $|C|_{vec}$ the matrix of the same size, obtained by replacing the entries of C by their absolute values. Finally, for $t \in \mathbb{Z}$, let \mathcal{F}_t be the sigma-field generated by

the random vectors (U_s, X_s) , $s \leq t$.

A1. The process $((U_t, X_t))_{t \in \mathbb{Z}}$ is stationary, ergodic and for any $t \in \mathbb{Z}$, U_t is independent from \mathcal{F}_{t-1} .

A2. There exists $s \in G := G_1 \times \dots \times G_k$ such that :

$$\mathbb{E}[|g(s, F_s^{-1}(U_0), X_0)|_1] < \infty$$

A3. There exists a square matrix H of size k , with nonnegative elements and such that $\rho(H) < 1$ and a.s.

$$\forall (s_1, s_2) \in G^2, \quad \mathbb{E}[|g(s_1, F_{s_1}^{-1}(U_0), X_0) - g(s_2, F_{s_2}^{-1}(U_0), X_0)|_{vec} | \mathcal{F}_{-1}] \preceq H |s_1 - s_2|_{vec}$$

Theorem 10 *Let Assumptions **A1-A3** hold true. There then exists a unique stochastic process $((Y_t, \lambda_t))_{t \in \mathbb{Z}}$ solution of (3.3) with $g_{\theta_0} = g$ and which is stationary, $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ -adapted and such that $\mathbb{E}(|\lambda_0|_1) < \infty$. Moreover the process $((Y_t, \lambda_t, X_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.*

Notes

1. The previous result only guarantees existence of an integrable solution. For statistical inference, higher-order moment conditions for this solution are sometimes required. In 3.6, Lemma 11 gives a useful criterion to check existence of such moments for this unique solution.
2. Assumption **A3** can be seen as a conditional contraction property obtained equation by equation. The use of conditional expectation is necessary if we include exogenous covariates in the dynamic. This approach has been used recently by [32] who provided many sufficient conditions for existence of stationary solutions for stochastic recursions of this type. Let us emphasize that our vectorial approach avoids the delicate choice of the norm for getting contraction properties, as in [57] for studying multivariate count models. Indeed for multivariate time series, the choice of such a norm depends on the model and cannot be universal. On the other hand, contraction properties are often known for univariate dynamics. Our approach is suitable for exploiting these equation-by-equation contraction properties. Interestingly, one can obtain optimal results for studying some models with positive coefficients, such as for the GARCH-INGARCH model developed in the next sections.

3.2.2 Specific results for linear type dynamics

In this section, we consider that for $1 \leq i \leq k$, $G_i = G$ where G is either equal to \mathbb{R}_+ or to \mathbb{R} . We specify the previous results when the latent process follows the dynamic

$$\lambda_t = d + B\lambda_{t-1} + A\bar{Y}_{t-1} + \Gamma X_{t-1}, \quad t \in \mathbb{Z}, \quad (3.5)$$

with $d \in G^k$, A and B are square matrices of size k and with coefficients in G , Γ is a matrix of size $k \times m$ and with coefficients in G and $\bar{Y}_{i,t} = g_i(Y_{i,t})$ where for $1 \leq i \leq k$, $g_i : E_i \mapsto G$ is a measurable mapping. The case of linear type dynamics is already interesting as many interesting univariate observation-driven models are based on some similar linear properties. Though the practical implementation of our models with the dynamic (3.5) will be only considered when the matrix B is diagonal, we give below a set of sufficient conditions ensuring **A2-A3** for general matrices B . For a vector $c \in \mathbb{R}^k$, we denote by $diag(c)$ the diagonal matrix of size k with diagonal elements c_1, \dots, c_k .

L1. For $s \in G^k$ and $1 \leq i \leq k$, the application $g_i \circ F_{i,s_i}^{-1}$ is integrable with respect to the Lebesgue measure on $[0, 1]$ and X_1 is integrable.

L2. For any $1 \leq i \leq k$, there exists $c_i > 0$ such that for every $(s_i, s'_i) \in G^2$,

$$\int_0^1 \left| g_i \circ F_{i,s_i}^{-1}(u) - g_i \circ F_{i,s'_i}^{-1}(u) \right| du \leq c_i |s_i - s'_i|.$$

L3. The spectral radius $\rho(|A|_{vec}diag(c) + |B|_{vec})$ is less than one.

The following result is a straightforward corollary of Theorem 10. In particular the matrix H in **A3** is given by $|A|_{vec}diag(c) + |B|_{vec}$.

Corollary 1 *Let Assumptions **L1-L3** and Assumption **A1** hold true. The conclusions of Theorem 10 are then valid.*

3.2.3 Examples of linear dynamics

For defining multivariate stationary time series models of type (3.5), the most constraining assumption to check is Assumption **L2** which imposes, coordinatewise, a Lipschitz type property on the autoregressive function. In the literature, there exist many univariate dynamics satisfying such a property. A general class of univariate positive time series models for which such a property holds true has been considered in [26], using stochastic

ordering properties. In this latter case, g_i is simply the identity function and the distribution $P_i(\cdot|s_i)$, defined from an exponential family, has mean s_i . However, there also exist additional dynamics for which **L2** is satisfied and we provide a discussion below. In what follows, we denote by (U_1, \dots, U_k) an arbitrary random vector with uniform marginals.

1. For count data, a natural univariate dynamic is obtained from the Poisson distribution. A popular one is the linear dynamic, i.e. $g_i(y) = y$ and F_{i,s_i} is the cdf of the Poisson distribution with parameter $s_i > 0$. In this case, Assumption **L2** is satisfied with $c_i = 1$ from the stochastic ordering property, i.e. $F_{i,s_i}^{-1} \leq F_{i,s'_i}^{-1}$ if $s_i \leq s'_i$ and the fact that $\mathbb{E}F_{i,s_i}^{-1}(U_i) = s_i$. See in particular [26], Proposition 4 and its proof. To accommodate with negative correlations, one can define a log-linear model as in [54]. In this case, we set $g_i(y) = \log(1 + y)$ and F_{i,λ_i} denotes the Poisson distribution of parameter $\exp(s_i)$. In this case, **L2** is satisfied. A proof can be found in [54], see the proof of their Lemma 2.1. For the reader convenience, we give a different proof here. From stochastic ordering and the monotone property of the logarithm function, if $s_i \leq s'_i$, we have

$$\begin{aligned} & \mathbb{E} \left| \log \left(1 + F_{i,s_i}^{-1}(U_i) \right) - \log \left(1 + F_{i,s'_i}^{-1}(U_i) \right) \right| \\ &= \mathbb{E} \log \left(1 + F_{i,s'_i}^{-1}(U_i) \right) - \mathbb{E} \log \left(1 + F_{i,s_i}^{-1}(U_i) \right) \leq s'_i - s_i. \end{aligned} \quad (3.6)$$

The last inequality can be obtained from the mean value theorem, by noticing that if X_μ follows a Poisson distribution of parameter $\exp(\mu)$, then $f : \mu \rightarrow \mathbb{E} \log(1 + X_\mu)$ has a derivative given by

$$f'(\mu) = \sum_{k \geq 0} \log \left(1 + \frac{1}{k+1} \right) e^{-e^\mu} \frac{e^{\mu(k+1)}}{k!} < e^{-e^\mu} \left(1 + \sum_{k \geq 0} \frac{e^{\mu k}}{k!} \right) = 1,$$

if we use the inequality $\log(1 + x) \leq x$ for $x \geq 0$. It is also possible to include others models such as the count time series model based on the negative binomial distribution, as in [26]. In this case parameter s_i is the mean $r(1-p)/p$ of the binomial distribution with parameters r and p and **L2** is always satisfied with $c_i = 1$.

2. Let us next discuss the case of binary time series by assuming that $F_{i,s_i}^{-1}(U_i) = \mathbb{1}_{U_i > 1 - F(s_i)}$ meaning that $P_i(\cdot|s_i)$ is the Bernoulli distribution on parameter $F(s_i)$ where F is a given cdf. See [106] for the stability property of such univariate dyna-

mics. We have here

$$\mathbb{E} \left| F_{i,s'_i}^{-1}(U_i) - F_{i,s_i}^{-1}(U_i) \right| \leq |F(s'_i) - F(s_i)| \leq c_i |s_i - s'_i|.$$

Here c_i denotes the Lipschitz constant of F . Two well-known cdf F are widely used in practice, the logistic $F(\mu) = (1 + e^{-\mu})^{-1}$, $\mu \in \mathbb{R}$, for which $c_i = 1/4$ and the cdf of the standard Gaussian distribution and for which $c_i = 1/\sqrt{2\pi}$ (probit model).

3. Finally, let us discuss the case of continuous components. For a GARCH component, $P_i(\cdot|s_i)$ is the probability distribution of $\sqrt{s_i}\epsilon$ where ϵ is a centered random variable with unit variance. We then have $F_{i,s_i}^{-1}(U_i) = \sqrt{s_i}F_\epsilon^{-1}(U_i)$ where F_ϵ is the cdf of ϵ . If $g_i(y) = y^2$, it is easily seen that **L2** is satisfied with $c_i = 1$. One can also consider the log-GARCH model which does not impose any positivity condition on lag parameters. Log-GARCH models are discussed in [60] and are the analogs of log-linear Poisson autoregressions for count data. With our formulation, $P_i(\cdot|s_i)$ is now the probability distribution of $\exp(s_i/2)\epsilon$ (a linear dynamic is specified on the logarithm of the conditional variance). Setting $g_i(y) = \log(y^2)$ and assuming that $\mathbb{E}|\log(\epsilon)|^2 < \infty$, **L2** is satisfied with $c_i = 1$.

Another interesting dynamic concerns the linear ARMA(1, 1) dynamic. As explained in the introduction, this dynamic is equivalent, up to a reparametrization, to the case where $P_i(\cdot|s_i)$ is the probability distribution of the sum $s_i + \epsilon$, with ϵ a centered random variable. It is then possible to check **L2** with $g_i(y) = y$ and $c_i = 1$.

Notes

1. From the previous discussion, it is possible to combine many univariate dynamics to define multivariate models for mixed data and for which the dynamic of the latent process is given by (3.5). However, we have to take care that some of these univariate models impose a sign restriction on the parameters. For instance, for the GARCH model and the linear INGARCH model, the autoregressive parameters have to be nonnegative and the latent processes $\lambda_{i,t}$ are required to take positive values as they represent the conditional standard deviation and the intensity respectively. On the other hand some univariate models such as log-linear Poisson autoregressions, log-GARCH, ARMA and binary time series do not impose a sign restriction for the autoregressive parameters. It is then not natural to combine models with and without sign restriction, though one can always impose an artificial positivity

condition for models without sign restriction. Technically, the more general model (3.1) can be used for combining any dynamics of the previous type (whatever the signs of the univariate latent processes). However in this case, specifying a function g preserving the sign constraints could appear to be quite arbitrary. It will be then implicit that a continuous/count bivariate time series model can be constructed by combining standard GARCH, INGARCH or negative binomial dynamics while a continuous/count/binary time series model is obtained by combining log-GARCH, ARMA, log-linear INGARCH and logistic/probit dynamics.

2. As explained above, a linear type equation (3.5) is already interesting for generalizing well-known univariate dynamics and Corollary 1 provides a result for stability for the model. Such a result will be applied to two examples studied in details in the rest of the paper. Its main interest is pedagogical as it illustrates that many classical univariate models can be combined together for defining a multivariate times series model. But we point out that Corollary 1 is not necessarily sharp with respect to Theorem 10. For instance, assume (3.5) with **L1-L2** satisfied and with a component, say i , defined from the log-GARCH model. In this case, we have

$$\log \left(F_{i,s_i}^{-1}(U_i)^2 \right) = s_i + \log \left(F_\epsilon^{-1}(U_i)^2 \right)$$

and one can directly check Assumption **A3** with a matrix H such that $H(\ell, i) = |A(\ell, i) + B(\ell, i)|$ and $H(\ell, j) = |A(\ell, j)|c_j + |B(\ell, i)|$ if $j \neq i$. Condition $\rho(H) < 1$ is less restrictive than **L3** in this case because of the inequalities

$$H(\ell, i) \leq (|A|_{vec}diag(c) + |B|_{vec})(\ell, i).$$

The same improvement can be obtained if we consider an ARMA component.

3. Our framework also includes some multivariate time series models for discrete data found in the literature. [95] considered a multivariate binary time series models with applications to electricity price spikes. The conditional distribution of each marginal can be logistic, Gaussian or of a more general form and the dynamic on the latent process is similar to (3.5). A copula structure is also used for modeling the simultaneous dependence for the multivariate time series. In [57], multivariate count autoregressions have been introduced. In these models, the conditional distribution of each marginal is Poisson and both the linear and the log-linear case are studied. The simultaneous dependence is also based on a copula. A main dif-

ference with our approach concerns the generations of univariate Poisson marginal distributions. While we use directly the inverse of the Poisson cdf to construct our model, [57] simulates several independent copies of the copula to generate exponential inter-arrival times of a Poisson process. However, both models have very similar properties.

3.2.4 Two specific examples

The model GAIN

The GARCH-INGARCH (abbreviated as GAIN) mixed model combines the dynamic of the univariate GARCH model of [9] and the Poisson autoregressive model called INGARCH in [47]. Here, $E_1 = \mathbb{R}$, $E_2 = \mathbb{N}$, $G_1 = G_1 = \mathbb{R}_+$ and we define the model as follows.

$$Y_t = \begin{cases} Y_{1,t} &= \lambda_{1,t}^{1/2} F_\epsilon^{-1}(U_{1,t}) \\ Y_{2,t} &= \inf \left\{ y \in \mathbb{N} : \sum_{j=0}^y e^{-\lambda_{2,t}} \frac{\lambda_{2,t}^j}{j!} \geq U_{2,t} \right\} \end{cases} \quad (3.7)$$

$$\lambda_t = d + B\lambda_{t-1} + A\bar{Y}_{t-1} + \Gamma X_{t-1}$$

where $\bar{Y}_t = (Y_{1,t}^2, Y_{2,t})'$ and F_ϵ^{-1} stands for the inverse of the cumulative probability function of a centered random variable ϵ with unit variance. The elements of d , A and B are assumed to be nonnegative. The following result gives a necessary and sufficient condition for the existence of some solutions.

Proposition 8 *Consider the model (3.7) and let Assumption A1 holds true with X_1 integrable.*

1. *If $\rho(A+B) < 1$, there exists a unique solution $(Y_t)_{t \in \mathbb{Z}}$ to (3.7) such that $((Y_t, \lambda_t))_{t \in \mathbb{Z}}$ is a stationary, $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ -adapted and integrable process. Moreover, the process $((Y_t, \lambda_t, X_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.*
2. *Conversely, assume that d has positive coordinates and suppose that $((Y_t, \lambda_t))_{t \in \mathbb{Z}}$ is a stationary, $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ -adapted and integrable process solution of (3.7). Then $\rho(A+B) < 1$.*

We next give a result for existence of higher-order moments.

Proposition 9 Consider the model (3.7) under the assumptions of Proposition 8 and assume that for some integer $r \geq 1$, $\mathbb{E}(|X_0|_1^r) < \infty$. If in addition $\mathbb{E}^{1/r}[\epsilon^{2r}] < \infty$ and $\rho(B + \text{Adiag}(\mathbb{E}^{1/r}[\epsilon^{2r}], 1)) < 1$, then

$$\mathbb{E}(|\bar{Y}_0|_1^r) < \infty \text{ and } \mathbb{E}(|\lambda_0|_1^r) < \infty.$$

Notes

1. Under the stationarity condition $\rho(A + B) < 1$ and if there exists $r' > 1$ such that $\mathbb{E}\epsilon^{2r'} < \infty$ and $\mathbb{E}|X_0|_1^{r'} < \infty$, one can always find $r > 1$ such that $\rho(B + \text{Adiag}(\mathbb{E}^{1/r}[\epsilon^{2r}], 1)) < 1$. Existence of a moment of order larger than 1 is then obtained without any restriction on the lag parameters A and B . This property will be particularly important for proving consistency and asymptotic normality of pseudo-likelihood estimators.
2. For this model, one can compute some auto-covariances. For simplicity, assume that $\Gamma = 0$ (no exogenous covariates are included in the model) and set $\Gamma(h) = \text{Cov}(\bar{Y}_t, \bar{Y}_{t-h})$ for $h \geq 1$, which is meaningful when $r = 2$ in Proposition 9. Setting $\varepsilon_t = \bar{Y}_t - \lambda_t$, note that we have the weak VARMA(1, 1) representation $\bar{Y}_t = d + (A + B)\bar{Y}_{t-1} + \varepsilon_t - B\varepsilon_{t-1}$. Hence $\Gamma(h)$ is proportional to $(A + B)^h$ and all the covariances are positive. Note also that from the GARCH structure for the first component, all the covariances between $Y_{1,t}$ and $Y_{2,t-h}$, $h \geq 1$ are automatically equal to 0. For other models, such as the BIP model discussed below, it is quite difficult to get similar recursive formula for the autocovariances, since in general, both $\mathbb{E}[Y_{i,t}|\mathcal{F}_{t-1}]$ or $\mathbb{E}[\bar{Y}_{i,t}|\mathcal{F}_{t-1}]$ will be given by a nonlinear function of $\lambda_{i,t}$.

The model BIP

In what follows, we consider a bivariate time series model compatible with sequences of binary/count data. This model, called Binary-Poisson (abbreviated as BIP) mixed model, combines an autoregressive logistic model with a log-linear Poisson autoregressive model. For the sake of simplicity, we consider a model with two coordinates but extensions including several binary/count time series is straightforward. Here $E_1 = \{0, 1\}$, $E_2 = \mathbb{N}$ and $G_1 = G_2 = \mathbb{R}$. The model writes as follows.

$$Y_t = \begin{cases} Y_{1,t} &= \mathbb{1}_{\{U_{1,t} \geq 1 - F(\lambda_{1,t})\}}, & F : s \mapsto \frac{1}{1 + \exp(-s)} \\ Y_{2,t} &= \inf \left\{ y \in \mathbb{N} : \sum_{j=0}^y e^{-\lambda_{2,t}} \frac{e^{j\lambda_{2,t}}}{j!} \geq U_{2,t} \right\} \end{cases} \quad (3.8)$$

$$\lambda_t = d + B\lambda_{t-1} + A\bar{Y}_{t-1} + \Gamma X_{t-1}$$

where $\bar{Y}_t = (Y_{1,t}, \log(1+Y_{2,t}))'$. Here the coefficients in d, A, B, Γ and the covariate process $(X_t)_{t \in \mathbb{Z}}$ can take arbitrary signs.

Proposition 10 *Consider the model (3.8) and suppose that Assumption A1 holds true with X_0 integrable. Assume furthermore that $\rho(|B|_{vec} + |A|_{vec} \text{diag}(1/4, 1)) < 1$. Then the conclusions of Corollary 1 are valid and we also have $\mathbb{E}(|\bar{Y}_0|_1) < \infty$.*

Note Contrarily to the model GAIN, our conditions for stationarity are not optimal. This is already the case for the univariate log-linear Poisson autoregressive model for which our stability condition is equivalent to $|A_{2,2}| + |B_{2,2}| < 1$. See [39], Proposition 17 for a sharper result. For univariate logistic autoregressions, our condition writes as $|A_{1,1}|/4 + |B_{1,1}| < 1$. This condition is similar to that of [106] but much more restrictive than the condition $|B_{1,1}| < 1$ given in [56] or in [128], Proposition 2. Our results seems to give optimal conditions for some dynamics with positive latent processes, such as for the model GAIN or for multivariate linear Poisson autoregressions. In the latter case, see [30], Theorem 4. On the other hand, we point out that our approach can be applied to many nonlinear multivariate dynamics and allows exogenous covariates not necessarily strictly exogenous (i.e. the noise process $(U_t)_{t \in \mathbb{Z}}$ and the covariate process $(X_t)_{t \in \mathbb{Z}}$ are not necessarily independent).

We now investigate existence of some higher-order moments that will be necessary for statistical inference. To this end, we denote by \bar{A} the matrix obtained by replacing the first column of A by the null vector $\mathbf{0}$. We recall that for a matrix C of size $k \times k$, its infinite norm, denoted by $|C|_\infty$, is defined by $|C|_\infty = \max_{1 \leq i \leq k} \sum_{j=1}^k |C(i, j)|$.

Proposition 11 *Consider the model (3.8) and assume that the assumptions of Proposition 10 are valid. Suppose furthermore that $\|\bar{A}\|_{vec} + |B|_{vec} < 1$ and that for any $r > 0$, $\mathbb{E}(\exp(r|X_0|_1)) < \infty$. Then, for any $r > 0$, we have*

$$\mathbb{E}(\exp(r|\lambda_0|_1)) < \infty, \quad \mathbb{E}(|Y_0|_1^r) < \infty.$$

3.3 Statistical inference

In this section, we detail our estimation procedures for the dynamic parameters as well as for the copula parameters. We first explain the main idea of these methods and

introduce our estimators in the two first subsections. Next, we provide some asymptotic results in the two last subsections.

3.3.1 Estimation of dynamic parameters

Going back to the dynamic (3.3), we now assume that the function g depends on a vector of parameters $\theta \in \Theta \subset \mathbb{R}^Q$ and we denote by θ_0 the vector of parameters associated to a given sample Y_1, \dots, Y_n generated by (3.3). Under the linear assumption (3.5), one can set $\theta = (d', \text{vec}(\Gamma)', \text{vec}(A)', \text{vec}(B)')'$ where for a matrix C of any size, $\text{vec}(C)$ denotes the usual vectorization of the matrix C .

Inference of parameter θ can be done by minimizing a suitable contrast. We construct such contrasts from univariate ones (e.g. log conditional densities). We then adopt a (conditional) pseudo-maximum likelihood approach by writing the likelihood function as if the coordinates of the Y_t 's were independent conditionally on their past values. However, the univariate contrasts are not necessarily defined from the conditional log-densities and we provide a more flexible approach by allowing more general univariate contrasts such as least squares or Gaussian quasi-maximum likelihood. We recall that E_1, \dots, E_k denote the state spaces of the different univariate time series, typically one of the sets $\{0, 1\}$, \mathbb{N} and \mathbb{R} . Notations G_1, \dots, G_k are used for the state spaces of the latent processes (typically $G_i = \mathbb{R}$ or $G_i = \mathbb{R}_+$). Finally, for $1 \leq i \leq k$, μ_i will denote either the Lebesgue measure on the real line or the counting measure on $\{0, 1\}$ or on \mathbb{N} .

To this end, for $1 \leq i \leq k$, let $(y, s_i) \mapsto h_{i,y}(s_i)$ be a measurable mapping defined on $E_i \times G_i$ and taking real values such that

$$\mathbb{E}(h_{i,Y}(s_i)) \geq \mathbb{E}(h_{i,Y}(\bar{s}_i)) \text{ and } \mathbb{E}(h_{i,Y}(s_i)) = \mathbb{E}(h_{i,Y}(\bar{s}_i)) \Rightarrow s_i = \bar{s}_i \quad (3.9)$$

whenever $Y \sim P_i(\cdot|s_i)$. Here are two important examples of such mappings.

1. If the distribution $P_i(\cdot|s_i)$ is absolutely continuous with respect to μ_i , i.e. $P_i(dy|s_i) = p_i(y|s_i)\mu_i(dy)$, one can use the opposite of logarithm of the density $h_{i,y}(s_i) = -\log p_i(y|s_i)$. This case is particularly important when $P_i(\cdot|s_i)$ is specified (e.g. Poisson distribution with parameter s_i).
2. Other standard objective functions such as $h_{i,y}(s_i) = (y - s_i)^2$ (least-squares esti-

mation) which is adapted to ARMA processes or

$$h_{i,y}(s_i) = \frac{y^2}{\ell(s_i)} + \log(\ell(s_i))$$

which corresponds in the context of GARCH type models to Gaussian Quasi-Maximum Likelihood Estimation. Here, the choice $\ell(s_i) = s_i$ corresponds to the standard GARCH model whereas $\ell(s_i) = \exp(s_i)$ corresponds to the log-GARCH model. Note that the Gaussian QMLE can be also used for ARMA type models, setting $h_{i,y}(s_{1,i}, s_{2,i}) = (y - s_{1,i})^2/s_{2,i} + \log(s_{2,i})$, where the additional parameter $s_{2,i}$ corresponds to the variance of the noise ε , see (3.2). Note that parameter $s_{2,i}$ is an additional parameter with respect to the autoregressive parameters involved in the dynamic of the latent process but it will be straightforward to adapt our results to cover this case as well as others two-parameters distributions such as the negative binomial distribution for count components.

If $(\lambda_t(\theta))_{t \in \mathbb{Z}}$ denotes the process defined recursively by

$$\lambda_t(\theta) = g_\theta(\lambda_{t-1}(\theta), Y_{t-1}, X_{t-1}), \quad \theta \in \Theta, \quad t \in \mathbb{Z},$$

we define an estimator of θ_0 by minimizing the criterion

$$\theta \mapsto \ell_n(\theta) := n^{-1} \sum_{t=1}^n \sum_{i=1}^k h_{i,Y_{i,t}}(\lambda_{i,t}(\theta)), \quad (3.10)$$

where for $t = 1, \dots, n$,

$$\lambda_t(\theta) = g_\theta(\lambda_{t-1}(\theta), Y_{t-1}, X_{t-1}).$$

However, as usual with observation-driven models, the previous estimator can not be computed using the available data. To get a feasible estimator, the dynamic of the latent process has to be initialized and we consider a process $(\bar{\lambda}_t(\theta))_{t \geq 0}$ defined by $\bar{\lambda}_0(\theta) = \bar{\lambda}_0$ for every θ in Θ , where $\bar{\lambda}_0$ is a deterministic, and then recursively by $\bar{\lambda}_t(\theta) = g_\theta(\bar{\lambda}_{t-1}(\theta), Y_{t-1}, X_{t-1})$ for $t \geq 1$. We then define the computable estimator

$$\hat{\theta} = \arg \min_{\theta \in \Theta} n^{-1} \sum_{t=1}^n \sum_{i=1}^k h_{i,Y_{i,t}}(\bar{\lambda}_{i,t}(\theta)), \quad (3.11)$$

Note that $\hat{\theta}_n$ can be obtained equation by equation for model (3.5) with a diagonal matrix

B. Indeed, in this case, parameter θ is composed of k sub-vectors $\theta^{(1)}, \dots, \theta^{(k)}$ and we have $\hat{\theta} = (\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(k)})$ with

$$\hat{\theta}^{(i)} = \arg \min_{\theta^{(i)}} n^{-1} \sum_{t=1}^n h_{i, Y_{i,t}}(\bar{\lambda}_{i,t}(\theta^{(i)})), \quad 1 \leq i \leq k.$$

In particular, for $1 \leq i \leq k$, $\theta^{(i)} = (d_i, \Gamma(i, 1), \dots, \Gamma(i, m), A(i, 1), \dots, A(i, k), B(i, i))$.

Note. Let us provide some guidelines for the choice of the univariate contrasts. From the equation-by-equation optimization property, fitting the multivariate model is equivalent to fitting k univariate models with exogenous regressors to the individual times series. For count or binary models, the opposite of the log conditional density is then a standard choice. For GARCH models, the Gaussian QML is the most popular and corresponds to the opposite of the log conditional density when the noise is Gaussian. For ARMA models, the least-square objective function is also natural and corresponds to the opposite of the log conditional density in the Gaussian case. For these two last examples, these choices are robust to the noise density misspecification (i.e. in the univariate case, they lead to consistent estimators even if the noise density is not Gaussian). Additionally, from this strategy, the estimation procedure (3.11) corresponds to conditional likelihood estimation when the $U'_{i,t}$ s are independent (independence copula) and the noise densities are Gaussian for the continuous components. We then adopt the aforementioned choices for investigating theoretical properties of the estimators and for practical implementation.

3.3.2 Estimation of copula parameters

Our aim here is to define an estimator of parameter R_0 obtained from an estimator of the dynamic parameters. Note that all the marginal conditional probability distribution $P_i(\cdot | s_i)$, $s_i \in G_i$ are known. We assume here that $h_{i, s_i} = -\log p_i(\cdot | s_i)$ for $1 \leq i \leq k$. We recall that $p_i(\cdot | s_i)$ denotes the probability density of $P_i(\cdot | s_i)$ with respect to the measure μ_i . An estimator of parameter θ_0 can be obtained as explained in the previous section. The model being parametric, likelihood inference is adapted for estimating R_0 . For simplicity, we assume that $F_{1, s_1}, \dots, F_{\ell, s_\ell}$ are diffeomorphisms (the continuous components) and $F_{\ell+1, s_{\ell+1}}, \dots, F_{k, s_k}$ are cdf corresponding to discrete distributions with a support included in $\{0, 1\}$ or \mathbb{N} (binary or count). Setting for $1 \leq t \leq n$ and $1 \leq i \leq k$, $Z_{i,t}(\theta) = F_{i, \bar{\lambda}_{i,t}(\theta)}(Y_{i,t})$ and $Z_{i,t}^-(\theta) = F_{i, \bar{\lambda}_{i,t}(\theta)}^-(Y_{i,t}^-)$, the approximated conditional log-

likelihood function for the model is defined by

$$\begin{aligned} \ell_n(\theta, R) &= \sum_{t=1}^n \log P_R(Y_t | \bar{\lambda}_t(\theta)) \\ &= \sum_{t=1}^n \sum_{i=1}^{\ell} \log p_i(Y_{i,t} | \bar{\lambda}_{i,t}(\theta)) \\ &\quad + \sum_{t=1}^n \log \left\{ \int_{Z_{\ell+1,t}^-(\theta)}^{Z_{\ell+1,t}(\theta)} \cdots \int_{Z_{k,t}^-(\theta)}^{Z_{k,t}(\theta)} c_R(Z_{1,t}(\theta), \dots, Z_{\ell,t}(\theta), u_{\ell+1}, \dots, u_k) du_{\ell+1} \cdots du_k \right\} \end{aligned}$$

Here, $p_R(y|s)$ denotes the conditional density of Y_t given $\lambda_t(\theta) = s$ when the copula parameter is R and the autoregressive parameters are given by θ .

We adopt a plug-in approach by first estimating θ_0 and then optimize the partial log-likelihood function. A possible estimator of R_0 can be then obtained by minimizing

$$R \mapsto -n^{-1} \ell_n(\hat{\theta}, R),$$

where $\hat{\theta}$ is the estimator obtained as explained in the previous section.

3.3.3 Asymptotic results for inference of autoregressive parameters

In this section, we give a simple set of sufficient conditions ensuring consistency and asymptotic normality of pseudo-likelihood estimators.

A4 For $1 \leq i \leq k$ and any $s_i^* \in G_i$, the mapping

$$\mu \mapsto \int h_{i,y}(s_i) P_i(dy | s_i^*)$$

is uniquely minimized at point $s_i = s_i^*$.

A5 For $1 \leq i \leq k$, we have

$$\mathbb{E} \int \sup_{\theta \in \Theta} |h_{i,y}(\lambda_{i,0}(\theta))| P_i(dy | \lambda_{i,0}(\theta)) < \infty.$$

A6 For $1 \leq i \leq k$,

$$\frac{1}{n} \sum_{t=1}^n \int \sup_{\theta \in \Theta} |h_{i,y}(\lambda_{i,t}(\theta)) - h_{i,y}(\bar{\lambda}_{i,t}(\theta))| P_i(dy | \lambda_{i,0}(\theta_0)) = o_{\mathbb{P}}(1).$$

A7 We have

$$\lambda_0(\theta) = \lambda_0(\theta_0) \text{ a.s.} \Rightarrow \theta = \theta_0.$$

The proof of the following result is straightforward and follows from standard arguments. See for instance [125], Theorem 5.3.1., for the Gaussian QMLE but the arguments used can be extended to this more general setup.

Theorem 11 *Let Assumptions A1-A7 hold true with Θ a compact subset of \mathbb{R}^Q . We then have $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0$ a.s.*

We now turn on the asymptotic normality of our estimator. In what follows, for a function $f : \Theta \rightarrow \mathbb{R}$, we denote by $\nabla f(\theta)$ the gradient vector (column vector of the partial derivatives) and $\nabla^{(2)} f(\theta)$ the Hessian matrix of f , evaluated at point $\theta \in \Theta$. If $f : \Theta \rightarrow \mathbb{R}^k$, we denote by $J_f(\theta)$ the Jacobian matrix of f at point $\theta \in \Theta$ (we recall that in term of partial derivatives, we have $J_f(\theta)_{i,j} = \frac{\partial f_i}{\partial \theta_j}(\theta)$). For a function f defined on a subset of the real line, we simply denote by \dot{f} and \ddot{f} its first and second derivatives.

A8 For $1 \leq i \leq k$ and $y \in E_i$, the mapping $h_{i,y}$ is two-times continuously differentiable.

Moreover, for any $s_i \in G_i$, we have $\int \ddot{h}_{i,y}(s_i) P_i(dy|s_i) > 0$ and $\int \dot{h}_{i,y}(s_i) P_i(dy|s_i) = 0$.

A9 For $1 \leq i \leq k$, the random mapping $\theta \mapsto \lambda_{i,0}(\theta)$ is almost surely two-times continuously differentiable and the following uniform integrability condition holds true :

$$\mathbb{E} \int \sup_{\theta \in \Theta} \left[\left| \ddot{h}_{i,y}(\lambda_{i,0}(\theta)) \right| \cdot \|\nabla \lambda_{i,0}(\theta)\|^2 + \|\dot{h}_{i,y}(\lambda_{i,0}(\theta)) \nabla^{(2)} \lambda_{i,0}(\theta)\| \right] P_i(dy|\lambda_{i,0}(\theta_0)) < \infty.$$

A10 For $1 \leq i \leq k$, we have

$$\mathbb{E} \int \|\dot{h}_{i,y}(\lambda_{i,0}(\theta_0)) \nabla \lambda_{i,0}(\theta_0)\|^2 P_i(dy|\lambda_{i,0}(\theta_0)) < \infty.$$

A11 For $1 \leq i \leq k$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \int \sup_{\theta \in \Theta} \|\nabla (h_{i,y} \circ \lambda_{i,t})(\theta) - \nabla (h_{i,y} \circ \bar{\lambda}_{i,t})(\theta)\| P_i(dy|\lambda_{i,t}(\theta_0)) = o_{\mathbb{P}}(1).$$

A12 If there exists $x \in \mathbb{R}^Q$ such that $J_{\lambda_0}(\theta_0)x = 0$ a.s. then $x = 0$.

As for consistency, we will not prove the following result. See for instance [125], Theorem 5.6.1., the same arguments can be used for proving Theorem 12 below.

