TI 2025-004/III
Tinbergen Institute Discussion Paper

Copula tensor count autoregressions for modeling multidimensional integer-valued time series

Mirko Armillotta[1]
Paolo Gorgi[2]
Andre Lucas[3]

1 University of Rome Tor Vergata

2 Vrije Universiteit Amsterdam, Tinbergen Institute

3 Vrije Universiteit Amsterdam, Tinbergen Institute

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Copula tensor count autoregressions for modeling multidimensional integer-valued time series*

Mirko Armillotta[1], Paolo Gorgi[2], and André Lucas[2]

[1]University of Rome Tor Vergata, Rome, Italy
[2]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

January 24, 2025

Abstract

This paper presents a novel copula-based autoregressive framework for multilayer arrays of integer-valued time series with tensor structure. It complements recent advances in tensor time series that predominantly focus on real-valued data and overlook the unique properties of integer-valued time series, such as discreteness and non-negativity. Our approach incorporates feedback effects for the time-varying parameters that describe the counts' temporal dynamics and introduces new identification constraints for parameter estimation. We provide an asymptotic theory for a Two-Stage Maximum Likelihood Estimator (2SMLE) tailored to the new tensor model. The estimator tackles the model's multidimensionality and interdependence challenges for large-scale count datasets, while at the same time addressing computational challenges inherent to copula parameter estimation. In this way it significantly advances the modeling of count tensors. An application to crime time series demonstrates the practical utility of the proposed methodology.

Keywords: INGARCH, tensor autoregression, parameter identification, quasi-likelihood, two-stage estimator.

*Corresponding author: André Lucas. Email address: a.lucas@vu.nl.

1 Introduction

The study of tensor time series has gained considerable attention in recent years due to its relevance in fields such as finance (Aldasoro and Alves, 2018; Chang et al., 2023), neuroscience (Zhou et al., 2013), meteorology and environmental science (Yuan et al., 2023; Chen et al., 2024), and many others. Traditional time series models are often inadequate for capturing the complex, multidimensional structures present in such data, as they are generally designed for univariate or multivariate (vector) observations. Recent advancements in the literature have addressed these limitations by developing approaches that better capture the intricate relationships and dependencies in multidimensional arrays of time series through order-p tensor autoregressive models (Chen et al., 2021; Billio et al., 2023; Samadi and Billard, 2024) or tensor factor models (Chen et al., 2022; Han et al., 2024). However, this work has thus far mainly focused on real-valued data. Little progress has been made in the context of integer-valued tensor time series, despite its practical importance in applications involving count data, such as financial transactions, crime reports, and biological data. Integer-valued time series carry unique properties, including discreteness and non-negativity, which typically make the tools developed for real-valued tensors unsuitable.

The literature on integer-valued time series is relatively well established, particularly in the univariate context (Davis et al., 2021; Fokianos, 2022). Two common approaches, INteger-valued AutoRegressive (INAR) models (Al-Osh and Alzaid, 1987; Drost et al., 2009) and INteger-valued GARCH models (INGARCH) (Heinen, 2003; Ferland et al., 2006) have been designed to accommodate the unique properties of count data. Their multivariate extensions are also available (Pedeli and Karlis, 2013a,b; Fokianos et al., 2020; Debaly and Truquet, 2023). Nevertheless, these methods fall short when applied to settings involving multilayer arrays of integer-valued data, as they are not equipped to handle the inherent complexities of tensor structures and the associated abundance of parameters to be estimated.

In this paper, we aim to bridge this gap by introducing a novel autoregressive copula-based modeling approach for multilayer arrays of integer-valued time series with tensor structure. To the best of our knowledge, this is the first study to introduce a model for tensor integer-valued time series. The joint dependence of the multidimensional array is accounted for by a probabilistic framework based on discrete marginals and a copula. Copulas provide a flexible way to capture the dependence structure between random variables (Nelsen, 2006; Joe, 2014), making them a suitable tool for handling the complex interactions present in different tensor dimensions. Our approach extends the applicability of copula-based models to the context of tensor count data, thus opening up new possibilities for analyzing multidimensional datasets with integer-valued observations.

An important distinction of our work is the incorporation of temporal feedback dependence into the count tensor autoregression. Previous tensor autoregressive models have primarily relied on temporal dependence limited to a finite lag (Li and Xiao, 2021; Billio et al., 2023). This fails to capture the long-range dependence that often exists in empirical datasets. Our model, in contrast, considers the entire historical path of the count tensor time series through the inclusion of lagged time-varying parameters in the autoregressive dynamics. This provides a more empirically congruent representation of the temporal dependencies in typical data sets. Moreover, we introduce new identification constraints for the autoregressive tensors, which enable the full identification of model parameters. These constraints address the limitations of existing estimation algorithms, which are often hindered by iterative procedures that can be computationally inefficient and prone to convergence issues (Chen et al., 2021; Hsu et al., 2021; Li and Xiao, 2021). Our approach not only simplifies the estimation process, but also ensures that the parameters are uniquely identified. The latter is crucial for the interpretability and reliability of the model.

We establish the asymptotic theory for a Two-Stage Maximum Likelihood Estimator (2SMLE) of the tensor model's parameters. In this approach, the parameters associated to the time-varying tensor means are estimated in a first stage by a Quasi-Maximum Likelihood Estimator (QMLE) (Gourieroux et al., 1984; Ahmad and Francq, 2016). Next, the copula parameters are estimated in a second stage by using the first-stage estimation of the time-varying means as a plug-in in the log-likelihood function. Although this approach is well-known in copula theory, the development of such an estimator represents a significant advancement, as it not only accommodates the complexities of multidimensional count data, but also extends existing asymptotic results to counts with tensor structures. The two-stage MLE framework allows one to handle the challenges posed by the multi-dimensionality and interdependencies in tensor data at the same time. As the estimation of copula parameters is often complicated by computational scalability issues as the dimensionality of the data increases (Panagiotelis et al., 2012), we propose practical solutions to these challenges for integer-valued tensor data by a suitable specification of copula parameters or by a likelihood approximation. This further enhances the feasibility of the model and the two-stage estimator for large multidimensional integer-valued datasets. Collectively, these advancements offer researchers powerful new tools for uncovering relationships in structured count datasets that were thus far not attainable through conventional univariate or multivariate analyses of integer-valued time series.

The rest of the paper is organized as follows. Section 2 gives a short introduction to tensor notation. Section 3 describes the general copula-based tensor count autoregressive model, discusses its identification and stationarity conditions, and introduces the two-stage estimation procedure. Section 4 provides a detailed treatment of the asymptotic theory for the first-stage QMLE. Large sample properties of the 2SMLE are established in section 5 and proposes an approximation to retain computational feasibility in large tensor dimensions. Section 6 introduces the special cases of linear and log-linear copula tensor INGARCH models, discusses their properties and derives related numerical results. An empirical application to crime time series is presented in section 7. The proofs of the main results are postponed to the appendix. All the other proofs are contained in an online appendix in the supplementary material together with further numerical results.

2 Tensor and matrix notation

An N-order real-valued *tensor* is an N-dimensional array $\mathcal{X} = (\mathcal{X}_{i_1\dots i_N}) \in \mathbb{R}^{n_1\times\cdots\times n_N}$ with entries $\mathcal{X}_{i_1\dots i_N}$, for $i_s = 1,\dots,n_s$ and $s = 1,\dots,N$. The *order* is the number of dimensions (also called modes). Vectors and matrices are examples of 1- and 2-order tensors, respectively. In the rest of the article, we use lower-case letters for vectors, capital letters for matrices, and calligraphic capital letters for tensors. The *mode-s product* $\mathcal{C} = \mathcal{X} \times_s B$ of a tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_N}$ and a matrix $B \in \mathbb{R}^{m_s\times n_s}$ yields a tensor $\mathcal{C} \in \mathbb{R}^{n_1\times\cdots\times n_{s-1}\times m_s\times n_{s+1}\times\cdots\times n_N}$ with entries

$$\mathcal{C}_{i_1\cdots i_{s-1} j_s i_{s+1}\cdots i_N} = \sum_{i_s=1}^{n_s} \mathcal{X}_{i_1\cdots i_{s-1} i_s i_{s+1}\cdots i_N} B_{j_s i_s} .$$

The compact notation $\mathcal{X} \star \boldsymbol{A} = \mathcal{X} \times_1 A_1 \times_2 \cdots \times_N A_N$ denotes the *multilinear product* between an N-order tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_N}$ and a set of matrices $\boldsymbol{A} = \{A_s\}_{s=1}^N$ such that $A_s \in \mathbb{R}^{m_s\times n_s}$, for $s = 1,\dots,N$. The result is a tensor of size $m_1 \times m_2 \times \cdots \times m_N$.

The *vectorization* operator $x = \mathrm{vec}(\mathcal{X}) \in \mathbb{R}^{n_1 n_2\cdots n_N}$ stacks all the elements of a tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_N}$ in a vector of length $n = \prod_{s=1}^N n_s$ with the entry at position $i_1 + \sum_{s=2}^N (i_s - 1) n_1 n_2 \cdots n_{s-1}$ equal to $\mathcal{X}_{i_1\dots i_N}$. The multilinear product can equivalently be defined in vector form as $\mathrm{vec}(\mathcal{X} \star \boldsymbol{A}) = (A_N \otimes A_{N-1} \otimes \cdots \otimes A_1)\,\mathrm{vec}(\mathcal{X})$. For more details on tensor operators see Cichocki et al. (2015).

The symbols I, 1 and O are the identity matrix, a vector of ones and a matrix of zero, respectively, all of suitable dimensions depending on the context. The notation $|A|_e$ for a matrix $A \in \mathbb{R}^{n_1\times n_2}$ denotes the elementwise absolute value. The quantity $\rho(A)$ stands for the spectral radius of A. Let $a \in \mathbb{R}^n$ be a vector. Then, $\|a\|_p$ denotes the l_p norm. The notation $\|a\|$ stands for l_1 norm. For the same reason, $|||A|||_p$ ($|||A|||$) stands for the generalized matrix norm induced by the l_p (l_1) vector norm. If $n_1 = n_2$, these norms are matrix norms. The inequality $x \le y$ between vectors $x, y \in \mathbb{R}^n$ denotes $x_i \le y_i$ for all $i = 1,\dots,n$.

3 Copula tensor count autoregressions

Consider a tensor time series of counts $\{\mathcal{Y}_t\}_{t\in\mathbb{Z}}$ of order N with $\mathcal{Y}_t \in \mathbb{N}^{n_1\times n_2\times\cdots\times n_N}$ and $\mathcal{F}_t$ denotes the σ-field that is generated by $\{\mathcal{Y}_\tau, \mathcal{X}_\tau, \tau \le t\}$ where $\{\mathcal{X}_t\}_{t\in\mathbb{Z}} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_N}$ is a tensor of exogenous covariates. Then, $\mathcal{Y}_t = (\mathcal{Y}_{i_1\dots i_N,t})$ is a N-dimensional array of time series of counts, where $\mathcal{Y}_{i_1\dots i_N,t}$ is the single element of the tensor. We assume that each element of the tensor is generated by the following quantile transformation, for $t \in \mathbb{Z}$,

$$\mathcal{Y}_{i_1\dots i_N,t} = F^{-1}_{i_1\dots i_N}(\mathcal{U}_{i_1\dots i_N,t}; \Lambda_{i_1\dots i_N,t})\,, \qquad 1 \le i_s \le n_s\,,\ 1 \le s \le N \tag{1}$$

where $\Lambda_{i_1\dots i_N,t} = \mathrm{E}(\mathcal{Y}_{i_1\dots i_N,t}|\mathcal{F}_{t-1})$, and $F^{-1}_{i_1\dots i_N}(\,\cdot\,; \Lambda_{i_1\dots i_N,t})$ is a shorthand for $F^{-1}_{i_1\dots i_N}(\,\cdot\,; \Lambda_{i_1\dots i_N,t}|\mathcal{F}_{t-1})$, which denotes the conditional (on the past) quantile function of the Poisson distribution with conditional mean $\Lambda_{i_1\dots i_N,t} \in \mathbb{R}_+$. The process $\{\mathcal{U}_{i_1\dots i_N,t}\}_{t\in\mathbb{Z}}$ is a sequence of random variables which are iid over time, follow the standard uniform distribution, $\mathcal{U}_{i_1\dots i_N,t} \sim U(0,1)$, and which are the single elements of the tensor $\mathcal{U}_t = (\mathcal{U}_{i_1\dots i_N,t}) \in [0,1]^{n_1\times\cdots\times n_N}$. Define $n = \prod_{s=1}^N n_s$. By vectorizing all the elements of $\mathcal{U}_t$, i.e., $u_t = \mathrm{vec}(\mathcal{U}_t) \in [0,1]^n$, the joint distribution of u_t conditional on the past is denoted by C_r. This distribution depends on a set of parameters r, which we write as $u_t|\mathcal{F}_{t-1} \sim C_r$. Then, C_r is a conditional copula function, i.e., a conditional joint probability distribution on $[0,1]^n$ with uniform marginals (Patton, 2002, 2006). For more details on copula functions see Nelsen (2006).

Let $\Lambda_t = (\Lambda_{i_1\dots i_N,t}) \in \mathbb{R}_+^{n_1\times\cdots\times n_N}$ denote the N-dimensional tensor of Poisson conditional means. Define some measurable one-to-one functions $g : \mathbb{R}_+^{n_1\times\cdots\times n_N} \to \mathbb{R}^{n_1\times\cdots\times n_N}$, $f : \mathbb{N}^{n_1\times\cdots\times n_N} \to \mathbb{R}^{n_1\times\cdots\times n_N}$ where $g(\mathcal{X})$ means that $g(\cdot)$ is applied elementwise for each element of the tensor. The same holds for $f(\cdot)$. Some examples of specifications of these functions are presented in section 6. Set $\mathcal{V}_t = g(\Lambda_t)$ and $\mathcal{Z}_t = f(\mathcal{Y}_t)$ for all $t \in \mathbb{Z}$. We propose to model the dynamics of the latent conditional means with the following tensor count autoregression. For $t \in \mathbb{Z}$,

$$\mathcal{V}_t = \mathcal{D} + \mathcal{V}_{t-1} \star \boldsymbol{A} + \mathcal{Z}_{t-1} \star \boldsymbol{B} + \mathcal{X}_{t-1} \star \boldsymbol{Q} \tag{2}$$

where $\mathcal{D} \in \mathbb{R}^{n_1\times\cdots\times n_N}$ is a tensor of intercepts. Moreover, we recall that $\mathcal{V}_{t-1} \star \boldsymbol{A} = \mathcal{V}_{t-1} \times_1 A_1 \times_2 \cdots \times_N A_N$ is the multilinear product and $\boldsymbol{A} = \{A_s\}_{s=1}^N$ is set of matrices such that $A_s \in \mathbb{R}^{n_s\times n_s}$ are coefficient matrices for $s = 1, \dots, N$. The same holds for the other tensors. Following Li and Xiao (2021, Sec. 2.4) it can be seen that each matrix A_s in the tensor model (2) captures the impact of $\mathcal{V}_{t-1}$ on $\mathcal{V}_t$ along the s-th direction of the tensor. Similar arguments hold for B_s and Q_s. Therefore, the matrices $\{A_s, B_s, Q_s\}_{s=1}^N$ will be the focus of the estimation. To further clarify the interpretation of the parameters in tensor matrices, we report in appendix A the description of model (1)-(2) and the interpretation of its coefficient matrices in the special case $N = 2$.

Note that equation (2) does not only depend on lag-one effects of the tensors $\{\mathcal{Y}_{t-1}, \mathcal{X}_{t-1}\}$, but on their whole past dynamics $\mathcal{F}_{t-1}$ through the inclusion of the lagged time-varying tensor parameter $\mathcal{V}_{t-1}$. Therefore, the stochastic process jointly defined by (1)-(2) will be called *copula tensor count autoregression* since it generates a tensor of count time series, where each univariate time series is marginally Poisson (conditional on the past), i.e. $\mathcal{Y}_{i_1\ldots i_N,t}|\mathcal{F}_{t-1} \sim Poisson(\Lambda_{i_1\ldots i_N,t})$ and the joint dependence between the counts is modeled through the conditional copula function C_r. For example, when the joint dependence is modelled through a Gaussian copula we have

$$C_r(u_t) = \Phi_r\left(\Phi^{-1}(u_{1,t}), \ldots, \Phi^{-1}(u_{n,t})\right) \tag{3}$$

where Φ_r denotes the joint cumulative distribution function (cdf) of the multivariate standard normal distribution with correlation matrix R such that $r = \text{vech}(R)$ holds the vectorized strictly lower triangular part of R, and Φ^{-1} denotes the inverse cdf of a standard normal distribution. The correlation matrix can be completely unstructured or take a specific form, e.g. a *tensor correlation* $R = R_N \otimes R_{N-1} \otimes \cdots \otimes R_1$ where R_s corresponds to the mode s correlations, for $s = 1, \ldots, N$. See section 6.1 for further details and examples.

3.1 Vectorization

Define the n-dimensional vectors $y_t = \text{vec}(\mathcal{Y}_t)$, $z_t = \text{vec}(\mathcal{Z}_t)$, $\lambda_t = \text{vec}(\Lambda_t)$, $\nu_t = \text{vec}(\mathcal{V}_t)$ and $x_t = \text{vec}(\mathcal{X}_t)$. We have $z_t = f(y_t)$ and $\nu_t = g(\lambda_t)$, where now the functions f and g are as defined in (2), but applied to each element of the vectorized tensors. Model (1)-(2) can be represented in vectorized form for $t \in \mathbb{Z}$ as

$$y_{h,t} = F_h^{-1}(u_{h,t}; \lambda_{h,t}), \quad 1 \leq h \leq n, \qquad \nu_t = d + A\nu_{t-1} + Bz_{t-1} + Qx_{t-1} \tag{4}$$

where $d = \text{vec}(\mathcal{D})$, $A = A_N \otimes A_{N-1} \otimes \cdots \otimes A_1$, $B = B_N \otimes B_{N-1} \otimes \cdots \otimes B_1$, and $Q = Q_N \otimes Q_{N-1} \otimes \cdots \otimes Q_1$. The dynamics of λ_t are modeled through their one-to-one transformation ν_t in the right-hand side equation of (4). The elementwise functions $F_h^{-1}(\cdot; \lambda_{h,t})$ are the same Poisson quantiles defined in (1), but applied after vectorization of the tensors. Therefore, from the properties of the quantile function we obtain that the joint conditional cdf of y_t takes the form

$$F(y_t; \lambda_t, r) = C_r\left(F_1(y_{1,t}; \lambda_{1,t}), \ldots, F_n(y_{n,t}; \lambda_{n,t})\right) \tag{5}$$

where $F_h(y_{h,t}; \lambda_{h,t})$ for $h = 1, \ldots, n$ are the cdfs of univariate conditional Poisson distributions.

From the vector form (4), tensor count autoregressions can be seen as multivariate count autoregressions with parameter constraints. In the usual specification of multivariate count autoregressions (Fokianos et al., 2020) the matrices A, B, Q would be $n \times n$ unconstrained matrices. Therefore, the total number of parameters to estimate is $M_1 = n + 3n^2 = n + 3\prod_{s=1}^{N} n_s^2$. Instead, in our specification such matrices are given by a set of Kronecker products, requiring the estimation of $m_1 = n + 3\sum_{s=1}^{N} n_s^2$ parameters. Even when the tensor order N and tensor dimensions n_s are small, it can easily be seen that $M_1 \gg m_1$. Therefore, multivariate count autoregressions are usually over-parametrized for tensor datasets, whereas model (1)-(2) is specifically suited for estimation with multidimensional integer-valued time series. This comes at the cost of identifiability issues in the elements of the coefficients matrices A_s (Chen et al., 2021; Li and Xiao, 2021). For example, it is clear from (4) that even though the product matrix A is identified, the model remains unchanged if two matrices A_s and A_l, with $s \neq l$, are divided and multiplied by the same nonzero constant, respectively.

3.2 Parameter identification and stationarity conditions

The parameters of model (1)-(2) are not univocally identifiable. In order to solve this problem, previous contributions have proposed relatively complex identifiability conditions on all the elements of the parameter matrices; see for example Li and Xiao (2021, Prop. 1). However, such conditions cannot easily be incorporated in Maximum Likelihood (ML) or Least Squares (LS) estimation. Iterative estimation algorithms have been proposed to overcome this problem, where ML/LS are sequentially applied on one parameter matrix at a time while keeping all other matrices fixed (Chen et al., 2021), leading to a cumbersome iterative updating scheme. Moreover, the proposed conditions typically only guarantee identification up to sign changes. Some authors (Billio et al., 2023) do not impose any identifiability constraints, arguing that the Kronecker product matrices in (4), like A, are identifiable. This approach implies renouncing the interpretation of the single matrices A_s and the computation of their standard errors, since the Fisher information matrix of ML/LS would be singular without identifiability constraints. We overcome these limitations by imposing the following identifying restrictions.

I For a set of matrices $\boldsymbol{A} = \{A_s\}_{s=1}^{N}$, define the single elements of $A_s = \left(A_{ij}^{(s)}\right)$ for $i, j = 1, \ldots, n_s$ and $s = 1, \ldots, N$, and let $A_{11}^{(N-1)} = \cdots = A_{11}^{(2)} = A_{11}^{(1)} = 1$.

Theorem 1. Consider model (1)-(2). If condition **I** holds for $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{Q}$, then the matrices $\{A_s, B_s, Q_s\}_{s=1}^{N}$ are identified.

This identifiability condition is conceptually simple and in line with similar restrictions employed in the literature, e.g. for factor models (Bai and Li, 2012). Moreover, it achieves full unequivocal identification as it is also sensitive to the sign of the parameters. Finally, it allows valid simultaneous inference for all the parameter matrices of the model, thereby avoiding time consuming iterative updating algorithms.

We develop sufficient stationarity conditions of the copula tensor count autoregressions (1)-(2) by employing the vector form (4). Let $u = (u_1, \dots, u_n)'$ denote an arbitrary random vector with standard uniform marginals and define $F_h^{-1}(u_h; \lambda_h) = F_h^{-1}(u_h; g^{-1}(\nu_h)) = F_{\nu_h}^{-1}(u_h)$.

S1 The process $\{u_t, x_t\}_{t\in\mathbb{Z}}$ is strictly stationary and ergodic, and the sequence u_t is independent from $\{u_\tau, x_\tau, \tau \leq t-1\}$.

S2 $f \circ F_{\nu_h}^{-1}(\cdot)$ and x_t are integrable.

S3 For $h = 1, \dots, n$, there exists a constant $s_h > 0$ such that for any $(\nu_h, \nu_h^*) \in \mathbb{R}^2$,

$$\mathrm{E}\left| f \circ F_{\nu_h}^{-1}(u_h) - f \circ F_{\nu_h^*}^{-1}(u_h) \right| \leq s_h |\nu_h - \nu_h^*| \, .$$

S4 $\rho(|A|_e S + |B|_e) < 1$ where $S = \mathrm{diag}(s_1, s_2, \dots, s_n)$.

Proposition 1. Let assumptions **S1**-**S4** hold. Then, there exists a unique solution $\{y_t, \nu_t\}_{t\in\mathbb{Z}}$ of (4) which is stationary and ergodic with $\mathrm{E}\,\|\nu_t\| < \infty$. Moreover, $\{y_t, \nu_t, x_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic.

The result follows from Debaly and Truquet (2023, Cor. 1). Condition **S1** imposes joint stationarity and ergodicity of $\mathcal{U}_t$ and $\mathcal{X}_t$. It implies that $\mathcal{X}_t$ is not necessarily strictly exogenous i.e. $\mathcal{U}_t$ and $\mathcal{X}_t$ are allowed to be dependent random variables. As a result, the tensor of interest $\mathcal{Y}_t$ is not necessarily independent of $\mathcal{X}_t$ conditional on the past $\mathcal{F}_{t-1}$, and hence contemporaneous dependence is allowed. A further assumption is that $\mathcal{U}_t$ is independent of the past. This condition guarantees that $\mathcal{Y}_{i_1 \dots i_N, t}$ conditional on the past has a Poisson distribution with mean $\Lambda_{i_1 \dots i_N, t}$. Conditions **S2**-**S4** depend on the function $f(\cdot)$. See section 6 below for some relevant examples. Finally, it is worth noting that the stationarity conditions of the tensor model do not depend on the copula, but only on the margins.

3.3 Two-stage inference

Here we discuss the estimation of the unknown parameters of the tensor model (1)-(2), denoted by $\psi = (\theta', r')' \in \mathbb{R}^m$, where $\theta = (d', \mathrm{vec}(A_1)', \dots, \mathrm{vec}(A_N)', \mathrm{vec}(B_1)', \dots, \mathrm{vec}(Q_N)')' \in \mathbb{R}^{m_1}$ are the tensor mean parameters, while $r \in \mathbb{R}^{m_2}$ are the copula parameters, such that $m = m_1 + m_2$. If we employ the Gaussian copula described in (3), the multivariate cdf of the vectorized tensor y_t takes the form (5) and its corresponding probability mass function (pmf) can be written as

$$p(y_t; \lambda_t, r) = \int_{\Phi^{-1}[F_1(y_{1,t-1};\lambda_{1,t})]}^{\Phi^{-1}[F_1(y_{1,t};\lambda_{1,t})]} \cdots \int_{\Phi^{-1}[F_n(y_{n,t-1};\lambda_{n,t})]}^{\Phi^{-1}[F_n(y_{n,t};\lambda_{n,t})]} \phi_r(x_1, \ldots, x_n) dx_1 \ldots dx_n, \tag{6}$$

where for brevity we let $\lambda_t = \lambda_t(\theta)$ and let $\phi_r(\cdot)$ denote the density of a multivariate standard normal distribution with correlation matrix R, such that $r = \text{vech}(R)$.

The full maximum likelihood estimation of ψ is obtained by maximizing the log-likelihood based on $\log[p(y_t; \lambda_t, r)]$ with respect to its m parameters. Note that with Gaussian copula $m_2 = n(n-1)/2$ and therefore m can grow at speed $\mathcal{O}(n^2)$ due to the copula parameters. This large dimension of the parameter vector would adversely affect the finite sample performance of the ML estimator. Hence, we propose a two-stage estimation, which in the literature is sometimes called Inference Functions for Margins (IFM); see Joe (2014). In this approach the parameters associated to the means of the process, θ, are estimated in a first stage by Quasi-Maximum Likelihood (QML) estimation employing the marginal distributions of the tensor. Next, the copula parameter r is estimated in a second stage by plugging the first-stage estimated means, say $\hat{\lambda}_t$, in (6) and maximizing the resulting log-likelihood function. This approach is popular for copula estimation and is conveniently adopted for numerical efficiency and feasibility. Indeed, a numerical optimization with many parameters is replaced by two smaller numerical optimization problems. Furthermore, the estimates of the mean parameters of the margins will be robust against misspecification of the copula function. The next two sections describe in detail the estimation procedure and corresponding asymptotic theory of the above two-stage procedure for the mean and copula related parameters, respectively, for the new model for integer-valued tensor time series.

4 Estimation of mean parameters

In the two-stage approach, the parameter $\theta \in \mathbb{R}^{m_1}$ associated to the means of the count tensor process, $\Lambda_t(\theta)$, are estimated in a first stage by QML estimation. In this section, we derive the asymptotic properties of the QML estimator for θ. For notational convenience, we use the vectorized notation described in section 3.1 where possible, but the results hold equivalently for model (1)-(2) in tensor form.

The conditional time-varying parameter of the tensor process in vectorized form, $\nu_t(\theta) = g(\lambda_t(\theta))$, can alternatively be defined by the following stochastic recurrence equation (SRE):

$$\nu_t(\theta) = h_\theta\big(y_{t-1}, x_{t-1}; \nu_{t-1}(\theta)\big), \qquad t \in \mathbb{Z}, \tag{7}$$

where $h_\theta(\cdot,\cdot,\cdot)$ is the parametric updating function defined in (4) that maps from $\mathbb{N}^n \times \mathbb{R}^n \times \mathbb{R}^n$ into $\mathbb{R}^n$. In the sample, only a subset of the realized path of the process (4), or equivalently (2), is observed, with true parameter value $\theta = \theta_0$. Put differently, we only know $\{y_t, x_t\}_{t=0}^T$ where T is the sample size. Therefore, estimation of the unknown parameters is performed under an approximation of the time-varying process, say $\tilde{\nu}_t(\theta)$, for $t \in [1, T]$, which is obtained recursively by using the observed tensor data $\{y_t, x_t\}_{t=0}^{T-1}$ as follows:

$$\tilde{\nu}_t(\theta) = h_\theta\big(y_{t-1}, x_{t-1}; \tilde{\nu}_{t-1}(\theta)\big), \tag{8}$$

where the recursion is initialized at a fixed point $\tilde{\nu}_0(\theta) \in \mathbb{R}^n$. We note that initializing the recursion in equation (8) is required, since the observed data start from $t = 0$. This is a standard procedure in the literature of observation-driven models. Consequently, the marginal log-quasi-likelihood employed for the estimation is defined as

$$\tilde{L}_T(\theta) = T^{-1} \sum_{t=1}^{T} \tilde{l}_t(\theta) = T^{-1} \sum_{t=1}^{T} \sum_{h=1}^{n} \log\left[p_h\big(y_{h,t}; \tilde{\lambda}_{h,t}(\theta)\big)\right], \tag{9}$$

where $p_h(y_{h,t}; \tilde{\lambda}_{h,t}(\theta))$ for $h = 1, \dots, n$ are the pmfs of conditional Poisson distributions with the mean process substituted by the proxy $\tilde{\lambda}_{h,t}(\theta) = g^{-1}(\tilde{\nu}_{h,t}(\theta))$. Then, the first-stage Quasi-Maximum Likelihood Estimator (QMLE) is defined as

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} \tilde{L}_T(\theta), \tag{10}$$

where $\Theta \subseteq \mathbb{R}^{m_1}$ is the mean parameter space.

4.1 Consistency of QMLE

In order to determine the strong consistency of the QMLE, we first need to derive the stochastic limit properties of the time-varying parameter $\tilde{\nu}_t$ defined in expression (8). Note that $\tilde{\nu}_t(\cdot)$ is a stochastic function that maps from Θ into $\mathbb{R}^n$. Due to the initialization, $\tilde{\nu}_t$ evaluated at the true parameter value θ_0, i.e., $\tilde{\nu}_t(\theta_0)$, generally does *not* correspond to the true $\nu_t = g(\lambda_t)$, where λ_t is the true conditional mean. However, we prove that $\{\tilde{\nu}_t(\theta)\}_{t\in\mathbb{N}}$ converges *exponentially fast almost surely (e.a.s.)* and uniformly in Θ to the stationary and ergodic sequence of functions $\{\nu_t(\theta)\}_{t\in\mathbb{N}}$ as defined in (7), such that $\nu_t(\theta_0) = \nu_t$ almost surely (a.s.). According to the definition of Straumann and Mikosch (2006), a sequence of random variables $\{\tilde{x}_t\}_{t\in\mathbb{N}}$ converges e.a.s. to another sequence $\{x_t\}_{t\in\mathbb{N}}$ if there is a constant $\delta > 1$ such that $\delta^t \|\tilde{x}_t - x_t\| \to 0$ almost surely as $t \to \infty$. For a vector function $f : \Theta \mapsto \mathbb{R}^n$, the supremum norm is defined as $\|f\|_\Theta = \sup_{\theta\in\Theta} \|f(\theta)\|$. Similarly, for a matrix function $F : \Theta \mapsto \mathbb{R}^{n_1 \times n_2}$, $|||F|||_\Theta = \sup_{\theta\in\Theta} |||F(\theta)|||$.

A1 The conditions **S1-S4** hold. Moreover, $\rho(A_N \otimes A_{N-1} \otimes \cdots \otimes A_1) < 1$ for any $\theta \in \Theta$, where Θ is compact and $\theta_0 \in \Theta$.

Proposition 2. Let Assumption **A1** hold. Then, the sequence $\{\tilde{\nu}_t(\theta)\}_{t \in \mathbb{N}}$ converges e.a.s. and uniformly over Θ to a unique stationary and ergodic sequence $\{\nu_t(\theta)\}_{t \in \mathbb{Z}}$, i.e.,

$$\|\tilde{\nu}_t - \nu_t\|_\Theta \to 0 \ \text{(e.a.s.)},$$

as $t \to \infty$, for any fixed initialization $\tilde{\nu}_0(\theta) \in \mathbb{R}^n$. Furthermore, $\{\nu_t(\theta)\}_{t \in \mathbb{Z}}$ is $\mathcal{F}_{t-1}$ measurable with a bounded uniform moment $\mathrm{E}\|\nu_t\|_\Theta < \infty$, and $\nu_t(\theta_0) = \nu_t$.

The property stated in Proposition 2 is sometimes called *invertibility*. It states that the difference between feasible approximate process (8) based on the arbitrary initialization on the one hand, and the original process (7) other hand, becomes asymptotically negligible at an exponential rate as we move away from $t = 0$. For a vector $x \in \mathbb{R}^p$, we write $x \nsim 0$ to denote that all elements of x are not 0.

A2 The map $g : \mathbb{R}^n_+ \mapsto \mathbb{R}^n$ is continuous and invertible. Moreover, $\|g^{-1}(\tilde{\nu}_t) - g^{-1}(\nu_t)\|_\Theta \to 0$, e.a.s. as $t \to \infty$, for any initialization $\tilde{\nu}_0(\theta) \in \mathbb{R}^n$.

A3 $\mathrm{E}\left\|\log^+(y_t)\right\| < \infty, \quad \mathrm{E}\|\lambda_t\|_\Theta < \infty, \quad \mathrm{E}\|y_t' \log(\lambda_t)\|_\Theta < \infty.$

A4 There exists a constant vector $\underline{\lambda} > 0$, such that the function $(\theta, \nu) \mapsto g^{-1} \circ h_\theta(y, x; \nu)$ is uniformly lower-bounded, i.e., $g^{-1} \circ h_\theta(y, x; \nu) \geq \underline{\lambda}$ for any $(\theta, y, x, \nu) \in \Theta \times \mathbb{N}^n \times \mathbb{R}^n \times \mathbb{R}^n$.

A5 Condition **I** holds for $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{Q}$ for any $\theta \in \Theta$. Furthermore, $|B_s|_e 1 \nsim 0$ or $|Q_s|_e 1 \nsim 0$ at $\theta = \theta_0$, for any $s = 1, \ldots, N$.

Condition **A2** is needed, together with the results of Proposition 2, to ensure that our computable log-likelihood (9) that depends on the approximate process $\tilde{\lambda}_t = g^{-1}(\tilde{\nu}_t)$, converges to a stationary and ergodic limit $L_T(\theta) = T^{-1}\sum_{t=1}^T l_t(\theta) = T^{-1}\sum_{t=1}^T \sum_{h=1}^n \log[p_h(y_{h,t}; \lambda_{h,t}(\theta))]$, where the latter depends on the limit sequence $\lambda_t = g^{-1}(\nu_t)$ rather than on its initialized counterpart. Assumption **A3** is a standard moment condition, which guarantees that the limiting log-likelihood $L_T(\theta)$ converges to the expected log-likelihood $L(\theta) = \mathrm{E}[l_t(\theta)]$, uniformly over the compact Θ. Assumption **A4** imposes a lower bound on the updating function, such that the vectorized tensor mean and its proxy are lower-bounded, i.e., $\lambda_{h,t}(\theta) \geq \underline{\lambda}_L$, $\tilde{\lambda}_{h,t}(\theta) \geq \underline{\lambda}_L$ for any $\theta \in \Theta$, and $t = 1, \ldots, T$, where $\underline{\lambda}_L = \min_{h=1,\ldots,n} \underline{\lambda}_h$ and $\underline{\lambda}_h$ is a single element of $\underline{\lambda}$. Finally, condition **A5** is an identification condition. The first part requires that the parameter matrices for the mean dynamics are identified. The second part implies that the time-varying parameter ν_t is not degenerate. Indeed, we will show that when ν_t is a degenerate random variable, then the set of matrices $\boldsymbol{A}$ in model (2) is not identifiable. Together, these two conditions determine the identifiability of the tensor mean, i.e., the tensor process $\Lambda_t(\theta)$ generates a different tensor path for each $\theta \neq \theta_0$ and $\Lambda_t(\theta) = \Lambda_t(\theta_0)$ if and only if $\theta = \theta_0$.

Theorem 2. Let Assumptions **A1**-**A5** hold true. Then the QMLE defined in (10) is strongly consistent, i.e., $\hat{\theta}_T \to \theta_0$ almost surely as $T \to \infty$.

4.2 Asymptotic normality of QMLE

In this section we derive the asymptotic normality of the QML estimator. For first and second derivatives of a scalar function $f(\cdot)$ with respect to a vector parameter γ, we use the notation $\nabla_\gamma f = \partial f(\gamma)/\partial\gamma$ and $\nabla_\gamma^2 f = \partial^2 f(\gamma)/\partial\gamma\partial\gamma'$.

A6 The map $g : \mathbb{R}_+^n \mapsto \mathbb{R}^n$ is twice continuously differentiable. Moreover, $\theta_0 \in \dot{\Theta}$ where $\dot{\Theta}$ is the interior of Θ.

A7 For any $h = 1, \dots, n$,

$$\|\nabla_{\tilde{\nu}_{h,t}} g^{-1} - \nabla_{\nu_{h,t}} g^{-1}\|_\Theta \to 0\,,$$

e.a.s. as $t \to \infty$, for any initialization $\tilde{\nu}_0(\theta) \in \mathbb{R}^n$. Moreover,

$$\mathrm{E}\left(\log^+ \|\nabla_\theta \lambda_{h,t}\|_\Theta\right) < \infty\,, \qquad \mathrm{E}\left(\log^+ \left\|\nabla_{\nu_{h,t}} g^{-1}\right\|_\Theta\right) < \infty\,.$$

A8 The elements of $(y_t', x_t')'$ are linearly independent random variables. Moreover, at $\theta = \theta_0$, $A_{11}^{(N)}$, $B_{11}^{(N)}$ and $Q_{11}^{(N)}$ are different from 0.

A9 For any $h, l = 1, \dots, n$, with $h \neq l$,

$$\mathrm{E}(y_{h,t}^2) < \infty\,,\ \ \mathrm{E}\left|\left|\left|\nabla_\theta \lambda_{h,t} \nabla_\theta \lambda_{h,t}'\right|\right|\right|_\Theta^2 < \infty\,,\ \ \mathrm{E}\left|\left|\left|\nabla_\theta^2 \lambda_{h,t}\right|\right|\right|_\Theta^2 < \infty\,,\ \ \mathrm{E}\left|\left|\left|\nabla_{\theta_0} \lambda_{h,t} \nabla_{\theta_0} \lambda_{l,t}'\right|\right|\right| < \infty\,.$$

Condition **A6** allows the tensor mean process $\lambda_t(\theta)$ to be differentiable up to second order. Moreover, it implies that the derivative of the limit objective function will be zero at θ_0. Assumption **A7** together with the results of Proposition 2 guarantees that for the score of the log-quasi-likelihood the starting value of the initialized series $\tilde{\nu}_0(\theta)$ is asymptotically negligible. Condition **A8** is the tensor analogue of the classical linear independence assumption, which is required to obtain a positive definite Hessian matrix. Finally, Condition **A9** is a standard moment condition that ensures the existence of the asymptotic Hessian and Fisher information matrices, $H(\theta_0)$ and $I(\theta_0)$, respectively, which are defined below.

Theorem 3. Let Assumptions **A1-A9** hold true. Then the QMLE defined in (10) is asymptotically normal, i.e.,

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N(0, \Sigma) , \tag{11}$$

as $T \to \infty$, where

$$\Sigma = H^{-1}(\theta_0) I(\theta_0) H^{-1}(\theta_0) , \qquad H(\theta_0) = -\mathrm{E}\left(\nabla^2_{\theta_0} l_t\right) , \qquad I(\theta_0) = \mathrm{E}\left(\nabla_{\theta_0} l_t \, \nabla_{\theta_0} l_t'\right) ,$$

and Σ is positive definite.

5 Estimation of copula parameters

5.1 Definition of the estimator and consistency result

In this section we prove the consistency of the second-stage estimator of the copula correlation parameter $r = \mathrm{vech}(R)$ based on the Gaussian copula pmf as defined in (6) and obtained by plugging-in the first-stage estimates of the vectorized tensor means $\hat{\lambda}_t = \lambda_t(\hat{\theta}_T)$, where $\hat{\theta}_T$ is the first-stage QMLE of θ. The resulting estimator is called the Two-Stage Maximum Likelihood Estimator (2SMLE). The corresponding second-stage log-likelihood is defined as

$$L_T(\hat{\theta}_T, r) = T^{-1} \sum_{t=1}^{T} l_t(\hat{\theta}_T, r) = T^{-1} \sum_{t=1}^{T} \log\left[p\big(y_t; \hat{\lambda}_t, r\big)\right] , \tag{12}$$

and the 2SMLE as

$$\hat{r}_T = \operatorname*{arg\,max}_{r \in \mathsf{P}^v} L_T(\hat{\theta}_T, r) , \tag{13}$$

where P is the space of all possible correlation matrices and $\mathsf{P}^v \subseteq \mathbb{R}^{m_2}$ is the space of vectorized (strictly lower triangular half) of the correlation matrices in P. Let R_0 denote the true correlation matrix of the Gaussian copula and $r_0 = \mathrm{vech}(R_0)$. Moreover, for a vector function $f : \Theta \mapsto \mathbb{R}^n$, let $\|f\|_{p,\Theta} = \sup_{\theta \in \Theta} \|f(\theta)\|_p$.

B1 P^v is compact and $r_0 \in \mathsf{P}^v$.

B2 $\mathrm{E}\|\lambda_t\|^2_{2,\Theta} < \infty$.

Theorem 4. Let assumptions **A1-A5** and **B1-B2** hold true. Then the 2SMLE defined in (13) is strongly consistent, i.e., $\hat{r}_T \to r_0$ almost surely as $T \to \infty$.

Remark 1. (Identification of R) Following Genest and Nešlehová (2007), when coupling a copula with discrete marginals, the copula is not unique, meaning that the same joint cdf can be generated using a different copula function with the same discrete marginals, but estimation of copula parameters remains possible if these are identifiable. In Lemma C.9 we show that the correlation parameters for the Gaussian copula (3) are identified in the tensor model (1)-(2), leading to consistent second-stage estimates of R.

Remark 2. (Interpretation of R) Note that when copulas are employed with discrete pmfs, the copula parameters can still be interpreted as dependence parameters (Genest and Nešlehová, 2007). In particular, the Gaussian copula (3) implies that the conditional distribution of $(u_{i,t}, u_{j,t})'$ follows a bivariate Gaussian copula with correlation parameter R_{ij} being the (i, j) entry of R. Therefore, by the results of Genest and Nešlehová (2007, Sec. 5.2) and the Hoeffding's covariance identity, we have that the pairwise (i, j) conditional covariance of the counts, say $\mathrm{Cov}(y_{i,t}, y_{j,t}|\mathcal{F}_{t-1})$, is increasing with respect to R_{ij}. Moreover, by the same arguments and Genest and Nešlehová (2007, Prop. 11), the sign of R_{ij} equals to the sign of the respective (i, j)th conditional covariance. Therefore, although R_{ij} does not correspond to the conditional pairwise correlation of $(y_{i,t}, y_{j,t})$, its sign and relative comparison can still be interpreted with respect to the true conditional correlations. By this we mean that if $R_{ij} < R_{ik}$, then the conditional correlation for pair (i, k) is higher than that of (i, j).

The results of Sections 4 and 5 fill a crucial gap in the existing literature by offering a theoretical foundation for tensor count time series models. To the best of our knowledge, this is the first asymptotic theory result established for integer-valued tensor time series processes. Additionally, our result covers an existing gap within the classical framework of multivariate time series. Previous literature has described the asymptotic theory regarding QMLE-2SMLE in the context of mixed multivariate time series models (Debaly and Truquet, 2023). However, the result for the QMLE was stated without proof and does not directly extend to tensors.

5.2 Feasibility in large tensor dimensions

The computation of the second-stage likelihood based on the pmf (6) requires the numerical evaluation of $n = n_1 n_2 \cdots n_N$ integrals and its optimization with respect to $n(n-1)/2$ copula parameters. The evaluation of multi-dimensional integrals is numerically time-consuming and may easily result in a numerically infeasible or impractical algorithm. In this section we propose two procedures to guarantee a feasible estimation of the correlation matrix R when the tensor dimensions are large.

Specification (i): Reduction of the parameter space. An immediate solution is to constrain the structure of R in order to decrease the dimension of the parameter vector r. For example, a simple and effective structure would be $R = R_N \otimes R_{N-1} \otimes \cdots \otimes R_1$ with $R_s = r_s I_{n_s} + (1 - r_s) J_{n_s}$, where $J_{n_s} \in \mathbb{R}^{n_s \times n_s}$ denotes the unit matrix. Indeed, according to (3) and by recalling that $u_t = \text{vec}(\mathcal{U}_t)$, we have $\Phi^{-1}(u_t)|\mathcal{F}_{t-1} \sim N(0, R_N \otimes R_{N-1} \otimes \cdots \otimes R_1)$ with the function $\Phi^{-1}(\cdot)$ applied elementwise. This is equivalent to assuming $\Phi^{-1}(\mathcal{U}_t) = \bar{\mathcal{Z}}_t \times_1 R_1^{1/2} \times_2 \cdots \times_N R_N^{1/2}$, where $\bar{\mathcal{Z}}_t$ is a tensor whose elements have zero mean, unit variance, and are uncorrelated, such that $\text{vec}(\bar{\mathcal{Z}}_t)|\mathcal{F}_{t-1} \sim N(0, I_n)$. This structure reproduces the data tensor structure in the correlation matrix. Intuitively, R_s corresponds to the mode s interactions, for $s = 1, \ldots, N$. Therefore, the proposed structure can be called a *tensor equicorrelation* model as it induces for each tensor dimension an equicorrelation structure with equicorrelation parameter $r_s \in (-1/(n_s - 1), 1)$ for $s = 1, \ldots, N$. Moreover, note that the set of matrices $\boldsymbol{R} = \{R_s\}_{s=1}^N$ satisfies condition **I** by definition. Therefore, the parameters r_s are identified. Finally, the additional constraint $r_s > -1/(n_s - 1)$ ensures the positive definiteness of R_s, for $s = 1, \ldots, N$, and therefore guarantees that R is a valid correlation matrix. We see in the application section 7 that the parameters r_s can be interpreted as s-mode dependence parameters. The tensor equicorrelation form allows us to reduce the number of copula parameters from $n(n-1)/2$ to only N. Alternatively, a further reduction of the copula parameter space to a single parameter $\bar{r}$ can be considered by specifying an equicorrelation structure $R = \bar{r} I_n + (1 - \bar{r}) J_n$, with $\bar{r} \in (-1/(n-1), 1)$.