Theorem 12 *Let Assumptions **A1-A12** hold true with θ_0 being located in the interior of the compact parameter space Θ . We then have*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow \mathcal{N}_Q(0, J^{-1}IJ^{-1}),$$

with

$$I = \sum_{i=1}^k \sum_{j=1}^k \mathbb{E} \left[\dot{h}_{i,Y_{i,0}}(\lambda_{i,0}(\theta_0)) \dot{h}_{j,Y_{j,0}}(\lambda_{j,0}(\theta_0)) \nabla \lambda_{i,0}(\theta_0) \nabla \lambda_{j,0}(\theta_0)' \right],$$

$$J = \sum_{i=1}^k \mathbb{E} \left[\ddot{h}_{i,Y_{i,0}}(\lambda_{i,0}(\theta_0)) \nabla \lambda_{i,0}(\theta_0) \nabla \lambda_{i,0}(\theta_0)' \right].$$

Let us note that Assumptions **A8** and **A10** ensure that the process $(M_t)_{t \in \mathbb{Z}}$ defined by $M_t = \sum_{i=1}^k \nabla h_{i,Y_{i,t}}(\lambda_{i,t}(\theta_0))$ is a square-integrable martingale difference. Moreover, Assumptions **A8-A9-A12** entail that the Hessian matrix $\mathcal{H}(\theta) = \sum_{i=1}^k \nabla^{(2)} h_{i,Y_{i,0}}(\lambda_{i,0}(\theta))$ is well defined, uniformly integrable with respect to $\theta \in \Theta$ and with an invertible expectation at point θ_0 . Assumption **A11** guarantees that initializing the latent process has no effect on the asymptotic distribution of the estimator.

3.3.4 Sufficient conditions for **A7** and **A12**

Here we exhibit a set of simple conditions ensuring both identification of autoregressive parameters and non-degeneracy of the derivative of the latent process. We will provide such conditions for the linear dynamic (3.5). Note that the two conditions **A7** and **A12** only involve the autoregressive latent process and not the contrast functions $h_{i,\mu}$. This is why we give a separate study of these two conditions making as few as possible assumptions on the conditional distribution of the multivariate time series model.

In the rest of this section, we assume that the trivariate process $((Y_t, X_t, U_t))_{t \in \mathbb{Z}}$ is stationary and then that Assumptions **L1-L3** are satisfied. We recall that $\theta_0 = (d_0', \text{vec}(\Gamma_0)', \text{vec}(A_0)', \text{vec}(B_0)')'$ denotes the true value of the parameter. For any $t \in \mathbb{Z}$, we also denote by \mathcal{F}_t the sigma-field generated by (U_j, X_j) , $j \leq t$. We will need the following set of assumptions :

I0 For any $\theta \in \Theta$, $\rho(B) < 1$,

I1 For any $v \in \mathbb{R}^m$, we have

$$v'X_1 \in \mathcal{F}_0 \vee \sigma(U_1) \Rightarrow v = 0.$$

I2 For $1 \leq i \leq k$, the function g_i is non-degenerate on the support of P_i and the density c_{R_0} of the copula is positive everywhere.

I3 If v is equal either to a column vector of A_0 or to a column vector of Γ_0 , the equalities $B^j v = B_0^j v$, $j \geq 1$, entail $B = B_0$.

Lemma 4 *Let Assumptions **I0-I3** hold true for model (3.5). Condition **A7** is then satisfied.*

Notes

1. For our main setup, the matrices B and B_0 are assumed to be diagonal. In this case, Assumption **I3** is satisfied as soon as all the rows of the concatenated matrix $C := [A_0, \Gamma_0]$ are non-null.
2. Assumption **I1** is more difficult to interpret. It means that any (non degenerate) linear combination of the covariate process at time t cannot be explained only by past information and the disturbance term U_t . For instance, assume that X_t writes as a square integrable infinite moving average expansion $\sum_{j \geq 0} c_j \varepsilon_{t-j}$ where $(c_j)_{j \geq 0}$ is a sequence of matrices and the random vectors (U_t, ε_t) , $t \in \mathbb{Z}$, are i.i.d. In this case, one can take $\mathcal{F}_t = \sigma((U_j, \varepsilon_j) : j \leq t)$ for $t \in \mathbb{Z}$. If $v'X_1$ is measurable with respect to $\mathcal{F}_0 \vee \sigma(U_1)$ then so is $v'c_0\varepsilon_1$ which has conditional variance $v'c_0\text{Var}(\varepsilon_1|U_1)c_0'v = 0$. When c_0 is invertible and $\text{Var}(\varepsilon_1|U_1)$ is invertible with positive probability, we automatically get $v = 0$ and Assumption **I1** is satisfied.

Under an additional condition on the covariates, called strict exogeneity, we give below an alternative condition to **I1**.

I1' The two processes $(U_t)_{t \in \mathbb{Z}}$ and $(X_t)_{t \in \mathbb{Z}}$ are independent and if $\sum_{j \geq 1} \Phi_j X_{-j} + c = 0$ a.s. then all the matrices Φ_j of size $p \times m$ and the vector c of length p are equal to zero.

The latter condition is satisfied for instance if a linear combination of the coordinates of X_0 cannot be equal to an element of $\sigma(X_{-j} : j \geq 1)$, except if the weights are vanishing. This condition is then the analogue of **I1**, when the noise process and the covariate process are independent. The latter independence condition is often called strict exogeneity in the time series literature.

Lemma 5 *Let Assumptions **I0-I1'-I2-I3** hold true for model (3.5). Condition **A7** is then satisfied.*

The validity of **A12** can be obtained under an additional condition.

I4 The rank of all the column vectors included in the matrices $B_0^j[A_0, \Gamma_0]$, $j \geq 0$, is equal to k .

Lemma 6 *Suppose that either Assumptions **I0-I4** or Assumptions **I0,I1',I2-I4** hold true. Condition **A12** is then satisfied.*

Note. Assumptions **I3-I4** are checked for instance when the block matrix $[A_0, \Gamma_0]$ is of full rank k . However, the latter condition is sufficient but not necessary. For instance if B is diagonal with distinct diagonal elements and the rows of the matrix $C = [A_0, \Gamma_0]$ are all non null, Assumptions **I3-I4** are also satisfied. Indeed in this case, if v_1, \dots, v_{k+m} denote the column vectors of C , the rank of the vectors $B^j v_i$, $0 \leq j \leq k$, $1 \leq i \leq k+m$, equals to the rank of the matrix $[\text{diag}(v_1), \dots, \text{diag}(v_k)] \times I_k \otimes V$, where for $w \in \mathbb{R}^k$, $\text{diag}(w)$ denotes the square diagonal matrix with diagonal elements w_1, \dots, w_k , V denotes the (invertible) Vandermonde matrix associated to $B_0(i, i)$, $1 \leq i \leq k$, I_k is the diagonal matrix of size k and \otimes denotes the Kronecker product.

3.3.5 Examples

In this section, we go back to our two examples of bivariate time series models.

Asymptotic results for the GAIN model

We recall that $\theta = (d', \text{vec}(\Gamma)', \text{vec}(A)', \text{vec}(B)')$ denotes the vector of parameters we have to estimate in the model (3.7). When B is assumed to be diagonal, we simply replace $\text{vec}(B)$ by $\text{diag}(B)$. Here we assume that

$$\Theta \subset \left\{ \theta \in \mathbb{R}_+^Q : \theta_i < 1, \quad Q-1 \leq i \leq Q, \quad \min(\theta_1, \theta_2) \geq d_- \right\},$$

with $Q = 2(m+4)$ and d_- being a positive constant. We combine the Gaussian quasi-likelihood and the Poisson likelihood to estimate the autoregressive parameters, that is $h_{1,y}(\mu_1) = \frac{y^2}{\mu_1} + \log(\mu_1)$ and $h_{2,y}(\mu_2) = \mu_2 - y \log(\mu_2)$.

Proposition 12 *Consider model (3.7) with $\Theta \ni \theta_0$ compact. Suppose that Assumption **A1** and Assumptions **I0, I1** or **I1', I2-I3** hold true. Suppose furthermore that*

$\rho(A_0 + B_0) < 1$ and that there exists $\delta > 0$ such that $\mathbb{E}|X_0|_1^{1+\delta} < \infty$ and $\mathbb{E}\varepsilon^{2(1+\delta)} < \infty$. The pseudo-maximum likelihood estimator is then strongly consistent, i.e.

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0 \text{ a.s.}$$

For asymptotic normality, our result writes as follows.

Proposition 13 *Suppose that all the assumptions of Proposition 12 hold true as well as Assumption I4. Suppose furthermore that θ_0 belongs to the interior of Θ and that $\mathbb{E}\varepsilon^4 < \infty$. We then have the convergence in distribution,*

$$\lim_{n \rightarrow \infty} n^{1/2}(\hat{\theta}_n - \theta_0) = \mathcal{N}_Q(0, J^{-1}IJ^{-1'}),$$

where I and J are given in the statement of Theorem 12.

Asymptotic results for the BIP model

Here, setting $Q = 2(m + 4)$, we assume that

$$\Theta \subset \{\theta \in \mathbb{R}^Q : |\theta_i| < 1, \quad Q - 1 \leq i \leq Q\}.$$

We use the pseudo-maximum approach with $h_{1,y}(s_1) = \log(1 + e^{s_1}) - ys_1$ and $h_{2,y}(s_2) = e^{s_2} - ys_2$.

Proposition 14 *Consider model (3.8) with $\Theta \ni \theta_0$ compact. Suppose that Assumption A1 and Assumptions I0, I1 or I1', I2-I3 hold true. Suppose furthermore that $\rho(|B_0|_{vec} + |A_0|_{vec} \text{diag}(1/4, 1)) < 1$, $\|\bar{A}_0\|_{vec} + \|B_0\|_{vec} < 1$ and that for any $r > 0$, $\mathbb{E}(\exp(r|X_0|_1)) < \infty$. The pseudo-maximum likelihood estimator is then strongly consistent, i.e.*

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0 \text{ a.s.}$$

Additionally, if θ_0 is located in the interior of Θ and if Assumption I4 holds true, we have asymptotic normality

$$\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{N}_Q(0, J^{-1}IJ^{-1'}),$$

where I and J are given in the statement of Theorem 12.

3.3.6 Guidelines for more general models

In this part, we provide some guidelines for checking our conditions on more general models of the form (3.5) when the parameter θ is in a compact set Θ such that $\rho(B) < 1$ for all matrix B in the parameter space. We restrict the study to some univariate models without sign restriction as they appear to be the most flexible for practical applications, though the standard GARCH or INGARCH models are widely used in the literature. Assumptions **A5-A6** and **A9-A10-A11**, which are specific to each univariate dynamic, generally involve some moment conditions on the coordinates of the latent process and then on the model. Lemmas 12, 13 and 14 given in section 3.6 give useful moment conditions and approximation results for the corresponding latent process $\lambda_t(\cdot)$ for models of the form (3.5). One can then give specific moment conditions that are required for each univariate dynamic. Some of them are listed below. First, let us notice that Assumptions **A6** and **A11** are satisfied as soon as for $1 \leq i \leq k$,

$$\begin{aligned} \sum_{t \geq 1} \int \sup_{\theta \in \Theta} |h_{i,y}(\lambda_{i,t}(\theta)) - h_{i,y}(\bar{\lambda}_{i,t}(\theta))| P_i(dy|\lambda_{i,0}(\theta_0)) &< \infty \text{ a.s.} \\ \sum_{t \geq 1} \int \sup_{\theta \in \Theta} \|\nabla(h_{i,y} \circ \lambda_{i,t})(\theta) - \nabla(h_{i,y} \circ \bar{\lambda}_{i,t})(\theta)\| P_i(dy|\lambda_{i,t}(\theta_0)) &< \infty \text{ a.s.} \end{aligned} \quad (3.13)$$

For simplicity of notations, for any function f defined on Θ and taking values in \mathbb{R}^d or in a space of real matrices, we denote by $|f|_\infty$ its supremum norm.

1. For the binary logistic model, we have $h_{i,y}(s) = y \log(1+e^{-s}) - (1-y) \log(1+e^s)$. A close inspection of the proof of Proposition 14 in 3.6 shows that the needed assumptions are satisfied as soon as

$$\mathbb{E} \left[|\lambda_{i,0}|_\infty + |\nabla \lambda_{i,0}|_\infty^2 + |\nabla^{(2)} \lambda_{i,0}|_\infty \right] < \infty. \quad (3.14)$$

See in particular the equations in the proof of Proposition 14 combined with Lemma 13 in section 3.6.

2. For the log-linear Poisson, we have $h_{i,y}(s) = e^s - sy$. Similarly to the previous point, one can see that the required condition is

$$\mathbb{E} \left[\exp(|\lambda_{i,0}|_\infty) \left(1 + |\lambda_{i,0}|_\infty + |\nabla \lambda_{i,0}|_\infty^2 + |\nabla^{(2)} \lambda_{i,0}|_\infty \right) \right] < \infty. \quad (3.15)$$

3. For an ARMA(1, 1) part, we have $h_{i,y}(s) = (y-s)^2$. When the ARMA noise is square

integrable, it is not difficult to check that the required conditions are satisfied as soon as the analogue of (3.14) holds true.

4. For a log-GARCH component, with a noise ϵ satisfying $\mathbb{E}(\epsilon) = 0$, $\mathbb{E}\epsilon^2 = 1$ and $\mathbb{E}|\log \epsilon^2| < \infty$, we have $h_{i,y}(s) = y^2 e^{-s} + s$. A careful inspection of conditions **A5-A6-A9-A10** and (3.13) combined with Lemma 12 shows that a sufficient condition is given by

$$\mathbb{E} \left[\exp(2|\lambda_{i,0}|_\infty) \left(1 + |\nabla \lambda_{i,0}|_\infty^2 + \left| \nabla^{(2)} \lambda_{i,0} \right|_\infty \right) \right] < \infty. \quad (3.16)$$

Let us point out that for univariate log-GARCH models, (3.16) is not sharp. Using more tedious arguments, [61] give much less restrictive conditions for studying the Gaussian QMLE. Using the same kind of conditions in our context would require to make more technical assumptions than that of Theorem 11 and Theorem 12. We prefer to avoid such additional technicalities in our general context.

Typically, the most restrictive moment conditions among the different coordinates will impose the moment condition for the multivariate model. Let us give an example of a multivariate mixed time series of the form (3.5) with ℓ_1 binary logistic components, ℓ_2 log-linear Poisson component and ℓ_3 log-GARCH components with $\ell_1 + \ell_2 + \ell_3 = k$. Remembering the discussion of Section 3.2.3, one can apply Corollary 1 if **L3** holds true with $c_i = 1/4$ if $1 \leq i \leq \ell_1$ and $c_i = 1$ otherwise. This gives existence of an integrable stationary solution. To check conditions (3.14), (3.15) and (3.16), a simple sufficient condition is the existence of all exponential moments for $|\lambda_{i,0}|_\infty$ and all the polynomial moments for \bar{Y}_0 . From Lemma 14 given in section 3.6, the latter condition implies the existence of all the polynomial moments for $|\nabla \lambda_{i,0}|_\infty$ and $\left| \nabla^{(2)} \lambda_{i,0} \right|_\infty$. If we assume that the log-GARCH noise component has a finite moment of any order and the covariates have an exponential moment of any order, one can proceed as in the proof of Proposition 11 or Proposition 14 and a sufficient condition for the three required moment conditions is then

$$\left| \bar{A}_0|_{vec} + |B_0|_{vec} \right|_\infty < 1,$$

where \bar{A}_0 has the same columns as A_0 except the ℓ_1 first columns which are equal to the null vector. In 3.6, Section 3.6.17, we give a more rigorous statement. Finally, note that existence of exponential moments for λ_t is not always necessary. For instance, if we combine an ARMA(1,1) dynamic with a square integrable noise with a logistic autoregression, condition **L3** is sufficient for checking the integrability conditions detailed above.

See Section 3.6.18 in 3.6.

3.3.7 Asymptotic results for inference of copula parameters

In this subsection, we consider a general parametric model for the copula density. For simplicity, we will only derive consistency results when the initialization of the latent process is ignored, i.e. we identify $\lambda_t(\theta)$ and $\bar{\lambda}_t(\theta)$. If we assume that the function g in (3.3) does not depend on its first component (in this case, we use the terminology "pure autoregressive processes"), both processes coincide and our consistency results apply. For non-pure autoregressive processes, deriving a result when the computable version of the latent process is used probably requires more tedious arguments. Note however that such a consistency result seems to be new even in the regression case (i.e. the function g only depends on the exogenous covariates) and it also gives positive results for fitting some existing models to multivariate binary or count times series ([95], [57]).

As pointed out in [63], it is hopeless to get a systematic identification of the copula parameters when the data are discrete. This is why, we will first state a result showing that one can always estimate consistently the conditional distribution $P_{R_0}(\cdot|s)$ even if identification of the parameter R is not possible. For Gaussian copulas, we next show that such identification is automatic, leading to the consistency of the MLE for the copula parameters.

For $t \in \mathbb{Z}$, we set

$$f_t(\theta, R) = \log p_R(Y_t | \lambda_t(\theta)),$$

where $p_R(\cdot|s)$, see (3.12) for an expression, denotes the density of the conditional distribution $Y_t | \lambda_t(\theta) = s$ for a copula parameter R and autoregressive parameters given by θ . Note that the conditional distribution of Y_t given $\lambda_t(\theta_0) = s$, denoted by $P_{R_0}(\cdot|s)$, is defined here by

$$P_{R_0}(A|s) = \int_A p_{R_0}(y|s) \mu(dy),$$

with μ being a product of measures with factors equal to either the Lebesgue measure or the counting measure over \mathbb{N} or $\{0, 1\}$. We make the following assumptions.

A13 The two parameters θ, R are contained in some compact sets denoted respectively by Θ, Γ .

A14 The mapping $(\theta, R) \mapsto f_1(\theta, R)$ is continuous over $\Theta \times \Gamma$ and we have

$$\mathbb{E} \left(\sup_{(\theta, R) \in \Theta \times \Gamma} |f_1(\theta, R)| \right) < \infty$$

A15 For any λ , the mapping $R \mapsto \int f \log(p_R(y|\lambda)) p_{R_0}(y|\lambda) \mu(dy)$ is continuous over Γ .

Finally let

$$\mathcal{I}_0 = \left\{ R \in \Gamma : \bar{f}(\theta_0, R) = \bar{f}(\theta_0, R_0) \right\},$$

$\hat{\theta}$ a strongly consistent estimator of θ_0 and

$$\hat{R} = \arg \max_{R \in \Gamma} \frac{1}{n} \sum_{t=1}^n f_t(\hat{\theta}, R).$$

In what follows, we denote by d_{TV} the total variation distance, i.e. for two probability measures ν and ν' defined on the same measurable space (F, \mathcal{F}) , $d_{TV}(\nu, \nu') = \sup_{A \in \mathcal{F}} |\nu(A) - \nu'(A)|$. Note that, in the case of existence of a density with respect to the same reference measure μ , i.e. $\nu = f \cdot \mu$ and $\nu' = f' \cdot \mu$, we have the alternative expression

$$d_{TV}(\nu, \nu') = \frac{1}{2} \int |f - f'| d\mu.$$

Proposition 15 *Suppose that Assumptions **A13-A16** hold true with $\hat{\theta}$ a strongly consistent estimator of θ_0 . We then have $\lim_{n \rightarrow \infty} d(\hat{R}, \mathcal{I}_0) = 0$. Moreover, there exists a Borel set Λ such that $\mathbb{P}(\lambda_0(\theta_0) \in \Lambda) = 1$ and for any $s \in \Lambda$,*

$$d_{TV}(P_{\hat{R}}(\cdot|s), P_{R_0}(\cdot|s)) \rightarrow 0 \text{ a.s.}$$

Notes

1. Existence of a strongly consistent estimator of θ_0 is of course guaranteed from Assumptions **A1-A7**.
2. For discrete components, identifiability of the parameter is not guaranteed and the copula is not unique in general. This problem is already known in the literature of copula for multivariate discrete distributions. See [63]. However, as explained in [63], Section 5, modeling discrete or mixed data with copula is still interesting because for many models, R can still be interpreted as a dependence parameter. Indeed, the

cdf of the conditional distribution of $Y_t|\mathcal{F}_{t-1}$ is given by

$$\mathbb{P}(Y_t \leq y|\mathcal{F}_{t-1}) = C_{R_0} \left(F_{1,\lambda_{1,t}(\theta_0)}(y), \dots, F_{k,\lambda_{k,t}(\theta_0)}(y) \right).$$

When R_0 is one-dimensional, we often have $R_0 \leq R \Rightarrow C_{R_0}(u) \leq C_R(u)$ which leads to a lower quadrant stochastic ordering for the multivariate conditional distribution. Moreover, remembering Hoeffding's covariance equality $\text{Cov}(W, Z) = \int \int [F_{W,Z}(w, z) - F_W(w)F_Z(z)] dw dz$ where $F_{W,Z}, F_W, F_Z$ denote the cdf of $(W, Z), W, Z$ and W, Z are some random variables, the stochastic ordering property entails that the conditional covariances will increase with R . For the Gaussian copula discussed below, we have $C_R = \Phi_R$, where Φ_R is the cdf of the Gaussian distribution with correlation matrix R and Slepian's lemma ensures that the multivariate cdf is increasing with respect to the entries of the correlation matrix. As a consequence, conditional covariances are increasing with respect to the copula parameters. Moreover, using again Hoeffding's formula, one can note that the sign of $R_{i,j}$ equals to the sign of the $\text{Cov}(Y_{i,t}, Y_{j,t}|\mathcal{F}_{t-1})$. Additionally, in a dynamic setting, Proposition (15) shows that one can consistently estimate the conditional distribution of the model, which is unique. For instance, one can get an estimation of the conditional covariances using a Monte-Carlo approach. To this end, one can sample using the representation (3.3), replacing the unknown parameters by their estimators.

3. As for standard conditional likelihood estimators, Assumptions **A14-A15** have to be checked model by model. In our setup, those conditions will crucially depend on the behavior of the univariate marginal cdf F_{i,s_i} near their left/right endpoints and of the shape of the copula density. It seems then difficult to exhibit more explicit sufficient conditions. In what follows, we carefully check these conditions for Gaussian copula. However in 3.6, Section 3.6.16, we show that these properties are also valid for some Archimedean copulas such as Clayton copula. We also think that our proofs, which consists in deriving lower and upper bounds for the multiple integrals involved in the likelihood expression, already gives some clues to check these assumptions for other parametric models.

We next study the case of Gaussian copula, an important parametric class which is often popular for modeling the joint dependence of continuous or discrete data. See for instance [96, 98, 57]. In this case, the parameters of the copula can be always identified, even if all the coordinates of the multivariate times series are binary. However, it is difficult to find

in the literature a mathematical study of consistency properties for the estimator of the correlation matrix associated to a Gaussian copula. We will provide directly such a result for the multivariate time series models considered in the present paper. Gaussian copula are defined by

$$c_R(u_1, \dots, u_k) = \frac{1}{\sqrt{\det(R)}} \exp\left(-\frac{1}{2}\Phi^{-1}(u)'(R^{-1} - I)\Phi^{-1}(u)\right),$$

where R is a correlation matrix and $\Phi^{-1}(u) = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_k))$ with Φ^{-1} being the quantile function of the standard Gaussian distribution.

We will need the following assumptions.

G1 The discrete components of the multivariate time series are either binary $\{0, 1\}$ or fully supported on \mathbb{N} . In the latter case, we assume that for any $s_i \in F_i$, $p_i(\cdot|s_i) > 0$ and

$$\sum_{y \in \mathbb{N}} \log(1 - F_{i,s_i}(y)) p_i(y|s_i) > -\infty, \quad \sum_{y \in \mathbb{N}} \log(p_i(y|s_i)) p_i(y|s_i) > -\infty.$$

G2 When $1 \leq i \leq \ell$ (continuous components), we assume that

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log F_{i,\lambda_{i,0}(\theta)}(Y_{i,0}) \right\} + \sup_{\theta \in \Theta} \left\{ -\log \left(1 - F_{i,\lambda_{i,0}(\theta)}(Y_{i,0}) \right) \right\} \right] < \infty.$$

G3 When $\ell + 1 \leq i \leq k$ (discrete components), we have

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log p_i(0|\lambda_{i,0}(\theta)) \right\} + \sup_{\theta \in \Theta} \left\{ -\log (1 - p_i(0|\lambda_{i,0}(\theta))) \right\} \right] < \infty, \quad (3.17)$$

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log p_i(Y_{i,0}|\lambda_{i,0}(\theta)) \right\} \right] < \infty \quad (3.18)$$

and for count marginal time series,

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log \left(1 - F_{i,\lambda_{i,0}(\theta)}(Y_{i,0}) \right) \right\} \right] < \infty. \quad (3.19)$$

Theorem 13 *Suppose that Assumptions **G1-G3** hold true with $\Theta \times \Gamma$ compact and that $\hat{\theta}$ is a strongly consistent estimator of θ_0 . Assume furthermore that **A5** holds true with $h_{i,y}(s_i) = -\log p_i(y|s_i)$ for $1 \leq i \leq \ell$. We then have strong consistency of the two-step*

estimator, i.e. $\lim_{n \rightarrow \infty} \hat{R} = R_0$ a.s.

Note. The additional assumptions **G1-G3** are not so restrictive. For instance, Poisson or logistic autoregressive models and GARCH or ARMA models will satisfy these conditions in general (up to some additional regularity conditions on the noise density). Below, we carefully check these assumptions for the GAIN and BIN model.

Consistency for the GAIN model

Here, the correlation matrix writes as $R_0 = \begin{pmatrix} 1 & r_0 \\ r_0 & 1 \end{pmatrix}$ and r_0 is the single parameter to estimate. For simplicity, we give below a consistency result when the noise density f_ϵ of the GARCH component is such that one can find positive constants c_1, c_2, c_3 such that

$$f_\epsilon(z) \geq c_1 \exp(-c_2 z^2), \quad |z| \geq c_3. \quad (3.20)$$

Note that condition (3.20) is satisfied for the Gaussian distribution as well as Student distributions.

Corollary 2 *Suppose that all the assumptions of Proposition 12 are valid and that the noise density of the GARCH component satisfies (3.20). We then have*

$$\lim_{n \rightarrow \infty} \hat{r} = r_0 \quad a.s.$$

Consistency for the BIP model

Consistency also holds for the BIP models when the assumptions ensuring consistency of the pseudo-maximum likelihood estimator are satisfied.

Corollary 3 *Suppose that all the assumptions of Proposition 14 are satisfied. We then have*

$$\lim_{n \rightarrow \infty} \hat{r} = r_0 \quad a.s.$$

3.4 Numerical experiments and real data applications

In this section, we discuss the implementation of our inference procedure for the GAIN and the BIP model. We only implement these models for a Gaussian copula. There also

exist many other interesting families of copula (Clayton, Gumbell...) and we defer the reader to [13] for an interesting survey about copulas properties and their use in finance. Throughout this section, the density of the GARCH noise is always assumed to be a standard Gaussian.

The main difficulty for fitting our models is the approximation of the likelihood function for estimating the correlation matrix R_0 . Pseudo-likelihood estimation of autoregressive parameters is straightforward. Note that the equation-by-equation estimation can be obtained from the standard software packages since it is equivalent to fit a standard time series model to one coordinate with past values of the other coordinates as covariates. Note that when $k \geq 3$, several iterated integrals have to be computed for approximating (3.12). This leads to an important computational problem. To solve this issue, one can use the importance sampling strategy considered in [98] which has been already applied to likelihood inference for Gaussian copula when the data are of mixed type. It is also possible to use pairwise composite likelihood methods as discussed in [98], which avoids the numerical computations of these integrals. However, in this case, the estimator of R_0 differs from the conditional likelihood estimator obtained from (3.12). In the present paper, we will not investigate such computational issues and their corresponding convergence properties. For our numerical experiments, we only consider the GAIN and the BIP model and the number of coordinates k is equal to 2. In this case, there is no gain in applying the aforementioned methods. When $k = 2$, let us first give a simpler formula for the likelihood function (3.12). The correlation matrix R_0 only involves one coefficient $r_0 \in (-1, 1)$ and using the properties of conditional distributions for Gaussian vectors, one can show that an estimation of parameter r_0 can be obtained by minimizing

$$r \mapsto \sum_{t=1}^n \log \left(\left\{ \Phi \left(\frac{\Phi^{-1}(Z_{i,t}) - r\Phi^{-1}(Z_{j,t})}{\sqrt{1-r^2}} \right) - \Phi \left(\frac{\Phi^{-1}(Z_{i,t}^-) - r\Phi^{-1}(Z_{j,t})}{\sqrt{1-r^2}} \right) \right\} \right)$$

for the GAIN model and

$$r \mapsto \sum_{t=1}^n \log \left(\int_0^1 \left\{ \Phi \left(\frac{\Phi^{-1}(Z_{i,t}) - r\epsilon_{j,t}(u)}{\sqrt{1-r^2}} \right) - \Phi \left(\frac{\Phi^{-1}(Z_{i,t}^-) - r\epsilon_{j,t}(u)}{\sqrt{1-r^2}} \right) \right\} du \right)$$

for the BIP model. Here,

$$Z_{i,t} = F_{i,\hat{\lambda}_{i,t}}(Y_{i,t}), \quad Z_{i,t}^- = F_{i,\hat{\lambda}_{i,t}}(Y_{i,t} - 1) \text{ and } \epsilon_{i,t}(u_i) = \Phi^{-1} \left(Z_{i,t} - u_i(Z_{i,t} - Z_{i,t}^-) \right),$$

where $\hat{\lambda}_t = \bar{\lambda}_t(\hat{\theta})$.

Note that for the GAIN model, the formula is explicit in term of the Gaussian cdf Φ whereas the formula for the BIP model involves the computation of one integral. Approximation for this integral can be obtained from Monte Carlo methods. In our simulations, we simply simulate a sample of size $N = 10^4$ of uniformly distributed random variables and approximate this integral by an empirical counterpart.

One can also compute standard errors for our estimators. For the autoregressive parameters, the asymptotic distribution of pseudo-likelihood estimators can be used. For the copula parameter, we did not investigate the asymptotic distribution of the likelihood estimator. However, one can simply use a parametric bootstrap : we simulate B paths of size n of the model using the estimated parameters $\hat{\theta}$ and \hat{r} and we compute the standard error from the sample $\hat{r}^{*,b}$ for $b = 1, \dots, B$. A theoretical justification of such a procedure is beyond the scope of this paper.

3.4.1 Numerical experiments

We fitted the GAIN and BIP models to simulated data. For the BIP model, we used an additional covariate process with $m = 1$ and defined by an AR(1) process, $X_t = -0.15 \times X_{t-1} + \xi_t$ where $(\xi_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. standard Gaussian random variables. This sequence is assumed to be independent of the sequence $(U_t)_{t \in \mathbb{Z}}$ used for the copula. For both models, Tables 3 to 6 in 3.6 give averages and standard deviations of $M = 500$ estimators and for two sample sizes, $n = 500$ and $n = 1000$. One can note that both estimation of autoregressive parameters and of the copula parameter work reasonably well whatever the values of r_0 which is allowed to vary from -0.9 to 0.9 . We found that $n = 500$ is a reasonable sample size to get an accurate estimation of all the parameters.

3.4.2 An application to sleep data

We use the data set already studied in [53], with sleep state measurements of a newborn infant together with his heart rate $Y_{2,t}$ (taking integer values) and temperature X_t sampled every 30 seconds. The sample size is $n = 1024$ and the sleep states are classified as : (1) quiet sleep, (2) indeterminate sleep, (3) active sleep, (4) awake. To define a binary time series, we aggregate States (1), (2) and (3) and we then set $Y_{1,t} = 1$ when the infant is awake and 0 if it is not. A BIP model is fitted to the time series $(Y_t)_{1 \leq t \leq n}$. It is quite intuitive to suspect a dependence between the heart rate and the sleep state and our aim

Log-Poisson (I)					Log-Poisson (II)				
d_1	$A(1, 1)$	$A(1, 2)$	$B(1, 1)$	$\Gamma(1, 1)$	d_1	$A(1, 1)$	$A(1, 2)$	$B(1, 1)$	$\Gamma(1, 1)$
-10.8984 (33.4872)	0.6929 (0.0428)	0.0068 (0.8275)	0.0289 (0.0907)	3.3907 (9.3608)	1.0859 (0.1611)	0.7437 (0.1380)	0.0108 (1.1646)	0.0331 (0.1328)	
Logit-Binary (I)					Logit-Binary (II)				
d_2	$A(2, 1)$	$A(2, 2)$	$B(2, 2)$	$\Gamma(2, 1)$	d_2	$A(2, 1)$	$A(2, 2)$	$B(2, 2)$	$\Gamma(2, 1)$
582.3259 (503.0512)	-1.5590 (3.0956)	13.9541 (1.4844)	-0.5054 (0.1090)	-161.0862 (138.3401)	7.1203 (16.4880)	-2.9145 (3.3355)	13.0971 (1.3732)	-0.4675 (0.1171)	
r (I)					r (II)				
estimate :	0.3337	sd :	0.1040		estimate :	0.2749	sd :	0.1058	
(I) : AIC = -535318.3					(II) : AIC = -535309.7				

TABLE 3.1 – Estimation of the parameters of the BIP model for sleep data. Standard errors are given in parenthesis.

is to analyze such a joint dynamic.

Results are displayed in Table 3.1. We consider two BIP models. Model (I) is unrestricted while in Model (II), we do not use the temperature as an exogenous variable. Using t -tests, one can note that the lag value of the temperature seems not to have a significant contribution to the dynamic, though the AIC is smaller when this covariate is incorporated in the model. Both lag values of the heart rate (the sleep state respectively) seem to have a negligible influence to the present value of the sleep state (the heart rate respectively). On the other hand, we get a positive coefficient \hat{r} for the copula and we then observe a positive association between the two time series at time t . Being awake is more likely associated with larger heart rates at the same time which seems to be quite logical. Note that such findings are compatible with that of the univariate modeling of [53] (see p. 372 of that paper) with a sleep state at time t which seems to depend on the current heart rate but less on its lag value.