Specification (ii): Numerical approximation of the likelihood. Although the previous approach allows us to reduce the parameters space, the optimization of the likelihood based on the multidimensional integral (6) may still be numerically unfeasible when N and/or n_s grow. To avoid the costly evaluations of the copula pmf in equation (6), we employ a fast numerical approximation. Using basic properties of integrals, the copula pmf can be rewritten as the integral of the copula density over a hyper-rectangular region:

$$p(y_t; \lambda_t, r) = \int_{F_1(y_{1,t}-1; \lambda_{1,t})}^{F_1(y_{1,t}; \lambda_{1,t})} \cdots \int_{F_n(y_{n,t}-1; \lambda_{n,t})}^{F_n(y_{n,t}; \lambda_{n,t})} c(u_1, \ldots, u_n; r) du_1 \ldots du_n,$$

where for brevity we again let $\lambda_t = \lambda_t(\theta)$, and where $c(u_1, \ldots, u_n; r)$ denotes the Gaussian copula density function with correlation parameters r. We approximate this integral by using the midpoint rule, obtaining the volume of the n-dimensional rectangular region times the value of the copula density evaluated at the midpoint of the same region, i.e.

$$\begin{aligned} p(y_t; \lambda_t, r) \approx \hat{p}(y_t; \lambda_t, r) &= c(u_{1,t}^*, \ldots, u_{n,t}^*; r) \prod_{h=1}^{n} \left[F_h(y_{h,t}; \lambda_{h,t}) - F_h(y_{h,t} - 1; \lambda_{h,t}) \right] \\ &= c(u_{1,t}^*, \ldots, u_{n,t}^*; r) \prod_{h=1}^{n} p_h(y_{h,t}; \lambda_{h,t}) \, , \end{aligned}$$

where $u_{h,t}^* = 0.5 \left[F_h(y_{h,t} - 1; \lambda_{h,t}) + F_h(y_{h,t}; \lambda_{h,t}) \right]$ are the midpoints for $h = 1, \ldots, n$. Therefore, the approximated second-stage log-likelihood with respect to the copula parameters r would be

$\hat{L}_T(\hat{\theta}_T, r) = T^{-1}\sum_{t=1}^{T}\log[\hat{p}(y_t; \hat{\lambda}_t, r)]$, but the univariate pmfs can be neglected since they do not contain the copula parameters. Therefore, by recalling that $r = \mathrm{vech}(R)$, we have

$$\hat{L}_T(\hat{\theta}_T, r) = T^{-1}\sum_{t=1}^{T}\log[c(\hat{u}^*_{1,t}, \ldots, \hat{u}^*_{n,t}; r)] = -\frac{1}{2}\log|R| + \frac{1}{2T}\sum_{t=1}^{T}\hat{\omega}_t'\left(I_n - R^{-1}\right)\hat{\omega}_t\,, \tag{14}$$

where $\hat{\omega}_t = (\hat{\omega}_{1,t}, \ldots, \hat{\omega}_{n,t})'$, $\hat{\omega}_{h,t} = \Phi^{-1}(\hat{u}^*_{h,t})$ and $\hat{u}^*_{h,t} = 0.5\left[F_h(y_{h,t}-1; \hat{\lambda}_{h,t}) + F_h(y_{h,t}; \hat{\lambda}_{h,t})\right]$. The approximation (14) leads to a substantial decrease in computational costs. Furthermore, in the case of (tensor) equicorrelation, it is possible to derive analytical expressions for the determinant and inverse of the correlation matrix. This yields further computational advantages. See Theorems 8.3.4 and 8.4.4 in Graybill (1983).

When the interest is in obtaining a feasible estimate of the full unconstrained correlation matrix R, the optimization problem is cumbersome and the number of correlation parameters to be estimated is too high. In these cases, it is possible to obtain an approximate solution that does not require any optimization. Indeed, if we maximize (14) with respect to the space of all covariance matrices, this maximization problem has the following analytical solution

$$\hat{r} = \mathrm{vech}(\hat{R})\,, \qquad \hat{R} = T^{-1}\sum_{t=1}^{T}\hat{\omega}_t\hat{\omega}_t', \tag{15}$$

which is the MLE of the covariance matrix R for the standard normal likelihood of the variables $\hat{\omega}_t$. Although the use of the estimator (15) does not guarantee a valid correlation matrix, in practice $\hat{R}$ is likely to be close to being a correlation matrix. Then, we can take the adjusted estimator

$$\tilde{r} = \mathrm{vech}(\tilde{R})\,, \qquad \tilde{R} = \hat{\Delta}^{-1}\hat{R}\hat{\Delta}^{-1} \tag{16}$$

where $\hat{\Delta} = \mathrm{diag}(\hat{R})^{1/2}$ is a diagonal matrix constructed by taking the square root of the elements in the main diagonal of $\hat{R}$. This approximate solution to the original problem is a convenient option for obtaining a feasible estimate of the whole correlation matrix without constraints. Moreover, the adjusted estimator is in general available for any arbitrary tensor order and dimension since it does not require any optimization.

In Section 6 we show by means of Monte Carlo simulation experiments that the approximation methods work well for linear and log-linear copula tensor count autoregressions. Moreover, in Section 7 we confirm using real empirical data that optimization based on the approximate likelihood produces estimates that are almost identical to those based on the true likelihood. To the best of our knowledge, we have not seen similar suggestions elsewhere in the literature to keep the computational burden of copula-based count tensor models feasible in high dimensions.

6 Copula tensor INGARCHX models

6.1 Examples of model designs

The dynamics described in the tensor model (1)-(2) specify a general class of tensor time series models through the choice of the link functions g, f. In this section we present some specific models of interest that are embedded in (2) and formulate their asymptotic theory results under low-level conditions.

An intuitive and simple way to specify the conditional tensor mean Λ_t is to consider a linear autoregressive process driven by past observations. Count processes with a linear specification of the conditional mean are often referred to in the literature as INGARCH models (Ferland et al., 2006). Let f and g be the identity mapping. We then define the copula tensor INGARCHX model by coupling the process (1) with the following dynamical mechanism,

$$\Lambda_t = \mathcal{D} + \Lambda_{t-1} \star \boldsymbol{A} + \mathcal{Y}_{t-1} \star \boldsymbol{B} + \mathcal{X}_{t-1} \star \boldsymbol{Q}, \tag{17}$$

for $t \in \mathbb{Z}$. This is a generalization of the multivariate INGARCH model (Fokianos et al., 2020) to multidimensional count time series, with the addition of exogenous regressors. Model (17) introduces linear dynamics in the time-varying mean parameters. In order to guarantee the almost sure positivity of conditional Poisson means, all the mean parameters and the exogenous regressors should be non-negative. The copula tensor INGARCHX model retains the vectorized notation

$$y_{h,t} = F_h^{-1}(u_{h,t}; \lambda_{h,t})\,, \quad 1 \leq h \leq n\,, \qquad \lambda_t = d + A\lambda_{t-1} + By_{t-1} + Qx_{t-1}\,, \tag{18}$$

for $t \in \mathbb{Z}$, which is a special case of (4) with the same Kronecker product-type restrictions on the coefficient matrices. Recall that the mean parameters are $\theta = (d', \mu')'$, with $\mu = (\mathrm{vec}(A_1)', \dots, \mathrm{vec}(Q_N)')'$, with true value $\theta_0 = (d_0', \mu_0')'$ and $\mu_0 = (\mathrm{vec}(A_{0,1})', \dots, \mathrm{vec}(Q_{0,N})')'$.

Theorem 5. Let the series $\{y_t\}_{t=0}^T$ be generated by model (18) with parameter value $\psi_0 = (\theta_0', r_0')'$, $\{u_t, x_t\}_{t\in\mathbb{Z}}$ satisfying **S1**, $x_t \geq 0$ a.s. for any $t \geq 0$, and $\mathrm{E}\|x_t\|^p < \infty$ for $p \geq 1$. Furthermore, let $\theta_0 \in \Theta$, where Θ is a compact parameter set such that $d > 0$, $\mu \geq 0$, $\rho(A) < 1$ for any $\theta \in \Theta$, condition **A5** is satisfied, and $\rho(A_0 + B_0) < 1$. Then, the QMLE is strongly consistent. Assume further that $r_0 \in \mathsf{P}^v$ where P^v is compact. Then, the 2SMLE is strongly consistent. Moreover, assume that that condition **A8** is satisfied, and $\theta_0 \in \dot{\Theta}$. Then, the QMLE is asymptotically normal.

The tensor model (17) does not allow one to estimate negative tensor coefficients and to introduce exogenous covariates that can also take negative values. This may be problematic for some empirical datasets. Therefore, we also propose a copula tensor log-linear INGARCHX model, which is specified by coupling (1) with

$$\log(\Lambda_t) = \mathcal{D} + \log(\Lambda_{t-1}) \star \boldsymbol{A} + \log(\mathcal{Y}_{t-1} + 1) \star \boldsymbol{B} + \mathcal{X}_{t-1} \star \boldsymbol{Q}\,, \tag{19}$$

for $t \in \mathbb{Z}$, which is obtained from (2) by setting $g(\cdot) = \log(\cdot)$ and $f(\cdot) = \log(\,\cdot + 1)$. Clearly, model (19) does not suffer from the limitations of the linear model and its vectorized version is defined for $t \in \mathbb{Z}$ as

$$y_{h,t} = F_h^{-1}(u_{h,t}; \lambda_{h,t})\,,\ 1 \le h \le n\,, \qquad \log(\lambda_t) = d + A\log(\lambda_{t-1}) + B\log(y_{t-1} + 1) + Qx_{t-1}\,. \tag{20}$$

Theorem 6. Let the series $\{y_t\}_{t=0}^T$ be generated by model (20) with parameter value $\psi_0 = (\theta_0', r_0')'$, $\{u_t, x_t\}_{t\in\mathbb{Z}}$ satisfying **S1**, and $\mathrm{E}\left[\exp(p\|x_t\|)\right] < \infty$ for $p \ge 1$. Furthermore, let $\theta_0 \in \Theta$, where Θ is a compact parameter set such that $\rho(A) < 1$ for any $\theta \in \Theta$, conditions **A4**-**A5** are satisfied, and $\||\,|A_0|_e + |B_0|_e|\||_\infty < 1$. Then, the QMLE is strongly consistent. Further assume that $r_0 \in \mathsf{P}^v$ where P^v is compact. Then, the 2SMLE is strongly consistent. Moreover, assume that condition **A8** is satisfied, and $\theta_0 \in \dot{\Theta}$. Then, the QMLE is asymptotically normal.

By employing a backward substitution argument, it is straightforward to prove that a sufficient condition for **A4** to hold for the log-linear model is that the parameter matrices $\{A, B, Q\}$ and the covariates x_t in (20) have non-negative entries.

6.2 Monte Carlo simulations

We investigate the finite sample properties of QMLE and 2SMLE by means of Monte Carlo simulation experiments. We generate 1000 samples from the copula tensor INGARCH model (1) plus (17) for several sample sizes ($T = 500, 1000, 2000$). The tensors are generated with order $N = 2$ and dimensions $n_1 = n_2 = 2$. The Gaussian copula is employed with an equicorrelation structure, where $\bar{r} = 0.5$. The tensor mean parameters are estimated using the QMLE (10). The first-stage estimated tensor means in $\hat{\Lambda}_t$ are subsequently plugged-in the 2SMLE (13) for the estimation of the copula parameter. In the second stage we optimize the approximated likelihood (14) to test its effectiveness in finite samples. The data are generated with the following set of true parameter values $d = (0.5, 0.667, 0.833, 1)'$,

$$A_1 = \begin{pmatrix} 1 & 0.2 \\ 0.15 & 0.3 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.3 & 0.2 \\ 0.175 & 0.2 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 0.2 \\ 0.15 & 0.3 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0.3 & 0.25 \\ 0.2 & 0.25 \end{pmatrix}.$$

The values of $A_{11}^{(1)}$ and $B_{11}^{(1)}$ are set to 1 to satisfy the identification constraint **I** and are thus not estimated.

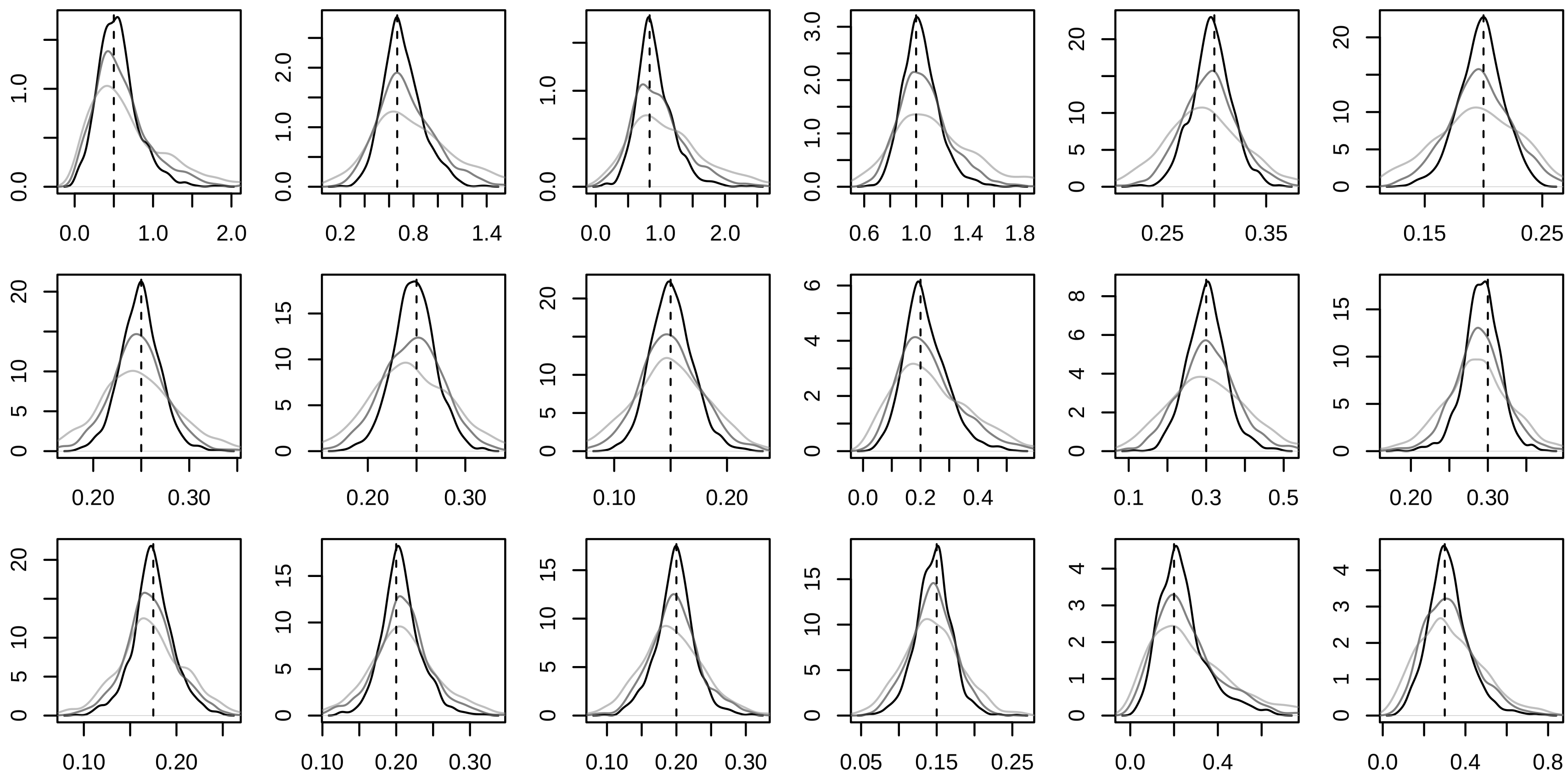

Figure 1: Kernel density estimator for the distribution of the QML estimates for the copula tensor INGARCH model with equicorrelation structure obtained from 1000 Monte Carlo replications for sample sizes 500 (——), 1000 (——) and 2000 (——), true parameter values (- - - -); the bandwidth of the kernel is selected by using Silverman's rule of thumb; the plots are reported rowwise following the order of parameter vector $\theta = (d', \mathrm{vec}(B_2)', \mathrm{vec}(B_1)' \, \mathrm{vec}(A_2)', \mathrm{vec}(A_1)')'$.

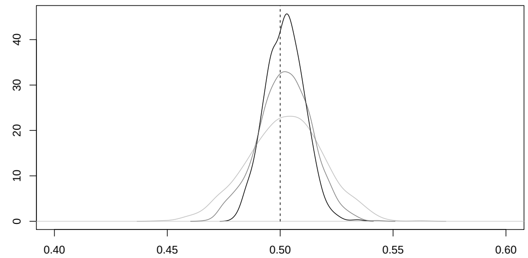

Figure 2: Kernel density estimator for the distribution of the 2SQML estimates for the copula tensor INGARCH model with equicorrelation structure obtained from 1000 Monte Carlo replications for sample sizes 500 (——), 1000 (——) and 2000 (——), true parameter value (- - - -); the bandwidth of the kernel is selected by using Silverman's rule of thumb.

Figure 1 reports the kernel density of the QML estimates for the remaining parameters. The results confirm that the estimates are consistent, as their distributions collapse towards the true parameter values as the sample size increases. Furthermore, the distributions of the QML estimators seem symmetric for almost all parameters, in particular at larger sample sizes. This suggests that the asymptotic normal distribution is an accurate approximation. Figure 2 shows the kernel density of the 2SQML estimate for the copula parameter. The empirical distribution shows a slight bias, which is expected due to the uncertainty of the first-stage estimates and the use of the approximate likelihood. Nevertheless, the bias is minimal and the consistency of the second-stage estimate is confirmed given the distribution collapses towards the true parameter value as the sample size increases. These conclusions are also supported by further simulation results given in the supplementary material, together with an analogous simulation study for the copula tensor log-linear INGARCH model.

7 Empirical application

In this section, we present an empirical application to a 2-order tensor of monthly police report counts for five types of theft-related crimes ($n_1 = 5$) in five different cities of Australia ($n_2 = 5$), for a total of $n = n_1 n_2 = 25$ time series. The tensor time series is from January 1995 to March 2024 (Figure 3) and is part of the New South Wales (NSW) data set of police reports[1] with geographic breakdown by local government areas (LGA). The cities were chosen based on similar population sizes. Together they form the most populated LGAs in the Sydney metropolitan area. Plots of the autocorrelation functions of the series (reported in the supplementary material) reveal high

[1]Data available at https://bocsar.nsw.gov.au/statistics-dashboards/open-datasets/criminal-offences-data.html

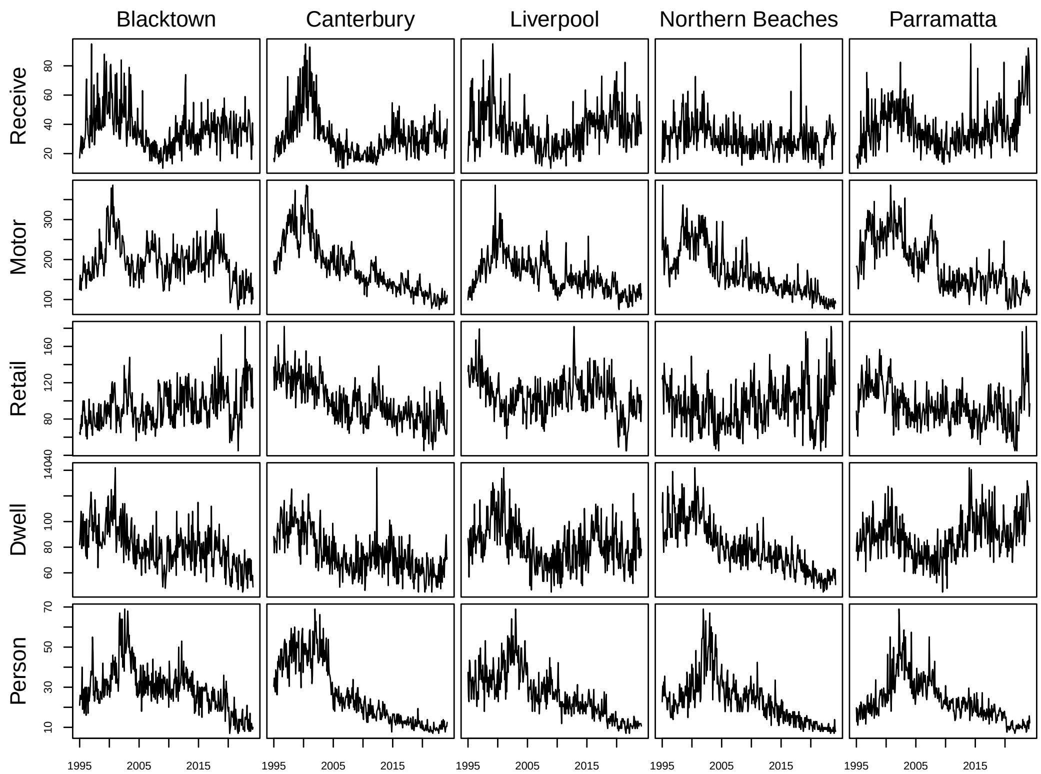

Figure 3: Monthly number of police reports for five theft-related crimes (Receive: Receiving or handling stolen goods; Motor: Steal from motor vehicle; Retail: Steal from retail store; Dwell: Steal from dwelling; Person: Steal from person) in five cities of Australia (Blacktown; Canterbury-Bankstown; Liverpool; Northern Beaches; Parramatta) from January 1995 to March 2024.

autocorrelations for the crime counts time-series. Therefore, tensor INGARCH dynamics seem particularly suited to describe the autocorrelation structure of the data.

Since the tensor order is $N = 2$, the modelling approach is encompassed by the family of copula tensor count autoregressions (A-1)-(A-2) described in Appendix A. In this case, recall that $\mathcal{Y}_t = (\mathcal{Y}_{ij,t})$ is an $n_1 \times n_2$ matrix of crime count time series and $\Lambda_t = (\Lambda_{ij,t})$ the corresponding matrix of conditional Poisson means. To further reduce the parameter dimensionality and simplify the model, we specify the tensor INGARCH dynamics as

$$\Lambda_t = D + \alpha \Lambda_{t-1} + B_1 \mathcal{Y}_{t-1} B_2' \,, \tag{21}$$

$$\log(\Lambda_t) = D + \alpha \log(\Lambda_{t-1}) + B_1 \log(\mathcal{Y}_{t-1} + 1) B_2' \,, \tag{22}$$

which are special cases of (A-2) with $A_1 = A_2 = \sqrt{\alpha} I_5$, where α is a scalar. Spillovers between the Poisson conditional means are now captured via the realized counts in $\mathcal{Y}_{t-1}$. By employing backward substitution, model (21) can be written as

$$\Lambda_t = \sum_{k=0}^{\infty} \alpha^k \left(D + B_1 \mathcal{Y}_{t-1-k} B_2'\right) .$$

The conditional tensor mean thus depends on the infinite past of all the series of the tensor $\mathcal{Y}_t$, with full matrices of coefficients B_1, B_2 whereas the coefficient α establishes the rate at which such temporal dependence vanishes over time. In this way, the model specified in (21) preserves the whole path of autoregressive tensor effects, while assuming that the time series under consideration have a common decay rate of the past dynamics, defined by α. A similar recursion applies to the log-linear model (22). Analogous simplifying assumptions are often encountered in empirical applications of multivariate time series models to further reduce the computational complexity of the model, especially when the number of series is large; see for example Engle (2002), Heinen and Rengifo (2007) and Opschoor et al. (2018), among others.

We use the data until December 2022 for the estimation of the model. The rest of the dataset is used for out-of-sample analysis. As we see later, the log-linear dynamic specification in (22) provides a better fit to the data than the linear specification. To save space, we therefore only present the full first-stage estimation results for the log-linear model and postpone the estimation results for the linear model to the supplementary material. From first-stage QML estimation of the linear model, the estimated tensor effects are
$\hat{\alpha} = 0.614^*$, and

$$\hat{B}_1 = \begin{array}{c} \\ \text{Receive} \\ \text{Motor} \\ \text{Retail} \\ \text{Dwell} \\ \text{Person} \end{array} \!\! \begin{array}{c} \begin{array}{ccccc} \text{Receive} & \text{Motor} & \text{Retail} & \text{Dwell} & \text{Person} \end{array} \\ \begin{pmatrix} 1.000 & -0.075 & 0.019 & 0.365^* & 0.005 \\ 0.005 & 1.190^* & 0.025 & 0.081^* & 0.037^* \\ 0.004 & -0.006 & 1.121^* & -0.080^* & -0.004 \\ 0.013 & 0.101^* & -0.009 & 1.160^* & -0.019 \\ -0.052 & 0.099^* & -0.078 & 0.023 & 1.271^* \end{pmatrix} \end{array}, \quad \hat{B}_2 = \begin{array}{c} \\ \text{Black.} \\ \text{Cant.} \\ \text{Liv.} \\ \text{North.} \\ \text{Parra.} \end{array} \!\! \begin{array}{c} \begin{array}{ccccc} \text{Black.} & \text{Cant.} & \text{Liv.} & \text{North.} & \text{Parra.} \end{array} \\ \begin{pmatrix} 0.246^* & 0.004 & 0.026^* & 0.005 & -0.014^* \\ 0.001 & 0.293^* & 0.016^* & -0.008 & 0.003 \\ 0.001 & 0.038^* & 0.231^* & -0.002 & 0.013^* \\ 0.019^* & 0.055^* & 0.000 & 0.234^* & 0.011 \\ 0.015^* & 0.042^* & 0.005 & -0.005 & 0.243^* \end{pmatrix} \end{array}$$

with the model allowing for negative coefficients. Asterisks denote all the coefficients significantly different from 0 at 5% level. The significance tests are based on standard errors obtained by computing the sample counterparts of the asymptotic covariance of the QMLE.[2] The significant interactions in the model can be interpreted as Granger-causality effects. The interpretation of the estimated matrices of coefficients is similar to the one described in Appendix A. Hence, $\hat{B}_1$ captures the row-wise dynamics, i.e., the interaction between different types of crimes, whereas $\hat{B}_2$

[2] In linear models the null hypothesis $H_0 : \theta_i = 0$ is defined at the boundary of the parameter space for the non-intercept coefficients. Therefore, the asymptotic normal distribution of the t test requires a correction. Following Ahmad and Francq (2016, Rem. 2.4) and Francq and Zakoian (2019, Sec. 8.3.3) this is equivalent to employing the rejection region $|t| > \Phi^{-1}(1-\alpha)$ instead of the usual $|t| > \Phi^{-1}(1-\alpha/2)$, where α is the significance level. We use this correction for testing significance of parameters in the tensor INGARCH model (21).

captures the column-wise dynamics, i.e., interactions between cities. The bilinear matrix product in model (21) then combines the row-wise and column-wise interactions. Note that $B_{11}^{(1)}$ is set to 1 for identification. All the other coefficients on the main diagonals of both $B^{(1)}$ and $B^{(2)}$ are significant and show the largest magnitude. This is reasonable. We expect the largest impact on the conditional mean of a particular crime count to be the most recent realized experience of that same crime type. The same holds for the effect of cities: the most important predictor of the conditional mean in a city is the recent crime experience in that same city. However, we also see several significant spillover effects by looking at the off-diagonal elements. Most of these are quite small compared to the diagonal elements. However, there is a strong positive spill-over from stealing from dwellings to crimes related to receiving and handling stolen goods. This makes intuitive sense. We also see several spillovers between cities, though these are typically quite modest with respect to the diagonal elements, i.e., the same-city feedback. Still, the biggest LGA for population, which is Canterbury-Bankstown, has an autoregressive impact on almost all the closer districts. Northern Beaches, by contrast, being the farthest district, only seems to have a dynamic impact on its own next-month crime count, without spillovers to the other districts. Finally, we note that the crime times series are significantly affected by the more remote past since the (autoregressive) decay coefficient $\hat{\alpha} = 0.614$ is large and significant.

To examine the model fit, we consider the Pearson residuals, defined by $\mathcal{E}_{ij,t} = (\mathcal{Y}_{ij,t} - \Lambda_{ij,t})/\sqrt{\Lambda_{ij,t}}$ for $i = 1, \ldots, n_1$ and $j = 1 \ldots, n_2$. Under the correct model, the sequence $\mathcal{E}_{ij,t}$ is a white noise sequence with constant variance. We substitute $\Lambda_{ij,t}$ by $\hat{\Lambda}_{ij,t}$ to obtain $\hat{\mathcal{E}}_{ij,t}$. We compute the Pearson residuals for both the linear and log-linear tensor model and examine their autocorrelation functions (ACF). This reveals that the models are generally able to capture the serial correlations of the time series simultaneously, leading to relatively clean ACF plots of the residuals (as reported in the supplementary material).

Using the parameters estimated in the first stage, we can use the fitted matrix conditional mean $\hat{\Lambda}_t$ to estimate the Gaussian copula parameters in a second-stage estimation. We specify three different structures for the Gaussian correlation matrix R: equicorrelation (e), tensor equicorrelation (te), and full unstructured (*full*). For the first two structures, the 2SMLE (13) is obtained by maximizing the true log-likelihood computed using the full pmf (6). We also compute estimates by optimizing the approximate log-likelihood (14). The results are almost identical, however, and therefore not reported. Since $n = 25$, the direct optimization for the full correlation case is unfeasible as it involves estimating $n(n-1)/2 = 300$ correlation parameters. We therefore estimate R for this case by using the adjusted estimator (16).

Table 1 reports the estimates of the correlation parameters. We see that the copula parameter

Table 1: Estimated correlation parameters for the Gaussian copula obtained from 2SMLE with first-stage estimates computed by QMLE of tensor INGARCH models (21)-(22). For the full correlation matrix the adjusted estimator was used and reported minimum, average and maximum for the lower triangle of the estimated matrix $\tilde{R}$.

Models	Equi.	Tensor equi.		Full		
	$\hat{\bar{r}}$	$\hat{r}_1$	$\hat{r}_2$	$\tilde{r}_{\min}$	$\tilde{r}_{\text{avg}}$	$\tilde{r}_{\max}$
Linear	0.024	0.036	0.065	-0.167	0.050	0.231
Log-linear	0.024	0.037	0.063	-0.170	0.049	0.238

estimates are quite close for the linear and the log-linear specification. To interprete the tensor equicorrelation coefficients, note that $R = R_2 \otimes R_1$. For space constraints, consider an example with only 2 crimes $(\mathtt{a}, \mathtt{b})$ and 2 cities $(\mathtt{A}, \mathtt{B})$. As in Figure 3, we put crimes in the row dimension, and cities in columns. Then, the vectorized tensor $y_t = (y_{\mathtt{aA},t}, y_{\mathtt{bA},t}, y_{\mathtt{aB},t}, y_{\mathtt{bB},t})'$, and its Probability Integral Transforms (PITs) have a Gaussian copula with (copula) correlation matrix R, where

$$R_1 = \begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1 & r_2 \\ r_2 & 1 \end{pmatrix}, \qquad R = R_2 \otimes R_1 = \begin{array}{c} \\ \mathtt{aA} \\ \mathtt{bA} \\ \mathtt{aB} \\ \mathtt{bB} \end{array} \begin{array}{c} \begin{array}{cccc} \mathtt{aA} & \mathtt{bA} & \mathtt{aB} & \mathtt{bB} \end{array} \\ \begin{bmatrix} 1 & r_1 & r_2 & r_1 r_2 \\ r_1 & 1 & r_1 r_2 & r_2 \\ r_2 & r_1 r_2 & 1 & r_1 \\ r_1 r_2 & r_2 & r_1 & 1 \end{bmatrix} \end{array}.$$

We see that r_1 corresponds to the conditional correlation of two different crimes occurring in the same city, while r_2 is the conditional correlation of the same crime committed in two different cities. The cross-product $r_1 r_2$ picks up the conditional correlation of two different crimes happening in two different cities, i.e., the conditional copula correlation between two time series is the cumulative product of the equicorrelations for the dimensions in which their coordinate differs. Table 1 shows that both the equicorrelations between crime types and between cities are positive, with the correlation of same crime type among different cities being larger than the correlation for different crimes in the same city. Comparing the tensor equicorrelation model to the equicorrelation model, we see that the copula correlations for the tensor equicorrelation model are higher within a crime type (for a given city) than in the equicorrelation model. The same holds for the copula correlations within a city (given the crime type). However, if both the crime type and the city differ, the copula correlations in the tensor equicorrelation model are actually on order of magnitude smaller $(0.036 \times 0.065 \approx 0.0023 < 0.024)$ than in the equicorrelation specification. Note that following Remark 2, the copula correlations shown in Table 1 are not equal to the conditional correlations for the original crime counts time series. Still, although the estimated equicorrelations $\hat{\bar{r}}$, $\hat{r}_1$, $\hat{r}_2$ and the average adjusted correlation $\tilde{r}_{\text{avg}}$ appear modest, we see later in Table 2 below that including the copula-based pmf as computed for the tensor model greatly improves both the in-sample fit and the density forecasts of crime tensor data compared to the case in which the likelihood is based on only the marginals. Hence, there appears to be a contemporaneous dependence among the crime times series that is successfully accounted for by the copula tensor model.

Table 2 reports several measures of in-sample fit and forecasting accuracy for the estimated models. The performances of the first-stage estimates of the tensor INGARCH models (21)-(22) are denoted by QMLE, whereas the accuracy of the second-stage estimates that use the Gaussian copula with three different specifications for the copula correlation matrix are denoted by 2SMLE_e, 2SMLE_{te}, and 2SMLE_{full}, respectively. As a further benchmark, we also perform an alternative first-stage estimate by specifying univariate INGARCH models of the form

$$\Lambda_{ij,t} = D_{ij} + \alpha_{ij}\Lambda_{ij,t-1} + \beta_{ij}\mathcal{Y}_{ij,t-1}\,, \qquad \log(\Lambda_{ij,t}) = D_{ij} + \alpha_{ij}\log(\Lambda_{ij,t-1}) + \beta_{ij}\log(\mathcal{Y}_{ij,t-1}+1)\,,$$

for each crime time series $i = 1, \dots, n_1$ in each city $j = 1, \dots, n_2$. In this case each series has its own autoregressive and feedback effect, but no interaction coefficients are estimated among the series. The Poisson likelihood is used for each univariate model, such that the total log-likelihood is equal the one used for the QMLE defined in (9). The accuracy measures of this alternative modelling procedure are denoted by QMLE_{uni}.

To evaluate the in-sample fit, we employ three indicators. The quantity maxL is the maximum average log-likelihood computed at the estimated values. The Akaike and Bayesian information criteria (AIC, BIC) are computed at the same optimum. To evaluate the out-of-sample forecast accuracy, we use the mean absolute error (MAE) and the log-score criterion (LS). The MAE is used to measure the accuracy of point forecasts. The log-score criterion is used to measure the accuracy of pmf forecasts (Geweke and Amisano, 2011). The latter two measures are obtained from a one-step-ahead forecast study where the observed data $\{\mathcal{Y}_t\}_{t=1}^T$ is split into an in-sample dataset from $t = 1$ to $t = T^*$, and an out-of-sample dataset from $t = T^* + 1$ to $t = T$. We set T^* to December 2022. The prediction of each observation in the out-of-sample dataset is obtained by estimating the models with all previous observations. In particular, the point prediction of the vectorized tensor $y_t = \text{vec}(\mathcal{Y}_t)$ at time $t = \tau$ is given by $\hat{\lambda}_\tau(\hat{\theta}_{\tau-1})$, where $\theta_{\tau-1}$ denotes the QML estimate of θ using the data from $t = 1$ to $t = \tau - 1$. The pmf prediction of y_t at time $t = \tau$ is given by $\prod_{h=1}^n p_h(y_h; \hat{\lambda}_{h,\tau}(\hat{\theta}_{\tau-1}))$, $y_h \in \mathbb{N}$, for the QMLE and by $p(y; \hat{\lambda}_\tau(\hat{\theta}_{\tau-1}), \hat{r}_{\tau-1})$, $y \in \mathbb{N}^n$, for the 2SMLE where $\hat{r}_{\tau-1}$ are the 2SQML estimates of copula parameter r using data up to time $t = \tau - 1$. Finally, the MAE and LS are obtained as

$$\text{MAE} = \frac{1}{T - T^*} \sum_{t=T^*+1}^{T} \sum_{h=1}^{n} |y_{h,t} - \hat{\lambda}_{h,t}(\hat{\theta}_{t-1})|\,,$$

Table 2: Performance measures of QML and 2SQML estimates of linear (top) and log-linear (bottom) models for the crime dataset. The columns contain the maximized average log-likelihood (maxL), AIC, BIC, MAE and log-score criterion (LS). Best results in bold.

Models	maxL	AIC	BIC	MAE	LS
QMLE	-106.602	71573.295	71859.355	**8.486**	-1436.365
QMLE$_{uni}$	-107.274	72023.450	72309.510	8.657	-1454.406
2SMLE$_{e}$	-106.012	71030.305	71034.119	-	-1336.548
2SMLE$_{te}$	**-105.751**	**70857.487**	**70865.116**	-	**-1328.743**
2SMLE$_{full}$	-105.766	71463.016	72607.255	-	-1353.276
QMLE	-106.300	71371.193	71657.253	**8.474**	-1437.837
QMLE$_{uni}$	-107.737	72333.883	72619.943	8.689	-1457.628
2SMLE$_{e}$	-105.700	70820.784	70824.598	-	-1336.619
2SMLE$_{te}$	-105.454	**70658.056**	**70665.684**	-	**-1328.605**
2SMLE$_{full}$	**-105.423**	71233.702	72377.941	-	-1350.611

and

$$\mathrm{LS}_{QMLE} = \sum_{t=T^*+1}^{T} \sum_{h=1}^{n} \log\left[p_h(y_{h,t}; \hat{\lambda}_{h,t}(\hat{\theta}_{t-1}))\right], \qquad \mathrm{LS}_{2SMLE} = \sum_{t=T^*+1}^{T} \log\left[p(y_t; \hat{\lambda}_t(\hat{\theta}_{t-1}), \hat{r}_{t-1})\right].$$

Regarding the QMLEs, the results show that the tensor INGARCH specifications give a better in-sample fit. The they have bigger values of maxL and lower values for AIC and BIC compared to the univariate INGARCH models. When the tensor INGARCH models are used as a first-stage plug-in, the best second-stage estimates are provided by 2SMLE with tensor equicorrelation structure as it has lower AIC and BIC. All the two-stage estimators show a better in-sample fit with respect to the first-stage QMLEs. More accurate point forecasts are obtained by the tensor model against the univariate specifications, whereas the best pmf forecast occurs with the 2SMLE$_{te}$, which has the largest log-score (LS).

In the density forecast study (LS column), all the two-stage estimators outperform the QMLEs by a wide margin. The resulting overall best model from both in-sample and out-of-sample analysis is the copula tensor log-linear INGARCH model with tensor equicorrelation. Finally, note that the tensor equicorrelation specification always seems to provide the best performance against alternative specifications. This can be due to a threefold effect: (i) ability to reproduce the data tensor structure in the correlation matrix with (ii) a more flexible, empirically required structure compared to the standard equicorrelation model, while (iii) retaining parsimony compared to the specification with a fully unstructured correlation matrix.

Summarizing, the empirical findings of this section indicate that count tensor models are powerful tools capable of modeling dynamic spillover effects and interconnections among multidimensional time series in a parsimoninous way (as the specifications (21)-(22) have the same number of parameters of the univariate INGARCH counterparts). Such an analysis is not achievable with existing univariate/multivariate modelling approaches. Moreover, by keeping the same number of parameters as in univariate specifications, these models are able to provide better in-sample fit and deliver more accurate out-of-sample predictions. Finally, the inclusion of copula-based joint distributions in tensor models further improves goodness-of-fit and density forecast accuracy. This suggests that the copula tensor count autoregressions proposed in this contribution are able to handle both contemporaneous and lagged dependence among a large set of count times series and successfully model complex inter-relations in multidimensional datasets.

Acknowledgments

M. Armillotta acknowledges financial support from the EU Horizon Europe programme under the Marie Skłodowska-Curie grant agreement No.101108797.

Appendix A Copula matrix count autoregressions

In the special case $N = 2$ the sequence of tensors introduced in Section 3 reduces to $\mathcal{Y}_t = (\mathcal{Y}_{ij,t})$ being a $n_1 \times n_2$ matrix time series of counts and $\mathcal{X}_t$ is a $n_1 \times n_2$ matrix of exogenous covariates. By (1) each element of the matrix satisfies, for $t \in \mathbb{Z}$,

$$\mathcal{Y}_{ij,t} = F_{ij}^{-1}(\mathcal{U}_{ij,t}; \Lambda_{ij,t}) \,, \qquad 1 \leq i \leq n_1 \,, \; 1 \leq j \leq n_2 \tag{A-1}$$

where $F_{ij}^{-1}(\cdot; \Lambda_{ij,t})$ is the quantile function of the conditional Poisson distribution with conditional mean $\Lambda_{ij,t}$, $\mathcal{U}_{ij,t} \sim U(0,1)$ are such that $\mathrm{vec}(\mathcal{U}_t)|\mathcal{F}_{t-1} \sim C_r$, and $\mathcal{U}_t = (\mathcal{U}_{ij,t})$ is a $n_1 \times n_2$ matrix. Define $\Lambda_t = (\Lambda_{ij,t})$ the $n_1 \times n_2$ matrix of Poisson conditional means. When $N = 2$ the multilinear products reduce to bilinear matrix products. In particular, following (2) the dynamics of the copula tensor count autoregression, for $t \in \mathbb{Z}$, boils down to

$$\mathcal{V}_t = D + A_1 \mathcal{V}_{t-1} A_2' + B_1 \mathcal{Z}_{t-1} B_2' + Q_1 \mathcal{X}_{t-1} Q_2' \tag{A-2}$$

where $D \in \mathbb{R}^{n_1 \times n_2}$ is a matrix of intercepts. First note that the left matrices A_1, B_1, Q_1 reflect row-wise interactions, and the right matrices A_2, B_2, Q_2 introduce column-wise dependence, and therefore the bilinear matrix products in (A-2) combine the row-wise and column-wise interactions.

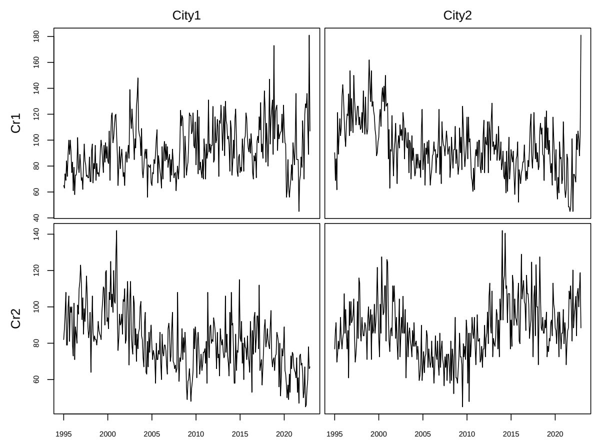

Figure A-1: Monthly number of police reports for 2 crimes (Cr1, Cr2) in 2 cities of Australia (City1, City2) from January 1995 to December 2022.

It is easier to see how the coefficient matrices reflect the row and column structures by looking at a few special cases. Define the single elements of a $n_1 \times n_2$ matrix $A_s = \left(A_{ij}^{(s)}\right)$ for $s = \{1, 2\}$, $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$. Consider the case without covariates and $A_1 = B_1 = I$. Then the model reduces to $\mathcal{V}_t = D + \mathcal{V}_{t-1} A_2' + \mathcal{Z}_{t-1} B_2'$. If we consider the crime dataset described in Figure A-1 the first column of the model is given by

$$\begin{array}{c} \text{City1} \\ \begin{pmatrix} \mathcal{V}_{Cr1} \\ \mathcal{V}_{Cr2} \end{pmatrix}_t \end{array} = \begin{pmatrix} D_{11} \\ D_{21} \end{pmatrix} + A_{11}^{(2)} \begin{array}{c} \text{City1} \\ \begin{pmatrix} \mathcal{V}_{Cr1} \\ \mathcal{V}_{Cr2} \end{pmatrix}_{t-1} \end{array} + A_{12}^{(2)} \begin{array}{c} \text{City2} \\ \begin{pmatrix} \mathcal{V}_{Cr1} \\ \mathcal{V}_{Cr2} \end{pmatrix}_{t-1} \end{array} + B_{11}^{(2)} \begin{array}{c} \text{City1} \\ \begin{pmatrix} \mathcal{Z}_{Cr1} \\ \mathcal{Z}_{Cr2} \end{pmatrix}_{t-1} \end{array} + B_{12}^{(2)} \begin{array}{c} \text{City2} \\ \begin{pmatrix} \mathcal{Z}_{Cr1} \\ \mathcal{Z}_{Cr2} \end{pmatrix}_{t-1} \end{array}$$

which means that at time t, the rescaled mean ($\mathcal{V}_t$) of a crime in one city, is a linear combination of rescaled means of the same crime from all the cities at $t-1$ and the past values of the same (rescaled) crime from all the cities at $t-1$. This linear combination is the same for different crimes. Therefore A_2, B_2 capture the column-wise interactions, i.e. interactions among cities. However, the interactions do not have any spillover, i.e. there are no interactions among the different crimes. Conversely, if we let $A_2 = B_2 = I$ in model (A-2), then a similar interpretation can be obtained, where the matrices A_1, B_1 reflect row-wise interactions, i.e. interactions among the crimes within each city. There are no interactions among the different cities.