3.4.3 An application to high-frequency transactions in finance

The data are downloaded from <http://www.nasdaqomxnordic.com> and represent the real-time transactions on Boliden, a metal exploring, extracting and processing firm. The count component, denoted by $Y_{2,t}$, is the number of transactions of this stock occurring in a time interval of two successive minutes. The continuous coordinate is the log-return $Y_{1,t} = \log(P_t) - \log(P_{t-1})$ of the transaction average price P_t . The transaction average price P_t is simply given by $\sum_{j=1}^m W_{j,t} P_{j,t} / \sum_{j=1}^m W_{j,t}$ where $W_{j,t}$ is the number of transactions at price $P_{j,t}$ occurring during this two minutes time interval. See also [54] who used a similar weighted average price. We model the dynamic of $(Y_t)_{1 \leq t \leq 467}$ with a GAIN process. The

GARCH			
d_1	$A(1, 1)$	$A(1, 2)$	$B(1, 1)$
0.0012 (0.0353)	0.7569 (0.3829)	0.00005 (0.0005)	0.1226 (0.0017)
INGARCH			
d_2	$A(2, 1)$	$A(2, 2)$	$B(2, 2)$
3.1988 (1.2767)	34.6621 (7.3539)	0.1236 (0.0406)	0.7719 (0.0702)
r			
estimate :	-0.011	sd :	0.0049
(I) : AIC = -40449.32			

TABLE 3.2 – Estimation of the parameters of the GAIN model for financial data. Standard errors are given in parenthesis.

data are collected for a two days time period between March 29th and March 30th , 2021. The result are given Table 3.2. For the autoregressive parameters, one can suspect that the lag value of the number of transactions has no effect on the volatility and then on the next log-return. We then test the hypothesis $H_0 : A(1, 2) = 0$ versus $H_1 : A(1, 2) > 0$. Since the parameter is on the boundary of the parameter set under the null hypothesis, the QMLE has not an asymptotic Gaussian distribution. We then used the corrected test given in [59] which consists in rejecting H_0 at level α if $\hat{A}(1, 2)^2/\hat{v}$ is larger than the quantile of order $1 - 2\alpha$ (instead of $1 - \alpha$ when the parameter is not on the boundary) of a χ^2 distribution with 1 degree of freedom. Here \hat{v} is simply an estimation of the asymptotic variance of $\hat{A}(1, 2)$ given in Proposition 13. We do not reject H_0 at level $\alpha = 5\%$. Moreover, \hat{r} is quite small and negative. Unfortunately, we did not derive the asymptotic distribution of this estimator to get a standard significance test. If the asymptotic distribution was Gaussian at the usual \sqrt{n} convergence rate, we would reject the hypothesis $H_0 : r = 0$, but further investigation is needed to make a rigorous conclusion. One can then conclude that past values of the log-returns seem to have an influence on the number of transactions at time t but not the inverse. Moreover a negative but very small association between the two time series at time t is possible.

3.5 Conclusion and perspectives

In this paper, we discussed a general modeling for multivariate time series with mixed data. Our approach is quite flexible since many popular univariate time series models

can be used to construct our multivariate models. Our contribution is also the first one dedicated to such a problem in the case of time series and we provided many asymptotic results for statistical inference in our models. Estimation of dynamic parameters is quite simple, using a pseudo-likelihood estimation method as if the coordinates at time t were independent conditionally on \mathcal{F}_{t-1} . A finer analysis of the simultaneous interactions between the components of the time series can be obtained from a fitting of a copula model and we also derived some asymptotic results, in particular for estimating the correlation matrix of Gaussian copulas. The latter result seems to be new, even in the regression case for i.i.d. observations and it also gives a basis for some existing models for multivariate count or binary time series.

Let us mention that our approach allows to include exogenous regressors, at least if the corresponding times series are stationary. This restriction is mostly interesting for technical reasons, as stationarity allows to use a battery of limit theorems. In practice, covariates might exhibit a non-stationary behavior. In this sense, extension of our results to covariates exhibiting a seasonal behavior could be possible, as stationary properties can be recovered blockwise. Including locally stationary covariates, as well as time-varying coefficients, could be also an interesting extension. Since we use contraction techniques to define our models, we believe that recent contributions such as [6, 25, 127] can be adapted to our framework. Finally, including covariates with a unit root, as in [70] for GARCH models, could be also interesting, though the asymptotic results will be quite different and more difficult to get.

A possible extension of our work concerns categorical time series with $d \geq 3$ modalities. Though the inverse of the cdf for a discrete distribution can also be used, the contraction condition needed for getting existence of a stationary solution could depend on the coding of the modalities. [106] used a coding with vectors of the canonical basis of \mathbb{R}^{d-1} for getting stability conditions but this coding already leads to a multivariate latent process and it cannot be covered by our results. However, adapting our results to cover such an extension could be possible. Extension of our results to autoregressive latent processes depending on $p \geq 2$ lag values could be also possible with appropriate technical modifications. Note also that some of our results such as Theorems 10, 15, 12 and 13 are valid for non-linear function g in (3.3). Checking the required assumptions for non-linear models such as some smooth threshold models available in the literature could be also interesting. Another important problem concerns model fitting. In particular, one could conduct a residual analysis and study Hausman specification tests as discussed in [98]. This needs

further investigation.

3.6 Supplementary material

3.6.1 Proof of Theorem 10

Define the mapping

$$f_t : s \mapsto g \left(s, F_s^{-1}(U_{t-1}), X_{t-1} \right).$$

From **A1-A3**, the assumptions of Theorem 4 in [32] are satisfied with $o = p = 1$ and $\zeta_t = (X_{t-1}, U_{t-1})$. In particular, there exists a unique stationary, integrable and $(\mathcal{F}_{t-1})_{t \in \mathbb{Z}}$ -adapted process $(\lambda_t)_{t \in \mathbb{Z}}$ such that $\lambda_t = f_t(\lambda_{t-1})$. Moreover, Theorem 2 in [32] guarantees the representation $\lambda_t = H \left((U_{t-j}, X_{t-j})_{j \geq 1} \right)$ for a suitable measurable function H defined on an infinite Cartesian product (Bernoulli shift representation with respect to the process $((X_t, U_t))_{t \in \mathbb{Z}}$). Setting $Y_t = F_{\lambda_t}^{-1}(U_t)$, we then get a stationary and ergodic process $((Y_t, \lambda_t, X_t))_{t \in \mathbb{Z}}$, as this process also has a Bernoulli representation with respect to the stationary and ergodic process $((X_t, U_t))_{t \in \mathbb{Z}}$. The uniqueness property easily follows. \square

3.6.2 Proof of Proposition 9

For a positive real s and an integer $r \geq 1$,

1. $\mathbb{E}^{1/r} \left[F_{1,s}^{-1}(U_{1,t})^{2r} \right] = \mathbb{E}^{1/r} \left[\epsilon^{2r} \right] s$
2. $\mathbb{E}^{1/r} \left[F_{2,s}^{-1}(U_{2,t})^r \right] \leq (1 + \delta)s + b_{r,\delta}$

for $\delta > 0$ arbitrarily small and $b_{r,\delta}$ a positive constant which depends on r and δ (see Lemma 9 below). We will apply Lemma 11. Since

$$\|g(s, F_s^{-1}(U_t), X_t)\|_{r,t-1,vec} \preceq (A\mathcal{H} + B)s + \Gamma \|X_t\|_{r,t-1,vec} + \bar{b}_{r,\delta}$$

where $\bar{b}_{r,\delta} = (0, b_{r,\delta})$ and $\mathcal{H} = \text{diag}(\mathbb{E}^{1/r}[\epsilon^{2r}], 1 + \delta)$, the result follows from Lemma 11. Indeed, if δ is small enough, we have $\rho(A\mathcal{H} + B) < 1$. \square

3.6.3 Proof of Proposition 8

1. The first point is a consequence of Corollary 1 with $g_1(y) = y^2$ and $g_2(y) = y$. Assumption **L1** is easy to check and from the discussion given in Section 2.3, Assump-

tion **L2** is also satisfied for both coordinates with $c_1 = c_2 = 1$. Straightforwardly, $\mathbb{E}(|\bar{Y}_0|_1) < \infty$ since $\mathbb{E}(|\bar{Y}_0|_1) = \mathbb{E}(|\lambda_0|_1)$.

2. Under the proposed assumptions, we have

$$\mathbb{E}\lambda_t = d + B\mathbb{E}\lambda_{t-1} + A\mathbb{E}\bar{Y}_{t-1} + \Gamma\mathbb{E}X_{t-1} = d + \Gamma\mathbb{E}X_{t-1} + (A + B)\mathbb{E}\lambda_{t-1}.$$

By stationarity, we get $m := \mathbb{E}\lambda_0 = c + (A + B)m$ where $c = d + \Gamma\mathbb{E}X_0$. Iterating the previous equality, we get $m = \sum_{i=0}^K (A + B)^i c$. By positivity of the components of c and non negativity of the matrices A and B , we deduce that the series $\sum_{i=0}^K (A + B)^i$ is converging term by term and then that $\lim_{i \rightarrow \infty} (A + B)^i = 0$. This automatically imply that $\rho(A + B) < 1$. \square

3.6.4 Proof of Proposition 10

Setting $g_1(y) = y$ and $g_2(y) = \log(1 + y)$, the discussion given in Section 2.3 shows that **L2** is satisfied with $c_1 = 1/4$ and $c_2 = 1$. Assumption **L1** is straightforward to show. The result then follows from an application of Corollary 1. The integrability condition follows from the fact that $(Y_{1,t})_{t \in \mathbb{Z}}$ is bounded and the inequality $\mathbb{E} \log(1 + Y_{2,t}) \leq \mathbb{E} |\lambda_{2,t}| < \infty$ which follows from the discussion of Section 2.3. \square

3.6.5 Proof of Proposition 11

We will apply the result of Lemma 11 with the $r = 1$ and the function $\phi(s) = \exp(\kappa|s|_{vec})$ where the exponential function is applied componentwise. Setting $\mu_i = \sum_{j=1}^2 (|\bar{A}(i, j)| + |B(i, j)|) < 1$, $C_i = |A(i, 1)|$ for $i = 1, 2$ and denoting by \odot the Hadamard product, we have

$$|g(s, F_s^{-1}(U_t), X_t)|_{vec} \preceq (1 - \mu) \odot \frac{|d + \Gamma X_t|_{vec} + C}{1 - \mu} + |\bar{A}|_{vec} \begin{pmatrix} 0 \\ g_2 \circ F_{2,s_2}^{-1}(U_{2,t}) \end{pmatrix} + |B|_{vec} |s|_{vec}.$$

Using convexity of the exponential function, we deduce that

$$\phi \left(g(s, F_s^{-1}(U_t), X_t) \right) \preceq (1 - \mu) \odot \phi \left(\frac{d + \Gamma X_t + C}{1 - \mu} \right) + |\bar{A}|_{vec} \begin{pmatrix} 1 \\ \exp(\kappa g_2 \circ F_{2,s_2}^{-1}(U_{2,t})) \end{pmatrix} + |B|_{vec} \phi(s). \quad (3.21)$$

Note that for any $\delta > 0$, there exists $d_{\kappa,\delta} > 0$ such that

$$\mathbb{E} \exp(\kappa g_2 \circ F_{2,s_2}^{-1}(U_{2,t})) = \mathbb{E} \left(1 + F_{2,s_2}^{-1}(U_{2,t}) \right)^\kappa \leq (1 + \delta) \exp(\kappa s_2) + d_{\kappa,\delta}.$$

The previous bound can be obtained from Lemma 9, using the convexity of power functions. In what follows, we denote by c_{t-1} the sum between the vector $|\bar{A}|_{vec} \begin{pmatrix} 0 \\ d_{\kappa,\delta} \end{pmatrix}$ and conditional expectation of $(1 - \mu) \odot \phi \left(\frac{d + \Gamma X_t + C}{1 - \mu} \right)$ with respect to \mathcal{F}_{t-1} . Taking the conditional expectation with respect to \mathcal{F}_{t-1} in (3.21), we deduce that

$$\|\phi(g(s, F_s^{-1}(U_t), X_t))\|_{t-1,1,vec} \preceq c_{t-1} + (|\bar{A}|_{vec} \text{diag}(1, 1 + \delta) + |B|_{vec}) \phi(s).$$

Taking δ small enough, our assumptions guarantee that the spectral radius of $|\bar{A}|_{vec} \text{diag}(1, 1 + \delta) + |B|_{vec}$ is less than 1 and Lemma 11 leads to $\mathbb{E}|\phi(\lambda_0)|_1 < \infty$. Finally, we use the bound

$$\mathbb{E}\phi(g_2 \circ Y_{2,0}) = \mathbb{E}(1 + Y_{2,0})^\kappa \leq (1 + \delta)\mathbb{E}\phi(\lambda_{2,0}) + d_{\kappa,\delta} < \infty. \square$$

3.6.6 Proof of Lemma 4

Suppose that $\lambda_t(\theta) = \lambda_t(\theta_0)$ a.s. Since $\forall B \in \Theta, \rho(B) < 1$, we get

$$\begin{aligned} \sum_{j=1}^{\infty} [B^{j-1}A - B_0^{j-1}A_0] \bar{Y}_{t-j} &= \sum_{j=1}^{\infty} [B^{j-1}\Gamma - B_0^{j-1}\Gamma_0] X_{t-j} \\ &+ \sum_{j=1}^{\infty} [B^{j-1}d - B_0^{j-1}d_0]. \end{aligned} \quad (3.22)$$

and consequently there exist a set of matrices $\Psi_j, \Phi_j, j \geq 1$ and a vector c of \mathbb{R}^p such that

$$\sum_{j \geq 1} \Psi_j \bar{Y}_{t-j} = c + \sum_{j \geq 1} \Phi_j X_{t-j} \text{ a.s.} \quad (3.23)$$

From **I1**, $\Phi_1 = 0$. Indeed, under our assumptions, if the previous equality is valid, the random vector $\Phi_1 X_{t-1}$ is measurable with respect to the sigma-field $\mathcal{F}_{t-2} \vee \sigma(U_{t-1})$.

Next, suppose that $\Psi_1 \neq 0$. There then exists a vector $v \in \mathbb{R}^p \setminus \{0\}$ such that $v' \bar{Y}_{t-1} = G_{t-2}$, where G_{t-2} is a random variable \mathcal{F}_{t-2} -measurable. For $1 \leq i \leq p$, set $H_i =$

$g_i \circ F_{i,\lambda_{i,t-1}}^{-1}(\theta_0)$. Note that

$$1 = \mathbb{P}\left(v'\bar{Y}_{t-1} = G_{t-2} | \mathcal{F}_{t-2}\right) = \int_{[0,1]^p} c_{R_0}(u_1, \dots, u_p) \mathbb{1}_{\sum_{i=1}^p v_i H_i(u_i) = G_{t-2}} du_1 \cdots du_p.$$

Since c_{R_0} is positive, we deduce that $\lambda_p(\{\sum_{i=1}^p v_i H_i(\cdot) = G_{t-2}\}) = 1$ a.s. where λ_p denotes the Lebesgue measure on $[0, 1]^p$. From **I2**, we automatically have $v = 0$ because otherwise, the value of one of the H_i 's is determined by the values of the others functions H_j . We then get $\Psi_1 = 0$. Recursively, we obtain $\Phi_j = \Psi_j = 0, \forall j \geq 2$ and finally $c = 0$.

Then, the equation (3.22) yields $B^j A = B_0^j A_0$ for any $j \in \mathbb{N}$ and then $A = A_0$. Moreover, $B^j \Gamma = B_0^j \Gamma_0$ for all $j \in \mathbb{N}$ entails $\Gamma = \Gamma_0$. From **I3**, we get $B = B_0$ and then $d = d_0$. \square

3.6.7 Proof of Lemma 5

The proof is similar to the proof of Lemma 1. The single difference concerns the treatment of equality (3.23). If the noise process and the covariate process are independent, then the conditional distribution of Y_t given \mathcal{F}_{t-1} is also the conditional distribution of Y_t given $\sigma(X_j : j \in \mathbb{Z}) \vee \sigma(U_{t-i} : i \geq 1)$. Assume an equality of the form (3.23). From **I2**, we obtain recursively $\Psi_j = 0$ for $j \geq 1$, using the same arguments as in the proof of Lemma 1. Hence, we get $\sum_{j \geq 1} \Phi_j X_{t-j} + c = 0$ a.s. From **I1'**, we get $\Phi_j = 0$ for $j \geq 1$ and then $c = 0$. The rest of the proof is identical to that of Lemma 1. \square

3.6.8 Proof of Lemma 6

For $1 \leq i, j \leq k, 1 \leq i' \leq k$ and $1 \leq j' \leq m$, let us denote by $E_{i,j}$ and $G_{i',j'}$ the matrices of size $k \times k$ and $k \times m$ respectively and with elements equal to 1 for the couple of indices (i, j) or (i', j') and 0 elsewhere. We also denote by $J_{\lambda_t}(\theta_0)$ the Jacobian matrix of λ_t at point θ_0 . Assume that there exists a vector x such that $J_{\lambda_t}(\theta_0)x = 0$ a.s. We have

$$J_{\lambda_t}(\theta_0) = [1|\bar{Y}_{t-1}|X_{t-1}|\lambda_{t-1}(\theta_0)] + B_0 J_{\lambda_{t-1}}(\theta_0),$$

where $[1|\bar{Y}_{t-1}|X_{t-1}|\lambda_{t-1}(\theta_0)]$ is a concatenated matrix with block elements

$$I_k, E(1, 1)\bar{Y}_{t-1}, \dots, E(k, k)\bar{Y}_{t-1}, G(1, 1)X_{t-1}, \dots, G(k, m)X_{t-1}, E(1, 1)\lambda_{t-1}(\theta_0), \dots, E(k, k)\lambda_{t-1}(\theta_0),$$

with I_k the identity matrix of size k . By stationarity, we also have $J_{\lambda_{t-1}}(\theta_0)x = 0$ a.s. and we then obtain $[1|\bar{Y}_{t-1}|X_{t-1}|\lambda_{t-1}(\theta_0)]x = 0$. We then deduce the existence of a vector ν in \mathbb{R}^k , two square matrices α and β of size k and a matrix γ of size $k \times m$ such that

$$\nu + \alpha\bar{Y}_{t-1} + \beta\lambda_{t-1}(\theta_0) + \gamma X_{t-1} = 0 \text{ a.s.}$$

Using the same kind of arguments as in the proof of Lemma 1 (see the implication of equality (3.23)), we get the equalities $\nu + \beta(I - B_0)^{-1}d_0 = 0$, $\alpha = 0$, $\gamma = 0$ and $\beta B_0^j[A_0, \Gamma_0] = 0$ for any $j \geq 0$. From **A4**, we get $\beta = 0$ and then $\nu = 0$. Since $x = (\nu', \text{vec}(\alpha)', \text{vec}(\gamma)', \text{vec}(\beta)')$, we get $x = 0$ which means that **A12** is satisfied. \square

3.6.9 Proof of Proposition 14

We check the assumptions of Theorem 2 and Theorem 3. Let us denote $h_{1,Y_{1,t}}(\lambda_t(\theta)) = \log(1 + e^{\lambda_{1,t}(\theta)}) - Y_{1,t}\lambda_{1,t}(\theta)$, $h_{2,Y_{2,t}}(\lambda_t(\theta)) = e^{\lambda_{2,t}(\theta)} - Y_{2,t}\lambda_{2,t}(\theta)$. **A4** and **A8** are straightforward to show or result from the Kullback-Leibler divergence properties. Moreover **A7** and **A12** follows from the results given in Subsection 3.4 of the paper. We next check **A5**. We have

$$\begin{aligned} \mathbb{E} \left(\sup_{\theta \in \Theta} |h_{1,Y_{1,t}}(\lambda_{1,t}(\theta))| \right) &\leq \log(2) + 2\mathbb{E} \left(\sup_{\theta \in \Theta} |\lambda_{1,t}(\theta)| \right), \\ \mathbb{E} \left(\sup_{\theta \in \Theta} |h_{2,Y_{2,t}}(\lambda_{1,t}(\theta))| \right) &\leq \exp \left(\sup_{\theta \in \Theta} |\lambda_{2,t}(\theta)| \right) + \exp(\lambda_{2,t}(\theta_0)) \sup_{\theta \in \Theta} |\lambda_{2,t}(\theta)|. \end{aligned} \quad (3.24)$$

By recursion, note that

$$\lambda_t(\theta) = (I - B)^{-1}(d + \Gamma X_{t-1}) + \sum_{j \geq 0} B^j A \bar{Y}_{t-j-1}$$

and then

$$\sup_{\theta} |\lambda_t(\theta)|_1 \leq \sup_{\theta} |(I - B)^{-1}d|_1 + \sup_{\theta} |(I - B)^{-1}\Gamma X_{t-1}|_1 + \sup_{\theta} |A|_1 \sum_{j \geq 0} |B^j|_1 |\bar{Y}_{t-j-1}|_1.$$

From lemma 10, there exists $\tau \in (0, 1)$ and $\kappa > 0$ such that :

$$\mathbb{E}(e^{\sup_{\theta} |\lambda_t(\theta)|_1}) \leq K(1 - \tau) \sum_{j \geq 0} \tau^j \mathbb{E} \left(e^{\frac{\kappa \sup_{\theta} |A|_1}{1 - \tau} |\bar{Y}_{t-j-1}|_1} \right) = K \mathbb{E} \left(e^{\frac{\kappa \sup_{\theta} |A|_1}{1 - \tau} |\bar{Y}_0|_1} \right) \quad (3.25)$$

with $K = e^{\sup_{\theta} |(I-B)^{-1}d|_1} \mathbb{E} \left(e^{\sup_{\theta} |(I-B)^{-1}\Gamma X_0|_1} \right)$. From Proposition 4,

$$\mathbb{E} \left(e^{\frac{\kappa \sup_{\theta} |A|_1}{1-\tau} |\bar{Y}_0|_1} \right) \leq \left(1 + e^{\frac{\kappa \sup_{\theta} |A|_1}{1-\tau}} \right) \mathbb{E} \left[(1 + Y_{2,0})^{\frac{\kappa \sup_{\theta} |A|_1}{1-\tau}} \right] < \infty. \quad (3.26)$$

Altogether, $\sup_{\theta \in \Theta} |\lambda_t(\theta)|_1$ admits a finite exponential moment of any order and $Y_{2,0}$ has all polynomial moments. On other hand, $\sup_{\theta} e^{\lambda_{2,t}(\theta)} \leq e^{\sup_{\theta} \lambda_{2,t}(\theta)}$ since exponential function is increasing. Then **A5** follows.

We next check **A6**. Using the Lipschitz property of the function $h_{1,t}, h_{2,t}$ and Lemma 13, we only have to show that

$$\sum_{t \geq 1} \left(\sup_{\theta \in \Theta} |\lambda_{1,t}(\theta) - \bar{\lambda}_{1,t}(\theta)| + \exp \left(\sup_{\theta \in \Theta} |\lambda_{2,t}(\theta)| \right) \sup_{\theta \in \Theta} |\lambda_{2,t}(\theta) - \bar{\lambda}_{2,t}(\theta)| \right) < \infty. \quad (3.27)$$

Then **A6** follows from the existence of exponential moments and Lemma 13.

Next, we check **A9-A10**. To this end, it is sufficient to show integrability of the random variables

$$\sup_{\theta \in \Theta} \|\nabla \lambda_{1,0}(\theta)\|^2, \quad \sup_{\theta \in \Theta} \|\nabla^{(2)} \lambda_{1,0}(\theta)\|, \quad \exp \left(\sup_{\theta \in \Theta} |\lambda_{2,0}(\theta)| \right) \times \sup_{\theta \in \Theta} \|\nabla \lambda_{2,0}(\theta)\|^2 \quad (3.28)$$

and

$$\exp \left(\sup_{\theta \in \Theta} |\lambda_{2,0}(\theta)| \right) \times \sup_{\theta \in \Theta} \|\nabla^{(2)} \lambda_{2,0}(\theta)\|. \quad (3.29)$$

These integrability conditions follows from the existence of exponential moments and Lemma 12.

Finally, we check **A11**. This condition will be satisfied as soon as

$$\begin{aligned} \sum_{t \geq 1} d_t < \infty \text{ a.s. with } d_t \text{ being equal to one of the following quantities : } & (3.30) \\ \sup_{\theta \in \Theta} \|\nabla \lambda_{1,t}(\theta) - \nabla \bar{\lambda}_{1,t}(\theta)\|, \quad \sup_{\theta \in \Theta} |\lambda_{1,t}(\theta) - \bar{\lambda}_{1,t}(\theta)| \times \sup_{\theta \in \Theta} \|\nabla \bar{\lambda}_{1,t}(\theta)\|, \\ \exp \left(\sup_{\theta \in \Theta} |\lambda_{2,t}(\theta)| \right) \times \sup_{\theta \in \Theta} \|\nabla \lambda_{2,t}(\theta) - \nabla \bar{\lambda}_{2,t}(\theta)\|, \\ \sup_{\theta \in \Theta} \|\nabla \bar{\lambda}_{2,t}(\theta)\| \times \exp \left(\sup_{\theta \in \Theta} |\lambda_{2,t}(\theta)| \right) \times \sup_{\theta \in \Theta} |\lambda_{2,t}(\theta) - \bar{\lambda}_{2,t}(\theta)|. \end{aligned}$$

Using Lemma 13 and the integrability conditions of Lemma 12 as well as the existence of all the exponential moments for $\sup_{\theta \in \Theta} |\lambda_{2,t}(\theta)|$ already justified above, we easily get the

result. \square

3.6.10 Proof of Proposition 12

We check the assumptions of Theorem 2. First, note that Assumption **A7** follows directly from our assumptions and Lemma 1. Moreover, checking Assumption **A4** is straightforward and follows from standard arguments. We then check **A5** and define $h_t(s) = h_{1,Y_{1,t}}(s_1) + h_{1,Y_{2,t}}(s_2)$. We have

$$\begin{aligned} \sup_{\theta} |h_t(\lambda_t(\theta))| &\leq d_-^{-1} \left(Y_{1,t}^2 + \sup_{\theta} |\lambda_{1,t}(\theta)| + 1 \right) + d_-^{-1} Y_{2,t} \\ &\quad + \sup_{\theta} |\lambda_{2,t}(\theta)| + Y_{2,t} \left(\log(d_-) + \log \left(1 + \sup_{\theta} |\lambda_{2,t}(\theta)| \right) \right) \end{aligned}$$

Since $\rho(B_0 + A_0) < 1$ and $\mathbb{E}(\epsilon^2) = 1$, one can find $r < \delta$ small enough such that $\rho(B_0 + A_0 \text{diag}(\mathbb{E}^{1/(1+r)}[\epsilon^{2(1+r)}], 1)) < 1$. Then, from Proposition 2, we have for δ small enough, $\mathbb{E}[|\bar{Y}_0|^{1+\delta}] < \infty$ and Lemma 12 yields to $\mathbb{E}[(\sup_{\theta} |\lambda_0(\theta)|_1)^{1+\delta}] < \infty$. Hence $\mathbb{E}[\log^{1+1/\delta}(1 + \sup_{\theta} |\lambda_{2,t}(\theta)|)] < \infty$ and consequently $\mathbb{E}[Y_{2,t} \log(1 + \sup_{\theta} |\lambda_{2,t}(\theta)|)] < \infty$. It follows that $\mathbb{E}[\sup_{\theta} |h_t(\lambda_t(\theta))|] < \infty$. We then conclude that $\mathbb{E}(\sup_{\theta \in \Theta} |h_t(\lambda_t(\theta))|) < \infty$ and **A5** follows. Finally, we check **A6**. Using the Lipschitz property of the function h_t and Lemma 13, we have

$$\begin{aligned} |h_t(\bar{\lambda}_t(\theta)) - h_t(\lambda_t(\theta))| &\leq 2 \left(d_-^{-2} Y_{1,t}^2 + Y_{2,t} + d_-^{-1} + 1 \right) |\lambda_t(\theta) - \bar{\lambda}_t(\theta)|_1 \\ &\leq C (\sup_{\theta \in \Theta} |\lambda_0(\theta)|_1 + |\bar{\lambda}_0|_1) \left(d_-^{-2} Y_{1,t}^2 + Y_{2,t} + d_-^{-1} + 1 \right) \tau^t \end{aligned}$$

for some constants $C > 0, \tau \in (0, 1)$. Since the logarithmic moment of

$$(\sup_{\theta \in \Theta} |\lambda_0(\theta)|_1 + |\bar{\lambda}_0|_1) \left(d_-^{-2} Y_{1,t}^2 + Y_{2,t} + d_-^{-1} + 1 \right)$$

is finite, then $\sum_{t \geq 1} \sup_{\theta \in \Theta} |h_t(\bar{\lambda}_t(\theta)) - h_t(\lambda_t(\theta))| < \infty$ and consequently, almost surely, as n tends to infinity,

$$n^{-1} \sum_{t=1}^n \sup_{\theta \in \Theta} |h_t(\lambda_t(\theta)) - h_t(\bar{\lambda}_t(\theta))| \rightarrow 0. \square$$

3.6.11 Proof of Proposition 13

We check the assumptions of Theorem 3, in particular **A8-A12**. Note that **A12** follows directly from Lemma 3. Moreover, checking **A8** is straightforward and then omitted. We next check **A9** and **A10**. Due to the specific form of $h_{1,y}$ and $h_{2,y}$, we only have to check the integrability of the following random variables.

$$\lambda_{1,0}(\theta_0) \sup_{\theta \in \Theta} \frac{\|\nabla^{(2)} \lambda_{1,0}(\theta)\|}{\lambda_{1,0}(\theta)^2}, \quad \sup_{\theta \in \Theta} \frac{\|\nabla^{(2)} \lambda_{1,0}(\theta)\|}{\lambda_{1,0}(\theta)}, \quad \lambda_{1,0}(\theta_0) \sup_{\theta \in \Theta} \frac{\|\nabla \lambda_{1,0}(\theta)\|^2}{\lambda_{1,0}(\theta)^3}, \quad \sup_{\theta \in \Theta} \frac{\|\nabla \lambda_{1,0}(\theta)\|^2}{\lambda_{1,0}(\theta)^2},$$

as well as

$$\sup_{\theta \in \Theta} \|\nabla^{(2)} \lambda_{2,0}(\theta)\|, \quad \lambda_{2,0}(\theta_0) \left(\sup_{\theta \in \Theta} \frac{\|\nabla \lambda_{2,0}(\theta)\|^2}{\lambda_{2,0}(\theta)^2} + \sup_{\theta \in \Theta} \frac{\|\nabla^{(2)} \lambda_{2,0}(\theta)\|}{\lambda_{2,0}(\theta)} \right).$$

All these integrability conditions follows from Lemma 14, Lemma 12 and Proposition 2 with $r = 1 + \delta$.

Finally, we check **A11**. To this end, it is sufficient to show that $\sum_{t=1}^{\infty} \sup_{\theta \in \Theta} \zeta_t(\theta) < \infty$, when $\zeta_t(\theta)$ is one of the following quantities.

$$\begin{aligned} & \|\nabla \lambda_{1,t}(\theta) - \nabla \bar{\lambda}_{1,t}(\theta)\| \times \lambda_{1,t}(\theta_0), \quad |\lambda_{1,t}(\theta) - \bar{\lambda}_{1,t}(\theta_0)| \times \|\nabla \bar{\lambda}_{1,t}(\theta)\|, \quad \frac{\|\nabla \bar{\lambda}_{1,t}(\theta)\|}{\lambda_{1,t}(\theta)} \times |\lambda_{1,t}(\theta) - \bar{\lambda}_{1,t}(\theta)| \times \lambda_{1,t}(\theta_0), \\ & \|\nabla \lambda_{2,t}(\theta) - \nabla \bar{\lambda}_{2,t}(\theta)\| \times \lambda_{2,t}(\theta_0), \quad \frac{\|\nabla \bar{\lambda}_{2,t}(\theta)\|}{\lambda_{2,t}(\theta)} \times |\lambda_{2,t}(\theta) - \bar{\lambda}_{2,t}(\theta)| \times \lambda_{2,t}(\theta_0). \end{aligned}$$

The result follows from the approximation results given in Lemma 13 and the integrability conditions given by Lemmas 14, 12 and Proposition 9. \square

3.6.12 Proof of Proposition 15

From **A13-A14** and the continuity assumption on f_1 , we have a uniform law of large numbers. In particular,

$$\sup_{R \in \Gamma} \left| \frac{1}{n} \sum_{t=1}^n f_t(\hat{\theta}, R) - \bar{f}(\theta_0, R) \right| \rightarrow 0 \text{ a.s.}$$

Next,

$$\frac{1}{n} \sum_{t=1}^n f_t(\hat{\theta}, \hat{R}) \geq \frac{1}{n} \sum_{t=1}^n f_t(\hat{\theta}, R_0).$$

Assume that \tilde{R} is a cluster point of the sequence $(\hat{R}(\omega))_n$ for an ω such that the previous uniform convergence holds true. Taking the limit in the previous equality, we get

$$\bar{f}(\theta_0, \tilde{R}) \geq \bar{f}(\theta_0, R_0).$$

Then $\tilde{R} \in \mathcal{I}_0$. Hence, $d(\hat{R}, \mathcal{I}_0) \rightarrow 0$ a.s.

We next study the convergence in total variation distance. To this end, we give another description of the set \mathcal{I}_0 . Denoting by

$$KL_\lambda(R_0, R) = \int \log \left(\frac{p_R(y|\lambda)}{p_{R_0}(y|\lambda)} \right) p_{R_0}(y|\lambda) \mu(dy)$$

the Kullback-Leibler divergence between $P_R(\cdot|\lambda)$ and $P_{R_0}(\cdot|\lambda)$ which is a non-negative quantity, we have

$$\bar{f}(R, \theta_0) - \bar{f}(R_0, \theta_0) = \mathbb{E}[KL_{\lambda_0}(R_0, R)].$$

Hence, if $R \in \mathcal{I}_0$, we have $KL_\lambda(R_0, R) = 0$ for every λ in an event of \mathbb{P}_{λ_0} -probability one. From **A15**, $R \mapsto KL_\lambda(R_0, R)$ is continuous for every λ . We then deduce the existence of a measurable set Λ such that $\mathbb{P}_{\lambda_0}(\Lambda) = 1$ and for every $\lambda \in \Lambda$ and $R \in \mathcal{I}_0$, $KL_\lambda(R_0, R) = 0$. Now let us show that almost surely,

$$KL_\lambda(R_0, \hat{R}) \rightarrow 0, \quad \lambda \in \Lambda. \tag{3.31}$$

Let $\lambda \in \Lambda$. Since any cluster point \tilde{R} of a sequence $\hat{R}(\omega)$ is in \mathcal{I}_0 , we have $KL_\lambda(R_0, \tilde{R}) = 0$ and then (3.31) follows. Using Pinsker's inequality (the total variation distance is bounded by the square root of one half of the Kullback-Leibler divergence), we also get the convergence in total variation distance. \square

3.6.13 Proof of Theorem 13

If R is in a compact set Γ , there exist some positive real numbers $\alpha_0, \beta_0, \alpha_1, \beta_1$ such that for any $R \in \Gamma$,

$$\alpha_0 \exp \left(-\frac{1}{2} (\beta_0 - 1) \Phi^{-1}(u)' \Phi^{-1}(u) \right) \leq c_R(u) \leq \alpha_1 \exp \left(-\frac{1}{2} (\beta_1 - 1) \Phi^{-1}(u)' \Phi^{-1}(u) \right). \tag{3.32}$$

In what follows, we will derive a lower and an upper bound for the integral

$$I = \int_{d_{\ell+1}}^{e_{\ell+1}} \cdots \int_{d_k}^{e_k} c_R(u_1, \dots, u_k) du_{\ell+1} \cdots du_k,$$

where $0 \leq d_i < e_i \leq 1$ for $\ell + 1 \leq i \leq k$.

Upper bound for I Getting an upper bound for I is straightforward. From (3.32), we have

$$I \leq I' \alpha_1 \prod_{i=\ell+1}^k \int_0^1 \exp\left(-\frac{1}{2}(\beta_1 - 1)\Phi^{-1}(u_i)^2\right) du_i = \alpha_1 J_1^{k-\ell},$$

where $I' = \prod_{i=1}^{\ell} \exp\left(-\frac{1}{2}(\beta_1 - 1)\Phi^{-1}(u_i)^2\right)$ and (after a change of variable $x = \Phi^{-1}(u_i)$),

$$J_1 = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\beta_1 x^2\right) dx = \sqrt{2\pi\beta_1^{-1}}.$$

Lower bound for I Setting

$$I'' = \prod_{i=1}^{\ell} \exp\left(-\frac{1}{2}(\beta_0 - 1)\Phi^{-1}(u_i)^2\right)^2$$

and using again (3.32), we have

$$I \geq I'' \alpha_0 \prod_{i=\ell+1}^k \int_{d_i}^{e_i} \exp\left(-\frac{1}{2}(\beta_0 - 1)\Phi^{-1}(u_i)^2\right) du_i.$$

It is then necessary to get a lower bound for

$$J = \int_d^e \exp\left(-\frac{1}{2}(\beta_0 - 1)\Phi^{-1}(u)^2\right) du,$$

for some real numbers $0 \leq d < e \leq 1$. We consider several cases.

1. Suppose first that $0 < d < e < 1$. In this case, we have

$$J \geq (e - d) \exp\left(-\frac{1}{2}(\beta_0 - 1)\Phi^{-1}(e)^2\right) \exp\left(-\frac{1}{2}(\beta_0 - 1)\Phi^{-1}(d)^2\right).$$

2. Assume now that $d = 0$ and $e < 1$. In this case

$$J = \int_{-\infty}^{\Phi^{-1}(e)} \exp\left(-\frac{1}{2}\beta_0 x^2\right) dx = \sqrt{2\pi\beta_0^{-1}} \Phi\left(\sqrt{\beta_0} \Phi^{-1}(e)\right).$$

Using the inequality $\Phi(x) + \Phi(-x) = 1$ and the inequality

$$1 - \Phi(x) \geq \frac{\exp\left(-\frac{x^2}{2}\right)}{2\sqrt{2\pi}x}, \quad x \geq 1, \quad (3.33)$$

we get for $x \geq 1$, $1 - \Phi(x) \geq f \exp(-x^2)$ for a suitable constant $1 \geq f > 0$ such that $f \leq \Phi(-1)$ and $f \leq \sqrt{2\pi\beta_0^{-1}}$. Then

$$\Phi(x) \geq \Phi(-1)\mathbf{1}_{x \geq -1} + (1 - \Phi(-x))\mathbf{1}_{x < -1} \geq f \exp(-x^2).$$

We then get

$$J \geq f^2 \exp\left(-\beta_0 \Phi^{-1}(e)^2\right).$$

3. Assume next that $e = 1$ and $d > 0$. We also get

$$\sqrt{2\pi\beta_0^{-1}} \left(1 - \Phi\left(\sqrt{\beta_0} \Phi^{-1}(d)\right)\right) \geq f^2 \exp\left(-\beta_0 \Phi^{-1}(d)^2\right).$$

4. Finally if $e = 1$ and $d = 0$, then $J \geq \sqrt{2\pi\beta_0^{-1}} \geq f \geq f^2$.

We then showed the following result.

Lemma 7 1. *There exist some real numbers f_1 and f_2 , not depending on the d_i 's and the e_i 's such that*

$$\log I \leq f_1 + f_2 \sum_{i=1}^{\ell} \Phi^{-1}(u_i)^2.$$

2. *There exist some real numbers f'_1, f'_2 and f'_3 , not depending on the d_i 's and the e_i 's such that*

$$\log I \geq f'_1 + f'_2 \sum_{i=1}^{\ell} \Phi^{-1}(u_i)^2 + \sum_{i=\ell+1}^k \left[\log(e_i - d_i) \mathbf{1}_{0 < d_i < e_i < 1} + f'_3 \Phi^{-1}(e_i)^2 \mathbf{1}_{e_i < 1} + f'_3 \Phi^{-1}(d_i)^2 \mathbf{1}_{d_i > 0} \right].$$

The next lemma will be also needed.

Lemma 8 1. *There exist $\delta_1 > 0$ such that for any $u \in (0, 1)$,*

$$\Phi^{-1}(u)^2 \leq \delta_1 (1 - \log(u) - \log(1 - u)).$$

2. *Let X be a random variable supported on the integers and such that $p_0 = \mathbb{P}(X =$*

$0) \in (0, 1)$. If F denotes the cdf of X , we have the bound

$$\mathbb{E} \left[\Phi^{-1} \left(F(X^-) \right)^2 \mathbf{1}_{X \geq 1} \right] \leq \delta_2 (1 - \log(p_0) - \log(1 - p_0)),$$

where δ_2 does not depend on F .

3. Let X be a random variable supported on the integers and such that $p_k = \mathbb{P}(X = k) \in (0, 1)$ for any $k \in \mathbb{N}$. If F denotes the cdf of X , we have the bound

$$\mathbb{E} \Phi^{-1} (F(X))^2 \leq \delta_1 (1 - \log(p_0) + \mathbb{E} \log(1 - F(X))),$$

where $\delta_3 > 0$ does not depend on F .

Proof of Lemma 8

1. Since $\Phi^{-1}(u) \sim \sqrt{-2 \log(u)}$ when $u \sim 0$ and $\Phi^{-1}(u) \sim \sqrt{-2 \log(1 - u)}$ when $u \sim 1$, the result is straightforward.
2. We represent X as $F^{-1}(U)$. On the event $\{X \geq 1\}$, we have $F(X^-) \geq p_0$ and

$$F(X^-) = \sum_{k=1}^{\infty} F(k-1) \mathbf{1}_{F(k-1) < U \leq F(k)} \leq U.$$

Since $\mathbb{E} \Phi^{-1}(U)^2 = 1$, the result follows from the bound given in the previous point.

3. Using the first point of the lemma, it is only necessary to bound $-\mathbb{E} \log(F(X)) \leq -\log(p_0)$. \square

We now go back to the proof of Theorem 4. It is only necessary to check **A14** and **A15**.

1. We first check **A15**. We use the fact that $R \mapsto c_R$ is continuous and we apply the dominated convergence theorem. To this end, we use the upper/lower bound on I given in Lemma 7 and it is necessary to check the following integrability conditions.

$$\int \Phi^{-1} (F_{i,\lambda_i}(y))^2 p_{i,\lambda_i}(y) dy < \infty, \quad 1 \leq i \leq \ell, \quad (3.34)$$

$$\int \Phi^{-1} (F_{i,\lambda_i}(y))^2 \mathbf{1}_{F_{i,\lambda_i}(y) < 1} p_{i,\lambda_i}(y) d\mu_i(y) < \infty, \quad \ell + 1 \leq i \leq k, \quad (3.35)$$

$$\int \Phi^{-1} (F_{i,\lambda_i}(y^-))^2 \mathbf{1}_{F_{i,\lambda_i}(y^-) > 0} p_{i,\lambda_i}(y) d\mu_i(y) < \infty, \quad \ell + 1 \leq i \leq k, \quad (3.36)$$

$$\int -\log(p_{i,\lambda_i}(y)) p_{i,\lambda_i}(y) d\mu_i(y) < \infty, \quad \ell + 1 \leq i \leq k. \quad (3.37)$$

Checking (3.34) is automatic by continuity of F_{i,λ_i} , since any integral of this form writes as $\mathbb{E}\Phi^{-1}(U)^2 = 1$ where U is a uniformly distributed over $[0, 1]$. (3.35) and (3.37) follow from Assumption **G1**. Moreover, it is easy to check that (3.36) is valid either for the Bernoulli distribution or for any distribution with full support \mathbb{N} .

2. We next check **A14**. From Lemma 7 and Lemma 8, we only have to check the following integrability conditions. When $1 \leq i \leq \ell$, we have to show that

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log F_{i,\lambda_{i,0}(\theta)}(Y_{i,0}) \right\} + \sup_{\theta \in \Theta} \left\{ -\log \left(1 - F_{i,\lambda_{i,0}(\theta)}(Y_{i,0}) \right) \right\} \right] < \infty. \quad (3.38)$$

When $\ell + 1 \leq i \leq k$, we have to show that

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log p_i(0|\lambda_{i,0}(\theta)) \right\} + \sup_{\theta \in \Theta} \left\{ -\log \left(1 - p_i(0|\lambda_{i,0}(\theta)) \right) \right\} \right] < \infty, \quad (3.39)$$

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log p_i(Y_{i,0}|\lambda_{i,0}(\theta)) \right\} \right] < \infty \quad (3.40)$$

and for count marginal time series,

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log \left(1 - F_{i,\lambda_{i,0}(\theta)}(Y_{i,0}) \right) \right\} \right] < \infty. \quad (3.41)$$

These conditions are precisely ensured by Assumptions **G2-G3**.

To end the proof of Theorem 4, it is necessary to check that parameter R can be identified, i.e. that $\mathcal{I}_0 = \{R_0\}$ where \mathcal{I}_0 is defined before the statement of Proposition 8. From the properties of Kullback-Leibler divergence, it is sufficient to show that if $p_R(\cdot|s) = p_{R_0}(\cdot|s)$, μ -almost everywhere for some $s \in F_1 \times \dots \times F_k$, then $R = R_0$. Such identification property is already known in the literature. See for instance [96], appendix A. For simplicity, we summarize the required arguments, using our notations. To show this, we first give an expression of the density $p_R(\cdot|s)$ which will be simply denoted $p_R(\cdot)$ here. Moreover, we simply denote by p_i the density $p_i(\cdot|s_i)$ and $F_i = F_{i,s_i}$. In what follows, for any value of k , we denote by Φ_R the Gaussian density with mean 0 and covariance matrix R and simply by ϕ the density of the standard Gaussian distribution on the real

line. We have

$$\begin{aligned} & \frac{p_R(\mathbf{y})}{\prod_{i=1}^{\ell} p_i(y_i)} \\ = & \int_{F_{\ell+1}(y_{\ell+1}^-)}^{F_{\ell+1}(y_{\ell+1})} \cdot \int_{F_k(y_k^-)}^{F_k(y_k)} \frac{\phi_R(\Phi^{-1}(F_1(\mathbf{y}_1)), \cdot, \Phi^{-1}(F_{\ell}(y_{\ell})), \Phi^{-1}(u_{\ell+1}), \cdot, \Phi^{-1}(u_k))}{\phi_{I_k}(\Phi^{-1}(F_1(\mathbf{y}_1)), \cdot, \Phi^{-1}(F_{\ell}(y_{\ell})), \Phi^{-1}(u_{\ell+1}), \cdot, \Phi^{-1}(u_k))} du_{\ell+1} \cdot du_k. \end{aligned}$$

Suppose that $\Phi_R = \Phi_{R_0}$, μ -a.e. For $1 \leq i < j \leq k$, let $R(i, j) = \begin{pmatrix} 1 & r_{ij} \\ r_{ij} & 1 \end{pmatrix}$, which is simply the (sub-)correlation matrix for components i and j . Finally, we denote by $p_{R(i,j)}$ the bivariate density corresponding to these components. We consider three cases.

1. Assume first that $1 \leq i \leq \ell$ and $\ell + 1 \leq j \leq k$. In this case, we have

$$\begin{aligned} & p_{R(i,j)}(y_i, y_j) \\ = & p_i(y_i) \int_{\Phi^{-1}(F_j(y_j^-))}^{\Phi^{-1}(F_j(y_j))} \frac{\phi_{R_{ij}}(\Phi^{-1}(F_i(y_i)), x_j)}{\phi_{I_1}(\Phi^{-1}(F_i(y_i)))} dx_j \\ = & p_i(y_i) \left\{ \Phi \left(\frac{\Phi^{-1}(F_j(y_j)) - r_{ij}\Phi^{-1}(F_i(y_i))}{\sqrt{1 - r_{ij}^2}} \right) - \Phi \left(\frac{\Phi^{-1}(F_j(y_j^-)) - r_{ij}\Phi^{-1}(F_i(y_i))}{\sqrt{1 - r_{ij}^2}} \right) \right\}. \end{aligned}$$

Then if $p_{R(i,j)} = p_{R_0(i,j)}$ almost everywhere, there exists $w \in \mathbb{R}$ such that

$$\Phi \left(\frac{w - r_{ij}\Phi^{-1}(F_i(y_i))}{\sqrt{1 - r_{ij}^2}} \right) = \Phi \left(\frac{w - r_{0ij}\Phi^{-1}(F_i(y_i))}{\sqrt{1 - r_{0ij}^2}} \right),$$

for almost every value of y_i (with respect to the Lebesgue measure). Since Φ is one-to-one and $F_i(y_i)$ can take arbitrary values between 0 and 1, it is easily seen that $r_{ij} = r_{0ij}$.

2. Assume now that $\ell + 1 \leq i < j \leq k$. In this case, we have

$$p_{R(i,j)}(y_i, y_j) = \int_{\Phi^{-1}(F_i(y_i^-))}^{\Phi^{-1}(F_i(y_i))} \int_{\Phi^{-1}(F_j(y_j^-))}^{\Phi^{-1}(F_j(y_j))} \phi_{R(i,j)} dx_i dx_j.$$

We use the expression,

$$\phi_{R(i,j)}(x_i, x_j) = \phi(x_i) \left(2\pi(1 - r_{ij}^2) \right)^{-1/2} \exp \left(-\frac{(x_j - r_{ij}x_i)^2}{2(1 - r_{ij}^2)} \right).$$

Whatever the cases (binary or count variables), if $p_{R(i,j)} = p_{R_0(i,j)}$ almost everywhere,

there exists two real numbers w_i and w_j such that we have the equality $f(r_{ij}) = f(r_{0ij})$ with

$$f(r) = \int_{-\infty}^{w_i} \Phi\left(\frac{w_j - rx_i}{\sqrt{1-r^2}}\right) \phi(x_i) dx_i.$$

However, since after some computations, the derivative of f can be written as

$$\dot{f}(r) = -(1+d^2)^{-1} \phi(c-dw_i) \phi(w_i),$$

with $c = (1-r^2)^{-1/2}w_j$ and $d = (1-r^2)^{-1/2}r$, we see that f is decreasing. Hence, $r_{ij} = r_{0ij}$.

3. Finally, if $1 \leq i < j \leq \ell$, we recover the identification problem for continuous margins with

$$p_{R_{ij}}(y_i, y_j) = p_i(y_i)p_j(y_j) \frac{\phi_{R_{ij}}(\Phi^{-1}(F_i(y_i)), \Phi^{-1}(F_j(y_j)))}{\phi(\Phi^{-1}(F_i(y_i)))\phi(\Phi^{-1}(F_j(y_j)))}.$$

Identification of r_{ij} is straightforward in this case.

We then deduce that the set \mathcal{I}_0 only contains R_0 and the consistency result now follows from Proposition 8. \square

3.6.14 Proof of Corollary 2

We check the assumptions of Theorem 4. When p_{i,s_i} is the Poisson distribution with parameter $\phi(s_i)$, with $\phi(s_i) = s_i$ or $\phi(s_i) = \exp(s_i)$, it is straightforward to check the first and the third conditions **G1**. It remains to check the second one. If X follows a Poisson distribution with parameter μ and denoting by F its cdf, we have

$$1 - F(k) \geq \exp(-\mu) \frac{\mu^{k+1}}{(k+1)!}. \quad (3.42)$$

The previous lower bound follows from a Taylor-Lagrange expansion. We then get

$$-\log(1 - F(k)) \leq \log(\mu) - (k+1)\log(\mu) + \sum_{i=1}^{k+1} \log(i) \leq -k\log(\mu) + \frac{k(k+1)}{2}.$$

In the last bound, we have simply used the bound $\log(i) \leq i - 1$ for $i \geq 1$. We then get

$$-\mathbb{E} \log(1 - F(X)) \leq C_1(1 + \mu^2),$$

where $C_1 > 0$ does not depend on F .

Next we check **G3**. It is easily seen that if p_{i,s_i} is the Poisson distribution with parameter s_i , then the three conditions in **G3** are satisfied as soon as $\mathbb{E} \sup_{\theta \in \Theta} \lambda_{i,0}(\theta)^{1+\delta} < \infty$, which is guaranteed from the assumptions of Proposition 5.

Finally, **G2** is satisfied for the GARCH component as soon as $\mathbb{E} \sup_{\theta \in \Theta} \lambda_{i,0}(\theta) < \infty$, which is also automatic under the assumptions of Proposition 5. Indeed, let us denote by s_- the smallest values taken by $\sqrt{\lambda_{i,t}(\theta)}$. If (20) holds true, we have on the even $\{\varepsilon_t < 0\}$ where ε_t denotes the noise component,

$$F_{i,\lambda_{i,t}(\theta)}(Y_{i,t}) \geq F_\epsilon(Y_{i,t}/s_-) \geq F_\epsilon(-c_3) \mathbb{1}_{Y_{i,t} \geq -c_3 s_-} + c_4 \exp(-c_5 Y_{i,t}^2) \mathbb{1}_{Y_{i,t} < -c_3 s_-}$$

for some positive constant c_4, c_5 . For the last inequality, we used the lower bound (20) on f_ϵ combined with an inequality similar to (3.33). Hence $\sup_{\theta \in \Theta} -\log(F_{i,\lambda_{i,t}(\theta)}(Y_{i,t}))$ is integrable as soon as $\mathbb{E}(Y_{i,t}^2) = \mathbb{E}\lambda_{i,t}(\theta_0) < \infty$. The second integrability condition required in **G2** is similar. The result then follows from Theorem 4. \square

3.6.15 Proof of Corollary 3

The proof is similar to that of Corollary 2. One can show that **G2** is satisfied for the Poissonian component as soon as $\mathbb{E} \exp((1+\delta) \sup_{\theta \in \Theta} |\lambda_{i,0}(\theta)|) < \infty$ for some $\delta > 0$, which is covered by our assumptions. Note that (19) is satisfied under the same type of conditions by using the lower bound (3.42) for the survival function of a Poisson distribution.

For the binary coordinate with $p_i(1|\lambda_i) = F(\lambda_i)$ where F is the logistic cdf. In this case, it is only necessary to check (17). The required conditions are satisfied if $\mathbb{E} \sup_{\theta \in \Theta} |\lambda_{i,0}(\theta)| < \infty$ which is the case under our assumptions. The other conditions in **G1** are trivial to show or has been discussed in the proof of Corollary 2. The result then follows from Theorem 4. \square

3.6.16 Checking Assumptions A14-A15 for Clayton copula

Here we assume $k = 2$ and for some R in a compact subset Γ of $(0, \infty)$,

$$c_R(u_1, u_2) = (R+1)^2 u_1^{-R-1} u_2^{-R-1} \left(u_1^{-R} + u_2^{-R} - 1 \right)^{-\frac{2R+1}{R}}.$$

We assume furthermore that the first component is continuous (such as ARMA or GARCH) and the second one discrete (non-degenerate binary or count distribution) and we will prove a result analogue to that of Theorem 4 of the paper. A similar result can be obtained when both components are discrete. More precisely, we prove the following result.

Proposition 16 *Let $\hat{\theta}$ be a consistent estimator of θ_0 . Assume that condition (18) in **G3** holds true as well as **A5** with $h_{1,y}(s_1) = -\log p_1(y|s_1)$,*

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \left\{ -\log F_{1,\lambda_{1,0}(\theta)}(Y_{1,0}) \right\} \right] < \infty$$

and if the discrete component follows a count distribution,

$$\sum_{y \in \mathbb{N}} \log(p_i(y|s_i)) p_i(y|s_i) > -\infty.$$

We then have $\lim_{n \rightarrow \infty} \hat{R} = R_0$ a.s.

Proof of Proposition 16 Let us first check **A14** and set

$$I = I_R = \int_{d_1}^{e_2} c_R(u_1, u_2) du_2, \quad \text{with } d_2 = Z_{2,0}^-(\theta), \quad e_2 = Z_{2,0}(\theta), \quad u_1 = Z_{1,0}(\theta).$$

We have

$$I = \frac{R+1}{u_1^{R+1}} \left[\left(e_2^{-R} + u_1^{-R} - 1 \right)^{-1-R^{-1}} - \left(d_2^{-R} + u_1^{-R} - 1 \right)^{-1-R^{-1}} \right]. \quad (3.43)$$

1. For the upper bound, we have

$$\log(I) \leq \log(R+1) - (R+1) \log(u_1) + (R+1) \log(e_2).$$

2. We also derive a lower bound. If $d_2 \neq 0$, setting $h = u_1^{-R} - 1$, and $\beta = (R+1)/R$ and using the mean value theorem, we have

$$\begin{aligned} & \log(I) \\ & \geq 2 \log(R+1) - (R+1) \log(u_1) - (R+1) \log(z) + \log(e_2 - d_2) - (\beta + 1) \log(z^{-R} + h), \end{aligned}$$

for some $z \in [d_2, e_2]$. Note also that

$$\log(z^{-R} + h) \leq \log(2) - R \log(z) - R \log(u_1) \leq \log(2) - R \log(d_2) - R \log(u_1).$$

Now, if $d_2 = 0$, we have

$$\log(I) \geq \log(R + 1) - (R + 1) \log(u_1) - \beta \log(2) + (R + 1) \log(e_2) + (R + 1) \log(u_1).$$

From these bounds, we deduce that **A14** is valid under our assumptions. Checking **A15** follows as for **A14**, using the lower and upper bounds for $\log(I)$ as well as Lebesgue's theorem. Finally, we check that $\mathcal{I}_0 = \{R_0\}$. Note first that from the Kullback-Leibler divergence properties, the equality $\bar{f}(\theta_0, R_0) = \bar{f}(\theta_0, R)$ entails that

$$p_R(Y_0 | \lambda_t(\theta_0)) = p_{R_0}(Y_0 | \lambda_t(\theta_0)) \text{ a.s.}$$

In particular, we get $I_R = I_{R_0}$ a.s. By taking summations for possible values for $Y_{2,0}$, we deduce that almost surely

$$\frac{R + 1}{u_1^{R+1}} (e_2^{-R} + u_1^{-R} - 1)^{-1-R^{-1}} = \frac{R_0 + 1}{u_1^{R_0+1}} (e_2^{-R_0} + u_1^{-R_0} - 1)^{-1-R_0^{-1}}$$

and taking $e_2 = 1$ in the previous equality, we deduce that $R = R_0$. \square

3.6.17 A multivariate binary/logINGARCH/logGARCH model

In this section, our aim is to check the integrability conditions needed for the model discussed in Section 3.6 of the paper. We assume that the ℓ_1 -first components are binary-logistic, the ℓ_2 next components are Poisson log-linear and the remaining ℓ_3 components are given by a log-GARCH. We assume that the noise of the log-GARCH component has finite polynomial moments, X_0 has exponential moments of any order,

$$|\bar{A}_0|_{vec} + |B_0|_{vec} < 1$$

and **L3** holds true with $c_i = 1/4$ if $1 \leq i \leq \ell_1$ and $c_i = 1$ otherwise. We first prove the analogue of Proposition 4. We will still apply the result of Lemma 11 with $r = 1$ and the function $\phi(s) = \exp(\kappa|s|_{vec})$ where the exponential function is applied componentwise. Here, note that $F_{i,s_i}^{-1}(U_{i,t}) = e^{s_i/2} F_\epsilon^{-1}(U_{i,t})$ for $\ell_1 + \ell_2 + 1 \leq i \leq k$. Setting

$$\mu_i = \sum_{j=1}^2 (|\bar{A}(i, j)| + |B(i, j)|) < 1,$$

$$C_i = \sum_{j=1}^{\ell_1} |A(i, j)| + \sum_{j=\ell_2+1}^k |A(i, j)| \log F_\epsilon^{-1}(U_{i,t})^2$$

for $i = 1, \dots, k$ and denoting by \odot the Hadamard product, we now have

$$|g(s, F_s^{-1}(U_t), X_t)|_{vec} \preceq (1 - \mu) \odot \frac{|d + \Gamma X_t|_{vec} + C}{1 - \mu} + |\bar{A}|_{vec} V_t + |B|_{vec} |s|_{vec},$$

where $V_{i,t} = 0$ if $1 \leq i \leq \ell_1$, $V_{i,t} = F_{i,s_i}^{-1}(U_{i,t})$ for $\ell_1 + 1 \leq i \leq \ell_1 + \ell_2$ and $V_{i,t} = s_i$ for $\ell_1 + \ell_2 + 1 \leq i \leq k$. Using convexity of the exponential function, we deduce that

$$\phi \left(g(s, F_s^{-1}(U_t), X_t) \right) \preceq (1 - \mu) \odot \phi \left(\frac{d + \Gamma X_t + C}{1 - \mu} \right) + |\bar{A}|_{vec} \phi(V_t) + |B|_{vec} \phi(s). \quad (3.44)$$

Note that for any $\delta > 0$, there exists $d_{\kappa,\delta} > 0$ such that for $\ell_1 + 1 \leq i \leq \ell_1 + \ell_2$,

$$\mathbb{E} \exp(\kappa g_i \circ F_{i,s_i}^{-1}(U_{i,t})) = \mathbb{E} \left(1 + F_{2,s_i}^{-1}(U_{i,t}) \right)^\kappa \leq (1 + \delta) \exp(\kappa s_i) + d_{\kappa,\delta}.$$

The previous bound can be obtained from Lemma 9, using the convexity of power functions. Now let v the vector given by $v_i = d_{\kappa,\delta}$ if $\ell_1 + 1 \leq i \leq \ell_1 + \ell_2$ and $v_i = 0$ otherwise. In what follows, we denote by c_{t-1} the sum between the vector $|\bar{A}|_{vec} v$ and conditional expectation of $(1 - \mu) \odot \phi \left(\frac{d + \Gamma X_t + C}{1 - \mu} \right)$ with respect to \mathcal{F}_{t-1} . Taking the conditional expectation with respect to \mathcal{F}_{t-1} in (3.44), we deduce that

$$\|\phi \left(g(s, F_s^{-1}(U_t), X_t) \right)\|_{t-1, vec} \preceq c_{t-1} + \left(|\bar{A}|_{vec} \widetilde{D} + |B|_{vec} \right) \phi(s),$$

where \widetilde{D} is a diagonal matrix of size k with $\widetilde{D}_{i,i} = 1 + \delta$ if $\ell_1 + 1 \leq i \leq \ell_1 + \ell_2$ and $\widetilde{D}_{i,i} = 1$ otherwise. Note first that the existence of all exponential moments for X_0 and all polynomial moments for ϵ entail that $\mathbb{E} c_0 < \infty$. Taking δ small enough, our assumptions guarantee that the spectral radius of $|\bar{A}|_{vec} \widetilde{D} + |B|_{vec}$ is less than 1 and Lemma 11 leads to $\mathbb{E} |\phi(\lambda_0)|_1 < \infty$. Finally, we use the bound

$$\mathbb{E} \phi(g_i \circ Y_{i,0}) = \mathbb{E} (1 + Y_{i,0})^\kappa \leq (1 + \delta) \mathbb{E} \phi(\lambda_{i,0}) + d_{\kappa,\delta} < \infty$$

for $\ell_1 + 1 \leq i \leq \ell_1 + \ell_2$ and

$$\mathbb{E}\phi(g_i \circ Y_{i,0}) \leq \mathbb{E}(\epsilon^{2\kappa})\mathbb{E}\phi(\lambda_{i,0})$$

for $\ell_1 + \ell_2 + 1 \leq i \leq k$ to conclude.

3.6.18 Combining a binary-logistic and an ARMA(1, 1) dynamic

We check the main condition given in the paper for getting consistency and asymptotic normality of the pseudo-likelihood estimators of the autoregressive parameters, when one component is given by a binary-logistic dynamic and the second one by an ARMA type dynamic. Here we assume that F_{1,s_1} is the cdf of a Bernoulli distribution with parameter $(1 + e^{-s_1})^{-1}$ and F_{2,s_2} is the cdf of $s_2 + \epsilon$ where ϵ is a centered random variable with finite variance. We also assume that X_0 has a finite second moment. Here the link functions g_1, g_2 are both the identity function. To check conditions (14) of the main document for both coordinates, we assume that

$$\rho(|A_0|_{vec}diag(1/4, 1) + |B|_{vec}) < 1. \quad (3.45)$$

From Corollary 1, this condition ensures the existence of a stationary solution with a finite first moment. We show that this solution has a finite second moment. To this end, we apply Lemma 11. From Minkowski's inequality for conditional expectations, we have

$$\mathbb{E}^{1/2} \left[|g(s, F_s^{-1}(U_t), X_t)|_{vec}^2 | \mathcal{F}_{t-1} \right] \preceq \mathbb{E} \left[|d + \Gamma X_t|_{vec}^2 | \mathcal{F}_{t-1} \right] + |B|_{vec} \cdot |s|_{vec} + |A|_{vec} \mathbb{E}^{1/2} \left[|F_s^{-1}(U_t)|_{vec}^2 \right].$$

Since

$$\mathbb{E}^{1/2} \left[F_{1,s_1}^{-1}(U_{1,t})^2 \right] \leq 1, \quad \mathbb{E}^{1/2} \left[F_{2,s_2}^{-1}(U_{2,t})^2 \right] \leq s_2 + \text{Var}(\epsilon),$$

it is not difficult to see that Lemma 11 can be applied as soon as $\rho(|\bar{A}|_{vec} + |B|_{vec}) < 1$. The latter condition is implied by (3.45). Indeed if C and D are square matrices with non-negative entries and $C \leq D$ componentwise, we have $\rho(C) \leq \rho(D)$. One then conclude to the existence of a second moment and applying Lemma 12, we get the result.

Appendix

3.6.19 Two useful lemmas

Lemma 9 *Let $\lambda > 0$ and X_λ Poisson variable with parameter λ . Then, $\forall r \geq 1$ and any $\delta \in (0, 1)$, there exists $b_{r,\delta}$, not depending on λ and such that*

$$\|X_\lambda\|_r \leq (1 + \delta)\lambda + b_{r,\delta}.$$

Proof of lemma 9 We have the equality $\mathbb{E}(X_\lambda^r) = \sum_{i=1}^r \lambda^i \left\{ \begin{matrix} r \\ i \end{matrix} \right\}$ with $\left\{ \begin{matrix} r \\ i \end{matrix} \right\}$ are the Sterling's numbers of second kind. See for instance [76].

Then

$$\mathbb{E}(X_\lambda^r) = \lambda^r + \sum_{i=1}^{r-1} \lambda^i \left\{ \begin{matrix} r \\ i \end{matrix} \right\} \leq \lambda^r + C_r(\lambda + \lambda^{r-1}),$$

where $C_r > 0$ only depends on r . But, we can notice that, for any $\delta > 0$, there exists $\bar{b}_{\delta,r} > 0$ such that for all $x \geq 0$: $x + x^{r-1} \leq \delta'x^r + \bar{b}_{\delta,r}$ with $\delta' = \frac{(1+\delta)^{r-1}}{C_r}$. Then $\mathbb{E}(X^r) \leq (1 + C_r\delta')\lambda^r + C_r\bar{b}_{\delta,r}$. Therefore $\|X\|_r \leq (1 + C_r\delta')^{1/r}\lambda + C_r^{1/r}\bar{b}_{\delta,r}^{1/r}$. Setting $b_{\delta,r} = C_r^{1/r}\bar{b}_{\delta,r}^{1/r}$, we get the result. \square

Lemma 10 *Let $B(\theta)$ be a matrix with entries depending continuously on a parameter $\theta \in \Theta$ and Θ is a compact set of \mathbb{R}^d , $d \in \mathbb{N}^*$. Suppose that $\rho(B(\theta)) < 1$ for any $\theta \in \Theta$. There then exist $C > 0$ and $\tau \in (0, 1)$ such that for all integer $j \geq 1$, $\sup_{\theta \in \Theta} |B(\theta)^j|_1 \leq C\tau^j$.*

Proof of Lemma 10 Let $\|\cdot\|$ be an arbitrary norm on \mathbb{R}^d . For any $\theta \in \Theta$, from Gelfand's formula $\lim_{n \rightarrow \infty} |B(\theta)^n|_1^{1/n} = \rho(B(\theta))$. Therefore there exists $\rho_\theta \in (0, 1)$ and $n_\theta \in \mathbb{N}$ such that $|B(\theta)^{n_\theta}|_1 < \rho_\theta$. By continuity of the function $\theta \mapsto B(\theta)$, we can found ϵ_θ such that $\forall \bar{\theta} \in \mathcal{B}(\theta, \epsilon_\theta) = \{\eta \in \Theta : \|\eta - \theta\| < \epsilon_\theta\}$, $|B^{n_\theta}(\bar{\theta})|_1 < \rho_\theta$. By compactness of Θ and Borel-Lebesgue property, $\Theta \subset \bigcup_{i=1}^N \mathcal{B}(\theta_i, \epsilon_{\theta_i})$ for $\theta_1, \dots, \theta_N \in \Theta$. Let us set $\bar{\rho} = \max_{1 \leq i \leq N} \rho_{\theta_i} \in (0, 1)$ and $n_0 = n_{\theta_1} \times \dots \times n_{\theta_N}$, it follows that

$$\sup_{\theta \in \Theta} |B(\theta)^{n_0}|_1 \leq \bar{\rho}.$$

If $n \geq n_0, n = kn_0 + r, k \geq 1, r \in \{0, \dots, n_0 - 1\}$, we will set $\bar{C} = \max_{\theta \in \Theta} (|B(\theta)|_1 + 1)^{n_0}$, and we obtain

$$\sup_{\theta \in \Theta} |B(\theta)^n|_1 \leq \bar{C}\bar{\rho}^k = \bar{C}\bar{\rho}^{\lfloor \frac{n}{n_0} \rfloor} \leq \bar{C}\bar{\rho}^{-1} \left(\bar{\rho}^{\frac{1}{n_0}} \right)^n := C\rho^n. \square$$

Sufficient conditions for finiteness of moments

The following result gives some sufficient conditions for existence of some moments for the stationary solution of the general model. For a random vector $Z = (Z_1, \dots, Z_d), d \in \mathbb{N}^*$, we define $\|Z\|_{t-1, r, vec} := (\mathbb{E}^{1/r} [|Z_1|^r | \mathcal{F}_{t-1}], \dots, \mathbb{E}^{1/r} [|Z_d|^r | \mathcal{F}_{t-1}])'$. In the following result, we consider a measurable mapping $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ and a real number $r \geq 1$. Typically, either $\phi(x) = |x|$ and $r > 1$ or $\phi(x) = \exp(\kappa|x|)$ for some $\kappa > 0$ and $r = 1$. If $x \in \mathbb{R}^d$, $\phi(x)$ is simply the vector $(\phi(x_i))_{1 \leq i \leq d}$.