We can also give a second interpretation. For example, in the case $A_2 = B_2 = I$ the model becomes $\mathcal{V}_t = D + A_1 \mathcal{V}_{t-1} + B_1 \mathcal{Z}_{t-1}$. By defining $\mathcal{V}_{j,t}$ the j column of $\mathcal{V}_t$, it is clear that each column of $\mathcal{V}_t$ follows $\mathcal{V}_{j,t} = D_j + A_1 \mathcal{V}_{j,t-1} + B_1 \mathcal{Z}_{j,t-1}$ for $j = 1, \ldots, n_2$. In other words, for each city, its crimes follow a multivariate count autoregression and the effects A_1, B_1 are the same for different cities. Finally, if $A_1 = B_1 = I$ each row of $\mathcal{V}_t$, say $\mathcal{V}_{i,t}$, would follow a multivariate count autoregression, and the coefficient matrices corresponding to different rows would be the same.

Appendix B Proofs of the main results

B.1 Proof of Theorem 1

Since the vectorized form (4) is equivalent to model (1)-(2) we can show the identification of parameters directly on (4). Consider the case $N = 2$. We need to prove that, under conditions **I**, for two different pairs of matrices $\{A_1, A_2\}$ and $\{\tilde{A}_1, \tilde{A}_2\}$ we have that $A_2 \otimes A_1 = \tilde{A}_2 \otimes \tilde{A}_1$ if and only if $A_1 = \tilde{A}_1$ and $A_2 = \tilde{A}_2$. The if statement follows trivially. For space convenience set $n_1 = n_2 = 2$ but the argument is valid for any arbitrary value of n_s. Then,

$$A_2 \otimes A_1 = \begin{pmatrix} A_{11}^{(2)} & A_{11}^{(2)} A_{12}^{(1)} & A_{12}^{(2)} & A_{12}^{(2)} A_{12}^{(1)} \\ A_{11}^{(2)} A_{21}^{(1)} & A_{11}^{(2)} A_{22}^{(1)} & A_{12}^{(2)} A_{21}^{(1)} & A_{12}^{(2)} A_{22}^{(1)} \\ A_{21}^{(2)} & A_{21}^{(2)} A_{12}^{(1)} & A_{22}^{(2)} & A_{22}^{(2)} A_{12}^{(1)} \\ A_{21}^{(2)} A_{21}^{(1)} & A_{21}^{(2)} A_{22}^{(1)} & A_{22}^{(2)} A_{21}^{(1)} & A_{22}^{(2)} A_{22}^{(1)} \end{pmatrix}$$

since $A_{11}^{(1)} = 1$. A similar formula holds for $\tilde{A}_2 \otimes \tilde{A}_1$ with $\tilde{A}_{ij}^{(1)}, \tilde{A}_{ij}^{(2)}$ instead of $A_{ij}^{(1)}, A_{ij}^{(2)}$. It is clear that to have $A_2 \otimes A_1 = \tilde{A}_2 \otimes \tilde{A}_1$ it is necessary that $a_2 = \tilde{a}_2$, where $a_2 = (A_{11}^{(2)}, A_{2,1}^{(2)}, A_{1,2}^{(2)}, A_{2,2}^{(2)})'$ and $\tilde{a}_2 = (\tilde{A}_{11}^{(2)}, \tilde{A}_{2,1}^{(2)}, \tilde{A}_{1,2}^{(2)}, \tilde{A}_{2,2}^{(2)})'$. However, $\mathrm{vec}(A_2) = a_2$ and $\mathrm{vec}(\tilde{A}_2) = \tilde{a}_2$ hence $A_2 = \tilde{A}_2$. Now for identification of A_1 we note that $A_2 \otimes A_1 = A_2 \otimes \tilde{A}_1$ is only possible if $a_1 = \tilde{a}_1$ where $a_1 = (A_{2,1}^{(1)}, A_{1,2}^{(1)}, A_{2,2}^{(1)})'$ and $\tilde{a}_1 = (\tilde{A}_{2,1}^{(1)}, \tilde{A}_{1,2}^{(1)}, \tilde{A}_{2,2}^{(1)})'$. Under the latter condition and since $\mathrm{vec}(A_1) = (1, a_1')'$, $\mathrm{vec}(\tilde{A}_1) = (1, \tilde{a}_1')'$ we have that $\mathrm{vec}(A_1) = \mathrm{vec}(\tilde{A}_1)$ and therefore $A_1 = \tilde{A}_1$.

For $N = 3$, the associative property of the Kronecker product guarantees that $A_3 \otimes A_2 \otimes A_1 = (A_3 \otimes A_2) \otimes A_1$. Then by previous arguments, since $A_{11}^{(2)} = 1$ we have that $A_3 \otimes A_2 = \tilde{A}_3 \otimes \tilde{A}_2$ if and only if (iff) $A_2 = \tilde{A}_2$ and $A_3 = \tilde{A}_3$. Now set $G_1 = (A_3 \otimes A_2)$. Since $A_{11}^{(1)} = 1$ it follows that $G_1 \otimes A_1 = \tilde{G}_1 \otimes \tilde{A}_1$ iff $G_1 = \tilde{G}_1$ and $A_1 = \tilde{A}_1$ but we proved that $G_1 = \tilde{G}_1$ iff $A_2 = \tilde{A}_2$ and $A_3 = \tilde{A}_3$. Therefore, wrapping everything together $A_3 \otimes A_2 \otimes A_1 = \tilde{A}_3 \otimes \tilde{A}_2 \otimes \tilde{A}_1$ iff $A_1 = \tilde{A}_1$, $A_2 = \tilde{A}_2$ and $A_3 = \tilde{A}_3$. For $N \geq 4$ the same argument applies iteratively to pairwise Kronecker products. $\square$

B.2 Proof of Theorem 2

Recall that $\tilde{L}_T(\theta)$ is the log-quasi-likelihood as defined in (9) and $L(\theta) = \mathrm{E}[l_t(\theta)]$ is the corresponding limit. In Lemma C.1 described in section C below we show that the convergence is uniform over Θ. Therefore, by the uniform limit theorem $L(\theta)$ is a continuous function and it attains at least a maximum in Θ since Θ is compact. We now prove that such maximum is unique so that it can be univocally identified. To this aim note that Lemma C.2 implies that $l_t(\theta) = l_t(\theta_0)$ a.s. if and only if $\theta = \theta_0$. By recalling that $\log(x) \leq x - 1$ for $x \in \mathbb{R}_+$ with equality only if $x = 1$ we have a.s.

$$l_t(\theta) - l_t(\theta_0) \leq \sum_{h=1}^{n} \frac{p_h(y_{h,t}; \lambda_{h,t}(\theta))}{p_h(y_{h,t}; \lambda_{h,t}(\theta_0))} - n\,.$$

Due to the previous results, the last inequality is strict for all $\theta \in \Theta$ with $\theta \neq \theta_0$. Therefore, $\forall \theta \neq \theta_0$

$$\mathrm{E}\left\{\mathrm{E}\left[l_t(\theta) - l_t(\theta_0) | \mathcal{F}_{t-1}\right]\right\} < \sum_{h=1}^{n} \mathrm{E}\left\{\mathrm{E}\left[\frac{p_h(y_{h,t}; \lambda_{h,t}(\theta))}{p_h(y_{h,t}; \lambda_{h,t}(\theta_0))}\middle| \mathcal{F}_{t-1}\right]\right\} - n = 0\,,$$

where the last equality holds since for $h = 1, \dots, n$, $p_h(y_{h,t}; \lambda_{h,t}(\theta_0))$ are the true conditional univariate pmfs of the count processes. Finally we have that

$$L(\theta) - L(\theta_0) = \mathrm{E}\left\{\mathrm{E}\left[l_t(\theta) - l_t(\theta_0) | \mathcal{F}_{t-1}\right]\right\} < 0\,, \qquad \forall \theta \neq \theta_0\,.$$

This proves that θ_0 is the unique maximizer of $L(\theta)$. The compactness of Θ and an application of White (1994, Thm. 3.5) provide the consistency of the estimator. □

B.3 Proof of Theorem 3

In order to show the asymptotic normality of the estimator define $S_T(\theta) = \nabla_\theta L_T$, $\tilde{S}_T(\theta) = \nabla_\theta \tilde{L}_T$ and $H_T(\theta) = T^{-1} \sum_{t=1}^{T} -\nabla_\theta^2 l_t$. Let $\rho_{h,l}(\theta_0) = \mathrm{E}[\varepsilon_{h,t}(\theta_0)\varepsilon_{l,t}(\theta_0)|\mathcal{F}_{t-1}]/[\sqrt{\lambda_{h,t}(\theta_0)}\sqrt{\lambda_{l,t}(\theta_0)}]$ be the conditional correlation of the vectorized time series with $\varepsilon_t(\theta_0) = y_t - \lambda_t(\theta_0)$. Note that

$$I(\theta_0) = \sum_{h=1}^{n} \sum_{l=1}^{n} \mathrm{E}\left[\frac{\rho_{h,l}(\theta_0)}{\sqrt{\lambda_{h,t}(\theta_0)}\sqrt{\lambda_{l,t}(\theta_0)}} \nabla_{\theta_0}\lambda_{h,t} \nabla_{\theta_0}\lambda'_{l,t}\right]$$

which is finite by **A4** and **A9**. Recall that $\sqrt{T} S_T(\theta_0) = T^{-1/2} \sum_{t=1}^{T} \nabla_t$ where $\nabla_t = \nabla_{\theta_0} l_t$. Note that $\{\nabla_t, \mathcal{F}_t\}$ is a stationary and ergodic martingale difference since $\mathrm{E}(\nabla_t | \mathcal{F}_{t-1}) = 0$, with a finite second moments matrix $I(\theta_0)$. Then, by the central limit theorem for martingales (Billingsley, 1961) it follows that $\sqrt{T}\, S_T(\theta_0) \xrightarrow{d} N(0, I(\theta_0))$ as $T \to \infty$.

For T large enough $\hat{\theta}_T \in \dot{\Theta}$ by **A6**, so the following derivatives exist almost surely

$$0 = \sqrt{T}\tilde{S}_T(\hat{\theta}_T) = \sqrt{T} S_T(\hat{\theta}_T) + o_p(1) = \sqrt{T} S_T(\theta_0) - H_T(\bar{\theta})\sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1),$$

where the first equality comes from the definition (10), the second equality holds by Lemma C.4, and the third equality is obtained by Taylor expansion at θ_0 with $\bar{\theta}$ lying between $\hat{\theta}_T$ and θ_0. The application of Lemmas C.6-C.7 establishes the asymptotic normality of the estimator $\hat{\theta}_T$ with covariance matrix Σ. □

B.4 Proof of Theorem 4

Analogously to the second-stage log-likelihood defined in (12), let $L_T(\theta, r) = T^{-1}\sum_{t=1}^{T} l_t(\theta, r) = T^{-1}\sum_{t=1}^{T}\log\left[p\big(y_t; \lambda_t(\theta), r\big)\right]$ and $L(\theta, r) = \mathrm{E}[l_t(\theta, r)]$ is the corresponding limit. Recall from section 3.3 that $\psi = (\theta', r')' \in \Psi$ where $\Psi = \Theta \times \mathsf{P}^v$. Define the compact notation $L_T(\theta, r) = L_T(\psi)$ and $L(\theta, r) = L(\psi)$. Lemma C.8 guarantees that $L_T(\psi)$ converges to $L(\psi) = \mathrm{E}[l_t(\psi)]$ uniformly and therefore the latter attains at least a maximum in Ψ since Ψ is compact. Consequently, also $L(\theta_0, r)$ attains at least a maximum in P^v. To prove the uniqueness of such a maximum note that Lemma C.9 entails $l_t(\theta_0, r) = l_t(\theta_0, r_0)$ a.s. if and only if $r = r_0$. By employing analogous arguments to the proof of Theorem 2 we shall have that

$$l_t(\theta_0, r) - l_t(\theta_0, r_0) \leq \frac{p(y_t; \lambda_t, r)}{p(y_t; \lambda_t, r_0)} - 1\,,$$

with $\lambda_t = \lambda_t(\theta_0)$ and strict inequality for all $r \in \mathsf{P}^v$ with $r \neq r_0$. Therefore, $\forall r \neq r_0$

$$\mathrm{E}\left\{\mathrm{E}\left[l_t(\theta_0, r) - l_t(\theta_0, r_0)|\mathcal{F}_{t-1}\right]\right\} < \mathrm{E}\left\{\mathrm{E}\left[\frac{p(y_t; \lambda_t, r)}{p(y_t; \lambda_t, r_0)}\middle|\mathcal{F}_{t-1}\right]\right\} - 1 = 0\,,$$

where the last equality holds since $p(y_t; \lambda_t, r_0) = p(y_t; \lambda_t(\theta_0), r_0)$ is the true conditional pmf of the vectorial count processes. Finally we have that

$$L(\theta_0, r) - L(\theta_0, r_0) = \mathrm{E}\left\{\mathrm{E}\left[l_t(\theta_0, r) - l_t(\theta_0, r_0)|\mathcal{F}_{t-1}\right]\right\} < 0\,, \qquad \forall r \neq r_0\,.$$

An application of White (1994, Thm. 3.10) provides the consistency of the 2SMLE. □

Appendix C Technical lemmas

Below we report the statements of the lemmas that were used in Appendix B. The proofs of the lemmas are given in the online supplementary material.

Lemma C.1. Let the assumptions of Theorem 2 hold. Then, $\|\tilde{L}_T - L\|_\Theta \to 0$ a.s., as $T \to \infty$.

Lemma C.2. Let the assumptions of Theorem 2 hold. Then, for any $\theta \in \Theta$ and all $y_h \in \mathbb{N}$

$$p_h(y_h; \lambda_{h,t}(\theta)) = p_h(y_h; \lambda_{h,t}(\theta_0)), \quad 1 \leq h \leq n, \quad a.s. \quad \text{if and only if} \quad \theta = \theta_0\,.$$

Lemma C.3. Let the assumptions of Theorem 3 hold. Then, $\|\|\nabla_\theta \tilde{\nu}_t - \nabla_\theta \nu_t\|\|_\Theta \to 0$ e.a.s., as $t \to \infty$, where $\nabla_\theta \nu_t$ is the stationary and ergodic derivative process of ν_t. Furthermore, $\{\nabla_\theta \nu_t\}_{t\in\mathbb{Z}}$ has a bounded uniform moment $\mathrm{E}\|\|\nabla_\theta \nu_t\|\|_\Theta < \infty$.

Lemma C.4. Let the assumptions of Theorem 3 hold. Then, $\sqrt{T}\|S_T - \tilde{S}_T\|_\Theta \to 0$ a.s., as $T \to \infty$.

Lemma C.5. Let the assumptions of Theorem 3 hold. Then, the columns of $\nabla_{\theta_0} \nu_t$ are linearly independent.

Lemma C.6. Let the assumptions of Theorem 3 hold. Then, $H(\theta_0)$ and Σ are positive definite.

Lemma C.7. Let the assumptions of Theorem 3 hold. Then, $|||H_T - H|||_\Theta \to 0$ a.s., as $T \to \infty$.

Lemma C.8. Let the assumptions of Theorem 4 hold. Then, $\|L_T - L\|_\Psi \to 0$ a.s., as $T \to \infty$.

Lemma C.9. Let the assumptions of Theorem 4 hold. Then, for any $r \in \mathsf{P}^v$ and all $y \in \mathbb{N}^n$

$$p(y; \lambda_t(\theta_0), r) = p(y; \lambda_t(\theta_0), r_0) \quad a.s. \quad \text{if and only if} \quad r = r_0 \,.$$

References

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

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

Aldasoro, I. and I. Alves (2018). Multiplex interbank networks and systemic importance: An application to European data. *Journal of Financial Stability 35*, 17–37.

Bai, J. and K. Li (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics 40*(1), 436 – 465.

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

Billio, M., R. Casarin, M. Iacopini, and S. Kaufmann (2023). Bayesian dynamic tensor regression. *Journal of Business & Economic Statistics 41*(2), 429–439.

Chang, J., J. He, L. Yang, and Q. Yao (2023). Modelling matrix time series via a tensor CP-decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(1), 127–148.

Chen, E. Y., D. Xia, C. Cai, and J. Fan (2024). Semi-parametric tensor factor analysis by iteratively projected singular value decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, qkae001.

Chen, R., H. Xiao, and D. Yang (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics 222*(1), 539–560.

Chen, R., D. Yang, and C.-H. Zhang (2022). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association 117*(537), 94–116.

Cichocki, A., D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine 32*(2), 145–163.

Davis, R. A., K. Fokianos, S. H. Holan, H. Joe, J. Livsey, R. Lund, V. Pipiras, and N. Ravishanker (2021). Count time series: A methodological review. *Journal of the American Statistical Association 116*, 1533–1547.

Debaly, Z.-M. and L. Truquet (2023). Multivariate time series models for mixed data. *Bernoulli 29*, 669–695.

Drost, F. C., R. Van den Akker, and B. J. Werker (2009). Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 71*(2), 467–485.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

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

Fokianos, K. (2022). Multivariate count time series modelling. *To appear in Econometrics and Statistics*.

Fokianos, K., B. Støve, D. Tjøstheim, and P. Doukhan (2020). Multivariate count autoregression. *Bernoulli 26*, 471–499.

Francq, C. and J.-M. Zakoian (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons.

Genest, C. and J. Nešlehová (2007). A primer on copulas for count data. *Astin Bull. 37*, 475–515.

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130–141.

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

Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Wadsworth Publishing Company.

Han, Y., D. Yang, C.-H. Zhang, and R. Chen (2024). Cp factor model for dynamic tensors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, qkae036.

Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. Technical Report MPRA Paper 8113, University Library of Munich, Germany. availabel at http://mpra.ub.uni-muenchen.de/8113/.

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

Hsu, N.-J., H.-C. Huang, and R. S. Tsay (2021). Matrix autoregressive spatio-temporal models. *Journal of Computational and Graphical Statistics 30*(4), 1143–1155.

Joe, H. (2014). *Dependence Modeling with Copulas*. CRC press.

Li, Z. and H. Xiao (2021). Multi-linear tensor autoregressive models. *arXiv preprint arXiv:2110.00928*.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer.

Opschoor, A., P. Janus, A. Lucas, and D. Van Dijk (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics 36*(4), 643–657.

Panagiotelis, A., C. Czado, and H. Joe (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association 107*(499), 1063–1072.

Patton, A. J. (2002). *Applications of copula theory in financial econometrics*. Ph. D. thesis, University of California, San Diego.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review 47*(2), 527–556.

Pedeli, X. and D. Karlis (2013a). On composite likelihood estimation of a multivariate INAR (1) model. *Journal of Time Series Analysis 34*, 206–220.

Pedeli, X. and D. Karlis (2013b). Some properties of multivariate INAR (1) processes. *Computational Statistics & Data Analysis 67*, 213–225.

Samadi, S. Y. and L. Billard (2024). On a matrix-valued autoregressive model. *Journal of Time Series Analysis jtsa.12748*.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge university press.

Yuan, C., Z. Gao, X. He, W. Huang, and J. Guo (2023). Two-way dynamic factor models for high-dimensional matrix-valued time series. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(5), 1517–1537.

Zhou, H., L. Li, and H. Zhu (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association 108*(502), 540–552.

Supplementary material of "Copula tensor count autoregressions for modeling multidimensional integer-valued time series"

Mirko Armillotta[1], Paolo Gorgi[2], and André Lucas[2]

[1]University of Rome Tor Vergata, Rome, Italy
[2]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

January 24, 2025

S1 Derivatives of copula tensor count autoregressions with identification constraints

In this section we compute the first derivatives of the tensor model (1)-(2) with respect to the mean parameters vector θ, under the identifiability constraints of Theorem 1. For convenience the computation is performed on the vectorized version of the model, defined in (4). The result of Theorem 1 induces the following constraints on the tensor model (4)

$$A_{11}^{(N-1)} = \cdots = A_{11}^{(1)} = B_{11}^{(N-1)} = \cdots = B_{11}^{(1)} = Q_{11}^{(N-1)} = \cdots = Q_{11}^{(1)} = 1\,.$$

Therefore, we set the following definitions: $a = \mathrm{vec}(A)$, $b = \mathrm{vec}(B)$, $q = \mathrm{vec}(Q)$, $a_N = \mathrm{vec}(A_N)$, $b_N = \mathrm{vec}(B_N)$ and $q_N = \mathrm{vec}(Q_N)$. Moreover, $(1, a_s')' = \mathrm{vec}(A_s)$, $(1, b_s')' = \mathrm{vec}(B_s)$, $(1, q_s')' = \mathrm{vec}(Q_s)$ for $s = 1, \dots, N-1$. Therefore, the mean parameters vector can be rewritten as $\theta = (d', a_1', \dots, a_N', b_1', \dots, q_N')' \in \mathbb{R}^{m_1}$ with $m_1 = n + 3(n_N^2 + \sum_{s=1}^{N-1}(n_s^2 - 1))$.