Lemma 11 *Assume that Assumptions **A1-A3** hold true. Assume that there exist a measurable mapping $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$, a real number $r \geq 1$, a matrix $D \in \mathcal{M}_k$ with nonnegative elements such that $\rho(D) < 1$ and a $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ -adapted and stationary process $(c_t)_{t \in \mathbb{Z}}$, taking values in \mathbb{R}_+^k such that for $s \in \mathbb{R}^k$ and $t \in \mathbb{Z}$,*

$$\left\| \phi \left(g(s, F_s^{-1}(U_t), X_t) \right) \right\|_{t-1, r, vec} \preceq c_{t-1} + D\phi(s).$$

Then $\mathbb{E} [|\phi(\lambda_0)|_1^r] < \infty$ provided that $\mathbb{E}^{1/r} (|c_0|_1^r) < \infty$,

Proof of Lemma 11 Setting $f_t(s) = g(s, F_s^{-1}(U_{t-1}), X_{t-1})$, as in the proof of Theorem 1, we have $\lambda_t = f_t(\lambda_{t-1})$ and Theorem 2 in [32] ensures that $\lambda_t = \lim_{m \rightarrow \infty} f_{t-m}^t(s)$ a.s. for any value of s . Here $f_{t-m}^t = f_t \circ f_{t-1} \circ \dots \circ f_{t-m+1}$. Since f_{t-m}^{t-1} is measurable with respect to \mathcal{F}_{t-2} , we have from our assumption,

$$\begin{aligned} \left\| \phi \left(f_{t-m}^t(s) \right) \right\|_{t-2, r, vec} &= \left\| \phi \left(g(f_{t-m}^{t-1}(s), F_{f_{t-m}^{t-1}(s)}^{-1}(U_{t-1}), X_{t-1}) \right) \right\|_{t-2, r, vec} \\ &\preceq c_{t-2} + D\phi \left(f_{t-m}^{t-1} \right). \end{aligned}$$

From the triangular inequality, we get $\|\phi \left(f_{t-m}^t(s) \right)\|_{r, vec} \preceq \|c_{t-2}\|_{r, vec} + D\|\phi \left(f_{t-m}^{t-1}(s) \right)\|_{r, vec}$. Setting $f = \|c_0\|_{r, vec}$, $h_m(s) = \|\phi \left(f_{-m}^0(s) \right)\|_{r, vec}$ and using stationarity, we get

$$h_m(s) \preceq f + Dh_{m-1}(s) \preceq \sum_{i=0}^{m-1} D^i f + D^m \phi(s).$$

Letting $m \rightarrow \infty$, the condition $\rho(D) < 1$ and Fatou's lemma leads to the result. \square

3.6.20 Approximation results for linear latent processes

In this section, we suppose that :

$$\lambda_t(\theta) = d + B\lambda_{t-1}(\theta) + A\bar{Y}_{t-1} + \Gamma X_{t-1}, \quad t \in \mathbb{Z} \text{ and } \theta \in \Theta, \quad (3.46)$$

The approximate latent process is then define as :

$$\bar{\lambda}_0(\theta) = \bar{\lambda}_0 ; \quad \bar{\lambda}_t(\theta) = d + B\bar{\lambda}_{t-1}(\theta) + A\bar{Y}_{t-1} + \Gamma X_{t-1}, \quad t > 0 \text{ and } \theta \in \Theta. \quad (3.47)$$

The approximate latent process is initialized by a given deterministic vector $\bar{\lambda}_0$. We introduce the following partial derivatives operators :

$$\partial_i = \frac{\partial}{\partial \theta_i} \quad \text{and} \quad \partial_{ij} = \frac{\partial^2}{\partial \theta_i \partial \theta_j}$$

where θ_i stands for the i -th component of the parameters vector θ . **In the whole subsection, we assume that the process $\left((\bar{Y}_t, X_t) \right)_{t \in \mathbb{Z}}$ is stationary, Θ is a compact set and for any $\theta \in \Theta$, $\rho(B) < 1$.**

Lemma 12 (Moments of latent process) *Suppose that there exists $r \geq 1$ such that $\mathbb{E}(|\bar{Y}_0|_1^r) < \infty$ and $\mathbb{E}(|X_0|_1^r) < \infty$. Then the mapping $\theta \mapsto \lambda_0(\theta)$ is almost surely two times continuously differentiable. Moreover,*

1. for $l = 1, \dots, k$, $\mathbb{E}[(\sup_{\theta} |\lambda_{l,0}(\theta)|)^r] < \infty$
2. for $l = 1, \dots, k, i = 1, \dots, Q$, $\mathbb{E}[(\sup_{\theta} |\partial_i \lambda_{l,0}(\theta)|)^r] < \infty$
3. for $l = 1, \dots, k, i, j = 1, \dots, Q$, $\mathbb{E}[(\sup_{\theta} |\partial_{ij} \lambda_{l,0}(\theta)|)^r] < \infty$

Proof of Lemma 12 First note that $\forall \theta$, $\lambda_t(\theta) = \sum_{j \geq 0} B^j (d + A\bar{Y}_{t-j-1} + \Gamma X_{t-j-1})$ is well defined and infinitely differentiable since $\sup_{\theta \in \Theta} \rho(B) < 1$. Moreover, from Lemma 10, there exists $C > 0$ and $\tau \in (0, 1)$ such that for any $\theta \in \Theta$ and any integer $j \geq 1$, $|B^j|_1 \leq C\tau^j$. Setting $D = \sup_{\theta \in \Theta} (|d|_1 + |A|_1 + |\Gamma|_1)$, we get

$$\mathbb{E}^{1/r}[(\sup_{\theta} |\lambda_t(\theta)|_1)^r] \leq CD \sum_{j \geq 0} \tau^j \left(\mathbb{E}^{1/r} (|\bar{Y}_{t-j-1}|_1^r) + \mathbb{E}^{1/r} (|X_{t-j-1}|_1^r) \right)$$

which is finite by stationarity and existence of the moment of order r . Next, for all possible indices, i and j , the partial derivatives of the latent process are given by :

$$\begin{aligned}
 \frac{\partial \lambda_t(\theta)}{\partial d_i} &= \iota_i + B \frac{\partial \lambda_{t-1}(\theta)}{\partial d_i} = \sum_{j \geq 0} B^j \iota_i = (I - B)^{-1} \iota_i, \quad i = 1, \dots, k; \\
 \frac{\partial \lambda_t(\theta)}{\partial A(i, j)} &= E(i, j) \bar{Y}_{t-1} + B \frac{\partial \lambda_{t-1}(\theta)}{\partial A(i, j)} = \sum_{l \geq 0} B^l E(i, j) \bar{Y}_{t-l-1}, \quad i, j = 1, \dots, k; \quad (3.48) \\
 \frac{\partial \lambda_t(\theta)}{\partial \Gamma(i, j)} &= \sum_{l \geq 0} B^l G(i, j) X_{t-l-1}, \quad i = 1, \dots, k, \quad j = 1, \dots, m; \\
 \frac{\partial \lambda_t(\theta)}{\partial B(i, j)} &= E(i, j) \lambda_{t-1}(\theta) + B \frac{\partial \lambda_{t-1}(\theta)}{\partial B(i, j)} = \sum_{l \geq 0} B^l E(i, j) \lambda_{t-l-1}(\theta), \quad i, j = 1, \dots, k
 \end{aligned}$$

where $\iota_i, i = 1, \dots, k$ is the vector of $\{0, 1\}^k$ with 1 at the position i and 0 elsewhere, $E(i, j), i, j = 1, \dots, k$ is the $k \times k$ matrix with 1 at the position (i, j) and 0 elsewhere, $G(i, j), i = 1, \dots, k, j = 1, \dots, m$ is the $k \times m$ matrix with 1 at the position (i, j) and 0 elsewhere and d_i is the i -th element of vector d .

For all possible indices i, j, l, v , the second-order partial derivatives of latent process are given by :

$$\begin{aligned}
 \frac{\partial^2 \lambda_t(\theta)}{\partial d_i \partial d_j} &= \frac{\partial^2 \lambda_t(\theta)}{\partial d_j \partial d_i} = 0 \\
 \frac{\partial^2 \lambda_t(\theta)}{\partial A(i, j) \partial A(u, l)} &= \frac{\partial^2 \lambda_t(\theta)}{\partial \Gamma(i, j) \partial \Gamma(u, l)} = 0 \\
 \frac{\partial^2 \lambda_t(\theta)}{\partial d_i \partial A(j, u)} &= \frac{\partial^2 \lambda_t(\theta)}{\partial A(j, u) \partial d_i} = \frac{\partial^2 \lambda_t(\theta)}{\partial A(i, j) \partial \Gamma(u, l)} = \frac{\partial^2 \lambda_t(\theta)}{\partial \Gamma(u, l) \partial A(i, j)} = \frac{\partial^2 \lambda_t(\theta)}{\partial \Gamma(j, u) \partial d_i} = \frac{\partial^2 \lambda_t(\theta)}{\partial d_i \partial \Gamma(j, u)} = 0 \\
 \frac{\partial^2 \lambda_t(\theta)}{\partial d_l \partial B(i, j)} &= \frac{\partial^2 \lambda_t(\theta)}{\partial B(i, j) \partial d_l} = E(i, j) \frac{\partial \lambda_{t-1}(\theta)}{\partial d_l} + B \frac{\partial^2 \lambda_{t-1}(\theta)}{\partial B(i, j) \partial d_l} = \sum_{u \geq 0} B^u E(i, j) \frac{\partial \lambda_{t-u-1}(\theta)}{\partial d_l} \quad (3.49) \\
 \frac{\partial^2 \lambda_t(\theta)}{\partial A(l, v) \partial B(i, j)} &= \frac{\partial^2 \lambda_t(\theta)}{\partial B(i, j) \partial A(l, v)} = E(i, j) \frac{\partial \lambda_{t-1}(\theta)}{\partial A(l, v)} + B \frac{\partial^2 \lambda_{t-1}(\theta)}{\partial B(i, j) \partial A(l, v)} = \sum_{u \geq 0} B^u E(i, j) \frac{\partial \lambda_{t-u-1}(\theta)}{\partial A(l, v)} \\
 \frac{\partial^2 \lambda_t(\theta)}{\partial \Gamma(l, v) \partial B(i, j)} &= \frac{\partial^2 \lambda_t(\theta)}{\partial B(i, j) \partial \Gamma(l, v)} = E(i, j) \frac{\partial \lambda_{t-1}(\theta)}{\partial \Gamma(l, v)} + B \frac{\partial^2 \lambda_{t-1}(\theta)}{\partial B(i, j) \partial \Gamma(l, v)} = \sum_{u \geq 0} B^u E(i, j) \frac{\partial \lambda_{t-u-1}(\theta)}{\partial \Gamma(l, v)} \\
 \frac{\partial^2 \lambda_t(\theta)}{\partial B(l, v) \partial B(i, j)} &= \frac{\partial^2 \lambda_t(\theta)}{\partial B(i, j) \partial B(l, v)} = \sum_{u \geq 0} B^u \left(E(i, j) \frac{\partial \lambda_{t-u-1}(\theta)}{\partial B(l, v)} + E(l, v) \frac{\partial \lambda_{t-u-1}(\theta)}{\partial B(i, j)} \right)
 \end{aligned}$$

Straightforwardly, from Lemma 10, the compactness of Θ and Minkowski's inequality, all these partial derivatives have the same polynomial moments than \bar{Y}_0 and X_0 . \square

Lemma 13 (Approximation of the derivatives of the latent process) *There exist*

$\tau \in (0, 1)$ and $C > 0$ such that

1. $\sup_{\theta} |\lambda_t(\theta) - \bar{\lambda}_t(\theta)|_1 < C\tau^t(\sup_{\theta \in \Theta} |\lambda_0(\theta)|_1 + |\bar{\lambda}_0|_1)$
2. for $i = 1, \dots, Q$, $\sup_{\theta} |\partial_i \lambda_t(\theta) - \partial_i \bar{\lambda}_t(\theta)|_1 < C^2 t \tau^{t-1} (\sup_{\theta \in \Theta} |\lambda_0(\theta)|_1 + |\bar{\lambda}_0|_1) + C\tau^t \sup_{\theta \in \Theta} \left| \frac{\partial \lambda_0}{\partial B_{i,j}}(\theta) \right|$.

Proof of Lemma 13 One can notice that

$$|\lambda_t(\theta) - \bar{\lambda}_t(\theta)|_{vec} \preceq |B^t|_{vec} |\lambda_0(\theta) - \bar{\lambda}_0|_{vec}$$

and the first result follows from Lemma 10, 1. For $i, j = 1, \dots, k$, we can write $\frac{\partial \lambda_t(\theta)}{\partial B(i,j)}$ as

$$\frac{\partial \lambda_t(\theta)}{\partial B(i,j)} = \sum_{l=1}^{t-1} B^l E(i,j) \lambda_{t-l-1}(\theta) + B^t \frac{\partial \lambda_0(\theta)}{\partial B(i,j)} \quad \text{and} \quad \frac{\partial \bar{\lambda}_t(\theta)}{\partial B(i,j)} = \sum_{l=1}^{t-1} B^l E(i,j) \bar{\lambda}_{t-l-1}(\theta) + B^t \frac{\partial \bar{\lambda}_0}{\partial B(i,j)}.$$

And then

$$\begin{aligned} \sup_{\theta} \left| \frac{\partial \lambda_t(\theta)}{\partial B(i,j)} - \frac{\partial \bar{\lambda}_t(\theta)}{\partial B(i,j)} \right|_1 &\leq \sum_{l=1}^{t-1} \sup_{\theta} |B^l|_1 |E(i,j)|_1 |\lambda_{t-l-1}(\theta) - \bar{\lambda}_{t-l-1}(\theta)|_1 + |B^t|_1 \sup_{\theta \in \Theta} \left| \frac{\partial \lambda_0(\theta)}{\partial B(i,j)} \right| \\ &\leq tC^2 \tau^{t-1} (|\lambda_0|_1 + |\bar{\lambda}_0|_1) + C\tau^t \sup_{\theta \in \Theta} \left| \frac{\partial \lambda_0(\theta)}{\partial B_{i,j}} \right|. \end{aligned}$$

The control of the difference between the other partial derivatives is similar. \square

Lemma 14 Suppose that all the parameters in (3.46) are positives and that the processes $(X_t)_{t \in \mathbb{Z}}$ and $(\bar{Y}_t)_{t \in \mathbb{Z}}$ take nonnegative values. Suppose furthermore that there exists some $\delta \in (0, 1)$ such that $\mathbb{E} \left(|\bar{Y}_0|_1^{\delta} \right) < \infty$ and $\mathbb{E} \left(|X_0|_1^{\delta} \right) < \infty$. Then for any $r \geq 1$,

$$\mathbb{E} \left(\sup_{\theta} \left| \frac{1}{\lambda_0(\theta)} \frac{\partial \lambda_0(\theta)}{\partial \theta_i} \right|_1^r \right) < \infty \quad \text{and} \quad \mathbb{E} \left(\sup_{\theta} \left| \frac{1}{\lambda_0(\theta)} \frac{\partial^2 \lambda_0(\theta)}{\partial \theta_i \partial \theta_j} \right|_1^r \right) < \infty, \quad i, j = 1, \dots, Q.$$

Proof of Lemma 14 Note that there exists $d_- > 0$ such that for any $\theta \in \Theta$, we have $d_i \geq d_-$. Here again, we will denote by ι_{ℓ} , $\ell = 1, \dots, k$ the vector of $\{0, 1\}^k$ with 1 at ℓ -th position and 0 elsewhere. We also set $a = \min_{\theta \in \Theta} \min_{1 \leq i, j \leq k} A(i, j)$, $\gamma = \min_{\theta \in \Theta} \min_{1 \leq i, j \leq k} \Gamma(i, j)$ and $b = \min_{\theta \in \Theta} \min_{1 \leq i, j \leq k} B(i, j)$ which are positive constant from the positivity assumption and the compactness of Θ . Note that by positivity, all the entries of a matrix of type $B^l A$ are greater than the entries of $aB^l E(i, j)$. From equations

(3.48), we have the bounds

$$\begin{aligned}
 \frac{1}{\lambda_{\ell,t}(\theta)} \frac{\partial \lambda_{\ell,t}(\theta)}{\partial d_i} &\leq \frac{\iota'_\ell (I - B)^{-1} \iota_i}{d_-}; \quad \ell, i = 1, \dots, k, \\
 \frac{1}{\lambda_{\ell,t}(\theta)} \frac{\partial \lambda_{\ell,t}(\theta)}{\partial A(i, j)} &\leq \frac{\iota'_\ell \sum_{l \geq 0} B^l E(i, j) \bar{Y}_{t-l-1}}{a \iota'_\ell \sum_{l \geq 0} B^l E(i, j) \bar{Y}_{t-l-1}} \\
 &\leq \frac{1}{a}; \quad \ell, i, j = 1, \dots, k, \\
 \frac{1}{\lambda_{\ell,t}(\theta)} \frac{\partial \lambda_{\ell,t}(\theta)}{\partial \Gamma(i, j)} &\leq \frac{\iota'_\ell \sum_{l \geq 0} B^l E(i, j) X_{t-l-1}}{\gamma \iota'_\ell \sum_{l \geq 0} B^l E(i, j) X_{t-l-1}} \\
 &\leq \frac{1}{\gamma}; \quad \ell, i = 1, \dots, k, j = 1, \dots, m.
 \end{aligned}$$

For $t \in \mathbb{Z}$, set $d_t = d + A\bar{Y}_{t-1} + \Gamma X_{t-1}$. Note that $\frac{\partial \lambda_{\ell,t}(\theta)}{\partial B(i, j)} = \sum_{h \geq 1} \sum_{u=1}^h B^{u-1} E(i, j) B^{h-u} d_{t-h}$ and that the entries of $bB^{u-1} E(i, j) B^{h-u}$ are smaller than that of B^h . We then obtain

$$\begin{aligned}
 \frac{1}{\lambda_{\ell,t}(\theta)} \frac{\partial \lambda_{\ell,t}(\theta)}{\partial B(i, j)} &\leq \sum_{h \geq 1} h \frac{\iota'_\ell B^h d_{t-h}}{b \iota'_\ell d + b \iota'_\ell B^h d_{t-h}} \tag{3.50} \\
 &\leq \sum_{h \geq 1} h \left(\frac{\iota'_\ell B^h d_{t-h}}{b d_-} \right)^s; \quad \ell, i, j = 1, \dots, k,
 \end{aligned}$$

for any $s \in (0, 1)$. From Lemma 10, there exist $C > 0$ and $\tau \in (0, 1)$ such that $|B^h|_1 \leq C\tau^h$ for any positive integer h . We then obtain the bound

$$\mathbb{E}^{1/r} \left[\sup_{\theta \in \Theta} \left| \frac{1}{\lambda_{\ell,t}(\theta)} \frac{\partial \lambda_{\ell,t}(\theta)}{\partial B(i, j)} \right|^r \right] \leq \sum_{h \geq 1} h \frac{C^s \tau^{hs}}{(b^2 d_-)^s} \mathbb{E}^{1/r} \left(\sup_{\theta \in \Theta} |d_{t-h}|_1^{rs} \right).$$

Taking $s = \delta/r$ and using the bound $E \left[(\sup_{\theta} |d_1|)^\delta \right] \leq \sup_{\theta} |d_1|^\delta + \sup_{\theta} |A|_1^\delta E \left[|\bar{Y}_0|^\delta \right] + \sup_{\theta} |\Gamma|_1^\delta E \left[|X_0|^\delta \right]$, we get the integrability conditions for the first-order partial derivatives.

For the second-order partial derivatives, one can use the expressions (3.49) and replace the partial derivatives in the series by the expressions given in (3.48). With more tedious computations, one can use similar arguments as above to get the required integrability conditions. Details are omitted. \square

3.6.21 Numerical experiments

TABLE 3.3 – Average and Mean Square Errors for the estimators of the BIP model ($n = 1000$)

n = 1000		Log INGARCH					Logit Binary				
		d_1	$A(1, 1)$	$A(1, 2)$	$B(1, 1)$	$\Gamma(1, 1)$	d_2	$A(2, 1)$	$A(2, 2)$	$B(2, 2)$	$\Gamma(2, 1)$
r		1	0.3	0.3	0.15	-0.1	-1	0.4	-0.6	0.2	0.1
-0.9	-0.8995	1.0748	0.3091	0.3047	0.1571	-0.1001	-1.0251	0.4092	-0.6038	0.1958	0.1001
	(0.0001)	0.0252	0.0020	0.0008	0.0065	0.0001	0.3877	0.0603	0.0278	0.0331	0.0045
-0.75	-0.7463	1.0649	0.3074	0.3036	0.1634	-0.1001	-1.0175	0.4058	-0.6060	0.1952	0.0995
	(0.0004)	0.0223	0.0016	0.0007	0.0057	0.0001	0.2818	0.0451	0.0229	0.0350	0.0046
-0.6	-0.6006	1.0578	0.3063	0.3032	0.1678	-0.1001	-1.0101	0.4023	-0.6062	0.1909	0.0986
	(0.0008)	0.0192	0.0014	0.0006	0.0050	0.0001	0.2407	0.0392	0.0220	0.0387	0.0047
-0.45	-0.4560	1.0521	0.3056	0.3027	0.1711	-0.1002	-1.0084	0.4014	-0.6062	0.1899	0.0994
	(0.0010)	0.0168	0.0013	0.0005	0.0044	0.0001	0.2130	0.0358	0.0206	0.0401	0.0047
-0.3	-0.3037	1.0455	0.3051	0.3025	0.1745	-0.1002	-0.9983	0.3966	-0.6053	0.1860	0.0988
	(0.0014)	0.0147	0.0012	0.0005	0.0039	0.0001	0.1991	0.0348	0.0203	0.0443	0.0047
-0.15	-0.1607	1.0404	0.3049	0.3026	0.1769	-0.1002	-1.0016	0.3973	-0.6014	0.1862	0.0991
	(0.0017)	0.0132	0.0012	0.0005	0.0036	0.0001	0.1775	0.0318	0.0199	0.0485	0.0048
0	0.0093	1.0356	0.3050	0.3024	0.1790	-0.1002	-0.9926	0.3925	-0.6007	0.1811	0.0985
	(0.0022)	0.0111	0.0011	0.0005	0.0032	0.0001	0.1654	0.0312	0.0201	0.0557	0.0048
0.15	0.1534	1.0324	0.3053	0.3017	0.1802	-0.1002	-0.9931	0.3924	-0.6001	0.1792	0.0984
	(0.0013)	0.0099	0.0012	0.0005	0.0030	0.0001	0.1592	0.0308	0.0205	0.0596	0.0047
0.3	0.2985	1.0290	0.3054	0.3013	0.1816	-0.1002	-0.9876	0.3900	-0.6004	0.1789	0.0984
	(0.0013)	0.0089	0.0012	0.0005	0.0028	0.0001	0.1576	0.0315	0.0215	0.0650	0.0046
0.45	0.4492	1.0271	0.3067	0.3004	0.1813	-0.1004	-0.9834	0.3875	-0.5990	0.1771	0.0978
	(0.0011)	0.0080	0.0014	0.0005	0.0028	0.0001	0.1521	0.0317	0.0226	0.0720	0.0046
0.6	0.5991	1.0251	0.3076	0.2994	0.1814	-0.1003	-0.9770	0.3839	-0.5947	0.1745	0.0979
	(0.0008)	0.0073	0.0015	0.0006	0.0028	0.0001	0.1489	0.0333	0.0255	0.0806	0.0046
0.75	0.7493	1.0236	0.3100	0.2977	0.1800	-0.1003	-0.9759	0.3825	-0.5921	0.1718	0.0984
	(0.0004)	0.0070	0.0018	0.0007	0.0032	0.0001	0.1528	0.0351	0.0273	0.0898	0.0046
0.9	0.9012	1.0237	0.3135	0.2955	0.1767	-0.1002	-0.9708	0.3798	-0.5872	0.1746	0.0991
	(0.0001)	0.0066	0.0022	0.0008	0.0038	0.0001	0.1557	0.0370	0.0295	0.0999	0.0047

TABLE 3.4 – Average and MSE for the estimators of the BIP model ($n = 500$)

n = 500		Log INGARCH					Logit Binary				
		d_1	$A(1, 1)$	$A(1, 2)$	$B(1, 1)$	$\Gamma(1, 1)$	d_2	$A(2, 1)$	$A(2, 2)$	$B(2, 2)$	$\Gamma(2, 1)$
r		1	0.3	0.3	0.15	-0.1	-1	0.4	-0.6	0.2	0.1
-0.9	-0.9001	1.1030	0.3193	0.3103	0.1329	-0.1002	-0.9645	0.3868	-0.6315	0.1896	0.1057
	(0.0002)	0.0479	0.0042	0.0017	0.0090	0.0002	0.7927	0.1230	0.0670	0.0707	0.0096
-0.75	-0.7494	1.0940	0.3157	0.3080	0.1410	-0.1002	-0.9683	0.3888	-0.6251	0.1980	0.1061
	((0.0008))	0.0410	0.0032	0.0013	0.0076	0.0002	0.5873	0.0923	0.0546	0.0720	0.0095
-0.6	-0.5955	1.0880	0.3134	0.3064	0.1465	-0.1003	-0.9697	0.3886	-0.6260	0.1932	0.1051
	(0.0017)	0.0368	0.0027	0.0011	0.0068	0.0002	0.4670	0.0769	0.0469	0.0753	0.0096
-0.45	-0.4482	1.0810	0.3125	0.3052	0.1507	-0.1001	-0.9730	0.3902	-0.6239	0.1958	0.1042
	(0.0026)	0.0322	0.0024	0.0009	0.0060	0.0002	0.3910	0.0655	0.0438	0.0789	0.0097
-0.3	-0.3059	1.0751	0.3121	0.3044	0.1538	-0.1000	-0.9747	0.3913	-0.6244	0.1958	0.1045
	(0.0025)	0.0286	0.0022	0.0009	0.0055	0.0002	0.3532	0.0611	0.0413	0.0830	0.0097
-0.15	-0.1513	1.06798	0.3125	0.3031	0.1568	-0.1001	-0.9826	0.3949	-0.6205	0.1992	0.1042
	(0.0034)	0.0253	0.0022	0.0008	0.0052	0.0002	0.3325	0.0583	0.0400	0.0876	0.0098
0	0.0001	1.0644	0.3127	0.3020	0.1584	-0.1001	-0.9851	0.3956	-0.6201	0.1979	0.1037
	(0.0032)	0.0225	0.0023	0.0008	0.0050	0.0002	0.3084	0.0567	0.0400	0.0935	0.0099
0.15	0.1519	1.0607	0.3140	0.3004	0.1590	-0.1000	-0.9833	0.3947	-0.6153	0.2024	0.1038
	(0.0039)	0.0204	0.0024	0.0008	0.0048	0.0002	0.3007	0.0560	0.0413	0.1059	0.0097
0.3	0.2978	1.0553	0.3154	0.2989	0.1602	-0.1000	-0.9928	0.3998	-0.6165	0.2045	0.1033
	(0.0027)	0.0186	0.0026	0.0009	0.0048	0.0002	0.2912	0.0557	0.0419	0.1109	0.0095
0.45	0.4550	1.0533	0.3179	0.2971	0.1589	-0.0998	-0.9873	0.3969	-0.6182	0.2024	0.1020
	(0.0025)	0.0173	0.0029	0.0010	0.0049	0.0002	0.2945	0.0584	0.0445	0.1205	0.0096
0.6	0.6021	1.0524	0.3218	0.2944	0.1559	-0.0999	-0.9883	0.3979	-0.6207	0.2031	0.1018
	(0.0015)	0.0161	0.0033	0.0012	0.0050	0.0002	0.3080	0.0622	0.04675	0.1220	0.0095
0.75	0.7476	1.0503	0.3263	0.2905	0.1528	-0.0997	-0.9706	0.3866	-0.6141	0.1863	0.1009
	(0.0009)	0.0153	0.0039	0.0014	0.0053	0.0002	0.3362	0.0726	0.0525	0.1386	0.0097
0.9	0.9008	1.048	0.3318	0.2856	0.1491	-0.0996	-0.9544	0.3819	-0.6128	0.1956	0.1004
	(0.0002)	0.0148	0.0049	0.0017	0.0058	0.0002	0.3423	0.0772	0.0595	0.1504	0.0095

TABLE 3.5 – Average and MSE for the estimators of the GAIN model ($n = 1000$)

n = 1000		GARCH				INGARCH			
		d_1	$A(1, 1)$	$A(1, 2)$	$B(1, 1)$	d_2	$A(2, 1)$	$A(2, 2)$	$B(2, 2)$
r		0.03	0.05	0.05	0.7	0.3	0.3	0.1	0.5
-0.9	-0.8995	0.0395	0.0514	0.0534	0.6593	0.3270	0.3088	0.0992	0.4703
	(0.0001)	0.0005	0.0008	0.0002	0.0101	0.0094	0.0058	0.0009	0.0133
-0.75	-0.7463	0.0405	0.0498	0.0540	0.6543	0.3291	0.3067	0.1008	0.4670
	(0.0004)	0.0006	0.0008	0.0002	0.01067	0.0098	0.0058	0.0009	0.0142
-0.6	-0.6006	0.0407	0.0504	0.0542	0.6539	0.3304	0.3123	0.0991	0.4669
	(0.0008)	0.0006	0.0008	0.0002	0.0105	0.0096	0.0058	0.0009	0.0137
-0.45	-0.4560	0.0415	0.0493	0.0537	0.6538	0.3275	0.3094	0.1016	0.4685
	(0.0010)	0.0006	0.0008	0.0002	0.0102	0.0099	0.0061	0.0009	0.0145
-0.3	-0.3037	0.0428	0.0514	0.0536	0.6462	0.3252	0.3107	0.1026	0.4693
	(0.0014)	0.0009	0.0008	0.0002	0.0151	0.0101	0.0063	0.0009	0.0147
-0.15	-0.1607	0.0416	0.0526	0.0525	0.6532	0.3340	0.3081	0.0988	0.4646
	(0.0017)	0.0007	0.0008	0.0002	0.0122	0.0118	0.0057	0.0009	0.0162
0	0.0093	0.0407	0.0523	0.0536	0.6532	0.3357	0.3096	0.1029	0.4578
	(0.0022)	0.0006	0.0008	0.0002	0.0112	0.0100	0.0058	0.0009	0.0146
0.15	0.1534	0.0411	0.0515	0.0527	0.6542	0.3295	0.3135	0.0982	0.4692
	(0.0013)	0.0006	0.0008	0.0002	0.0117	0.0109	0.0052	0.0009	0.0147
0.3	0.2985	0.04308	0.0518	0.0534	0.6475	0.3266	0.3121	0.0963	0.4729
	(0.0013)	0.0007	0.0008	0.0002	0.0119	0.0107	0.0057	0.0009	0.0154
0.45	0.4492	0.0404	0.0511	0.0531	0.6571	0.3371	0.3101	0.1008	0.4591
	(0.0011)	0.0007	0.0008	0.0002	0.0115	0.0110	0.0059	0.0010	0.0156
0.6	0.5991	0.0409	0.0526	0.0537	0.6511	0.3247	0.3092	0.1004	0.4712
	(0.0008)	0.0006	0.0008	0.0002	0.0116	0.0088	0.0053	0.0010	0.0120
0.75	0.7493	0.0413	0.0514	0.0532	0.6525	0.3378	0.3097	0.1009	0.4574
	(0.0004)	0.0007	0.0008	0.0002	0.0121	0.0110	0.0052	0.0010	0.0157
0.9	0.9012	0.0410	0.0530	0.0539	0.6510	0.3214	0.3016	0.1001	0.4777
	(0.0001)	0.0005	0.0008	0.0002	0.0098	0.0080	0.0057	0.0010	0.0115

TABLE 3.6 – Average and MSE for the estimators of the GAIN model ($n = 500$)

$n = 500$		GARCH				INGARCH			
		d_1	$A(1, 1)$	$A(1, 2)$	$B(1, 1)$	d_2	$A(2, 1)$	$A(2, 2)$	$B(2, 2)$
r		0.03	0.05	0.05	0.7	0.3	0.3	0.1	0.5
-0.9	-0.8841	0.0497	0.0530	0.0558	0.6173	0.3512	0.3090	0.0974	0.4472
	(0.0004)	0.0019	0.0016	0.0004	0.0297	0.0214	0.0125	0.0022	0.0291
-0.75	-0.7213	0.0517	0.0539	0.0575	0.6022	0.3658	0.3204	0.1029	0.4223
	(0.0013)	0.0019	0.0018	0.0004	0.0346	0.0247	0.0113	0.0021	0.0348
-0.6	-0.5731	0.0539	0.0511	0.0563	0.6045	0.3627	0.3135	0.0980	0.4342
	(0.0015)	0.0022	0.0017	0.0004	0.0333	0.0260	0.0118	0.0021	0.0333
-0.45	-0.4238	0.0514	0.0541	0.0585	0.6040	0.3654	0.3139	0.1026	0.4261
	(0.0019)	0.0021	0.0017	0.0004	0.0349	0.0260	0.0118	0.0019	0.0335
-0.3	-0.2828	0.0532	0.0508	0.0564	0.6077	0.3589	0.3139	0.1003	0.4373
	(0.0020)	0.0022	0.0015	0.0003	0.0321	0.0263	0.0110	0.0020	0.0337
-0.15	-0.1544	0.0529	0.0535	0.0559	0.6058	0.35308	0.3188	0.1003	0.4391
	(0.0017)	0.0022	0.0018	0.0004	0.0328	0.0231	0.0110	0.0021	0.0323
0	-0.0030	0.0540	0.0571	0.0555	0.6010	0.3611	0.3129	0.1005	0.4351
	(0.0020)	0.0024	0.0019	0.0004	0.0364	0.0260	0.0126	0.0019	0.0330
0.15	0.1469	0.0528	0.0508	0.0555	0.6110	0.3733	0.3106	0.0969	0.4263
	(0.0016)	0.0024	0.0015	0.0003	0.0325	0.0309	0.0105	0.0019	0.0393
0.3	0.2866	0.0531	0.0486	0.0543	0.6135	0.3479	0.3148	0.0964	0.4477
	(0.0017)	0.0023	0.0016	0.0003	0.0331	0.0237	0.0117	0.0018	0.0308
0.45	0.4233	0.0533	0.0548	0.0559	0.6032	0.3641	0.3186	0.0985	0.4288
	(0.0019)	0.0025	0.0018	0.0004	0.0366	0.0261	0.0117	0.0019	0.0353
0.6	0.5699	0.0489	0.0548	0.0567	0.6149	0.3537	0.3099	0.0988	0.4442
	(0.0017)	0.0017	0.0017	0.0004	0.0285	0.0222	0.0113	0.0019	0.0296
0.75	0.7210	0.0520	0.0551	0.0557	0.6075	0.3692	0.3188	0.1014	0.4226
	(0.0012)	0.0022	0.0018	0.0004	0.0341	0.0257	0.0115	0.0023	0.0344
0.9	0.8835	0.0509	0.0568	0.0554	0.6099	0.3489	0.3168	0.0997	0.4444
	(0.0004)	0.0018	0.0020	0.0004	0.0301	0.0205	0.0136	0.0019	0.0294

AUTOREGRESSIVE MODELS FOR TIME SERIES OF RANDOM SUMS OF POSITIVE VARIABLES : APPLICATION TO TREE GROWTH AS A FUNCTION OF CLIMATE AND INSECT OUTBREAKS

Abstract. We present a broad class of semi-parametric models for time series of random sums of positive variables. Our methodology allows the number of terms inside the sum to be time-varying and is therefore well suited to many examples encountered in the natural sciences. We study the stability properties of the models and provide a valid statistical inference procedure to estimate the model parameters. It is shown that the proposed quasi-maximum likelihood estimator is consistent and asymptotically gaussian distributed. This work is complemented by simulation results and applied to time series representing growth rates of white spruce (*Picea glauca*) trees from a few dozen sites in Quebec spanning 41 years, including one major spruce budworm (*Choristoneura fumiferana*) outbreak between 1968 and 1991. We found significant growth reductions due to budworm-induced defoliation up to two years in the past. Our results also revealed positive effects of maximum temperature, precipitation and the climate moisture index in the summer. Negative effects of the climate moisture index in the spring and the maximum temperature in the previous summer are also revealed. However, considering the interaction between climate and defoliation on growth did not improve the model's performance on this dataset. This study represent a major advance and our result represent useful tool in understanding the combined effects of climate and insect defoliation on tree growth in the face of climate change, where the frequency and the severity of outbreaks coupled with an increase of temperature is expected.

Long version of [34], Submitted for publication in Ecological Modelling

4.1 Introduction

In many studies in ecology, we measure positive dependent variables of random numbers of statistical individuals sampled over time [104]. There are two main reasons for this : first, researchers cannot observe the whole population ; second, the individuals that researchers can observe depends on their time-varying resources. Examples range from species behaviour to ecological services. For instance, in forestry, one can be interested in time series representing the mass or size of certain species, taking a random sample of trees each year and observing the corresponding quantity (see Vourlitis et al. [130] for instance). Another example consists in the occupied land area of colonies in relationship with the available resources across time [85]. In fisheries, scientists often investigate the temporal changes in the weight of fish caught, see Chan et al. [20] for more details.

In this paper, we contribute to the ongoing ecological study of the impact of climate change and insect outbreaks on tree growth measured by growth rings. Spruce budworm (*Choristoneura fumiferana* ; SBW) outbreak is the most important defoliator of conifer trees in the North American boreal forest [103]. Only in the province of Quebec (Canada), the forest surface affected by this species of Lepidoptera over the last century is twice the size of Ukraina [108]. At the epidemic stage, massive populations of larvae cause widespread damage to tree foliage [87]. SBW affect the main conifer boreal species in Canada, balsam fir (*Abies balsamea*), white spruce (*Picea glauca*) and black spruce (*Picea Mariana*). For this reason, SBW has a major impact in the regeneration and dynamics of boreal forest [97]. However, SBW outbreaks periods not only have major ecological impact but in addition produce important economic consequences due to the loss of forest productivity.

Previous works in this field have studied the changes of forest composition following insect outbreaks (see for instance Morin et al. [105]), the response of SBW outbreaks to climate change (Fleming and Volney [49] and Berguet et al. [8]) and the demography *i.e* the rate of mortality of spruce during outbreaks [62]. However, even with the major implication stemming from climate change, there is limited knowledge regarding the combined effects of outbreaks and climate on tree growth. Given that temperature variation and precipitation affect organisms' survival, reproduction cycles and spatial dispersion [1], it is critical to understand the links between past SBW outbreaks, the climate and tree

growth to understand how future climate change scenarios would impact forest productivity during outbreaks [81]. This is a major concern due to the increase in the severity and the frequency expected in the future for SBW outbreaks [108, 117].

In this paper, we contribute to filling this gap by proposing a broad class of semi-parametric models for positive-valued time series. Indeed, even though this type of data is common in forestry, the statistical approaches commonly used suffer from several drawbacks. These approaches range from descriptive exploratory techniques to linear mixed-effect models with time-varying variables on transformed data (see Montoro Girona et al. [101] and Boulanger and Arseneault [12] for example) and correlated error terms [64]. Whereas the former (exploratory techniques) do not allow us to draw inferences from the data, we can note at least two limits for the latter approach. First, as pointed out by several papers (see for example Chou et al. [21]) specifying a linear model on transformed data often leads to bad performance in term of prediction. Indeed, even though applying the log transformation for instance makes the positive-valued data more normal, the predicted value obtained this way underestimates the expected value due to Jensen's inequality. Second, models with autocorrelated error terms do not take into account the complex dependent structure of tree-ring growth. The class of semi-parametric autoregressive models we present here will be applied to investigate the relationship between climate, insect outbreaks and growth of white spruce. It also presents the advantage of accommodating the repeated measures design.

Many previous works have focused on modelling non-Gaussian time series, such as positive-valued processes. Indeed, Gaussian processes can be represented as linear models, whereas time series of count or binary data are modelled by non-linear dynamics, see for example Sim [118], Weiß [132] or Davis et al. [27] and references therein. For positive-valued time series data, the range volatility model was proposed by Engle and Russell [45] as an alternative for *garch* models in finance and its use has been rapidly expanding due to its various applications. We refer the interested reader to the review by Chou et al. [21]. Recently, Aknouche and Francq [3] have considered a positive-valued time series whose conditional distribution has a time-varying mean that can depend on exogenous variables. Our approach here is slightly different from theirs, since the positive process under consideration is itself the sum of a random number of other positive variables. It is strongly driven by the data we have to deal with. Indeed, these data consist of multiple time series collected over several ecological sites, where the number of individuals sampled changes over time as well as across sites. Hence, considering an aggregate value like the sum

or the mean of growth rings lead to the loss of variability linked to the sampling scheme. Moreover, one can note that in different fields like finance, some modelling strategies consisting in considering empirical quantities such as the realized volatility are employed. Historical returns of investment products within a defined time period are then analysed (see for example Allen et al. [4]). However, unlike our framework which is typical in ecological studies, all transactions on investment products are recorded (*i.e* the whole statistical population is observed).

The rest of this paper is organised as follows. In Section 4.2, we define the model used throughout this paper and discuss our modelling choice. Time-series properties of the models are also studied in that section. Maximum-likelihood based inference and its asymptotic properties are presented in Section 4.3. Section 4.4 contains a small simulation study and an application to empirical data on the growth of white spruce. All auxiliary lemmas and mathematical proofs are contained in Appendix.

4.2 Models and stability results

We introduce here a generalized linear dynamic model for time series of random sums of positive variables, motivated by our empirical application where we analyze the annual growth of spruce trees subject to climate variation and outbreaks of SBW. In this case, growth is measured by taking cores at 1.30 m height from the trunk of a sample of trees in a forest [102]. The samples were prepared, measured and analyzed conforming to standard dendroecological protocol [83]. Cores were air-dried, mounted on wood boards and sanded before tree rings were measured with a WinDendro system [67] or a manual Henson micrometer with an accuracy of 0.01 mm. The tree-ring series measurements covered the last 41 years, and were cross-dated using TSAP-Wi (Rinntech, Heidelberg, Germany).

We denote by $Y_{k,t}, t \in \mathbb{Z}, k = 1, \dots, K$ the time series of the total basal area increment related to the k -th observational site, *i.e.* the sum of increases in the trunk cross-sectional area for the $n_{k,t}$ trees sampled for site k on year t . We aim to model the dynamics of this process both in terms of its own past and in presence of m additional covariates $X_{k,t} \in \mathbb{R}^m$. In the empirical application presented in section 4.4, the covariate process encompasses climate variables such as temperature and precipitation, as well as the level of defoliation due to SBW in previous years. Our model is given by :

$$Y_{k,t} = \sum_{l=1}^{n_{k,t}} \zeta_{l,k,t} \quad (4.1)$$

where conditionally on $n_{k,t}, X_{k,t}, n_{k,t}^- = (n_{k,t-s}, s \geq 1)$ and $Y_{k,t}^- = (Y_{k,t-s}, s \geq 1)$, the variables $\zeta_{l,k,t}$, $1 \leq l \leq n_{k,t}$, representing the basal area increments of individual sampled trees, are identically distributed as a random variable $\zeta_{k,t}$ of mean $\lambda_{k,t}$. Moreover, $(n_{k,t})_{t \in \mathbb{Z}}$ is a sequence of *i.i.d* random variables where conditionally on $n_{k,t}^-$, the variable $n_{k,t}$ is independent from $X_{k,t}$ and $Y_{k,t}^-$. The mean process is given by

$$\varphi_\delta(\lambda_{k,t}) =: \eta_{k,t} = \omega_k + \sum_{j=1}^p \alpha_j \frac{Y_{k,t-j}}{n_{k,t-j}} + \beta^\top X_{k,t}, \quad k = 1, \dots, K \text{ and } t = 1, \dots, T, \quad (4.2)$$

such that $\omega_k \in \mathbb{R}, \alpha_j \in \mathbb{R}, \beta = (\beta_1, \dots, \beta_m) \in \mathbb{R}^m$ and φ_δ is a real-valued function defined on \mathbb{R}_+ that can depend on a parameter δ . It is worth mentioning, without loss of generality, that the covariate process considered at time t is included in the specification of $\lambda_{k,t}$ since multiple lags of a given set of variables can be included by simply stacking them into a vector. An example is the case of defoliation levels as in our application, since growth can be affected by defoliation up to 5 years prior (from $t - 5$ to $t - 1$).

The variables $\zeta_{k,t}$ will be referred to as the unity random variables. We do not make any assumption about the distribution of the variables $\zeta_{k,t}$. Any distribution on $(0, +\infty)$ can be chosen. Some examples includes Exponential distribution with parameter $1/\lambda_{k,t}$, log-Normal distribution with parameters $\log \lambda_{k,t} - \sigma^2/2$ and σ or a Gamma distribution with parameters $\alpha \lambda_{k,t}$ and α , to name a few. Whatever the distributions of unity random variables are, the conditional expectation of $Y_{k,t}$ is $n_{k,t} \lambda_{k,t}$. However, under the assumption of the independence of $\zeta_{l,k,t}$, $1 \leq l \leq n_{k,t}$, if they are exponentially distributed, the conditional variance is $n_{k,t} \lambda_{k,t}^2$ *i.e* a quadratic function of $\lambda_{k,t}$. In our example of Gamma-distributed unity random variables, the conditional variance is $n_{k,t} \lambda_{k,t} / \alpha$, *i.e.* a linear function of $\lambda_{k,t}$. But in the case of the log-Normal distribution, the conditional variance is $n_{k,t} \lambda_{k,t}^4 (\exp \sigma^2 - 1)$. With our semi-parametric framework, we will only focus on the estimation of regression parameters $\theta = (\delta, \omega_1, \dots, \omega_K, \alpha_1, \dots, \alpha_p, \beta^\top)^\top$ without the need to perform any distributional goodness of fit test.

Copies of unity variables. In our general set up, the copies $\zeta_{l,k,t}$, $1 \leq l \leq n_{k,t}$ of the unity random variables $\zeta_{k,t}$ are not required to be independent. In practical applications where $\zeta_{l,k,t}$ represents the measure of annual growth for a sampled tree for example, the

general assumption of being identical distributed can be thought as a local stationary condition inside the site k at time t .

Marginal stationary distributions. Note from equations (4.1)-(4.2),

$Y_{k,t} = f_{\theta_k}(X_{k,t-s}, n_{k,t-s}, \zeta_{\ell,k,t-s}, s \geq 0, \ell \geq 1)$ for $\theta_k = (\delta, \omega_k, \alpha_1, \dots, \alpha_p, \beta^\top)^\top$. Then for $1 \leq k_1 \neq k_2 \leq K$, the distributions of $(Y_{k_1,0}, n_{k_1,0}, X_{k_1,0})$ and $(Y_{k_2,0}, n_{k_2,0}, X_{k_2,0})$ are not equal unless $\omega_{k_1} = \omega_{k_2}$ and $(X_{k_1,0}, \zeta_{\ell,k_1,0}, \ell \geq 1)$ is equal in distribution to $(X_{k_2,0}, \zeta_{\ell,k_2,0}, \ell \geq 1)$. We will investigate in section 4.4 the consequences of the latter conditions on the proposed estimation procedure.

Regression function (4.2). Note that $\eta_{k,t}$ in (4.2) does not depend linearly on $Y_{k,t-i}, i = 1, \dots, p$, but on $Y_{k,t-i}/n_{k,t-i}, i = 1, \dots, p$. In fact, through (4.2), we make a link between the underlying mean process and the empirical estimate of the past mean process. Even for a constant size process, *i.e.* $n_{k,t} = n_k, \forall t$, since the regression parameter $\alpha_i, i = 1, \dots, p$ is free of k , we still cannot yet express $\eta_{k,t}$ as a linear combination of $Y_{k,t-i}, i = 1, \dots, p$. Moreover, one can expect $Y_{k,t-i}/n_{k,t-i} - \lambda_{k,t-i}, i = 1, \dots, p$ or more generally $h(Y_{k,t-i}/n_{k,t-i}, \lambda_{k,t-i}), i = 1, \dots, p$ for some mapping h such that $\mathbb{E}h(Y_{k,t-i}/n_{k,t-i}, \lambda_{k,t-i}) = 0$ in (4.2) at the place of $Y_{k,t-i}/n_{k,t-i}, i = 1, \dots, p$. Indeed, with the latter two mentioned specifications, (4.1)-(4.2) define the so-called *GLARMA* model (see for example, [132] for more details). In the present form, (4.1)-(4.2) has some similarities with the well known *ARCH* model [9] and *INARCH* models [132]. We leave the topic of *GLARMA* specification for further work.

Contrast with the non-linear mixed model . The model (4.1)-(4.2) has some similarities with the well-known mixed models. Indeed, as for mixed models, ω_k stands for the site fixed effect where the random effect is embedded in the distribution of unity variables. The simple example of $\zeta_{l,k,t} = \lambda_{k,t} \epsilon_{l,k,t}$, where $(\epsilon_{l,k,t})_{l \geq 1}$ is a sequence of identically distributed random variables of mean 1, fit with the so-called class of multiplicative random effect models [18]. But more complex random effects can be handled. However, the model (4.1)-(4.2) is more general since it allows the individuals sampled over time to change. Indeed, as we will see in section 4.3, the individual measurements are no longer needed when the sequence $(Y_{k,t}, n_{k,t})$ is available. Also, in terms of the application to resource management, it is often of interest to model and predict a population quantity like the sum of basal area growth in a forest.

Choice of the link function φ

The logarithmic link function is often applied and coincides with the well known log-linear model, see for example Cameron and Trivedi [18] for models for count data. This link function assumes a linear relationship between the logarithm of the mean process and the covariates. However, there exist some other link functions that preserve the linear correlation at least on the positive part of \mathbb{R} . Consider for example, the threshold mapping $x \mapsto \max(x, 0)$. This mapping is not smooth and most of the time, one makes some restrictions on model parameters to directly obtain the positiveness of the mean. Here, we will apply the inverse of the so-called softplus function as a link function. Indeed, the softplus function (see Glorot et al. [65]) is interesting for two reasons. The first one related to modelling is that it preserves the linearity on the positive part of real line. As shown by Weiß et al. [133] for count time series modeling, the models defined with the softplus link function are quite close to the truly linear model. This is also pertinent for our biological application, as we expect a linear effect of covariates on growth above a certain threshold representing the minimal favorable conditions for growth. The minimum growth expected may not be exactly zero, which is why we will later consider a slightly different version of softplus that we will refer to as softplus_δ for $\delta > 0$ defined as $\text{softplus}_\delta(x) = \log(1 + \delta + \exp(x))$. The second one and technical advantage is that the mapping softplus_δ is infinitely differentiable. Figure 4.1 in the Appendix shows the difference between the softplus_δ link function and $\max(x, 0)$ where softplus stands for softplus_0 . One can note that softplus_δ is lower bounded by $\log(1 + \delta)$. It is also worth noting that our generalization of softplus function differs from that of Mei and Eisner [99] and Weiß et al. [133] (eq. 3.2). As noted by the latter, it is possible to mimic the behavior of the Tobit model with the softplus generalization of Mei and Eisner [99]. In contrast, we aim here to lower bound the softplus link function by a non-zero constant since we request a minimum basal area increment at any time.

Model Interpretation. Obviously, with the softplus_δ link function, the mean process increases with the j -th covariate process if $\beta_j > 0$ and decreases with this one when $\beta_j < 0$. Since $\text{softplus}_\delta(x) \sim_\infty x$, the mean process can be approximated by the identity mapping. Therefore all other things remaining equal, the regression function is similar to $\beta_j X_{j,t}$ for large values of $X_{j,t}$ and $\beta_j > 0$ and then increases by $\beta_j \alpha$ for increasing values α of $X_{j,t}$. Let us denote by $\text{RG}(x, y)$, the relative rate of growth of the mean process between x and y *i.e.* $\text{RG}_\delta(x, y) = \gamma_\delta(x)/\gamma_\delta(y)$ where γ_δ is the derivative function of

softplus $_{\delta}$. For $\beta < 0$, $\lim_{x \rightarrow \infty} \text{RG}_{\delta}(\beta(x+\alpha), \beta x) = e^{\beta\alpha}$. Therefore, the rate toward $\log(1+\delta)$ driven by $X_{j,t}$ is given by $e^{\beta_j\alpha}$ when $\beta_j < 0$. Moreover, When $\delta \sim 0$, by l'Hôpital's rule $\lim_{x \rightarrow \infty} \text{RG}_0(x, y) = \lim_{x \rightarrow \infty} \text{softplus}_0(\beta(x+\alpha))/\text{softplus}_0(\beta x) = e^{\beta\alpha}$. Therefore, all other things remaining equal, the mean process will be divided by $e^{-\beta\alpha}$ when $X_{j,t}$ increases by α for large values of $X_{j,t}$ and $\beta_j < 0$.

Theorem 14 gives some stability conditions of model (4.1)-(4.2) with the inverse of the softplus function as the link whereas Lemma 15 in the Appendix stands for a general result for φ .

Theorem 14 *Under the assumptions (ST.1)-(ST.2) in Appendix and $\sum_{j=1}^p |\alpha_j| < 1$, there exists a unique set of K stationary, ergodic sequences $(Y_{k,t}, n_{k,t}, X_{k,t}), k = 1, \dots, K$ that are the solution of equations (4.1)-(4.2) with $\mathbb{E}|\eta_{k,0}| < \infty, k = 1, \dots, K$.*

4.3 Estimation and asymptotics properties

This section is devoted to the estimation of the conditional mean parameters by the Quasi-Maximum Likelihood Estimator (QMLE) based on a member of the exponential family. We consider the Exponential QMLE (EQMLE) because this estimator coincides with the Maximum Likelihood Estimator (MLE) when the unity random variables follow the Exponential $\Gamma(1, \lambda_{k,t}^{-1})$ distribution and the copies $\zeta_{l,k,t}, 1 \leq l \leq n_{k,t}$ are independent.

For our application, the K time series are observed between the time points 1 and T . We provide an asymptotic theory for the estimated parameters and present the results of a small simulation study investigating the finite-sample properties of the estimator. In the following, we will make $\lambda_{k,t}$ depend on the parameter $\theta (\in \Theta$ a compact set); that is

$$\log(\exp \circ \lambda_{k,t}(\theta) - 1 - \delta) = \omega_k + \sum_{j=1}^p \alpha_j \frac{Y_{k,t-j}}{n_{k,t-j}} + \beta^{\top} X_{k,t} =: \eta_{k,t}(\theta), \quad k = 1, \dots, K \text{ and } t = 1, \dots, T,$$

where $\delta \geq \delta_{-} > 0$. Let us denote the true, data-generating parameter value by θ_0 .

The loss function from the Exponential quasi-maximum likelihood is given by :

$$r_T(\theta) = \sum_{k=1}^K T^{-1} \sum_{t=1}^T \left(\frac{Y_{k,t}}{\lambda_{k,t}(\theta)} + n_{k,t} \log \circ \lambda_{k,t}(\theta) \right) =: \sum_{k=1}^K T^{-1} \sum_{t=1}^T \ell_{k,t}(\theta) =: \sum_{k=1}^K \ell_k(\theta) \quad (4.3)$$

and

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmin}} r_T(\theta). \quad (4.4)$$

The derivative of $\lambda_{k,t}(\theta)$ with respect to θ is given by :

$$\begin{aligned} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta} &=: \dot{\lambda}_{k,t}(\theta) \\ &= \left(\frac{1}{1 + \delta + e^{\eta_{k,t}(\theta)}}, \frac{e^{\eta_{k,t}(\theta)}}{1 + \delta + e^{\eta_{k,t}(\theta)}} \left(\iota_k, \frac{Y_{k,t-1}}{n_{k,t-1}} \dots \frac{Y_{k,t-p}}{n_{k,t-p}}, X_{k,t}^\top \right) \right)^\top. \end{aligned}$$

where ι_k is a vector of size K with 1 at the k -th position and 0 elsewhere. We will denote by $\dot{\lambda}_{k,t}$ (resp. $\lambda_{k,t}$) the vector $\dot{\lambda}_{k,t}(\theta)$ (resp. $\lambda_{k,t}(\theta)$) evaluated at the point $\theta = \theta_0$.

We will study the asymptotic properties of the QMLE estimator (4.4). To do so, we employ Taniguchi and Kakizawa [126] (Thm 3.2.23), which was extended in Klimko and Nelson [82]. The lemmas in our Appendix give the general result for the asymptotic properties of QMLE (4.4). The following theorem stands for the consistency and the asymptotic normality of (4.4) for softplus $_\delta$ link function. Let us set

$$V_k = \mathbb{E} \left[\frac{1}{\lambda_{k,0}^2} \left(n_{k,0} - \frac{Y_{k,0}}{\lambda_{k,0}} \right)^2 \dot{\lambda}_{k,0} \dot{\lambda}_{k,0}^\top \right] \text{ and } J_k = \mathbb{E} \left[n_{k,0} \frac{1}{\lambda_{k,0}^2} \dot{\lambda}_{k,0} \dot{\lambda}_{k,0}^\top \right].$$

Theorem 15 *Suppose that the assumptions (C.1)-(C.4) in Appendix are met. Then, almost surely,*

$$\lim_{T \rightarrow \infty} \hat{\theta}_T = \theta_0.$$

If in addition (AN.1)-(AN.3) hold true and θ_0 is located in the interior of Θ ,

$$\lim_{T \rightarrow \infty} \sqrt{T}(\hat{\theta}_T - \theta_0) = \mathcal{N}(0, J^{-1} V J^{-\top})$$

where $J = \sum_{k=1}^K J_k$ and $V = \sum_{k=1}^K V_k$.

Portmanteau-type tests for diagnostic checking

In this section, we will test the adequacy of the model (4.1) and (4.2) and the set of assumptions of the theorem 15. To do so, one will look at the residual autocovariances

$$\hat{\rho}_{k,h} = \frac{1}{T} \sum_{t=h+1}^T \hat{s}_{k,t} \hat{s}_{k,t-h}, \quad \hat{s}_{k,t} = \frac{Y_{k,t}}{\lambda_{k,t}(\hat{\theta}_T)} - n_{k,t}$$

where $h < T$.

For a fixed integer $q, 1 \leq q < T$, consider the statistic $\hat{\rho}_{1:q} = (\hat{\rho}_{1,1:q}, \dots, \hat{\rho}_{K,1:q})$ and for $k = 1, \dots, K, \hat{\rho}_{k,1:q} = (\hat{\rho}_{k,1}, \dots, \hat{\rho}_{k,q})$. Let \hat{V} and \hat{J} the empirical counterparts of the V and J . We also define

$$\hat{\eta}_k = T^{-1} \sum_{t=1}^T n_{k,t}, \quad \hat{\mu}_k = T^{-1} \sum_{t=1}^T \left(\frac{Y_{k,t}}{\lambda_{k,t}(\hat{\theta}_T)} - n_{k,t} \right)^2, \quad \hat{c}_{k,h} = - \sum_{t=1}^T \left(\frac{\hat{n}_{k,t}}{\lambda_{k,t}(\hat{\theta}_T)} \frac{\partial \lambda_{k,t}(\hat{\theta}_T)}{\partial \theta} \hat{s}_{k,t-h} \right)$$

$$\hat{C}_{k,1:q} = (\hat{c}_{k,1}, \dots, \hat{c}_{k,q}), \quad \hat{C}_{1:q} = (\hat{C}_{1,1:q}, \dots, \hat{C}_{K,1:q}), \quad \hat{\Sigma}_{\hat{\theta}_T, \rho_{k,1:q}} = -\frac{1}{\hat{\eta}_k} \hat{\mu}_k^2 \hat{J}^{-1} \hat{C}_{k,1:q},$$

$$\hat{\Sigma}_{\hat{\theta}_T, \rho_{1:q}} = \begin{pmatrix} \Sigma_{\hat{\theta}_T, \rho_{1,1:q}} & \cdots & \Sigma_{\hat{\theta}_T, \rho_{K,1:q}} \\ \hat{\mu}_1^2 I_q & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & \hat{\mu}_K^2 I_q \end{pmatrix}, \quad \hat{\Sigma}_{\rho_{1:q}} = \begin{pmatrix} \hat{\mu}_1^2 I_q & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & \hat{\mu}_K^2 I_q \end{pmatrix}$$

$$\text{and } \hat{B} = \hat{C}_{1:q}^\top \hat{J}^{-1} \hat{V} \hat{J}^{-\top} \hat{C}_{1:q} + \hat{\Sigma}_{\rho_{1:q}} + \hat{C}_{1:q}^\top \hat{\Sigma}_{\hat{\theta}_T, \rho_{1:q}} + \hat{\Sigma}_{\hat{\theta}_T, \rho_{1:q}}^\top \hat{C}_{1:q}.$$

Theorem 16 Consider the model (4.1)-(4.2) and assume the additional conditions (PM.1)-(PM.3) in Appendix. Under the assumptions of theorem 15 that yield the asymptotic normality of QMLE, as T tends to ∞ ,

$$T \hat{\rho}_{1:q}^\top \hat{B}^{-1} \hat{\rho}_{1:q} \rightarrow \chi_{Kq}^2.$$

The adequacy of the model (4.1)-(4.2) is then rejected at the asymptotic level α if

$$T \hat{\rho}_{1:q}^\top \hat{B}^{-1} \hat{\rho}_{1:q} > \chi_{Kq}^2(1 - \alpha).$$

4.4 Application

4.4.1 Simulation

We examined the finite-sample performance of the QMLE presented in the previous section through a small simulation study. We present the result for QMLE under two different data generating processes referred to as *scenario 1* and *scenario 2* with $m = 10$ covariates. For the first one, $X_{k,t}$ does not depend on k and is a sequence of *i.i.d* random variables distributed as exponential random variables with means $\lambda_1, \dots, \lambda_m$. For the second one, for a fixed k , $X_{k,t}$ is independently sampled from exponential distributions of

mean $0.4k\lambda_1, \dots, 0.4k\lambda_m$. For the two data generating processes, for a fixed k , the process $(n_{k,t})_{t \geq 1}$ is independently sampled from a Poisson distribution of mean τ_k as follows : for a fixed K, τ_1, \dots, τ_K are independent and distributed as an exponential random variable of mean K . Moreover, we take $p = 1$ and $\delta = 0.5, \beta = (0, 1, -1, 0.5, -0.5, -1.5, 1.5, -2, 2, 0), \alpha_1 = 0.6$ and $\omega_1, \dots, \omega_K$ independently uniformly sampling in the range $(-0.5K, 0.5K)$ for a fixed K . We sequentially choose $K = 5, 10, 15, 20$ and $T = 50, 100$. The samples are nested, *i.e* the sample for the first scenario and $K = 5, T = 50$ is a subset of that of $K = 5, T = 100$. Indeed, our aim here is to evaluate the consequences of increasing K and T on the performance of our estimator. For each sample, we compute the estimator (4.4) and the corresponding theoretical standard errors (TSE) given by Gaussian limit distribution. We replicate $B = 100$ times the experiment. Table 4.1 presents the simulation results. The line EQML refers to the average value of estimation of parameters and TSE refers to the average value of the estimation of the theoretical standard errors :

$$\text{EQML} = B^{-1} \sum_{b=1}^B \hat{\theta}_T^{(b)} \text{ and TSE} = B^{-1} \sum_{b=1}^B \text{diag} \left\{ \hat{J}^{-1(b)} \hat{V}^{(b)} \hat{J}^{-\top(b)} \right\}^{1/2}$$

where the superscript b stands for the index of replication and $\text{diag}M$ for a matrix M is the diagonal elements of M . It appears that the model parameters are well estimated except for the $\omega_k, k = 1, \dots, K$ when K is very small compared to T , which coincides here with $K = 5, T = 50, 100$. We leave deep simulation studies for further works.

4.4.2 Application to the white spruce growth series

Dendrochronology, *i.e.* the studies of the time series of tree growth rings, is a powerful tool to reconstruct past natural and anthropic disturbances (Montoro Girona et al. [101], Boulanger and Arseneault [12] and Labrecque-Foy et al. [85]). Tree-rings are hard disks of information, able to record each environmental change, thus having a strong potential to understand complex phenomena such as disturbance ecology. Many previous studies used dendrochronological data to better understand insect outbreak dynamics (Navarro et al. [108], Camarero et al. [17] and Speer and Kulakowski [121]).

In this research, we used the dendroecological series from the study by Jardon et al. [75], which includes annual tree-ring width measurements for 631 white spruce (*Picea glauca*) trees distributed across 45 sites in southwestern Quebec, Canada, with 1 to 23 trees per site. These time series comprise between 63 and 247 rings according to the tree's

age. We converted the ring width increments to basal area increments (BAI) using the full series, but due to covariate availability, we limit our analysis to the 1955-1995 time period (41 years) to study only one insect outbreak period (see fig 4.2 in Appendix).

We interpolated climate variables at the study sites for the 41-year period using BioSIM [113], a software package that interpolates daily climate station data based on latitudinal and elevational climate gradients, as well the spatial correlations estimated from 30-year climate normals. We computed the following climate summaries from daily data for the spring (April to June) and summer (July to September) seasons separately : mean of daily maximum temperatures, total precipitation, and the climate moisture index (CMI) equal to the difference between precipitation and potential evapotranspiration (PET). Daily PET values were estimated by the Penman-Monteith equation as implemented in the SPEI package [7] in R, based on BioSIM-interpolated values of the minimum and maximum temperature, wind speed at 2 m, solar radiation, dew point temperature and atmospheric pressure, using the "tall" crop model in SPEI.

One major SBW outbreak occurred in Quebec during the study period, which ranges from 1967 to 1991. We obtained annual estimates of the severity of SBW outbreaks at the location of each study site from defoliation maps produced by the Quebec Ministry of Forests, Wildlife and Parks [100]. These maps are digitized versions of hand-drawn outlines of defoliated areas produced by aerial surveys of the affected regions. The defoliation level for each area is classified on a scale of 1 to 3 corresponding to a low (approx. 1 – 35%), moderate (36 – 70%) or high (71 – 100%) fraction of the year's foliage defoliated by the SBW. We note that these defoliation levels mainly reflect the status of balsam fir (*Abies balsamea*) trees, which is the main SBW host and is generally more severely affected than white spruce. Therefore, these defoliation levels are a proxy for the outbreak severity, i.e. the potential herbivory pressure exerted by the budworm on spruce trees at the site.

Since tree growth and its vulnerability to both climate and defoliation depends on the tree-age, we split the dataset and separately fit our models for the following five age classes : ≤ 75 , 75 – 100, 100 – 125, 125 – 150 and ≥ 150 years. We include as covariates the mean daily maximum of temperature, the total precipitation and the mean CMI for the current and previous spring and summer. Only one of precipitation and CMI appears in a given model version due to the correlation between those two variables. We also include as covariates the defoliation levels for the five previous years, a delay which estimates the time needed to fully regrow the lost foliage after an outbreak. Note that we do not expect defoliation to have a marked effect on the same year's growth ring [84]. Finally, we

consider models with interaction effects of the previous year’s defoliation level and climate variables, representing the possibility that climate conditions can increase or decrease the tree’s sensitivity to SBW outbreaks.

Data processing and analyses were performed in R Core Team [112] with the package `dplR` [16] used to process tree-ring data. We minimize the criterion (4.3) with the R command `nlm` (Dennis and Schnabel [35]). All the developed software are made available under the Creative Commons (CC) license (see data availability statements). The model selection was carried out through the the QAIC criterion. The primary analysis based on partial autocorrelation plots leads us to select $p = 1$.

According to the QAIC, the best models were those without an interaction between climate and defoliation. Our model results (Figures 4.3 and 4.4) reveal that higher defoliation levels leads to reduced tree-ring growth, but this effect vanishes after two years; however, note that while the direct effect vanishes, expected growth will remain lower in successive years due to the large estimated first-order autocorrelation coefficient (0.8 to 0.9, depending on age class). Moreover, there is no significant effect of defoliation on the next year’s growth for the youngest and oldest trees, even though it produces an effect two years following the defoliation. The result are quite different for middle aged trees, which are significantly affected one year following the defoliation but not in the second year. For the climate variables, high maximum temperatures in the summer produce increased growth, with up to 5.6 square centimetre increase in basal area from a 10 degree Celsius increase in summer maximum temperature. However, the previous summer’s temperature has a negative effect on growth. Finally, the spring CMI is negatively correlated with tree-ring growth whereas the summer CMI has a positive effect. However, both the CMI and precipitation in the previous spring increase tree-ring growth of the current year : an increase of 100 millimetres in precipitation leads to at least a 6.8 square centimetre increase in basal area growth.

4.5 Proofs for the main results

Throughout this section, we will denote by $\zeta_{k,t}^\infty = (\zeta_{k,t,l})_{l \geq 1}$, the sequence of copies of the unity random variables $\zeta_{k,t}$. Moreover $\zeta_{k,t}$ can be decomposed into two components : its mean $\lambda_{k,t}$ function of $X_{k,t}$ and a free random variable ζ_t . For example, $\zeta_{k,t} = \lambda_{k,t} \zeta_t$ for a positive random variable ζ_t of mean 1. We will write $\zeta_{k,t} := \zeta_{k,t}(\lambda_{k,t}, \zeta_t)$ to denote the relationship between $\zeta_{k,t}$ and $\lambda_{k,t}$ and ζ_t . Accordingly, $\zeta_{k,t,l} = \lambda_{k,t} \zeta_{t,l}$ with $\zeta_{t,l}, l \geq 1$

i.i.d with mean 1 or in general $\zeta_{k,t,l} := \zeta_{k,t,l}(\lambda_{k,t}, \zeta_{t,l})$ with $\mathbb{E}\zeta_{k,t,l} = \lambda_{k,t}$. Let $\mathcal{F}_{k,t}$ denote the σ -algebra generated by $\zeta_s, X_{k,s+1}, s \leq t$ and $\mathcal{F}_{k,t,n}$ generated by $n_{k,s}, \zeta_s, X_{k,s+1}, s \leq t$. Finally, we will denote by ϕ_δ the inverse of $\varphi_\delta : \phi_\delta(x) = \varphi_\delta^{-1}(x)$. For stability, we will consider the following set of assumptions :

(A.1) The function ϕ_δ is v -Lipschitz and $v \sum_{i=1}^p |\alpha_i| < 1$.

(ST.1) For $k = 1, \dots, K, (n_{k,t-1}, \zeta_{k,t-1}^\infty, X_{k,t})_{t \in \mathbb{Z}}$ is stationary, ergodic, $(n_{k,t}, \zeta_{k,t}^\infty)$ is independent from $\mathcal{F}_{k,t-1,n}$ and $\mathbb{E}|X_{k,0}|_1 < \infty$.