Define I_x the identity matrix of dimension $x \times x$, and $O_{x\times y}$ a zero matrix of dimension $x \times y$. Furthermore, for the first derivative matrix of a vectorial function $f(\cdot)$ with respect to a vectorial parameter γ we denote the compact notation $\nabla_\gamma f = \partial f(\gamma)/\partial \gamma'$. By the chain rule $\nabla_{b_1}\nu_t = \nabla_b \nu_t \nabla_{b_1} b$. Moreover, by Seber (2008, 17.35) we have that $\nabla_b B z_{t-1} = (z_{t-1}' \otimes I_n)$. Therefore,

$\nabla_b \nu_t = (z'_{t-1} \otimes I_n) + A\nabla_b \nu_{t-1}$ and $\nabla_{b_1} \nu_t = (z'_{t-1} \otimes I_n)\nabla_{b_1} b + A\nabla_b \nu_{t-1} \nabla_{b_1} b$. The same holds for the other derivatives. By stacking all the derivatives with respect to θ in a matrix we obtain

$$\nabla_\theta \nu_t = \Delta_{t-1}\Xi + A\nabla_\theta \nu_{t-1}\,, \tag{S-1}$$

where $\Delta_{t-1} = [I_n, (\nu'_{t-1} \otimes I_n), (z'_{t-1} \otimes I_n), (x'_{t-1} \otimes I_n)]$ is a $n \times p^*$ block matrix, $p^* = n + 3n^2$ and

$$\Xi = \begin{pmatrix} I_n & O_{n\times n_1^2-1} & \dots & O_{n\times n_N^2} & O_{n\times n_1^2-1} & \dots & O_{n\times n_N^2} & O_{n\times n_1^2-1} & \dots & O_{n\times n_N^2} \\ O_{n^2\times n^2} & \nabla_{a_1} a & \dots & \nabla_{a_N} a & O_{n^2\times n_1^2-1} & \dots & O_{n^2\times n_N^2} & O_{n^2\times n_1^2-1} & \dots & O_{n^2\times n_N^2} \\ O_{n^2\times n^2} & O_{n^2\times n_1^2-1} & \dots & O_{n^2\times n_N^2} & \nabla_{b_1} b & \dots & \nabla_{b_N} b & O_{n^2\times n_1^2-1} & \dots & O_{n^2\times n_N^2} \\ O_{n^2\times n^2} & O_{n^2\times n_1^2-1} & \dots & O_{n^2\times n_N^2} & O_{n^2\times n_1^2-1} & \dots & O_{n^2\times n_N^2} & \nabla_{q_1} q & \dots & \nabla_{q_N} q \end{pmatrix}$$

is a $p^* \times m_1$ block matrix. Set $N = 2$. Then, following Magnus and Neudecker (2019, p. 55) some matrix manipulation entails

$$\nabla_{a_1} a = \nabla_{a_1} \operatorname{vec}(A_2 \otimes A_1) = \nabla_{a_1} (\bar{A}_2 \otimes I_{n_1}) \operatorname{vec}(A_1) = (\bar{A}_2 \otimes I_{n_1})\bar{I}_{n_1^2-1}\,, \quad \text{where}$$

$$\bar{A}_2 = (I_{n_2} \otimes K_{n_1,n_2})(a_2 \otimes I_{n_1})\,, \quad \bar{I}_{n_1^2-1} = \begin{pmatrix} O_{1\times n_1^2-1} \\ I_{n_1^2-1} \end{pmatrix}\,,$$

$$\nabla_{a_2} a = \nabla_{a_2} \operatorname{vec}(A_2 \otimes A_1) = \nabla_{a_2} (I_{n_2} \otimes \bar{A}_1) a_2 = I_{n_2} \otimes \bar{A}_1\,, \qquad \bar{A}_1 = (K_{n_1,n_2} \otimes I_{n_1})(I_{n_2} \otimes \operatorname{vec}(A_1))\,,$$

and K_{n_1,n_2} is the commutation matrix of dimension $n_1 \times n_2$. Similar results hold for the other blocks of Ξ. Analogous arguments apply by chain rule when $N \geq 3$. As an example consider $N = 3$. Recall that $a = \operatorname{vec}(A_3 \otimes A_2 \otimes A_1)$ and define $G_{3,2} = A_3 \otimes A_2$ with $g_{3,2} = \operatorname{vec}(G_{3,2})$. Set $j = \{2, 3\}$. Then, by chain rule $\nabla_{a_j} a = \nabla_{g_{3,2}} a \nabla_{a_j} g_{3,2}$ and the two derivatives have the same structure of the ones above.

S2 Proofs

S2.1 Proof of Proposition 2

The results of Proposition 2 follow the fashion of Straumann and Mikosch (2006, Prop. 3.12) by applying Bougerol (1993, Thm. 3.1) in the space of continuous functions equipped with the uniform norm $\|\cdot\|_\Theta$. The definition of the SRE (8) guarantees that the function $(\theta, \nu) \mapsto h_\theta(y, x; \nu)$ is continuous in $\Theta \times \mathbb{R}^n$ for any $y \in \mathbb{N}^n$, $x \in \mathbb{R}^n$, so the function $\theta \mapsto \tilde{\nu}_t(\theta)$ is almost surely continuous in the compact Θ. Moreover, the sequence $\{h_\theta(y_t, x_t; \nu)\}$ is stationary and ergodic for any $\nu \in \mathbb{R}^n$, by assumption **A1** and Proposition 1. Therefore, the result follows by showing conditions C1-C2 in Bougerol (1993, Thm. 3.1). First note that

$$\mathrm{E}\left(\log^+ \|h_\theta(y_t, x_t; \nu)\|_\Theta\right) \leq c_0 + c_1 \|\nu\| + c_2 \mathrm{E}\, \|f(y_t)\| + c_3 \mathrm{E}\, \|x_t\| < \infty$$

where $c_0 = \|d\|_\Theta$, $c_1 = \|A\|_\Theta$, $c_2 = \|B\|_\Theta$, $c_3 = \|Q\|_\Theta$ and $\mathrm{E}\,\|f(y_t)\| < \infty$, $\mathrm{E}\,\|x_t\| < \infty$ by condition **S2** so C1 is satisfied. Moreover, set the simplified notation $h_\theta(y_t, x_t; \nu) = h_t(\nu)$ and let define $h_\theta^{(k)}(y_{t-k+1}, x_{t-k+1}; \nu) = h_t \circ h_{t-1} \circ h_{t-k+1}(\nu)$ the k-fold convolution of the function h_t. Then, by (4)

$$\left\| h_\theta^{(k)}(y_{t-k+1}, x_{t-k+1}; \nu) - h_\theta^{(k)}(y_{t-k+1}, x_{t-k+1}; \nu^*) \right\|_\Theta \leq \left|\left|\left| A^k \right|\right|\right|_\Theta \|\nu - \nu^*\|$$

almost surely, for some $k \geq 1$. To prove C2 we need to verify the contraction condition $\left|\left|\left| A^k \right|\right|\right|_\Theta < 1$. Assumption **A1** provides $\rho(A) < 1$ for all $\theta \in \Theta$. Therefore, by recalling the definition of spectral radius $\rho(A) = \lim_{k\to\infty} \left|\left|\left| A^k \right|\right|\right|^{1/k}$ we have that for some $k \geq 1$, $\left|\left|\left| A^k \right|\right|\right|^{1/k} < 1$ and $\left|\left|\left| A^k \right|\right|\right| < 1$ for all $\theta \in \Theta$. The contraction condition follows by the compactness of Θ. Finally, we show the boundedness of the uniform moment $\mathrm{E}\,\|\nu_t\|_\Theta$. We can rewrite (4) by backward substitution

$$\nu_t = \sum_{k=0}^{t-1} A^k \left(d + B z_{t-1-k} + Q x_{t-1-k}\right) + A^t \nu_0 \,, \quad \text{almost surely}\,,$$

by using the convention $A^0 = I$. Therefore, with probability 1 for t sufficiently large

$$\begin{aligned}
\|\nu_t\|_\Theta &\leq \sum_{k=0}^{\infty} \left|\left|\left| A^k \right|\right|\right|_\Theta \left(c_0 + c_2 \|z_{t-1-k}\| + c_3 \|x_{t-1-k}\|\right) + 1 \\
&\leq \sum_{k=0}^{\infty} C_A \rho^k \left(c_0 + c_2 \|z_{t-1-k}\| + c_3 \|x_{t-1-k}\|\right) + 1
\end{aligned}$$

with some constants $C_A \geq 1$ and $\rho \in (\rho(A), 1)$. The last inequality follows by recalling that $\rho(A) = \lim_{k\to\infty} \left|\left|\left| A^k \right|\right|\right|^{1/k}$ and $\rho(A) < \rho$ therefore there exists an k_0 such that for $k \geq k_0$, $\left|\left|\left| A^k \right|\right|\right|_\Theta < \rho^k$. It suffices now to take $C_A = \max\left\{1, \left|\left|\left| A^k \right|\right|\right|_\Theta / \rho^k : k = 0, 1, \ldots, k_0\right\}$ to guarantee that $\left|\left|\left| A^k \right|\right|\right|_\Theta \leq C_A \rho^k$ for any $k \geq 0$. By stationarity $\mathrm{E}\,\|\nu_t\|_\Theta < \infty$ holds for any t and the result follows. □

S2.2 Proof of Lemma C.1

The lemma is verified by proving the uniform convergence of the two summands of equation (S-2) below.

$$|\tilde{L}_T(\theta) - L(\theta)| \leq |\tilde{L}_T(\theta) - L_T(\theta)| + |L_T(\theta) - L(\theta)| \,. \tag{S-2}$$

Regarding the first term in (S-2) note that $\tilde{l}_t(\theta) = \sum_{h=1}^n y_{h,t} \log(\tilde{\lambda}_{h,t}(\theta)) - \tilde{\lambda}_{h,t}(\theta)$ and by applying the inequality $|\log(x/y)| \leq |x - y| / \min\{x, y\}$ we obtain

$$\|\tilde{l}_t - l_t\|_\Theta \leq \sum_{h=1}^n \left(\frac{y_{h,t}}{\underline{\lambda}_L} + 1\right) \|\tilde{\lambda}_{h,t} - \lambda_{h,t}\|_\Theta$$

due to assumption **A4**. Moreover, Straumann and Mikosch (2006, Lem. 2.1), assumption **A2** and the first moment of **A3** imply that $\|\tilde{l}_t - l_t\|_\Theta \xrightarrow{e.a.s.} 0$, as $t \to \infty$. Furthermore, $\|\tilde{L}_T - L_T\|_\Theta \leq T^{-1} \sum_{t=1}^T \|\tilde{l}_t - l_t\|_\Theta \to 0$, almost surely as $T \to \infty$ by Cesaro's lemma.

To prove uniform converge of the second term in (S-2) note that by the stationarity conditions in **A1** and the continuity of g^{-1} in **A2** the log-quasi-likelihood contribution $l_t(\theta)$ is a stationary ergodic continuous sequence in the compact space Θ. Moreover, by **A3**

$$\mathrm{E}\|l_t\|_\Theta \leq \mathrm{E}\left\|\sum_{h=1}^{n} y_{h,t}\log(\lambda_{h,t})\right\|_\Theta + \mathrm{E}\left(\sum_{h=1}^{n}\|\lambda_{h,t}\|_\Theta\right) < \infty\,.$$

Therefore, Straumann and Mikosch (2006, Thm. 2.7) applies providing $\|L_T - L\|_\Theta \to 0$ almost surely, as $T \to \infty$. □

S2.3 Proof of Lemma C.2

For a random variable $y \sim Pois(\lambda)$ with pmf $p(y;\lambda)$ we have that for any random variable $\lambda \in (0,\infty)$

$$p(y;\lambda) = p(y;\lambda_0) \quad a.s. \quad \forall y \in \mathbb{N} \quad \text{if and only if} \quad \lambda = \lambda_0 \quad a.s. \tag{S-3}$$

The if statement is trivial. To prove the only if part we have from the property of Poisson distribution

$$\sum_{y=0}^{\infty} y\; p(y,\lambda) = \lambda \quad a.s.$$

Therefore, we obtain that (S-3) holds only if $\lambda = \lambda_0$ with probability 1. By the previous argument it immediately follows that for all $h = 1,\ldots,n$ and all $y_h \in \mathbb{N}$

$$p_h(y_h;\lambda_{h,t}(\theta)) = p_h(y_h;\lambda_{h,t}(\theta_0)) \quad a.s. \quad \text{if and only if} \quad \lambda_{h,t}(\theta) = \lambda_{h,t}(\theta_0) \quad a.s.$$

with probability 1. Therefore, the desired result follows by showing that for any $\theta \in \Theta$ and all $h = 1,\ldots,n$, $\lambda_{h,t}(\theta) = \lambda_{h,t}(\theta_0)$ a.s. if and only if $\theta = \theta_0$. However, by **A2** g^{-1} is bijective therefore it suffices to show that for any $\theta \in \Theta$

$$\nu_t(\theta) = \nu_t(\theta_0) \quad a.s. \quad \text{if and only if} \quad \theta = \theta_0\,. \tag{S-4}$$

By Proposition 2, $\{\nu_t(\theta)\}_{t\in\mathbb{Z}}$ is a stationary sequence and therefore $\{\nu_t(\theta) - \nu_t(\theta_0)\}_{t\in\mathbb{Z}}$ is also stationary for any $\theta \in \Theta$. Hence, (S-4) should be verified for any $t \in \mathbb{Z}$. The result is proved by contradiction. Assuming that $\nu_{t-1}(\theta) = \nu_{t-1}(\theta_0) = \nu_{t-1}$ a.s., we have $\nu_t(\theta) - \nu_t(\theta_0) = d - d_0 + (A - A_0)\nu_{t-1} + (B - B_0)z_{t-1} + (Q - Q_0)x_{t-1}$. If $\nu_t(\theta) = \nu_t(\theta_0)$ a.s. with $d \neq d_0$ then $0 \neq d_0 - d = (A - A_0)\nu_{t-1} + (B - B_0)z_{t-1} + (Q - Q_0)x_{t-1}$ a.s. and the equality will be possible only if at least one of the following is satisfied: i) $A - A_0 \neq O$ and ν_{t-1} has at least one element equal to a non-zero constant. ii) $B - B_0 \neq O$ and z_{t-1} has at least one element equal to a non-zero constant. iii) $Q - Q_0 \neq O$ and x_{t-1} has at least one element equal to a non-zero constant. However, z_{t-1} and

x_{t-1} are non-constant vectors and by the conditions defined in **A5**, ν_{t-1} is non-degenerate. Then, it is a function of random variables in all its elements and is a non-constant vector. Therefore, if $\nu_t(\theta) = \nu_t(\theta_0)$ a.s. then $d = d_0$ and a.s. $0 = (A - A_0)\nu_{t-1} + (B - B_0)z_{t-1} + (Q - Q_0)x_{t-1}$. Now to have $\nu_t(\theta) = \nu_t(\theta_0)$ a.s. with $A \neq A_0$, $B \neq B_0$ and $Q \neq Q_0$ we shall have that at least one element of ν_{t-1}, z_{t-1} and x_{t-1} should equal 0 a.s., but this is impossible since they are non-constant vectors. Therefore, if $\nu_t(\theta) = \nu_t(\theta_0)$ a.s. then $A = A_0$, $B = B_0$ and $Q = Q_0$, and by the identification assumption in **A5** and Theorem 1 this implies that $A_s = A_{0,s}$, $B_s = B_{0,s}$ and $Q_s = Q_{0,s}$, for $s = 1, \ldots, N$. The result (S-4) follows. □

S2.4 Proof of Lemma C.3

The result is proved by applying Straumann and Mikosch (2006, Thm. 2.10) to the derivative function defined in (S-1). To this aim recall from section S1 that the derivatives of the vectorized tensor model under the identifiability constraints **A5** are $\nabla_\theta \nu_t = \Delta_{t-1}\Xi + A\nabla_\theta \nu_{t-1}$. Set the simplified notation $\xi_t(x) = \Delta_t \Xi + Ax$ so $\xi_t(0) = \Delta_t \Xi$. Therefore, the result follows by showing conditions S.1-S.3 in Straumann and Mikosch (2006, Thm. 2.10) for ξ_t equipped with the uniform norm $|||\cdot|||_\Theta$. Recall that $\Delta_t = [I_n, (\nu_t' \otimes I_n), (z_t' \otimes I_n), (x_t' \otimes I_n)]$ is a $n \times p^*$ block matrix, with $p^* = n + 3n^2$, and Ξ is a $p^* \times m_1$ matrix of parameters. From basic properties of norms on block matrices it follows

$$\begin{aligned}
\mathrm{E}|||\Delta_t \Xi|||_\Theta &\leq \mathrm{E}\left(|||I_n||| + |||\nu_t' \otimes I_n|||_\Theta + |||z_t' \otimes I_n||| + |||x_t' \otimes I_n|||\right) |||\Xi|||_\Theta \\
&\leq \left(1 + \mathrm{E}\,\|\nu_t\|_\Theta + \mathrm{E}\,\|z_t\| + \mathrm{E}\,\|x_t\|\right) c_\Xi \,,
\end{aligned}$$

which is finite by condition **S2** and Proposition 2, and the second inequality follows from $|||z' \otimes I||| = \max_h z_h \leq \|z\|$ with $c_\Xi = |||\Xi|||_\Theta$. Therefore S.1 holds. To prove S.2, define $\xi_t^{(k)}(x) = \xi_t \circ \xi_{t-1} \circ \xi_{t-k+1}(x)$ the k-fold convolution of the function ξ_t. Then,

$$\left|\left|\left|\xi_t^{(k)}(x) - \xi_t^{(k)}(y)\right|\right|\right|_\Theta \leq \left|\left|\left|A^k\right|\right|\right|_\Theta |||x - y|||$$

almost surely, for some $k \geq 1$. Following the proof of Proposition 2, we can conclude that $\left|\left|\left|A^k\right|\right|\right|_\Theta < 1$, by **A1**, therefore the contraction condition S.2 is verified. Finally, let $\tilde{\xi}_t(x) = \tilde{\Delta}_t \Xi + Ax$ where $\tilde{\Delta}_t$ is defined as Δ_t by substituting ν_t with $\tilde{\nu}_t$. Then, $\tilde{\xi}_t(0) - \xi_t(0) = [O_{n\times n}, (\tilde{\nu}_t - \nu_t)' \otimes I_n, O_{n\times 2n^2}]\Xi$. Moreover, almost surely

$$\left|\left|\left|\tilde{\xi}_t(0) - \xi_t(0)\right|\right|\right|_\Theta \leq |||(\tilde{\nu}_t - \nu_t)' \otimes I_n|||_\Theta c_\Xi \leq c_\Xi \,\|\tilde{\nu}_t - \nu_t\|_\Theta \xrightarrow{e.a.s.} 0\,, \qquad \text{as } t \to \infty\,,$$

where the convergence follows from Proposition 2. By further noting that $\nabla_x \tilde{\xi}_t - \nabla_x \xi_t = A - A = 0$, S.3 is satisfied. Finally, we prove that $\mathrm{E}|||\nabla_\theta \nu_t|||_\Theta < \infty$. Following (S-1) we can rewrite the derivative by backward substitution

$$\nabla_\theta \nu_t = \sum_{k=0}^{t-1} A^k \Delta_{t-1-k} + A^t \nabla_\theta \nu_0 \,, \quad \text{almost surely} \,.$$

Therefore, with probability 1 for t sufficiently large

$$|||\nabla_\theta \nu_t|||_\Theta \leq \sum_{k=0}^{\infty} |||A^k|||_\Theta |||\Delta_{t-1-k}\Xi|||_\Theta + 1 \leq \sum_{k=0}^{\infty} C_A \rho^k |||\Delta_{t-1-k}\Xi|||_\Theta + 1$$

where the constants $C_A \geq 1$ and $\rho \in (\rho(A), 1)$ are defined analogously to the proof of Proposition 2. The result follows by stationarity and $\mathrm{E}|||\Delta_t \Xi|||_\Theta < \infty$. □

S2.5 Proof of Lemma C.4

First note that $\nabla_\theta \tilde{l}_t = \sum_{h=1}^n (y_{h,t}/\tilde{\lambda}_{h,t}(\theta) - 1)\nabla_\theta \tilde{\lambda}_{h,t}$ and

$$\begin{aligned}
\nabla_\theta l_t - \nabla_\theta \tilde{l}_t &= \sum_{h=1}^n y_{h,t} \left[\frac{\tilde{\lambda}_{h,t}(\theta)\nabla_\theta \lambda_{h,t} - \lambda_{h,t}(\theta)\nabla_\theta \tilde{\lambda}_{h,t}}{\lambda_{h,t}(\theta)\tilde{\lambda}_{h,t}(\theta)} \right] + \nabla_\theta \tilde{\lambda}_{h,t} - \nabla_\theta \lambda_{h,t} \\
&= \sum_{h=1}^n y_{h,t} \left[\frac{\left(\tilde{\lambda}_{h,t}(\theta) - \lambda_{h,t}(\theta)\right) \nabla_\theta \lambda_{h,t} + \lambda_{h,t}(\theta) \left(\nabla_\theta \lambda_{h,t} - \nabla_\theta \tilde{\lambda}_{h,t}\right)}{\lambda_{h,t}(\theta)\tilde{\lambda}_{h,t}(\theta)} \right] + \nabla_\theta \tilde{\lambda}_{h,t} - \nabla_\theta \lambda_{h,t} \,.
\end{aligned}$$

Therefore,

$$\|\nabla_\theta l_t - \nabla_\theta \tilde{l}_t\|_\Theta \leq \sum_{h=1}^n \underline{\lambda}_L^{-2} y_{h,t} \|\nabla_\theta \lambda_{h,t}\|_\Theta \|\tilde{\lambda}_{h,t} - \lambda_{h,t}\|_\Theta + \left(1 + \underline{\lambda}_L^{-1} y_{h,t}\right) \|\nabla_\theta \tilde{\lambda}_{h,t} - \nabla_\theta \lambda_{h,t}\|_\Theta \,. \tag{S-5}$$