(ST.2) For $k = 1, \dots, K,$

$$E(|\zeta_{k,t}(\lambda_{k,t}, \zeta_t) - \zeta_{k,t}(\bar{\lambda}_{k,t}, \zeta_t)| | \mathcal{F}_{k,t-1,n}) \leq |\lambda_{k,t} - \bar{\lambda}_{k,t}|.$$

It is worth noting that the example $\zeta_{k,t} = \lambda_{k,t} \zeta_t$ for a positive random variable ζ_k of mean 1 verifies the condition (ST.2).

Lemma 15 *Under the assumptions (A.1)-(ST.1), there exists a unique set of K stationary, ergodic sequences $(Y_{k,t}, n_{k,t}, X_{k,t}), k = 1, \dots, K$ that are a solution of equations (4.1)-(4.2) with $\mathbb{E}|\eta_{k,0}| < \infty, k = 1, \dots, K$.*

The proof of lemma 15 uses the techniques of iterated random maps. We refer the interested readers to [32] theorem 2 and 4 which investigated the problem of the solution of recursive stochastic equations with covariates or [31] in the case where no covariates are included in the dynamic.

Proof of lemma 15 From (4.2),

$$\eta_{k,t} = \omega_k + \sum_{j=1}^p \alpha_j \frac{1}{n_{k,t-j}} \sum_{\ell=1}^{n_{k,t-j}} \zeta_{k,t-j,\ell}(\phi_\delta(\eta_{k,t-j}), \zeta_{t-j,\ell}) + \beta^\top X_{k,t}.$$

Then under the condition (ST.1), the processes $(\eta_{k,t} = \varphi_\delta(\lambda_{k,t})_{t \in \mathbb{Z}}, k = 1, \dots, K$ obey some recursive stochastic equations

$$\eta_{k,t} = f(\eta_{k,t-1}, \dots, \eta_{k,t-p}; n_{k,t-1}, \dots, n_{k,t-p}, \zeta_{k,t-1}^\infty, \dots, \zeta_{k,t-p}^\infty, X_{k,t}).$$

and with (A.1), for $k = 1, \dots, K, (x, y) \in \mathbb{R}^{2p}$,

$$E(|f(x; n_{k,t-1}, \dots, n_{k,t-p}, \zeta_{k,t-1}^\infty, \dots, \zeta_{k,t-p}^\infty, X_{k,t}) - f(y; n_{k,t-1}, \dots, n_{k,t-p}, \zeta_{k,t-1}^\infty, \dots, \zeta_{k,t-p}^\infty, X_{k,t})| | \mathcal{F}_{k,t-1,n}) \leq v \alpha^\top |x-y|$$

with $\alpha = (\alpha_1, \dots, \alpha_p)$. Moreover $E|f(x; n_{k,-1}, \dots, n_{k,-p}, \zeta_{k,-1}^\infty, \dots, \zeta_{k,-p}^\infty, X_{k,0})| < \infty$. Then, from [32] theorem 4, we get the stationary and ergodic solution with $\mathbb{E}|\eta_{k,0}| < \infty, k = 1, \dots, K$. \square

The theorem 14 is a straight consequence of lemma 15 and follows the Lipschitz property of $x \mapsto \log(\exp(x) + 1 + \delta)$ for any $\delta > 0$. For the asymptotic results for $\hat{\theta}_T$, the following assumptions will be needed.

(A.2) The conditons (A.1) and (ST.1) are met and θ_0 verifies : $v \sum_{i=1}^p |\alpha_{i,0}| < 1$.

(A.3) For $k = 1, \dots, K, \mathbb{E}n_{k,0} < \infty$ and

$$\mathbb{E} \sup_{\theta} \left(\frac{\phi_\delta(\eta_{k,0}(\theta_0))}{\phi_\delta(\eta_{k,0}(\theta))} + |\log \circ \phi_\delta(\eta_{k,0}(\theta))| \right) < \infty.$$

(A.4) For $(\delta, \bar{\delta}) \in [\delta_-, \infty)^2, (\eta, \bar{\eta}) \in \mathbb{R}^2$,

$$\phi_\delta(\eta) = \phi_{\bar{\delta}}(\bar{\eta}) \Rightarrow (\delta = \bar{\delta}, \eta = \bar{\eta}).$$

(C.1) The conditons (ST.1) and (ST.2) are met and θ_0 verifies : $\sum_{i=1}^p |\alpha_{i,0}| < 1$.

(C.2) For $k = 1, \dots, K, \mathbb{E}n_{k,0} < \infty$.

(C.3) For $k = 1, \dots, K$, conditionally on $X_{k,0}$, the distribution of $\left(\frac{Y_{k,-1}}{n_{k-1}}, \dots, \frac{Y_{k,-p}}{n_{k-p}}\right)$ is not supported by an hyperplan of \mathbb{R}^p .

(C.4) For $k = 1, \dots, K$, the distribution of $X_{k,0}$ is not degenerate.

Lemma 16 *Suppose that the assumptions (A.2)-(A.4) and (C.3)-(C.4) are met. Then, almost surely,*

$$\lim_{T \rightarrow \infty} \hat{\theta}_T = \theta_0.$$

We do not prove the lemma 16. Similar results for time-series models can be found in [37], [3] or [33] among others.

Proof of consistency part of theorem 15 We will check (A.2) to (A.4).

— (A.2) comes from (C.1).

— One can note that here $\phi_\delta(x) = \log(1 + \delta + \exp(x))$ and $\phi_\delta(x) \geq \log(1 + \delta), \phi_\delta(x) \leq \kappa_1(\theta)(1 + |x|)$ and $|\log \circ \phi_\delta(x)| \leq \kappa_2(\theta)(1 + |x|) + \kappa_3(\theta)$, where $\kappa_i, i = 1, 2, 3$ are continuous functions of θ . Then (A.3) holds since $\mathbb{E} \sup_{\theta} |\eta_{k,0}(\theta)| < \infty$. Indeed $\mathbb{E}Y_{k,0}/n_{k,0} = \phi_\delta(\eta_{k,0}) < \infty$ since $\mathbb{E}|\eta_{k,0}| < \infty$.

— For (A.4), we note that

$$\phi_\delta(\eta) = \phi_{\bar{\delta}}(\bar{\eta}) \Rightarrow \delta - \bar{\delta} = \exp \bar{\eta} - \exp \eta$$

and $0 = \lim_{\eta \rightarrow -\infty, \bar{\eta} \rightarrow -\infty} \exp \bar{\eta} - \exp \eta = \delta - \bar{\delta}$. Then $\delta = \bar{\delta}$ and $\eta = \bar{\eta}$. \square

Let us set $\sigma_{k,0}^2 = \text{Var} \left(\frac{Y_{k,0}}{\lambda_{k,0}} \mid \mathcal{F}_{k,-1,n} \vee n_{k,0} \right)$, $\partial_\delta \phi_\delta$ the derivative of ϕ_δ with respect to δ , $\theta_{-\delta}$ the vector of parameters without δ . We will consider the following assumptions for the asymptotic distribution of $\hat{\theta}_T$.

(A.5) The function ϕ_δ is twice continuously differentiable and for $k = 1, \dots, K$,

$$\begin{aligned} \mathbb{E} \frac{\sigma_{k,0}^2}{\phi_\delta^2(\eta_{k,0})} \left[\partial_\delta \phi_\delta(\eta_{k,0}(\theta_0))^2 + \phi_\delta'(\eta_{k,0}(\theta_0))^2 \|\nabla_{\theta_{-\delta}} \eta_{k,0}(\theta_0)\|_2^2 \right] < \infty \text{ and} \\ \mathbb{E} \frac{1}{\phi_\delta^2(\eta_{k,0})} \left[\partial_\delta \phi_\delta(\eta_{k,0}(\theta_0))^2 + \phi_\delta'(\eta_{k,0}(\theta_0))^2 \|\nabla_{\theta_{-\delta}} \eta_{k,0}(\theta_0)\|_2^2 \right] < \infty. \end{aligned}$$

(A.6) For $k = 1, \dots, K$, the distribution of $(\partial_\delta \phi_\delta(\eta_{k,0}(\theta_0)), \phi_\delta'(\eta_{k,0}(\theta_0)) \nabla_{\theta_{-\delta}} \eta_{k,0}(\theta_0))$ is not degenerate.

(A.7) For $k = 1, \dots, K$, $\mathbb{E} \sup_\theta |W_{k,0}^{i,j}(\theta)| < \infty$, where $W_{k,0}^{i,j}(\theta)$ is one of the following quantities for all pairs i, j

$$\begin{aligned} \frac{1}{\phi_\delta^2(\eta_{k,0}(\theta))} \left(\frac{\phi_\delta(\eta_{k,0}(\theta_0))}{\phi_\delta(\eta_{k,0}(\theta))} + 1 \right) \frac{\partial \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_i} \frac{\partial \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_j} \\ \frac{1}{\phi_\delta^2(\eta_{k,0}(\theta))} \frac{\phi_\delta(\eta_{k,0}(\theta_0))}{\phi_\delta(\eta_{k,0}(\theta))} \frac{\partial \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_i} \frac{\partial \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_j} \\ \frac{1}{\phi_\delta(\eta_{k,0}(\theta))} \left(\frac{\phi_\delta(\eta_{k,0}(\theta_0))}{\phi_\delta(\eta_{k,0}(\theta))} + 1 \right) \frac{\partial^2 \phi_\delta(\eta_{k,0}(\theta))}{\partial \theta_i \partial \theta_j}. \end{aligned}$$

(AN.1) The K stationary sequences solution of (4.1)-(4.2) are independent of each other.

(AN.2) For $k = 1, \dots, K$, $\mathbb{E} n_{k,0}^2 < \infty$

$$\mathbb{E} \sigma_{k,0}^4 < \infty$$

(AN.3) For $k = 1, \dots, K$,

$$\mathbb{E} |X_{k,0}|_1^4 < \infty \text{ and } \mathbb{E} Y_{k,0}^4 < \infty.$$

Lemma 17 *Under the assumptions of lemma 15 and if (A.5)-(A.7) and (AN.1) hold, then*

$$\lim_{T \rightarrow \infty} \sqrt{T}(\hat{\theta}_T - \theta_0) = \mathcal{N}(0, J^{-1} V J^{-\top})$$

where $J = \sum_{k=1}^K J_k$ and $V = \sum_{k=1}^K V_k$,

$$V_k = \mathbb{E} \left[\frac{1}{\lambda_{k,0}^2} \left(n_{k,0} - \frac{Y_{k,0}}{\lambda_{k,0}} \right)^2 \dot{\lambda}_{k,0} \dot{\lambda}_{k,0}^\top \right], J_k = \mathbb{E} \left[n_{k,0} \frac{1}{\lambda_{k,0}^2} \dot{\lambda}_{k,0} \dot{\lambda}_{k,0}^\top \right] \text{ and}$$

$$\dot{\lambda}_{k,0} = (\partial_\delta \phi_\delta(\eta_{k,0}(\theta_0)), \phi'_\delta(\eta_{k,0}(\theta_0)) \nabla_{\theta-\delta} \eta_{k,0}(\theta_0))^\top.$$

As for lemma 16, we do not prove the lemma 17. We refer the interested reader to [37], [3] or [33] among others.

Proof of asymptotic normality part of theorem 15 For the proof of asymptotic normality part of theorem 15, one can note that in the single framework ($k = 1$), assumptions (AN.2) yield the asymptotic normality of $\sqrt{T} \nabla \ell_k(\theta_0)$ using the central limit theorem for difference martingale. Next,

$$\begin{aligned} \frac{\partial^2 \ell_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} &= \frac{1}{\lambda_{k,t}^2(\theta)} \left(\frac{Y_{k,t}}{\lambda_{k,t}(\theta)} - n_{k,t} \right) \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_i} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_j} \\ &+ \frac{1}{\lambda_{k,t}^2(\theta)} \frac{Y_{k,t}}{\lambda_{k,t}(\theta)} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_i} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_j} \\ &- \frac{1}{\lambda_{k,t}(\theta)} \left(\frac{Y_{k,t}}{\lambda_{k,t}(\theta)} - n_{k,t} \right) \frac{\partial^2 \lambda_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} =: I_{k,t}(\theta) + II_{k,t}(\theta) + III_{k,t}(\theta). \end{aligned}$$

For the first term,

$$\sup_\theta |I_{k,t}(\theta)| \leq n_{i,t} \left(\frac{\lambda_{i,t}}{\log(1 + \delta_-)} + 1 \right) \sup_\theta \frac{1}{\lambda_{k,t}^2(\theta)} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_i} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_j}$$

and

$$\frac{1}{\lambda_{k,t}(\theta)} \dot{\lambda}_{k,t}(\theta) \preceq \kappa_{\delta_-} \left(1, \iota_k, \frac{Y_{k,t-1}}{n_{k,t-1}} \dots \frac{Y_{k,t-p}}{n_{k,t-p}}, X_{k,t}^\top \right)^\top$$

where for $x = (x_1, \dots, x_d), y = (y_1, \dots, y_d), x \preceq y$ means $x_i \leq y_i, i = 1, \dots, d$ and κ_{δ_-} a function of δ_- . Then, $\mathbb{E} \sup_\theta |I_{k,t}(\theta)| < \infty$ under the assumption (AN.3). It can be shown similarly that $\mathbb{E} \sup_\theta |II_{k,t}(\theta)| < \infty$ and $\mathbb{E} \sup_\theta |III_{k,t}(\theta)| < \infty$. By the Taylor expansion

of $r_T(\cdot)$ between $\hat{\theta}_T$ and θ ,

$$\begin{aligned} 0 = \sqrt{T}\nabla r_T(\hat{\theta}_T) &= \sum_{k=1}^K \sqrt{T}\nabla \ell_k(\hat{\theta}_T) \\ &= \left(\sum_{k=1}^K \sqrt{T}\nabla \ell_k(\theta_0) \right) + \left(\sum_{k=1}^K \nabla^2 \ell_k(\theta_0) \right) \sqrt{T}(\hat{\theta}_T - \theta_0) + o_{\mathbb{P}}(1) \end{aligned}$$

The independence condition on paths (AN.1), assumption (AN.2) and central limit theorem for difference martingale allow us to conclude $\sum_{k=1}^K \sqrt{T}\nabla \ell_k(\theta_0)$ converges in distribution to central Gaussian vector of variance V as T tends to infinity. The assumption (AN.3) and ergodic theorem entails that $\sum_{k=1}^K \nabla^2 \ell_k(\theta_0)$ converges to J . Moreover, condition (AN.1), (C.3) and (C.4) ensure that the matrix J is invertible. \square

For the Portmanteau test, we will need the following additional conditions

(PM.1) For $k = 1, \dots, K$,

$$\mathbb{E}|X_{k,0}|_1^5 < \infty \text{ and } \mathbb{E}Y_{k,0}^5 < \infty.$$

(PM.2) For $k = 1, \dots, K$, the sequence

$$\left(s_{k,t} = \frac{Y_{k,t}}{\lambda_{k,t}} - n_{k,t} \right)_{t \in \mathbb{Z}}$$

is a strong white noise.

(PM.3) For any k , the random variables

$$(1 + \delta + e^{\eta_{k,0}}) \log(1 + \delta + e^{\eta_{k,0}}) s_{k,-i}, i = 1, \dots, q, e^{\eta_{k,0}} \frac{Y_{k,-j}}{n_{k,-j}}, j = 1, \dots, p, e^{\eta_{k,0}} X_{k,0}$$

are linearly independent.

One can note that in the case of $\zeta_{l,k,t} = \lambda_{k,t} \epsilon_{l,k,t}$ with $(\epsilon_{l,k,t})_{t \in \mathbb{Z}}$ is a sequence of independent identically distributed random variables of mean 1, $Y_{k,t}/\lambda_{k,t} - n_{k,t} = \sum_{l=1}^{n_{k,t}} \epsilon_{l,k,t} - n_{k,t}$ verifies (PM.2) when (ST.1) is met. The same kind of assumption is usually made for standard GARCH model. The last assumption is needed to show the invertibility of matrix B .

Proof of theorem 16 Let us set $\rho_{1:q}(\theta) = (\rho_{1,1:q}(\theta), \dots, \rho_{K,1:q}(\theta))$, $\rho_{k,1:q}(\theta) = (\rho_{k,1}(\theta), \dots, \rho_{k,q}(\theta))$, with

$$\rho_{k,h}(\theta) = \frac{1}{T} \sum_{t=h+1}^T s_{k,t}(\theta) s_{k,t-h}(\theta), \quad s_{k,t}(\theta) = \frac{Y_{k,t}}{\lambda_{k,t}(\theta)} - n_{k,t}$$

where we drop the dependence on θ of these quantities when there are evaluated at θ_0 and when these one are evaluated on $\hat{\theta}_T$, we put a hat over them. For example, $\hat{s}_{k,t}$ stands for $s_{k,t}(\hat{\theta}_T)$ Straightforwardly,

$$\begin{aligned} \frac{\partial s_{k,t}(\theta)}{\partial \theta_i} &= -\frac{Y_{k,t}}{\lambda_{k,t}^2(\theta)} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_i} \\ \frac{\partial^2 s_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} &= -\frac{Y_{k,t}}{\lambda_{k,t}^2(\theta)} \frac{\partial^2 \lambda_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} + \frac{Y_{k,t}}{\lambda_{k,t}^3(\theta)} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_i} \frac{\partial \lambda_{k,t}(\theta)}{\partial \theta_j} \\ \frac{\partial \rho_{k,h}(\theta)}{\partial \theta_i} &= \frac{1}{T} \sum_{t=h+1}^T \frac{\partial s_{k,t}(\theta)}{\partial \theta_i} s_{k,t-h}(\theta) + s_{k,t}(\theta) \frac{\partial s_{k,t-h}(\theta)}{\partial \theta_i} \\ \frac{\partial^2 \rho_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} &= \frac{1}{T} \sum_{t=h+1}^T \left(\frac{\partial^2 s_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} s_{k,t-h}(\theta) + \frac{\partial s_{k,t}(\theta)}{\partial \theta_j} \frac{\partial s_{k,t-h}(\theta)}{\partial \theta_i} \right. \\ &\quad \left. + s_{k,t}(\theta) \frac{\partial^2 s_{k,t-h}(\theta)}{\partial \theta_i \partial \theta_j} + \frac{\partial s_{k,t}(\theta)}{\partial \theta_i} \frac{\partial s_{k,t-h}(\theta)}{\partial \theta_j} \right). \end{aligned}$$

It follows that, $\lim_{T \rightarrow \infty} \mathbb{E} \sup_{\theta} \left| \frac{\partial^2 \rho_{k,t}(\theta)}{\partial \theta_i \partial \theta_j} \right| < \infty$ if

$$\mathbb{E} \frac{Y_{k,0}^5}{n_{k,0}^5} < \infty, \mathbb{E} |X_{k,0}|_1^5 < \infty \text{ and } \mathbb{E} Y_{k,0}^5 < \infty.$$

Then, by a Taylor expansion of $\rho_{1:q}(\cdot)$ around θ_0 and $\hat{\theta}_T$

$$\sqrt{T} \hat{\rho}_{1:q} = \sqrt{T} \rho_{1:q} + \frac{\partial \rho_{1:q}(\theta_0)}{\partial \theta^\top} \sqrt{T} (\hat{\theta}_T - \theta_0) + o_{\mathbb{P}}(1).$$

Setting $c_{k,h} = -\mathbb{E} \left(\frac{n_{k,t}}{\lambda_{k,t}} \frac{\partial \lambda_{k,t}(\theta_0)}{\partial \theta} s_{k,t-h} \right)$, $C_{k,1:q} = (c_{k,1} \cdots c_{k,q})$ and $C_{1:q} = (C_{1,1:q} \cdots C_{K,1:q})$, one can note that, as $T \rightarrow \infty$,

$$\frac{\partial \rho_{1:q}(\theta_0)}{\partial \theta^\top} \rightarrow C_{1:q}^\top.$$

We will now investigate the distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0, \rho_{1:q})$. First note that :

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = -\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\sum_{k=1}^K J^{-1} \frac{1}{\lambda_{k,t}} s_{k,t} \frac{\partial \lambda_{k,t}(\theta_0)}{\partial \theta} \right) + o_{\mathbb{P}}(1).$$

Let us denote by $S_{k,t}^d$ the squared matrix of d rows with $s_{k,t}$ as diagonal elements and 0 elsewhere. Straightforwardly,

$$\sqrt{T} \begin{pmatrix} \theta_0 - \hat{\theta}_T \\ \rho_{1:m} \end{pmatrix} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{pmatrix} S_{1,t}^Q & \cdots & S_{K,t}^Q & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & S_{1,t}^q & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & S_{2,t}^q & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & 0 & & & \vdots \\ 0 & \cdots & \vdots & \vdots & 0 & 0 & \cdots & S_{K,t}^q \end{pmatrix} \begin{pmatrix} J^{-1} \frac{1}{\lambda_{1,t}} \frac{\partial \lambda_{1,t}(\theta_0)}{\partial \theta} \\ \vdots \\ J^{-1} \frac{1}{\lambda_{K,t}} \frac{\partial \lambda_{K,t}(\theta_0)}{\partial \theta} \\ s_{1,t-1:t-q} \\ \vdots \\ s_{K,t-1:t-q} \end{pmatrix} + o_{\mathbb{P}}(1)$$

where $Q = 1+K+p+m$ is the total number of parameters and $s_{k,t-1:t-q} = (s_{k,t-1}, \dots, s_{k,t-q})^\top$. Let us set $\mu_k = Es_{k,0}^2$. From central limit theorem for martingale difference,

$$\sqrt{T} \begin{pmatrix} \theta_0 - \hat{\theta}_T \\ \rho_{1:m} \end{pmatrix} \Rightarrow \mathcal{N} \left(0, \begin{pmatrix} J^{-1} V J^{-\top} & \Sigma_{\hat{\theta}_T, \rho_{1,1:q}} & \cdots & \Sigma_{\hat{\theta}_T, \rho_{K,1:q}} \\ \Sigma_{\hat{\theta}_T, \rho_{1,1:q}}^\top & \mu_1^2 I_q & \cdots & 0 \\ \vdots & & & \\ \Sigma_{\hat{\theta}_T, \rho_{K,1:q}}^\top & 0 & \cdots & \mu_K^2 I_q \end{pmatrix} \right)$$

where

$$\Sigma_{\hat{\theta}_T, \rho_{k,1:q}} = -\frac{1}{\eta_k} \mu_k^2 J^{-1} C_{k,1:q}$$

with $\eta_k = \mathbb{E}n_{k,0}$. Let us set $B = C_{1:q}^\top J^{-1} V J^{-\top} C_{1:q} + \Sigma_{\rho_{1:q}} + C_{1:q}^\top \Sigma_{\hat{\theta}_T, \rho_{1:q}} + \Sigma_{\hat{\theta}_T, \rho_{1:q}}^\top C_{1:q}$.

We now prove that B is invertible. Indeed, $B = \mathbb{E}MM^\top$ with

$$M = \begin{pmatrix} S_{1,t}^q & 0 & \cdots & 0 \\ 0 & S_{2,t}^q & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ \vdots & 0 & 0 & S_{K,t}^q \end{pmatrix} \begin{pmatrix} s_{1,t-1:t-q} \\ \vdots \\ s_{K,t-1:t-q} \end{pmatrix} + C_{1:q}^\top \begin{pmatrix} S_{1,t}^Q & \cdots & S_{K,t}^Q \end{pmatrix} \begin{pmatrix} J^{-1} \frac{1}{\lambda_{1,t}} \frac{\partial \lambda_{1,t}(\theta_0)}{\partial \theta} \\ \vdots \\ J^{-1} \frac{1}{\lambda_{K,t}} \frac{\partial \lambda_{K,t}(\theta_0)}{\partial \theta} \end{pmatrix}$$

From independency assumptions (AN.1), B is invertible if for any $k = 1, \dots, K$, $\forall f =$

$(f_1, \dots, f_q),$

$$f^\top s_{k,-1:-q} + f^\top C_{k,1:q}^\top J^{-1} \frac{1}{\lambda_{k,0}} \frac{\partial \lambda_{k,0}(\theta_0)}{\partial \theta} = 0 \quad (4.5)$$

entails $f = 0$. But $f^\top C_{k,1:q}^\top J^{-1} \neq 0$ otherwise $f^\top s_{k,-1:-q} = 0$ and the assumption (PM.2) will be violated. Hence, (4.5) becomes

$$\lambda_{k,0} f^\top s_{k,-1:-q} + g_k^\top \frac{\partial \lambda_{k,0}(\theta_0)}{\partial \theta} = 0$$

where $g_k^\top = f^\top C_{k,1:q}^\top J^{-1}$. The assumption (PM.3) entails $f = 0$. It follows that $T\rho_{1:q}^\top B^{-1} \rho_{1:q}$ is distributed as Chi-square random variable of Kq degree of freedom. Almost \hat{B} converges almost surely to B . The result then follows. \square

Acknowledgements Funding was provided by the Contrat de service de recherche forestière number 3329-2019-142332177 obtained by PM and MMG from the Ministère des Forêts, de la Faune et des Parcs (Quebec, Canada), the Doctoral scholarships from GENES (ENSAE/ENSAI) obtained by ZMD. We thank G. Tougas for its data management work and A. Subedi for the map of study sites. All the authors acknowledge the Quebec Ministry of Forests, Wildlife and Parks (MFFP), H. Morin and for providing them with data and for the support to this project.

Data availability statements All softwares and data used in this paper are available at the following public zenodo repository : <https://doi.org/10.5281/zenodo.6340148>.

Author Contributions ZMD, PM and M.M.G. conceived and designed the study. ZMD and PM analyzed the data. ZMD wrote the first draft. PM and MMG contributed to the funding, and provided statistical and interpretation inputs. All authors contributed to the revision of the manuscript.

Appendix

TABLE 4.1 – Estimation results for the quasi maximum likelihood estimation

K	T	Scenario	α_1	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	
			0.6	0	1	-1	0.5	-0.5	-1.5	1.5	-2	2	0	
5	50	1	EQMLE	0.567	-0.018	0.985	-1.051	0.631	-0.247	-1.565	1.445	-2.165	1.673	0.011
			TSE	0.060	0.086	0.185	0.286	0.178	0.309	0.332	0.219	0.298	0.211	0.104
		2	EQMLE	0.517	0.019	0.862	-0.772	0.580	-0.493	-1.345	1.138	-1.709	1.908	-0.015
			TSE	0.053	0.097	0.176	0.217	0.193	0.252	0.313	0.214	0.280	0.211	0.104
	100	1	EQMLE	0.362	0.035	0.743	-0.632	0.457	-0.324	-1.219	1.099	-1.655	1.521	-0.080
			TSE	0.065	0.101	0.169	0.303	0.218	0.379	0.401	0.219	0.333	0.232	0.093
		2	EQMLE	0.361	-0.027	0.656	-0.678	0.294	-0.411	-0.956	1.148	-1.237	1.397	0.097
			TSE	0.085	0.114	0.215	0.362	0.232	0.388	0.450	0.268	0.376	0.271	0.121
10	50	1	EQMLE	0.304	-0.048	0.600	-0.487	0.476	-0.460	-0.844	1.012	-1.048	1.311	0.023
			TSE	0.043	0.063	0.122	0.177	0.134	0.199	0.229	0.148	0.146	0.147	0.069
		2	EQMLE	0.311	0.098	0.550	-0.547	0.243	-0.486	-0.762	1.015	-1.233	1.061	0.061
			TSE	0.046	0.084	0.153	0.220	0.175	0.258	0.256	0.183	0.235	0.174	0.079
	100	1	EQMLE	0.328	0.001	0.686	-0.507	0.222	-0.293	-0.838	0.903	-1.091	1.173	0.050
			TSE	0.035	0.056	0.112	0.153	0.107	0.184	0.164	0.108	0.133	0.119	0.059
		2	EQMLE	0.285	0.033	0.676	-0.651	0.258	-0.072	-0.603	0.826	-1.199	1.101	0.042
			TSE	0.039	0.063	0.142	0.175	0.119	0.249	0.205	0.158	0.160	0.160	0.076
15	50	1	EQMLE	0.546	-0.004	0.865	-0.734	0.462	-0.467	-1.418	1.280	-1.854	1.874	0.010
			TSE	0.041	0.069	0.124	0.186	0.125	0.220	0.219	0.140	0.222	0.156	0.062
		2	EQMLE	0.531	0.002	0.894	-0.985	0.405	-0.614	-1.342	1.228	-1.845	1.692	0.0163
			TSE	0.039	0.060	0.119	0.165	0.120	0.190	0.195	0.138	0.180	0.144	0.057
	100	1	EQMLE	0.384	-0.014	0.816	-0.549	0.160	-0.546	-1.226	0.987	-1.314	1.447	0.044
			TSE	0.028	0.040	0.075	0.105	0.080	0.147	0.134	0.096	0.115	0.088	0.042
		2	EQMLE	0.387	0.003	0.740	-0.675	0.471	-0.356	-0.751	1.104	-1.540	1.394	0.058
			TSE	0.053	0.077	0.160	0.230	0.151	0.242	0.273	0.166	0.241	0.168	0.078
20	50	1	EQMLE	0.370	0.018	0.613	-0.612	0.316	-0.337	-0.787	0.915	-1.223	1.277	0.003
			TSE	0.031	0.048	0.092	0.118	0.094	0.151	0.145	0.102	0.115	0.097	0.046
		2	EQMLE	0.369	-0.004	0.616	-0.745	0.350	-0.265	-1.022	0.949	-1.266	1.282	-0.010
			TSE	0.033	0.063	0.116	0.174	0.137	0.228	0.205	0.147	0.186	0.142	0.052
	100	1	EQMLE	0.339	0.002	0.534	-0.491	0.282	-0.393	-0.802	0.862	-1.183	1.199	0.021
			TSE	0.024	0.042	0.076	0.116	0.085	0.138	0.130	0.083	0.098	0.085	0.044
		2	EQMLE	0.311	-0.017	0.636	-0.524	0.294	-0.173	-0.996	1.016	-1.144	1.192	0.039
			TSE	0.025	0.052	0.104	0.128	0.109	0.190	0.151	0.131	0.130	0.122	0.055

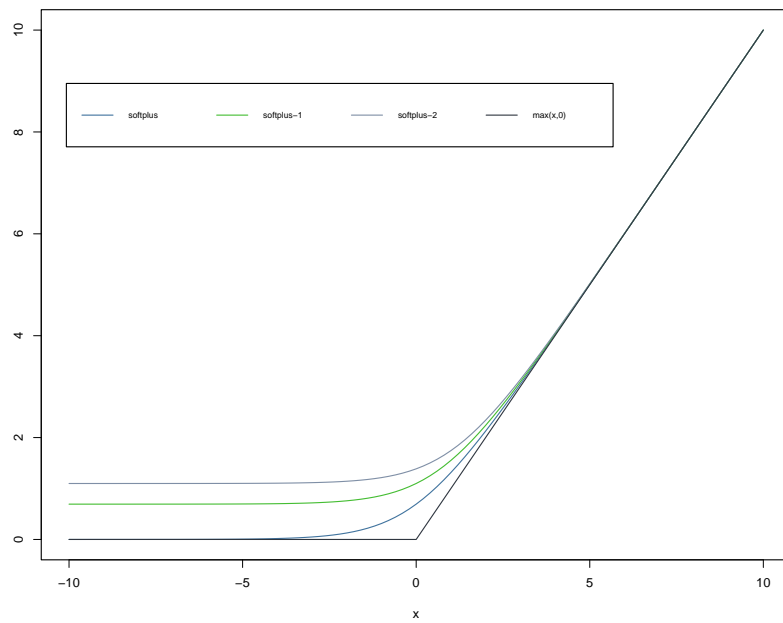


FIGURE 4.1 – Comparison between softplus and $\text{max}(x,0)$

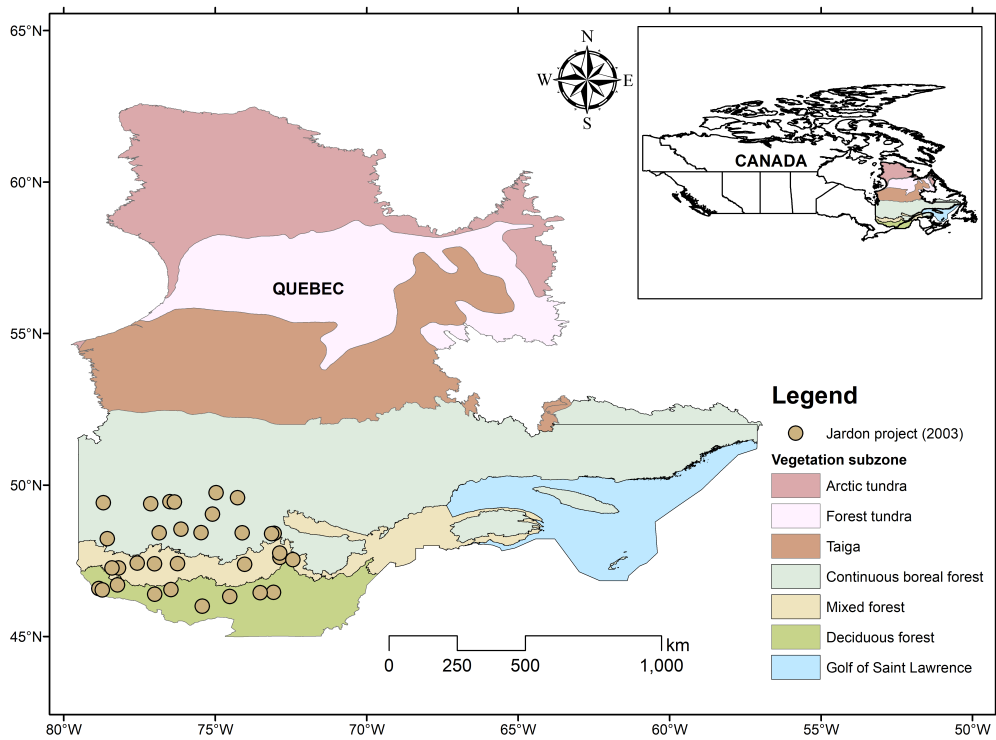


FIGURE 4.2 – Location of study sites from Jardon Project (2003) in Canadian boreal ecoregions.

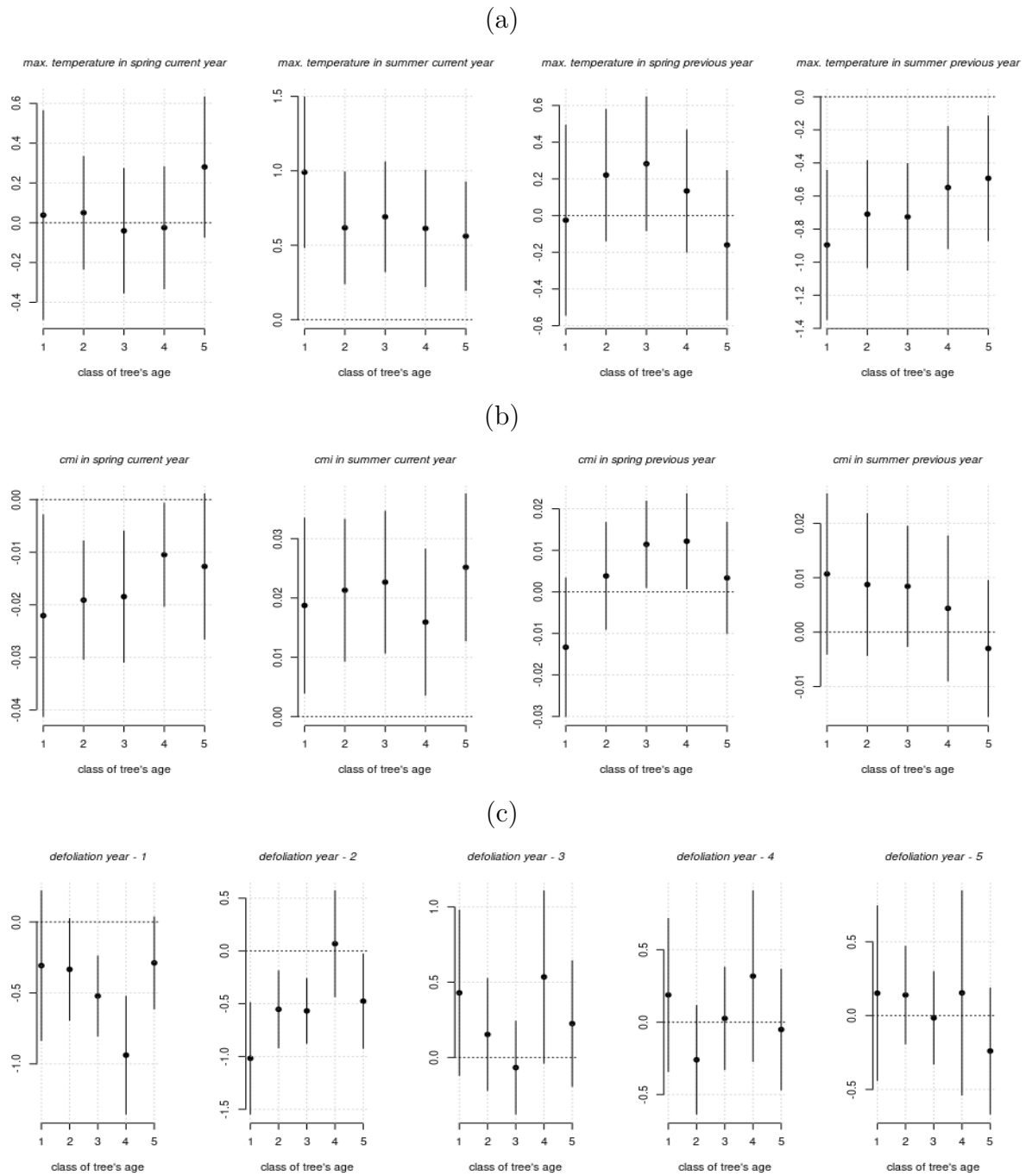


FIGURE 4.3 – Model with Temperature + CMI + Defoliation. classes of age 1 : < 75, 2 : 75 – 100, 3 : 100 – 125, 4 : 125 – 150 and 5 : > 150 years (a) effects of maximum temperature in spring and summer in current and previous year ; (b) effects of cmi index in spring and summer in current and previous year and (c) delayed effect of level of defoliation. The dashed horizontal line corresponds to zero.

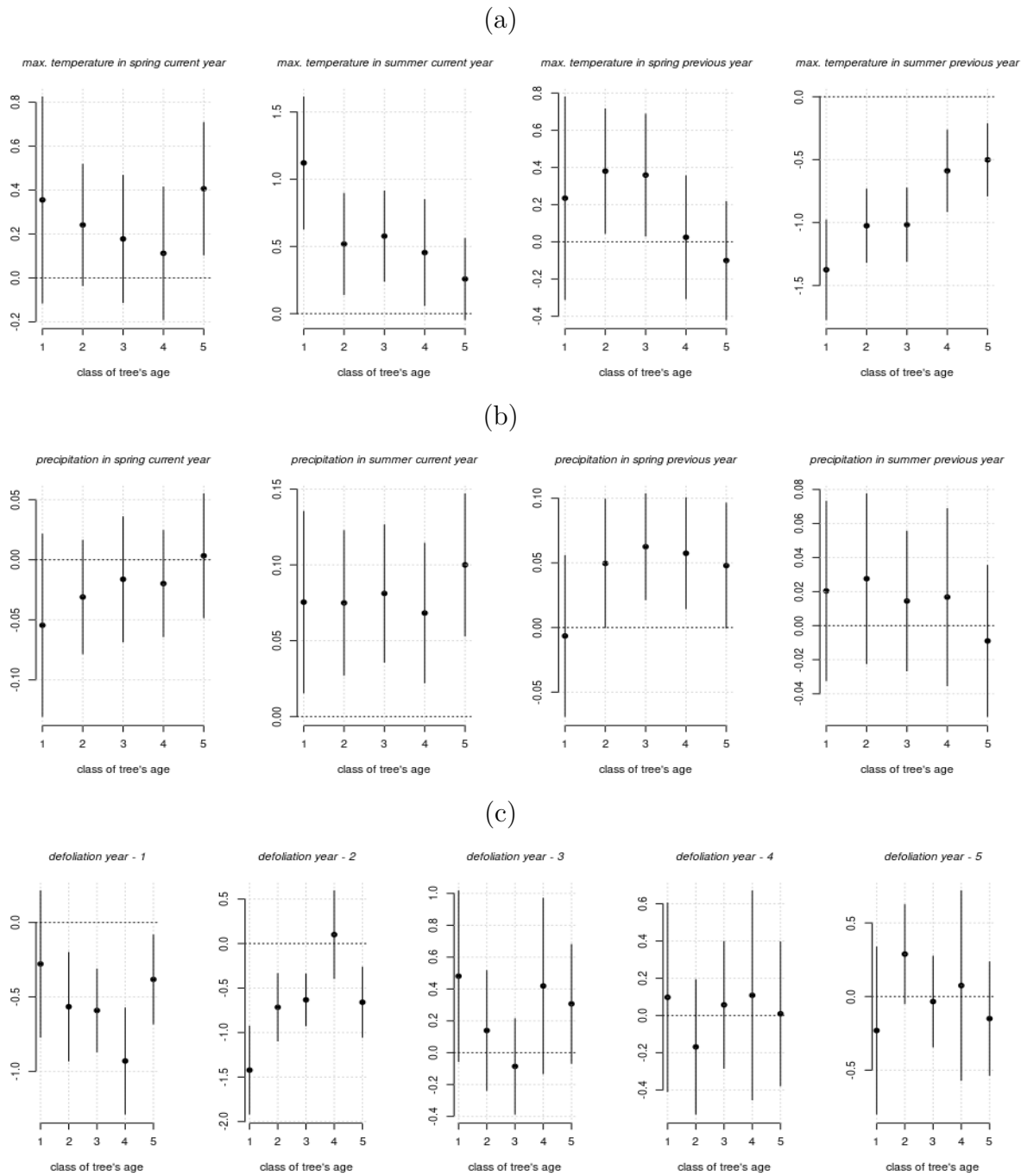


FIGURE 4.4 – Model with Temperature + Precipitation + Defoliation. classes of age 1 : < 75 , 2 : $75 - 100$, 3 : $100 - 125$, 4 : $125 - 150$ and 5 : > 150 years (a) effects of maximum temperature in spring and summer in current and previous year ; (b) effects of precipitation index in spring and summer in current and previous year and (c) delayed effect of level of defoliation. The dashed horizontal line corresponds to zero.

BIBLIOGRAPHIE

- [1] John Aber, Ronald P Neilson, Steve McNulty, James M Lenihan, Dominique Bachelet, and Raymond J Drapek. Forest processes and global environmental change : predicting the effects of individual and multiple stressors : we review the effects of several rapidly changing environmental drivers on ecosystem function, discuss interactions among them, and summarize predicted changes in productivity, carbon storage, and water balance. *BioScience*, 51(9) :735–751, 2001.
- [2] Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen, and Anders Rahbek. Modeling corporate defaults : Poisson autoregressions with exogenous covariates (parx). *Journal of Empirical Finance*, 38 :640–663, 2016.
- [3] Abdelhakim Aknouche and Christian Francq. Count and duration time series with equal conditional stochastic and mean orders. *Econometric Theory*, page 1–33, 2020. doi : 10.1017/S0266466620000134.
- [4] D.E. Allen, M. McAleer, M. Scharth, University of Canterbury. Department of Economics, and Finance. *Realized Volatility Risk*. Working paper (University of Canterbury. Department of Economics and Finance). Department of Economics and Finance, College of Business and Economics, University of Canterbury, 2010. URL <https://books.google.fr/books?id=YD-RtgEACAAJ>.
- [5] Donald WK Andrews. Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21(4) :930–934, 1984.
- [6] J-M Bardet, Paul Doukhan, and Olivier Wintenberger. Contrast estimation of general locally stationary processes using coupling. *arXiv preprint arXiv :2005.07397*, 2020.
- [7] Santiago Beguería and Sergio M. Vicente-Serrano. *SPEI : Calculation of the Standardised Precipitation-Evapotranspiration Index*, 2017. URL <https://CRAN.R-project.org/package=SPEI>. R package version 1.7.

- [8] Cassy Berguet, Maxence Martin, Dominique Arseneault, and Hubert Morin. Spatiotemporal dynamics of 20th-century spruce budworm outbreaks in eastern canada : Three distinct patterns of outbreak severity. *Frontiers in Ecology and Evolution*, 8, 2021. ISSN 2296-701X. doi : 10.3389/fevo.2020.544088. URL <https://www.frontiersin.org/article/10.3389/fevo.2020.544088>.
- [9] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31 :307–327, 1986.
- [10] Aleksandr Alekseevich Borovkov. *Ergodicity and stability of stochastic processes*. J. Wiley, 1998.
- [11] Philippe Bougerol and Nico Picard. Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, pages 1714–1730, 1992.
- [12] Yan Boulanger and Dominique Arseneault. Spruce budworm outbreaks in eastern quebec over the last 450 years. *Canadian Journal of Forest Research*, 34(5) :1035–1043, 2004.
- [13] Eric Bouyé, Valdo Durrleman, Ashkan Nikeghbali, Gaël Riboulet, and Thierry Roncalli. Copulas for finance-a reading guide and some applications. *Available at SSRN 1032533*, 2000.
- [14] Andreas Brandt. The stochastic equation $y_{n+1} = a_n y_n + b_n$ with stationary coefficients. *Advances in Applied Probability*, 18(1) :211–220, 1986.
- [15] Peter J Brockwell, Richard A Davis, and Stephen E Fienberg. *Time series : theory and methods : theory and methods*. Springer Science & Business Media, 1991.
- [16] Andrew G Bunn. A dendrochronology program library in r (dplr). *Dendrochronologia*, 26(2) :115–124, 2008.
- [17] JJ Camarero, E Martín, and E Gil-Pelegrín. The impact of a needleminer (epinotia subsequana) outbreak on radial growth of silver fir (abies alba) in the aragón pyrenees : a dendrochronological assessment. *Dendrochronologia*, 21(1) :3–12, 2003.
- [18] A Colin Cameron and Pravin K Trivedi. *Regression analysis of count data*, volume 53. Cambridge university press, 2013.

-
- [19] Gary Chamberlain. The general equivalence of granger and sims causality. *Econometrica : Journal of the Econometric Society*, pages 569–581, 1982.
- [20] Bunyeth Chan, Peng Bun Ngor, Zeb S. Hogan, Nam So, Sébastien Brosse, and Sovan Lek. Temporal dynamics of fish assemblages as a reflection of policy shift from fishing concession to co-management in one of the world’s largest tropical flood pulse fisheries. *Water*, 12(11), 2020. ISSN 2073-4441. doi : 10.3390/w12112974. URL <https://www.mdpi.com/2073-4441/12/11/2974>.
- [21] Ray Yeutien Chou, Hengchih Chou, and Nathan Liu. Range volatility : a review of models and empirical studies. *Handbook of financial econometrics and statistics*, pages 2029–2050, 2015.
- [22] D. R. Cox and E. J. Snell. *The Analysis of Binary Data*. Chapman & Hall, London, 1970.
- [23] Yan Cui and Fukang Zhu. A new bivariate integer-valued garch model allowing for negative cross-correlation. *Test*, 27(2) :428–452, 2018.
- [24] Yan Cui, Qi Li, and Fukang Zhu. Flexible bivariate poisson integer-valued garch model. *Annals of the Institute of Statistical Mathematics*, 72(6) :1449–1477, 2020.
- [25] Rainer Dahlhaus, Stefan Richter, and Wei Biao Wu. Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, 25(2) :1013–1044, 2019.
- [26] Richard A Davis and Heng Liu. Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, pages 1673–1707, 2016.
- [27] Richard A Davis, Scott H Holan, Robert Lund, and Nalini Ravishanker. *Handbook of discrete-valued time series*. CRC Press, 2016.
- [28] Robert M de Jong and Tiemen Woutersen. Dynamic time series binary choice. *Econometric Theory*, 27(4) :673–702, 2011.
- [29] Alexander R De Leon and Keumhee Carrière Chough. *Analysis of mixed data : methods & applications*. CRC Press, 2013.
- [30] Zinsou Max Debaly and Lionel Truquet. A note on the stability of multivariate non-linear time series with an application to time series of counts. *Statistics & Probability Letters*, 179 :109196, 2021.

- [31] Zinsou Max Debaly and Lionel Truquet. A note on the stability of multivariate non-linear time series with an application to time series of counts. *Statistics & Probability Letters*, 179 :109196, 2021.
- [32] Zinsou Max Debaly and Lionel Truquet. Iterations of dependent random maps and exogeneity in nonlinear dynamics. *Econometric Theory*, page 1–38, 2021. doi : 10.1017/S0266466620000559.
- [33] Zinsou Max Debaly and Lionel Truquet. Multivariate time series models for mixed data. *to appear in Bernoulli*, 202x.
- [34] Zinsou Max Debaly, Philippe Marchand, and Miguel Montoro Girona. Autoregressive models for time series of random sums of positive variables : application to tree growth as a function of climate and insect outbreaks. *arXiv preprint arXiv :2203.04926*, 2022.
- [35] JE Dennis and RB Schnabel. Numerical methods for unconstrained optimization and nonlinear equations prentice-hall. *Inc. : Englewood Clis*, 1983.
- [36] Persi Diaconis and David Freedman. Iterated random functions. *SIAM review*, 41 (1) :45–76, 1999.
- [37] Mamadou Lamine Diop and William Kengne. Inference and model selection in general causal time series with exogenous covariates. *arXiv preprint arXiv :2102.02870*, 2021.
- [38] Joseph L Doob. *Measure theory*, volume 143. Springer Science & Business Media, 2012.
- [39] R. Douc, P. Doukhan, and E. Moulines. Ergodicity of observation-driven time series models and consistency of the maximum-likelihood estimator. *Stochastic Processes and their Applications*, 123 :2620–2647, 2013.
- [40] Randal Douc, Eric Moulines, and David Stoffer. *Nonlinear time series : Theory, methods and applications with R examples*. CRC press, 2014.
- [41] P. Doukhan. *Mixing : properties and examples*. Number 85 in Lecture Notes in Statistics. Springer-Verlag, New York, 1994.

-
- [42] Paul Doukhan, Michael H Neumann, and Lionel Truquet. Stationarity and ergodic properties for some observation-driven models in random environments. *arXiv preprint arXiv :2007.07623*, 2020.
- [43] John H Elton. A multiplicative ergodic theorem for lipschitz maps. *Stochastic Processes and their Applications*, 34(1) :39–47, 1990.
- [44] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50 :987–1007, 1982.
- [45] Robert F. Engle and Jeffrey R. Russell. Autoregressive conditional duration : A new model for irregularly spaced transaction data. *Econometrica*, 66(5) :1127–1162, 1998. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/2999632>.
- [46] Robert F Engle, David F Hendry, and Jean-Francois Richard. Exogeneity. *Econometrica : Journal of the Econometric Society*, pages 277–304, 1983.
- [47] René Ferland, Alain Latour, and Driss Oraichi. Integer-valued garch process. *Journal of Time Series Analysis*, 27(6) :923–942, 2006.
- [48] Amanda Fernández-Fontelo, Alejandra Cabaña, Pedro Puig, and David Moriña. Under-reported data analysis with inar-hidden markov chains. *Statistics in Medicine*, 35(26) :4875–4890, 2016.
- [49] Richard A Fleming and W Jan A Volney. Effects of climate change on insect defoliator population processes in canada’s boreal forest : some plausible scenarios. *Water, Air, and Soil Pollution*, 82(1) :445–454, 1995.
- [50] K. Fokianos and L. Truquet. On categorical time series models with covariates. *arXiv :1709.09372v2*, 2017.
- [51] K. Fokianos and L. Truquet. On categorical time series with covariates. *Stochastic processes and their applications*, 2020.
- [52] K. Fokianos, A. Rahbek, and D. Tjøstheim. Poisson autoregression. *J. Amer. Statist. Assoc.*, 104 :1430–1439, 2009.
- [53] Konstantinos Fokianos and Benjamin Kedem. Regression theory for categorical time series. *Statistical science*, 18(3) :357–376, 2003.

- [54] Konstantinos Fokianos and Dag Tjøstheim. Log-linear Poisson autoregression. *J. Multivariate Anal.*, 102 :563–578, 2011.
- [55] Konstantinos Fokianos and Dag Tjøstheim. Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64(6) :1205–1225, 2012.
- [56] Konstantinos Fokianos and Lionel Truquet. On categorical time series models with covariates. *Stochastic Processes and their Applications*, 2018.
- [57] Konstantinos Fokianos, Bard Stove, Dag Tjostheim, and Paul Doukhan. Multivariate count autoregression. *Bernoulli*, 2020.
- [58] C. Francq and J.-M. Zakoïan. *GARCH models : Structure, Statistical Inference and Financial Applications*. Wiley, United Kingdom, 2010.
- [59] Christian Francq and Le Quyen Thieu. Qml inference for volatility models with covariates. *Econometric Theory*, 35(1) :37–72, 2019.
- [60] Christian Francq and Jean-Michel Zakoïan. *GARCH models : structure, statistical inference and financial applications*. Wiley, 2019.
- [61] Christian Francq, Olivier Wintenberger, and Jean-Michel Zakoïan. GARCH models without positivity constraints : Exponential or log garch ? *Journal of Econometrics*, 177(1) :34–46, 2013.
- [62] Sylvie Gauthier, Patrick Bernier, T Kuuluvainen, AZ Shvidenko, and DG Schepaschenko. Boreal forest health and global change. *Science*, 349(6250) :819–822, 2015.
- [63] Christian Genest and Johanna NeVaršlehová. A primer on copulas for count data. *ASTIN Bulletin : The Journal of the IAA*, 37(2) :475–515, 2007.
- [64] Martin P. Girardin, Olivier Bouriaud, Edward H. Hogg, Werner Kurz, Niklaus E. Zimmermann, Juha M. Metsaranta, Rogier de Jong, David C. Frank, Jan Esper, Ulf Büntgen, Xiao Jing Guo, and Jagtar Bhatti. No growth stimulation of canada’s boreal forest under half-century of combined warming and co2 fertilization. *Proceedings of the National Academy of Sciences*, 113(52) :E8406–E8414, 2016. ISSN 0027-8424. doi : 10.1073/pnas.1610156113. URL <https://www.pnas.org/content/113/52/E8406>.

-
- [65] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.
- [66] Paolo Gorgi and Siem Jan Koopman. Beta observation-driven models with exogenous regressors : A joint analysis of realized correlation and leverage effects. *Journal of Econometrics*, 2021.
- [67] Régent Guay, Réjean Gagnon, and Hubert Morin. A new automatic and interactive tree ring measurement system based on a line scan camera. *The Forestry Chronicle*, 68(1) :138–141, 1992.
- [68] A.D. Hall and D.F. Nicholls. The evaluation of exact maximum likelihood estimates for varma models. *Journal of Statistical Computation and Simulation*, 10(3-4) :251–262, 1980. doi : 10.1080/00949658008810373. URL <https://doi.org/10.1080/00949658008810373>.
- [69] Jamie Halliday and Georgi N Boshnakov. Poarx modelling for multivariate count time series. *arXiv preprint arXiv :1806.04892*, 2018.
- [70] Heejoon Han and Dennis Kristensen. Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of business & economic statistics*, 32(3) :416–429, 2014.
- [71] Wolfgang Härdle and Alexandre Tsybakov. Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of econometrics*, 81(1) :223–242, 1997.
- [72] Wolfgang Härdle, Alexandre Tsybakov, and Lijian Yang. Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, 68(2) :221–245, 1998.
- [73] EE Holmes, EJ Ward, and MD Scheuerell. Analysis of multivariate time-series using the marss package. *NOAA Fisheries, Northwest Fisheries Science Center*, 2725 : 98112, 2014.
- [74] Marius Iosifescu. *Iterated function systems : A critical survey*. Univ., 2009.

- [75] Yves Jardon, Hubert Morin, and Pierre Dutilleul. Périodicité et synchronisme des épidémies de la tordeuse des bourgeons de l'épinette au québec. *Canadian Journal of Forest Research*, 33(10) :1947–1961, 2003.
- [76] Norman L Johnson, Adrienne W Kemp, and Samuel Kotz. *Univariate discrete distributions*, volume 444. John Wiley & Sons, 2005.
- [77] O Kallenberg. Foundations of modern probability. 1997. *Collection : Probability and its applications Springer*, 1997.
- [78] Heikki Kauppi and Pentti Saikkonen. Predicting us recessions with dynamic binary response models. *The Review of Economics and Statistics*, 90(4) :777–791, 2008.
- [79] Tara Kessaram, James Stanley, and Michael G Baker. Estimating influenza-associated mortality in new zealand from 1990 to 2008. *Influenza and other respiratory viruses*, 9(1) :14–19, 2015.
- [80] John Frank Charles Kingman. *Poisson processes*, volume 3. Clarendon Press, 1992.
- [81] Maartje J Klapwijk, György Csóka, Anikó Hirka, and Christer Björkman. Forest insects and climate change : Long-term trends in herbivore damage. *Ecology and evolution*, 3(12) :4183–4196, 2013.
- [82] Lawrence A Klimko and Paul I Nelson. On conditional least squares estimation for stochastic processes. *The Annals of statistics*, pages 629–642, 1978.
- [83] Cornelia Krause and Hubert Morin. Changes in radial increment in stems and roots of balsam fir [*abies balsamea* (l.) mill.] after defoliation spruce budworm. *The forestry chronicle*, 71(6) :747–754, 1995.
- [84] Cornelia Krause, F Gionest, Hubert Morin, and David A MacLean. Temporal relations between defoliation caused by spruce budworm (*choristoneura fumiferana* clem.) and growth of balsam fir (*abies balsamea* (l.) mill.). *Dendrochronologia*, 21 (1) :23–31, 2003.
- [85] Julie-Pascale Labrecque-Foy, Hubert Morin, and Miguel Montoro Girona. Dynamics of territorial occupation by north american beavers in canadian boreal forests : A novel dendroecological approach. *Forests*, 11(2) :221, 2020.

- [86] Alain Latour. The multivariate ginar (p) process. *Advances in Applied Probability*, 29(1) :228–248, 1997.
- [87] Janie Lavoie, Miguel Montoro Girona, and Hubert Morin. Vulnerability of conifer regeneration to spruce budworm outbreaks in the eastern canadian boreal forest. *Forests*, 10(10) :850, 2019.
- [88] Sang-Won Lee and Bruce E Hansen. Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric theory*, 10(1) :29–52, 1994.
- [89] Gérard Letac. A contraction principle for certain markov chains and its applications. *Contemp. Math*, 50 :263–273, 1986.
- [90] Weidong Liu and Wei Biao Wu. Asymptotics of spectral density estimates. *Econometric Theory*, 26(4) :1218–1245, 2010.
- [91] Weidong Liu, Wei Biao Wu, et al. Simultaneous nonparametric inference of time series. *The Annals of Statistics*, 38(4) :2388–2421, 2010.
- [92] Shaowen Lu and Shuyu Huang. Segmentation of multivariate industrial time series data based on dynamic latent variable predictability. *IEEE Access*, 8 :112092–112103, 2020.
- [93] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.
- [94] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.
- [95] Hans Manner, Dennis Türk, and Michael Eichler. Modeling and forecasting multivariate electricity price spikes. *Energy Economics*, 60 :255–265, 2016.
- [96] Matthieu Marbac, Christophe Biernacki, and Vincent Vandewalle. Model-based clustering of gaussian copulas for mixed data. *Communications in Statistics-Theory and Methods*, 46(23) :11635–11656, 2017.
- [97] Maxence Martin, Miguel Montoro Girona, and Hubert Morin. Driving factors of conifer regeneration dynamics in eastern canadian boreal old-growth forests. *PLoS One*, 15(7) :e0230221, 2020.

- [98] Guido Masarotto and Cristiano Varin. Gaussian copula marginal regression. *Electronic Journal of Statistics*, 6 :1517–1549, 2012.
- [99] Hongyuan Mei and Jason M Eisner. The neural Hawkes process : A neurally self-modulating multivariate point process. *Advances in neural information processing systems*, 30, 2017.
- [100] MFFP. Données sur les perturbations naturelles - insecte : Tordeuse des bourgeons de l'épinette. URL <https://www.donneesquebec.ca/recherche/fr/dataset/donnees-sur-les-perturbations-naturelles-insecte-tordeuse-des-bourgeons-de-lepinet>. Accessed : 2019-05-19.
- [101] Miguel Montoro Girona, Hubert Morin, Jean-Martin Lussier, and Denis Walsh. Radial growth response of black spruce stands ten years after experimental shelterwoods and seed-tree cuttings in boreal forest. *Forests*, 7(10) :240, 2016.
- [102] Miguel Montoro Girona, Sergio Rossi, Jean-Martin Lussier, Denis Walsh, and Hubert Morin. Understanding tree growth responses after partial cuttings : A new approach. *PLoS One*, 12(2) :e0172653, 2017.
- [103] Miguel Montoro Girona, Lionel Navarro, and Hubert Morin. A secret hidden in the sediments : Lepidoptera scales. *Frontiers in Ecology and Evolution*, 6 :2, 2018.
- [104] Miguel Montoro Girona, Hubert Morin, Jean-Martin Lussier, and Jean-Claude Ruel. Post-cutting mortality following experimental silvicultural treatments in unmanaged boreal forest stands. *Frontiers in Forests and Global Change*, page 4, 2019.
- [105] Hubert Morin, Réjean Gagnon, Audrey Lemay, and Lionel Navarro. Chapter thirteen - revisiting the relationship between spruce budworm outbreaks and forest dynamics over the holocene in eastern north america based on novel proxies. In Edward A. Johnson and Kiyoko Miyaniishi, editors, *Plant Disturbance Ecology (Second Edition)*, pages 463–487. Academic Press, San Diego, second edition edition, 2021. ISBN 978-0-12-818813-2. doi : <https://doi.org/10.1016/B978-0-12-818813-2.00013-7>. URL <https://www.sciencedirect.com/science/article/pii/B9780128188132000137>.
- [106] T. Moysiadis and K. Fokianos. On binary and categorical time series models with feedback. *J. Multivariate Anal.*, 131 :209–228, 2014.

-
- [107] Theodoros Moysiadis and Konstantinos Fokianos. On binary and categorical time series models with feedback. *Journal of Multivariate Analysis*, 131 :209–228, 2014.
- [108] Lionel Navarro, Hubert Morin, Yves Bergeron, and Miguel Montoro Girona. Changes in spatiotemporal patterns of 20th century spruce budworm outbreaks in eastern canadian boreal forests. *Frontiers in Plant Science*, 9 :1905, 2018.
- [109] Nooshin Omranian, Bernd Mueller-Roeber, and Zoran Nikoloski. Segmentation of biological multivariate time-series data. *Scientific reports*, 5(1) :1–6, 2015.
- [110] Rasmus Søndergaard Pedersen and Anders Rahbek. Testing garch-x type models. *Econometric Theory*, pages 1–36, 2018.
- [111] James Gary Propp and David Bruce Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9(1-2) :223–252, 1996.
- [112] R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.
- [113] Jacques Régnière, Rémi Saint-Amant, Ariane Béchar, and Ahmed Moutaoufik. *BioSIM 10 : User’s manual*. Laurentian Forestry Centre Québec, QC, Canada, 2014.
- [114] S. Robin, F. Rodolphe, and S. Schbath. *DNA, words and models : statistics of exceptional words*. Cambridge University Press, 2005.
- [115] Andréa V Rocha and Francisco Cribari-Neto. Beta autoregressive moving average models. *Test*, 18(3) :529–545, 2009.
- [116] Jeffrey R Russell and Robert F Engle. A discrete-state continuous-time model of financial transactions prices and times : The autoregressive conditional multinomial–autoregressive conditional duration model. *Journal of Business & Economic Statistics*, 23(2) :166–180, 2005.
- [117] Rupert Seidl, Dominik Thom, Markus Kautz, Dario Martin-Benito, Mikko Peltoniemi, Giorgio Vacchiano, Jan Wild, Davide Ascoli, Michal Petr, Juha Honkaniemi, et al. Forest disturbances under climate change. *Nature climate change*, 7(6) : 395–402, 2017.

- [118] Chiaw-Hock Sim. First-order autoregressive models for gamma and exponential processes. *Journal of Applied Probability*, 27(2) :325–332, 1990.
- [119] Christopher A Sims. Money, income, and causality. *The American economic review*, 62(4) :540–552, 1972.
- [120] Peter X-K Song, Mingyao Li, and Ying Yuan. Joint regression analysis of correlated data using gaussian copulas. *Biometrics*, 65(1) :60–68, 2009.
- [121] James H Speer and Dominik Kulakowski. Creating a buzz : insect outbreaks and disturbance interactions. In *Dendroecology*, pages 231–255. Springer, 2017.
- [122] Örjan Stenflo. Markov chains in random environments and random iterated function systems. *Transactions of the American Mathematical Society*, 353(9) :3547–3562, 2001.
- [123] Örjan Stenflo. A survey of average contractive iterated function systems. *Journal of Difference Equations and Applications*, 18(8) :1355–1380, 2012.
- [124] Daniel Straumann and Thomas Mikosch. *Quasi-maximum-likelihood estimation in heteroscedastic time series : a stochastic recurrence equations approach*. Citeseer, 2003.
- [125] Daniel Straumann, Thomas Mikosch, et al. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series : a stochastic recurrence equations approach. *The Annals of Statistics*, 34(5) :2449–2495, 2006.
- [126] Masanobu Taniguchi and Yoshihide Kakizawa. *Asymptotic theory of statistical inference for time series*. Springer Science & Business Media, 2002.
- [127] Lionel Truquet. Local stationarity and time-inhomogeneous markov chains. *The Annals of Statistics*, 47(4) :2023–2050, 2019.
- [128] Lionel Truquet. Coupling and perturbation techniques for categorical time series. *Bernoulli*, 26(4) :3249–3279, 2020.
- [129] Ruey S Tsay. *Multivariate time series analysis : with R and financial applications*. John Wiley & Sons, 2013.

-
- [130] George L. Vourlitis, Osvaldo Borges Pinto, Higo J. Dalmagro, Paulo Enrique Zanelle de Arruda, Francisco de Almeida Lobo, and José de Souza Nogueira. Tree growth responses to climate variation in upland and seasonally flooded forests and woodlands of the cerrado-pantanal transition of Brazil. *Forest Ecology and Management*, 505 :119917, 2022. ISSN 0378-1127. doi : <https://doi.org/10.1016/j.foreco.2021.119917>. URL <https://www.sciencedirect.com/science/article/pii/S0378112721010082>.
- [131] Christian H Weiß. *An introduction to discrete-valued time series*. John Wiley & Sons, 2018.
- [132] Christian H Weiß. *An introduction to discrete-valued time series*, chapter 4. John Wiley & Sons, 2018.
- [133] Christian H Weiß, Fukang Zhu, and Aisouda Hoshiyar. Softplus ingarch models. *Statistica Sinica*, 32(3), 2022.
- [134] Wei Biao Wu. Nonlinear system theory : Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40) :14150–14154, 2005.
- [135] Wei Biao Wu and Xiaofeng Shao. Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2) :425–436, 2004.
- [136] Wei Biao Wu, Yinxiao Huang, and Yibi Huang. Kernel estimation for time series : An asymptotic theory. *Stochastic Processes and their Applications*, 120(12) :2412–2431, 2010.
- [137] Wei-Biao Wu, Ying Nian Wu, et al. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1) :352–379, 2016.
- [138] Wei Biao Wu et al. Strong invariance principles for dependent random variables. *The Annals of Probability*, 35(6) :2294–2320, 2007.
- [139] Han Xiao, Wei Biao Wu, et al. Covariance matrix estimation for stationary time series. *The Annals of Statistics*, 40(1) :466–493, 2012.
- [140] Ying Yang and Jian Kang. Joint analysis of mixed poisson and continuous longitudinal data with nonignorable missing values. *Computational statistics & data analysis*, 54(1) :193–207, 2010.

BIBLIOGRAPHIE

- [141] Scott L. Zeger and Bahjat Qaqish. Markov regression models for time series : a quasi-likelihood approach. *Biometrics*, 44 :1019–1031, 1988. ISSN 0006-341X. doi : 10.2307/2531732. URL <http://dx.doi.org/10.2307/2531732>.

Titre : Modèles de séries temporelles multivariées non-linéaires avec régresseurs exogènes

Mot clés : Contraction d'applications aléatoires, copule, épidémies forestières, pseudo-vraisemblance, stationnarité et ergodicité, variables exogènes

Résumé : Dans cette thèse, on s'intéresse aux propriétés probabilistes et statistiques de modèles de séries temporelles non-linéaires qui prennent en compte des covariables exogènes. Les séries temporelles de comptage ou catégorielles sont en particulier considérées ainsi que la modélisation de données mixtes en multivarié. Des propriétés de stationnarité sont établies pour ces modèles à partir de techniques d'itérations d'application aléatoires dépendantes. Dans le cas multivarié, des approches par pseudo-vraisemblance et/ou utilisation de copules sont utilisées pour l'inférence statistique. Enfin, une application de certaines de ces méthodes dans le cadre de l'écologie est présentée.

Title: Non-linear multivariate time series models with exogenous regressors

Keywords: contraction random mapping, copula, exogenous covariates, forest epidemic, pseudo-likelihood, stationarity and ergodicity.

Abstract: In this dissertation, we are interested in the probabilistic and statistical properties of non-linear time series models with exogenous covariates. In particular, count and categorical time series data are considered as well as the multivariate models for mixed data. Stationarity properties are established for these models using the techniques of iterations of dependent random maps. In the multivariate case, pseudo-likelihood and/or copula approaches are used for statistical inference. Finally, an application of some of these methods in the context of ecology is presented.