For $h = 1, \ldots, n$, we have that $\mathrm{E}[\log^+(y_{h,t}\|\nabla_\theta \lambda_{h,t}\|_\Theta)] \leq \mathrm{E}[\log^+(y_{h,t})] + \mathrm{E}(\log^+ \|\nabla_\theta \lambda_{h,t}\|_\Theta) < \infty$, by **A3** and **A7**, respectively. Moreover, by **A2** and Straumann and Mikosch (2006, Lem. 2.1) the first summand of (S-5) converges e.a.s. to 0, as $t \to \infty$. Regarding the second summand of (S-5), analogously as before $\mathrm{E}[\log^+(y_{h,t})] < \infty$, by **A3**. Furthermore,

$$\begin{aligned}
\|\nabla_\theta \tilde{\lambda}_{h,t} - \nabla_\theta \lambda_{h,t}\|_\Theta &= \|\nabla_{\tilde{\nu}_{h,t}} g^{-1} \nabla_\theta \tilde{\nu}_{h,t} - \nabla_{\nu_{h,t}} g^{-1} \nabla_\theta \nu_{h,t}\|_\Theta \\
&\leq \|\nabla_{\tilde{\nu}_{h,t}} g^{-1}\|_\Theta \|\nabla_\theta \tilde{\nu}_{h,t} - \nabla_\theta \nu_{h,t}\|_\Theta + \|\nabla_\theta \nu_{h,t}\|_\Theta \|\nabla_{\tilde{\nu}_{h,t}} g^{-1} - \nabla_{\nu_{h,t}} g^{-1}\|_\Theta \,,
\end{aligned}$$

where for large enough t

$$\|\nabla_\theta \tilde{\lambda}_{h,t} - \nabla_\theta \lambda_{h,t}\|_\Theta \leq \left(1 + \|\nabla_{\nu_{h,t}} g^{-1}\|_\Theta\right) \|\nabla_\theta \tilde{\nu}_{h,t} - \nabla_\theta \nu_{h,t}\|_\Theta + \|\nabla_\theta \nu_{h,t}\|_\Theta \|\nabla_{\tilde{\nu}_{h,t}} g^{-1} - \nabla_{\nu_{h,t}} g^{-1}\|_\Theta \,, \tag{S-6}$$

by **A6**-**A7**. The results of Lemma C.3 in Appendix C entail that $\|\nabla_\theta \tilde{\nu}_{h,t} - \nabla_\theta \nu_{h,t}\|_\Theta$ converges e.a.s to 0, as $t \to \infty$ and $\mathrm{E}(\log^+ \|\nabla_\theta \nu_{h,t}\|_\Theta) < \infty$. Therefore, Straumann and Mikosch (2006, Lem. 2.1) and **A7** establish that (S-6) converges e.a.s to 0, as $t \to \infty$. This implies that also (S-5) converges e.a.s to 0, as $t \to \infty$. An application of Straumann and Mikosch (2006, Lem. 2.1) provides

$$\lim_{T\to\infty} T\|S_T - \tilde{S}_T\|_\Theta \le \lim_{T\to\infty} \sum_{t=1}^{T} \|\nabla_\theta l_t - \nabla_\theta \tilde{l}_t\|_\Theta = \sum_{t=1}^{\infty} \|\nabla_\theta l_t - \nabla_\theta \tilde{l}_t\|_\Theta < \infty\,, \quad a.s.$$

which entails the result. □

S2.6 Proof of Lemma C.5

Recall that from (S-1), $\nabla_{\theta_0}\nu_t = \Delta_{t-1}\Xi + A\nabla_{\theta_0}\nu_{t-1}$, where all the parameters matrices Ξ and A are evaluated at $\theta = \theta_0$. We first prove that $\forall\eta \in \mathbb{R}^{m_1}$, $\Xi\eta = 0$ only if $\eta = 0$. Consider $N = 2$. For space convenience set $n_1 = n_2 = 2$ but the argument is valid for any arbitrary value of n_s. Then, following the notation of section S1, the matrix Ξ is a block matrix and the second block is related to the derivatives of $a = \mathrm{vec}(A)$:

$$(\nabla_{a_1} a, \nabla_{a_2} a)\,\eta_a = \begin{pmatrix}
0 & 0 & 0 & A_{11}^{(2)} & 0 & 0 & 0 \\
A_{11}^{(1)} & 0 & 0 & A_{21}^{(2)} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & A_{11}^{(2)} & 0 & 0 \\
A_{21}^{(1)} & 0 & 0 & 0 & A_{21}^{(2)} & 0 & 0 \\
0 & A_{11}^{(1)} & 0 & A_{12}^{(2)} & 0 & 0 & 0 \\
0 & 0 & A_{11}^{(1)} & A_{22}^{(2)} & 0 & 0 & 0 \\
0 & A_{21}^{(1)} & 0 & 0 & A_{12}^{(2)} & 0 & 0 \\
0 & 0 & A_{21}^{(1)} & 0 & A_{22}^{(2)} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & A_{11}^{(2)} & 0 \\
A_{12}^{(1)} & 0 & 0 & 0 & 0 & A_{21}^{(2)} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & A_{11}^{(2)} \\
A_{22}^{(1)} & 0 & 0 & 0 & 0 & 0 & A_{21}^{(2)} \\
0 & A_{12}^{(1)} & 0 & 0 & 0 & A_{12}^{(2)} & 0 \\
0 & 0 & A_{12}^{(1)} & 0 & 0 & A_{22}^{(2)} & 0 \\
0 & A_{22}^{(1)} & 0 & 0 & 0 & 0 & A_{12}^{(2)} \\
0 & 0 & A_{22}^{(1)} & 0 & 0 & 0 & A_{22}^{(2)}
\end{pmatrix} \eta_a = \begin{pmatrix}
A_{11}^{(2)}\eta_{a,4} \\
A_{11}^{(1)}\eta_{a,1} + A_{21}^{(2)}\eta_{a,4} \\
A_{11}^{(2)}\eta_{a,5} \\
A_{21}^{(1)}\eta_{a,1} + A_{21}^{(2)}\eta_{a,5} \\
A_{11}^{(1)}\eta_{a,2} + A_{12}^{(2)}\eta_{a,4} \\
A_{11}^{(1)}\eta_{a,3} + A_{22}^{(2)}\eta_{a,4} \\
A_{21}^{(1)}\eta_{a,2} + A_{12}^{(2)}\eta_{a,5} \\
A_{21}^{(1)}\eta_{a,3} + A_{22}^{(2)}\eta_{a,5} \\
A_{11}^{(2)}\eta_{a,6} \\
A_{12}^{(1)}\eta_{a,1} + A_{21}^{(2)}\eta_{a,6} \\
A_{11}^{(2)}\eta_{a,7} \\
A_{22}^{(1)}\eta_{a,1} + A_{21}^{(2)}\eta_{a,7} \\
A_{12}^{(1)}\eta_{a,2} + A_{12}^{(2)}\eta_{a,6} \\
A_{12}^{(1)}\eta_{a,3} + A_{22}^{(2)}\eta_{a,6} \\
A_{22}^{(1)}\eta_{a,2} + A_{12}^{(2)}\eta_{a,7} \\
A_{22}^{(1)}\eta_{a,3} + A_{22}^{(2)}\eta_{a,7}
\end{pmatrix} \tag{S-7}$$

where $\eta_a = (\eta_{a,1}, \eta_{a,2}, \ldots, \eta_{a,7})'$ is the subvector of η associated to the block of Ξ related to a. By **A8**, $A_{11}^{(2)} \neq 0$. Therefore if $(\nabla_{a_1} a, \nabla_{a_2} a)\eta_a = 0$ then the subvector $(\eta_{a,4}, \eta_{a,5}, \eta_{a,6}, \eta_{a,7})'$ needs to be zero. In such a case, the vector (S-7) reduces to

$$\left(0, \eta_{a,1}, 0, A_{21}^{(1)}\eta_{a,1}, \eta_{a,2}, \eta_{a,3}, A_{21}^{(1)}\eta_{a,2}, A_{21}^{(1)}\eta_{a,3}, 0, A_{12}^{(1)}\eta_{a,1}, 0, A_{22}^{(1)}\eta_{a,1}, A_{12}^{(1)}\eta_{a,2}, A_{12}^{(1)}\eta_{a,3}, A_{22}^{(1)}\eta_{a,2}, A_{22}^{(1)}\eta_{a,3}\right)', \tag{S-8}$$

where we have used the fact that $A_{11}^{(1)} = 1$, so also $(\eta_{a,1}, \eta_{a,2}, \eta_{a,3})' = 0$ whenever $(\nabla_{a_1} a, \nabla_{a_2} a)\eta_a = 0$. Analogous arguments apply to the other blocks of Ξ and by chain rule to the case $N \geq 3$ by following the notation described at the end of section S1.

Recall that $\Delta_{t-1} = [I_n, (\nu'_{t-1} \otimes I_n), (z'_{t-1} \otimes I_n), (x'_{t-1} \otimes I_n)]$ is a $n \times p^*$ block matrix. Hence, the columns of Δ_{t-1} are made by a single element of one of the vectors $\{1, \nu_{t-1}, y_{t-1}, x_{t-1}\}$. By **A8**, the elements of $(y'_{t-1}, x'_{t-1})'$ are linearly independent and are non-constant. Moreover, by the definition in (4) and **A5** the elements of ν_{t-1} are non-constant and are not linear combinations of each others or of $(y'_{t-1}, x'_{t-1})'$. Therefore, the columns of Δ_{t-1} are linearly independent implying that $\forall \pi \in \mathbb{R}^{p^*}$, $\Delta_{t-1}\pi = 0$ a.s. only if $\pi = 0$.

Now set $\pi = \Xi\eta$. By previous arguments, $\forall \eta \in \mathbb{R}^{m_1}$, $\pi = 0$ only if $\eta = 0$ and in turn $\Delta_{t-1}\pi = 0$ a.s. only if $\pi = 0$. Therefore we obtain that $\forall \eta \in \mathbb{R}^{m_1}$, $\Delta_{t-1}\Xi\eta = 0$ a.s. only if $\eta = 0$. Finally, we conclude the proof by contradiction. Note that $\nabla_{\theta_0}\nu_t\eta = \Delta_{t-1}\Xi\eta + A\nabla_{\theta_0}\nu_{t-1}\eta$. If $\nabla_{\theta_0}\nu_t\eta = 0$ a.s. for some $\eta \neq 0$ then also $\nabla_{\theta_0}\nu_{t-1}\eta = 0$ a.s. by stationarity. Then, we will have $\Delta_{t-1}\Xi\eta = 0$ a.s. for some $\eta \neq 0$ but this is impossible since we showed that $\Delta_{t-1}\Xi\eta = 0$ a.s. only if $\eta = 0$. Therefore, we conclude that $\forall \eta \in \mathbb{R}^{m_1}$, $\nabla_{\theta_0}\nu_t\eta = 0$ a.s. only if $\eta = 0$. □

S2.7 Proof of Lemma C.6

Note that $H(\theta_0) = \sum_{h=1}^{n} \mathrm{E}\left(\lambda_{h,t}^{-1}(\theta_0)\nabla_{\theta_0}\lambda_{h,t}\nabla_{\theta_0}\lambda'_{h,t}\right)$. Therefore, for each $\eta \in \mathbb{R}^{m_1}$, with $\eta \neq 0$, $\eta' H(\theta_0)\eta = \sum_{h=1}^{n} \mathrm{E}\left[\lambda_{h,t}^{-1}(\theta_0)(\eta'\nabla_{\theta_0}\lambda_{h,t})^2\right] \geq 0$, and to obtain positive definiteness we need to show that there exists at least one h such that $\eta'\nabla_{\theta_0}\lambda_{h,t} \neq 0$, for each $\eta \in \mathbb{R}^{m_1}$, with $\eta \neq 0$. Note that $\eta'\nabla_{\theta_0}\lambda_{h,t} = \eta'\nabla_{\theta_0}\nu_{h,t}\nabla_{\nu_{h,t}}g^{-1}$ and $\nabla_{\nu_{h,t}}g^{-1} \neq 0$ since $g(\cdot)$ is a non-constant invertible function. Finally, the result of Lemma C.5 in Appendix C implies that there exists at least one h such that $\eta'\nabla_{\theta_0}\nu_{h,t} \neq 0$, for each $\eta \in \mathbb{R}^{m_1}$, with $\eta \neq 0$. Therefore, $H(\theta_0)$ is positive definite. Recall that $I(\theta_0) = \mathrm{E}(\nabla_{\theta_0}l_t\, \nabla_{\theta_0}l'_t)$ where $\nabla_{\theta_0}l_t = \sum_{h=1}^{n}(y_{h,t}/\lambda_{h,t}(\theta_0) - 1)\nabla_{\theta_0}\lambda_{h,t} = \sum_{h=1}^{n} m_{h,t}\nabla_{\theta_0}\lambda_{h,t}$. Therefore, for each $\eta \in \mathbb{R}^{m_1}$, with $\eta \neq 0$, $\eta' I(\theta_0)\eta = \mathrm{E}\left[(\eta'\nabla_{\theta_0}l_t)^2\right] \geq 0$ and the strict inequality follows if $\eta'\nabla_{\theta_0}l_t \neq 0$, for each $\eta \in \mathbb{R}^{m_1}$, with $\eta \neq 0$. The last condition is satisfied if there exists at least one h such that $\eta'\nabla_{\theta_0}\lambda_{h,t} \neq 0$, for each $\eta \in \mathbb{R}^{m_1}$, with $\eta \neq 0$, because a.s. $m_{h,t} \neq 0$ and the equality $m_{h,t}\eta'\nabla_{\theta_0}\lambda_{h,t} = -\sum_{l\neq h} m_{l,t}\eta'\nabla_{\theta_0}\lambda_{l,t}$ is impossible. Indeed, in $\eta'\nabla_{\theta_0}\lambda_{h,t} = -m_{h,t}^{-1}\sum_{l\neq h} m_{l,t}\eta'\nabla_{\theta_0}\lambda_{l,t}$ the left-hand side is $\mathcal{F}_{t-1}$ measurable whereas the right-hand side is not since it depends on y_t. Hence, $I(\theta_0)$ is positive definite. The positive definiteness of the matrix Σ follows since for all $\delta \in \mathbb{R}^{m_1}$, with $\delta \neq 0$, we have $H(\theta_0)^{-1}\delta \neq 0$ since $H(\theta_0)^{-1}$ is full rank. Now by setting $\eta = H(\theta_0)^{-1}\delta$ we proved that $\eta' I(\theta_0)\eta > 0$. Therefore, it follows that $\delta' H(\theta_0)^{-1} I(\theta_0) H(\theta_0)^{-1}\delta > 0$. □

S2.8 Proof of Lemma C.7

Recall that the second derivative matrix of the log-quasi-likelihood contribution is

$$\nabla_\theta^2 l_t = \sum_{h=1}^{n} -\frac{y_{h,t}}{\lambda_{h,t}^2(\theta)} \nabla_\theta \lambda_{h,t} \nabla_\theta \lambda_{h,t}' + \left(\frac{y_{h,t}}{\lambda_{h,t}(\theta)} - 1 \right) \nabla_\theta^2 \lambda_{h,t} \,.$$

Furthermore, by **A6**, $\nabla_\theta^2 l_t(\theta)$ is a stationary ergodic continuous sequence in the compact space Θ. Moreover,

$$\mathrm{E} |||\nabla_\theta^2 l_t|||_\Theta \leq \sum_{h=1}^{n} \mathrm{E} \left(y_{h,t} |||\lambda_{h,t}^{-2} \nabla_\theta \lambda_{h,t} \nabla_\theta \lambda_{h,t}' |||_\Theta \right) + \mathrm{E} \left(y_{h,t} |||\lambda_{h,t}^{-1} \nabla_\theta^2 \lambda_{h,t}|||_\Theta \right) + \mathrm{E} \, |||\nabla_\theta^2 \lambda_{h,t}|||_\Theta$$

is finite by **A4**, **A9** and Cauchy-Schwarz inequality. Therefore, Straumann and Mikosch (2006, Thm. 2.7) applies providing the result. □

S2.9 Proof of Lemma C.8

To prove uniform converge note that by (12) the log-likelihood contribution $l_t(\psi)$ is continuous in ψ since it is a function composition of continuous functions. Moreover, by the stationarity conditions in **A1** and the compactness conditions in **A1**,**B1**, $l_t(\psi)$ is a stationary ergodic continuous sequence in the compact space Ψ. Furthermore, an application of Debaly and Truquet (2023, Lem. 6-7) provides

$$\begin{aligned} \mathrm{E} \|l_t\|_\Psi \leq \gamma_1 + \gamma_2 \sum_{h=1}^{n} \mathrm{E} \| \log\left[p_h(0;\lambda_{h,t})\right] \|_\Theta + \gamma_3 \sum_{h=1}^{n} \mathrm{E} \| \log\left[1 - p_h(0;\lambda_{h,t})\right] \|_\Theta \\ + \gamma_4 \sum_{h=1}^{n} \mathrm{E} \| \log\left[1 - F_h(y_{h,t};\lambda_{h,t})\right] \|_\Theta + \gamma_5 \sum_{h=1}^{n} \mathrm{E} \| \log\left[p_h(y_{h,t};\lambda_{h,t})\right] \|_\Theta \,, \end{aligned} \tag{S-9}$$

where $\gamma_j \geq 0$, for $j = 1, \dots, 5$ are non-negative constants. We bound the single summands of (S-9). First note that $\mathrm{E} \| \log\left[p_h(0;\lambda_{h,t})\right] \|_\Theta = \mathrm{E} \| \log\left[\exp(-\lambda_{h,t})\right] \|_\Theta = \mathrm{E} \|\lambda_{h,t}\|_\Theta < \infty$ by **A3**. Regarding the second term by using the elementary inequalities $-\log(1-x) \leq x/(1+x)$, for $x < 1$, $x \neq 0$, and $e^{-y}/(1+e^{-y}) \leq e^{-x}/(1+e^{-x})$, for $y \geq x$, we obtain that $\mathrm{E} \| \log\left[1 - p_h(0;\lambda_{h,t})\right] \|_\Theta = \mathrm{E} \sup_{\theta \in \Theta} \left\{ -\log\left[1 - p_h(0;\lambda_{h,t}(\theta))\right] \right\} = \mathrm{E} \sup_{\theta \in \Theta} \left\{ -\log\left[1 - \exp(-\lambda_{h,t}(\theta))\right] \right\} \leq e^{-\lambda_L}/(1+e^{-\lambda_L}) < \infty$ by **A4**. The last two terms can be bounded by employing the following inequalities $\log(y!) = \sum_{j=1}^{y} \log(j) \leq y(y-1)/2$, since $\log(j) \leq j - 1$; $1 - F_h(y_{h,t};\lambda_{h,t}) \geq p_h(y_{h,t}+1;\lambda_{h,t})$, since $1 = \sum_{j=0}^{\infty} p_h(j;\lambda_{h,t})$, and

$$-\log\left[p_h(y_{h,t};\lambda_{h,t})\right] \leq \lambda_{h,t} - y_{h,t} \log(\lambda_{h,t}) + \sum_{j=1}^{y_{h,t}} \log(j) \leq \lambda_{h,t} - y_{h,t} \log(\lambda_{h,t}) + \frac{y_{h,t}(y_{h,t}-1)}{2} \,.$$

Furthermore, recall that $\mathrm{E}(y_{h,t}^2|\mathcal{F}_{t-1}) = \lambda_{h,t}^2 + \lambda_{h,t}$ where we use the shorthand $\lambda_t = \lambda_t(\theta_0)$. Therefore, $\mathrm{E}\| \log [p_h(y_{h,t}; \lambda_{h,t})] \|_\Theta \leq C_0 + C_1 \mathrm{E}\|\lambda_{h,t}\|_\Theta + C_2 \mathrm{E}\|\lambda_{h,t}^2\|_\Theta < \infty$ by **B2**, where C_0, C_1, C_2 are positive constants. Following an analogous argument $\mathrm{E}\| \log [1 - F_h(y_{h,t}; \lambda_{h,t})] \|_\Theta < \infty$. Hence, Straumann and Mikosch (2006, Thm. 2.7) applies providing the result. $\square$

S2.10 Proof of Lemma C.9

The if statement is trivial. Set the simplified notation $p(r) = p(y_t; \lambda_t, r)$ with $\lambda_t = \lambda_t(\theta_0)$ and the single elements of $R = (R_{ij})$, $R_0 = (R_{0,ij})$. For $i, j = 1, \ldots, n$ define by $p(R_{ij}) = p(y_{i,t}, y_{j,t}; \lambda_{i,t}, \lambda_{j,t}, R_{ij})$ the conditional bivariate distribution of $(y_{i,t}, y_{j,t})'$. Since the process $y_t|\mathcal{F}_{t-1}$ follows the Gaussian copula-based distribution (5), by the marginalization property of the multivariate Gaussian distribution we have that $(y_{i,t}, y_{j,t})'|\mathcal{F}_{t-1}$ follows the bivariate Gaussian copula-based distribution whose pmf can be written as

$$p(y_{i,t}, y_{j,t}; \lambda_{i,t}, \lambda_{j,t}, R_{ij}) = \int_{\Phi^{-1}[F_i(y_{i,t}-1;\lambda_{i,t})]}^{\Phi^{-1}[F_i(y_{i,t};\lambda_{i,t})]} \int_{\Phi^{-1}[F_j(y_{j,t}-1;\lambda_{j,t})]}^{\Phi^{-1}[F_j(y_{j,t};\lambda_{j,t})]} \phi_{R_{ij}}(x_i, x_j) dx_i dx_j \tag{S-10}$$

where $\phi_{R_{ij}}$ is the bivariate standard normal distribution with correlation parameter R_{ij}. Therefore, if $p(r) = p(r_0)$ then $p(R_{ij}) = p(R_{0,ij})$ for all $i, j = 1, \ldots, n$. Moreover, define the function

$$\tilde{\Phi}(\sigma) = \int_{-\infty}^{w_i} \int_{-\infty}^{w_j} \phi_\sigma(x_i, x_j) dx_i dx_j = \int_{-\infty}^{w_i} \Phi\left(\frac{w_j - \sigma x_i}{\sqrt{1-\sigma^2}}\right) \phi(x_i) dx_i$$

for some real values w_i and w_j, where ϕ and Φ are the standard normal pdf and cdf, respectively. The second equality comes from the application of a well-known integral reduction formula (Curnow and Dunnett, 1962, eq. (2.5)). After some algebra one finds that

$$\nabla_\sigma \tilde{\Phi} = -\frac{1}{1+\bar{\sigma}^2} \phi(\underline{\sigma}_j - \bar{\sigma} w_i) \phi(w_i)$$

where $\bar{\sigma} = \sigma(1-\sigma^2)^{-1/2}$ and $\underline{\sigma}_j = w_j(1-\sigma^2)^{-1/2}$. We see that $\nabla_\sigma \tilde{\Phi}$ is negative therefore $\tilde{\Phi}(\cdot)$ is decreasing. Finally, for all $i, j = 1, \ldots, n$ it follows that $\tilde{\Phi}(R_{ij}) = \tilde{\Phi}(R_{0,ij})$ by (S-10), and since $\tilde{\Phi}(\cdot)$ is one-to-one we conclude that $R_{ij} = R_{0,ij}$. Hence, $p(r) = p(r_0)$ implies $r = r_0$. $\square$

S2.11 Proof of Theorem 5

The result follows by proving the assumptions of Theorems 2-4 for model (18). The conditions **S2**-**S3** depend on the function $f(\cdot)$. When such a function is the identity, as in the linear model (17) it is straightforward to see that the conditions hold with $s_h = 1$ for $h = 1, \ldots, n$. Since here A_0, B_0 are non-negative matrices by definition the stationarity condition of Proposition 1 simplifies to $\rho(A_0 + B_0) < 1$. Therefore, **A1** follows. Since g is the identity function **A2** is verified by Proposition 2. By the Poisson property (Debaly and Truquet, 2023, Lem. 8,10) it follows that $\mathrm{E}\|y_t\|^p < \infty$ for any $p \geq 1$. Moreover, by following the same arguments of the proof of Proposition 2 we obtain that for t sufficiently large

$$(\mathrm{E}\,\|\lambda_t\|_\Theta^p)^{\frac{1}{p}} \leq \sum_{k=0}^{\infty} C_A \rho^k \left[c_0 + c_2 \left(\mathrm{E}\,\|y_{t-1-k}\|^p\right)^{\frac{1}{p}} + c_3 \left(\mathrm{E}\,\|x_{t-1-k}\|^p\right)^{\frac{1}{p}}\right] + 1 < \infty\,,$$

by Minkowski inequality, with some constants $c_0, c_2, c_3 \geq 0$, $C_A \geq 1$ and $\rho \in (\rho(A), 1)$. This fact entails that $\mathrm{E}\|\lambda_t\|_\Theta^p < \infty$ for any $p \geq 1$. Therefore, **A3** follows. By the non-negativity of the summands in (18) we shall see that $\lambda_t(\theta) \geq d$ and $\tilde{\lambda}_t(\theta) \geq d$ for all $\theta \in \Theta$, all $t \geq 1$ and any fixed initialization $\tilde{\lambda}_0 \in \mathbb{R}_+^n$. The condition **A4** holds by the compactness of Θ. This proves the almost sure consistency of the QMLE. Moreover, **B2** is verified hence the 2SMLE is strongly consistent.

The convergence condition in **A7** holds trivially since $\nabla_{\nu_{h,t}} g^{-1} = \partial\lambda_{h,t}/\partial\nu_{h,t} = \partial\tilde{\lambda}_{h,t}/\partial\tilde{\nu}_{h,t} = 1$. For the same reason the second expectation is 0. To prove the finiteness of the first expectation of **A7** note that by arguments analogous to the proof of Lemma C.3 we have that for t sufficiently large

$$(\mathrm{E}|||\nabla_\theta \lambda_t|||_\Theta^p)^{\frac{1}{p}} \leq \sum_{k=0}^{\infty} C_A \rho^k \,(\mathrm{E}|||\Delta_{t-1-k}\Xi|||_\Theta^p)^{\frac{1}{p}} + 1\,,$$

with Δ_{t-1-k} and Ξ defined as in section S1, and following the proof of Lemma C.3

$$(\mathrm{E}|||\Delta_t \Xi|||_\Theta^p)^{\frac{1}{p}} \leq \left[1 + (\mathrm{E}\,\|\lambda_t\|_\Theta^p)^{\frac{1}{p}} + (\mathrm{E}\,\|y_t\|^p)^{\frac{1}{p}} + (\mathrm{E}\,\|x_t\|^p)^{\frac{1}{p}}\right] |||\Xi|||_\Theta < \infty\,.$$

Therefore $\mathrm{E}|||\nabla_\theta \lambda_t|||^p < \infty$ for any $p \geq 1$ and **A7** holds. Analogous recursions hold for the second derivatives since for $i, j = 1, \ldots, m_1$ all the terms $\partial^2\lambda_t/\partial\theta_i\partial\theta_j$ can be bounded suitably along the arguments of Fokianos et al. (2009, Supp. Mat., Proof of Lem. 3.4). We omit the details. Therefore $\mathrm{E}\,|||\nabla_\theta^2 \lambda_{h,t}|||_\Theta^p < \infty$ for any $p \geq 1$ and $h = 1, \ldots, n$. This fact and Cauchy-Schwarz inequality entail **A9** providing the asymptotic normality of the QMLE. □

S2.12 Proof of Theorem 6

The result follows by proving the assumptions of Theorems 2-4 for model (20). By Debaly and Truquet (2023, eq. (6)) conditions **S2-S3** hold for $f(\cdot) = \log(\cdot + 1)$ with $s_h = 1$ for $h = 1, \ldots, n$. Moreover, the implied stationarity condition is verified since $\rho(|A|_e + |B|_e) \leq |||\,|A_0|_e + |B_0|_e|||_\infty < 1$. Therefore, **A1** follows. Following the same argument of Debaly and Truquet (2023, Prop. 4) we have that $\mathrm{E}\|y_t\|^p < \infty$ for any $p \geq 1$. Moreover, by using the same backward recursion arguments of the proof of Proposition 2 we obtain that with probability 1 for t sufficiently large

$$\exp(p\|\nu_t\|_\Theta) \le \exp(p)\exp\left(p\sum_{k=0}^{\infty} C_A\rho^k\delta_{k,t}\right) = \exp(p)\exp\left(\frac{pC_A}{1-\rho}\sum_{k=0}^{\infty}(1-\rho)\rho^k\delta_{k,t}\right)$$
$$\le \exp(p)\sum_{k=0}^{\infty}(1-\rho)\rho^k\exp\left(K\delta_{k,t}\right) ,$$

where $\delta_{k,t} = c_0 + c_2\left\|z_{t-1-k}\right\| + c_3\left\|x_{t-1-k}\right\|$, $K = pC_A/(1-\rho)$, and the last inequality holds by the convexity of the function $\exp(Kx)$ since $(1-\rho)\rho^k > 0$ and $\sum_{k=0}^{\infty}(1-\rho)\rho^k = 1$. Therefore, we shall have $\mathrm{E}\left[\exp(p\|\nu_t\|_\Theta)\right] \le \exp(p)\sum_{k=0}^{\infty}(1-\rho)\rho^k\mathrm{E}\left[\exp\left(K\delta_{k,t}\right)\right]$ and

$$\mathrm{E}\left[\exp\left(K\delta_{k,t}\right)\right] = \exp(Kc_0)\mathrm{E}\left[\exp\left(Kc_2\left\|z_{t-1-k}\right\|\right)\exp\left(Kc_3\left\|x_{t-1-k}\right\|\right)\right]$$
$$\le \exp(Kc_0)\left\{\mathrm{E}\left[\exp\left(2Kc_2\left\|z_t\right\|\right)\right]\right\}^{\frac{1}{2}}\left\{\mathrm{E}\left[\exp\left(2Kc_3\left\|x_t\right\|\right)\right]\right\}^{\frac{1}{2}}$$

which is finite since $\mathrm{E}\left[\exp\left(p\left\|z_t\right\|\right)\right] = \mathrm{E}\left[(y_{1,t}+1)^p(y_{2,t}+1)^p\cdots(y_{n,t}+1)^p\right]$ is finite for every $p \ge 1$ by Holder's inequality and the fact that $\mathrm{E}\|y_t\|^p < \infty$ for every $p \ge 1$. This establishes that $\mathrm{E}\left[\exp(p\|\nu_t\|_\Theta)\right] < \infty$ and $\mathrm{E}\|\exp(p\nu_{h,t})\|_\Theta < \infty$ for all $p \ge 1$. Therefore, $\mathrm{E}\|\exp(p\nu_t)\|_\Theta < \infty$ for all $p \ge 1$, where the exponential is applied elementwise, and **A3** follows.

Recall that $g(\cdot) = \log(\cdot)$ and that for $x, y \in \mathbb{R}$, $|e^x - e^y| = e^y|e^{x-y} - 1| \le |x-y|e^{|x-y|}e^y$. Then, with probability 1

$$\|\exp(\tilde{\nu}_{h,t}) - \exp(\nu_{h,t})\|_\Theta \le \|\tilde{\nu}_{h,t} - \nu_{h,t}\|_\Theta\exp(\|\tilde{\nu}_{h,t} - \nu_{h,t}\|_\Theta)\exp(\|\nu_{h,t}\|_\Theta) ,$$

where for t large enough $\exp(\|\tilde{\nu}_{h,t} - \nu_{h,t}\|_\Theta) \le \exp(1)$ a.s. by Proposition 2. Straumann and Mikosch (2006, Lem. 2.1) together with $\mathrm{E}\{\log^+[\exp(\|\nu_{h,t}\|_\Theta + 1)]\} \le \mathrm{E}\left\|\nu_{h,t}\right\|_\Theta + 1 < \infty$ establishes **A2** and proves the almost sure consistency of the QMLE. Moreover, **B2** is verified hence the 2SMLE is strongly consistent.

The convergence condition in **A7** holds here since it is equivalent to **A2**. Moreover,

$$\mathrm{E}\left(\log^+\left\|\nabla_\theta\lambda_{h,t}\right\|_\Theta\right) = \mathrm{E}\left(\log^+\left\|\exp(\nu_{h,t})\nabla_\theta\nu_{h,t}\right\|_\Theta\right) \le \mathrm{E}\left(\log^+(\exp\left\|\nu_{h,t}\right\|_\Theta)\right) + \mathrm{E}\left(\log^+\left\|\nabla_\theta\nu_{h,t}\right\|_\Theta\right) .$$

The last two moments are finite since $\mathrm{E}\|\nu_t\|_\Theta < \infty$ and by following arguments on the same line of the proof of Theorem 5, we have $\mathrm{E}\|\|\nabla_\theta\nu_t\|\|_\Theta^p < \infty$ and $\mathrm{E}\|\|\nabla_\theta^2\nu_{h,t}\|\|_\Theta^p < \infty$ for any $p \ge 1$ and $h = 1, \dots, n$. The remaining moment in **A7** is bounded by $\mathrm{E}\|\nu_t\|_\Theta$ and therefore the assumption holds. Similarly, condition **A9** can be verified by noting that

$$\nabla_\theta^2\lambda_{h,t} = \exp(\nu_{h,t}(\theta))\nabla_\theta\nu_{h,t}\nabla_\theta\nu_{h,t}' + \exp(\nu_{h,t}(\theta))\nabla_\theta^2\nu_{h,t}$$

and all the elements possess uniform moments of any order. This fact and Cauchy-Schwarz inequality entail **A9** providing the asymptotic normality of the QMLE. □

S3 Further numerical and empirical results

We provide additional results for the simulation study in section 6.2. We perform an analogous simulation study for the copula tensor log-linear INGARCH model (1),(19) with true parameters $d = (0.1, 0.4, 0.7, 1.0)'$,

$$A_1 = \begin{pmatrix} 1 & 0.125 \\ 0.1 & 0.15 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.1 & 0.15 \\ 0.125 & 0.15 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 0.15 \\ 0.1 & 0.2 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0.2 & 0.2 \\ 0.15 & 0.3 \end{pmatrix},$$

and identical simulation setup. Figure S-1-S-2 report the kernel density of the QML and 2SQML estimates for the parameters of the log-linear model. Analogous comments to the linear model case apply and therefore the estimators work satisfactorily.

In Table S-1 we consider further results for the QMLE and 2SMLE for the simulation study described in section 6.2. The numerical optimization of the likelihood is initialized using the true parameter value. We can see that the RMSE of the estimators decreases as the sample size increases for all the parameters. This further confirms the consistency of the QML/2SQML estimates. The estimation bias for the parameters is quite moderate for all sample sizes since the standard deviation is close to the RMSE. Moreover, the small-sample bias in the estimates tends to become negligible in larger samples for almost all the parameters.

We repeated the same simulation study by using a full copula correlation matrix R where the single element is generated as $R_{ij} = 0.35^{|i-j|}$. For the 2SMLE of the correlation matrix the adjusted estimator (16) is employed. The kernel densities of all estimators are reported in Figures S-3-S-6 and suggest conclusions analogous to the equicorrelation case.

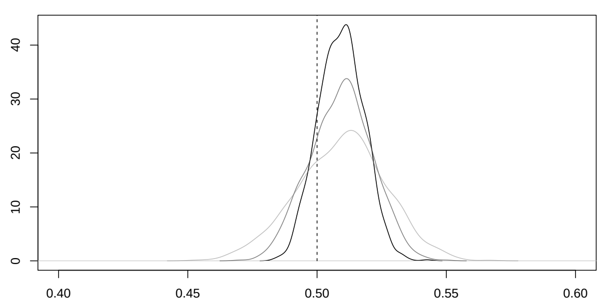

Figure S-1: Kernel density estimator for the distribution of the 2SQML estimates for the log-linear copula tensor INGARCH model with equicorrelation structure obtained from 1000 Monte Carlo replications for sample sizes 500 (——), 1000 (——) and 2000 (——), true parameter value (- - - -); the bandwidth of the kernel is selected by using Silverman's rule of thumb.

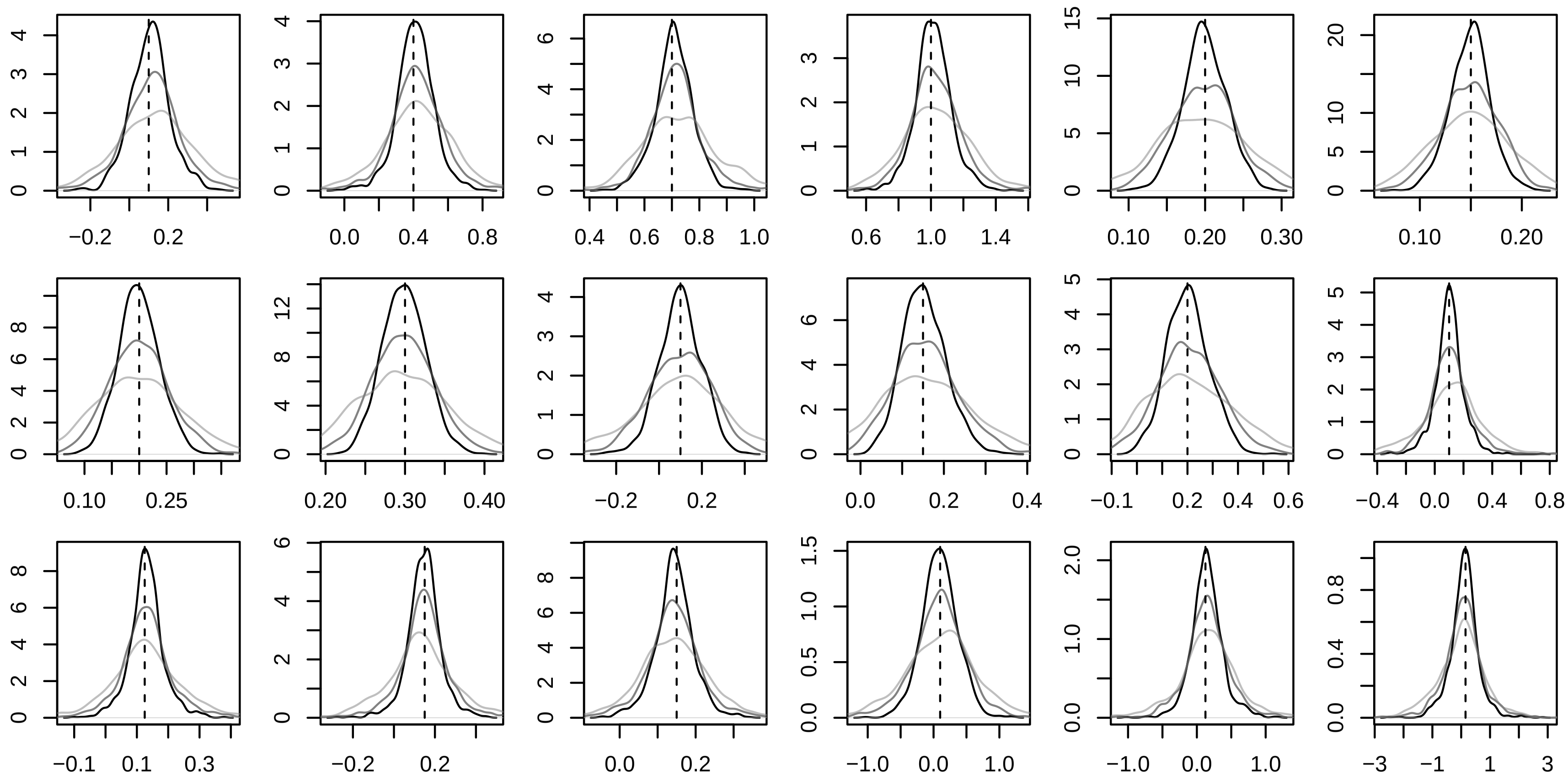

Figure S-2: Kernel density estimator for the distribution of the QML estimates for the copula tensor log-linear INGARCH model with equicorrelation structure obtained from 1000 Monte Carlo replications for sample sizes 500 (light grey line), 1000 (grey line) and 2000 (black line), true parameter values (- - - -); the bandwidth of the kernel is selected by using Silverman's rule of thumb; the plots are reported rowwise following the order of parameter vector $\theta = (d', \text{vec}(B_2)', \text{vec}(B_1)' \, \text{vec}(A_2)', \text{vec}(A_1)')'$.

Table S-1: Simulation results for the QML and 2SQML estimators obtained from 1000 Monte Carlo replications for linear (top) and log-linear (bottom) copula tensor INGARCH model with equicorrelation structure. The mean, the standard deviation (SD) and the root mean square error (RMSE) are reported for different sample sizes.

		d_1	d_2	d_3	d_4	$B_{11}^{(2)}$	$B_{21}^{(2)}$	$B_{12}^{(2)}$	$B_{22}^{(2)}$	$B_{21}^{(1)}$	$B_{12}^{(1)}$	$B_{22}^{(1)}$	$A_{11}^{(2)}$	$A_{21}^{(2)}$	$A_{12}^{(2)}$	$A_{22}^{(2)}$	$A_{21}^{(1)}$	$A_{12}^{(1)}$	$A_{22}^{(1)}$	$\bar{r}$
True values		0.500	0.667	0.833	1.000	0.300	0.200	0.250	0.250	0.150	0.200	0.300	0.300	0.175	0.200	0.200	0.150	0.200	0.300	0.500
$T = 500$	Mean	0.734	0.834	1.166	1.151	0.290	0.195	0.247	0.244	0.149	0.245	0.307	0.286	0.173	0.211	0.194	0.144	0.296	0.327	0.502
	SD	0.615	0.399	0.688	0.341	0.039	0.037	0.042	0.043	0.035	0.140	0.104	0.046	0.036	0.049	0.046	0.039	0.217	0.156	0.017
	RMSE	0.658	0.432	0.764	0.373	0.040	0.038	0.042	0.043	0.035	0.147	0.104	0.048	0.036	0.050	0.046	0.039	0.237	0.158	0.017
$T = 1000$	Mean	0.603	0.750	0.976	1.069	0.295	0.197	0.247	0.246	0.150	0.232	0.305	0.289	0.173	0.209	0.198	0.144	0.261	0.325	0.502
	SD	0.364	0.248	0.437	0.207	0.027	0.027	0.029	0.032	0.027	0.109	0.076	0.034	0.029	0.038	0.036	0.029	0.154	0.128	0.012
	RMSE	0.378	0.261	0.459	0.218	0.027	0.027	0.029	0.032	0.027	0.114	0.076	0.036	0.029	0.039	0.036	0.030	0.166	0.130	0.012
$T = 2000$	Mean	0.550	0.711	0.913	1.037	0.297	0.198	0.249	0.248	0.151	0.218	0.300	0.292	0.175	0.207	0.198	0.146	0.239	0.318	0.502
	SD	0.257	0.162	0.285	0.142	0.019	0.019	0.021	0.022	0.019	0.076	0.050	0.024	0.022	0.027	0.028	0.024	0.111	0.101	0.009
	RMSE	0.262	0.168	0.296	0.147	0.020	0.019	0.021	0.022	0.019	0.078	0.050	0.025	0.022	0.028	0.028	0.024	0.118	0.102	0.009
True values		0.100	0.400	0.700	1.000	0.200	0.150	0.200	0.300	0.100	0.150	0.200	0.100	0.125	0.150	0.150	0.100	0.125	0.150	0.500
$T = 500$	Mean	0.132	0.429	0.726	1.036	0.195	0.149	0.199	0.296	0.092	0.162	0.210	0.108	0.129	0.136	0.141	0.088	0.173	0.004	0.509
	SD	0.219	0.222	0.142	0.224	0.061	0.040	0.081	0.058	0.213	0.115	0.178	0.229	0.123	0.185	0.102	0.600	0.443	0.897	0.017
	RMSE	0.222	0.224	0.144	0.227	0.062	0.040	0.081	0.058	0.214	0.116	0.178	0.229	0.123	0.186	0.102	0.600	0.446	0.908	0.019
$T = 1000$	Mean	0.110	0.423	0.714	1.026	0.198	0.150	0.199	0.298	0.097	0.158	0.206	0.097	0.131	0.150	0.142	0.081	0.153	0.074	0.509
	SD	0.154	0.166	0.100	0.158	0.042	0.028	0.055	0.040	0.149	0.079	0.124	0.158	0.085	0.130	0.074	0.418	0.336	0.690	0.012
	RMSE	0.154	0.167	0.101	0.160	0.042	0.028	0.054	0.040	0.149	0.079	0.125	0.158	0.086	0.130	0.074	0.418	0.337	0.694	0.015
$T = 2000$	Mean	0.107	0.410	0.708	1.015	0.200	0.149	0.200	0.299	0.099	0.153	0.202	0.101	0.131	0.146	0.144	0.094	0.146	0.105	0.509
	SD	0.102	0.109	0.070	0.113	0.028	0.020	0.038	0.027	0.098	0.053	0.083	0.100	0.056	0.083	0.051	0.275	0.235	0.499	0.009
	RMSE	0.102	0.110	0.070	0.114	0.028	0.020	0.038	0.027	0.098	0.053	0.083	0.100	0.056	0.083	0.051	0.275	0.235	0.501	0.013

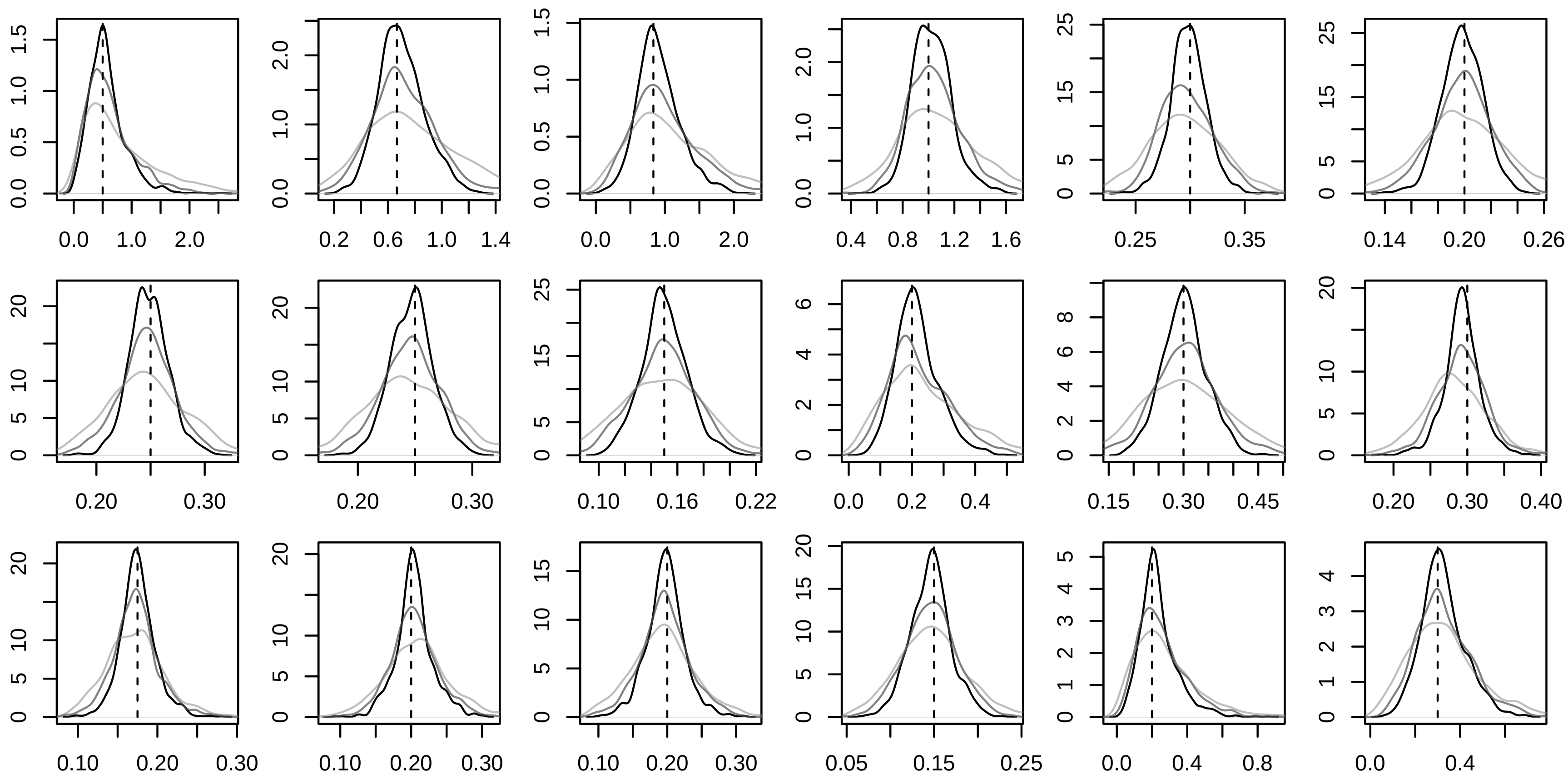

Figure S-3: Kernel density estimator for the distribution of the QML estimates for the copula tensor INGARCH model with full correlation structure obtained from 1000 Monte Carlo replications for sample sizes 500 (——), 1000 (——) and 2000 (——), true parameter values (- - - -); the bandwidth of the kernel is selected by using Silverman's rule of thumb; the plots are reported rowwise following the order of parameter vector $\theta = (d', \mathrm{vec}(B_2)', \mathrm{vec}(B_1)' \, \mathrm{vec}(A_2)', \mathrm{vec}(A_1)')'$.

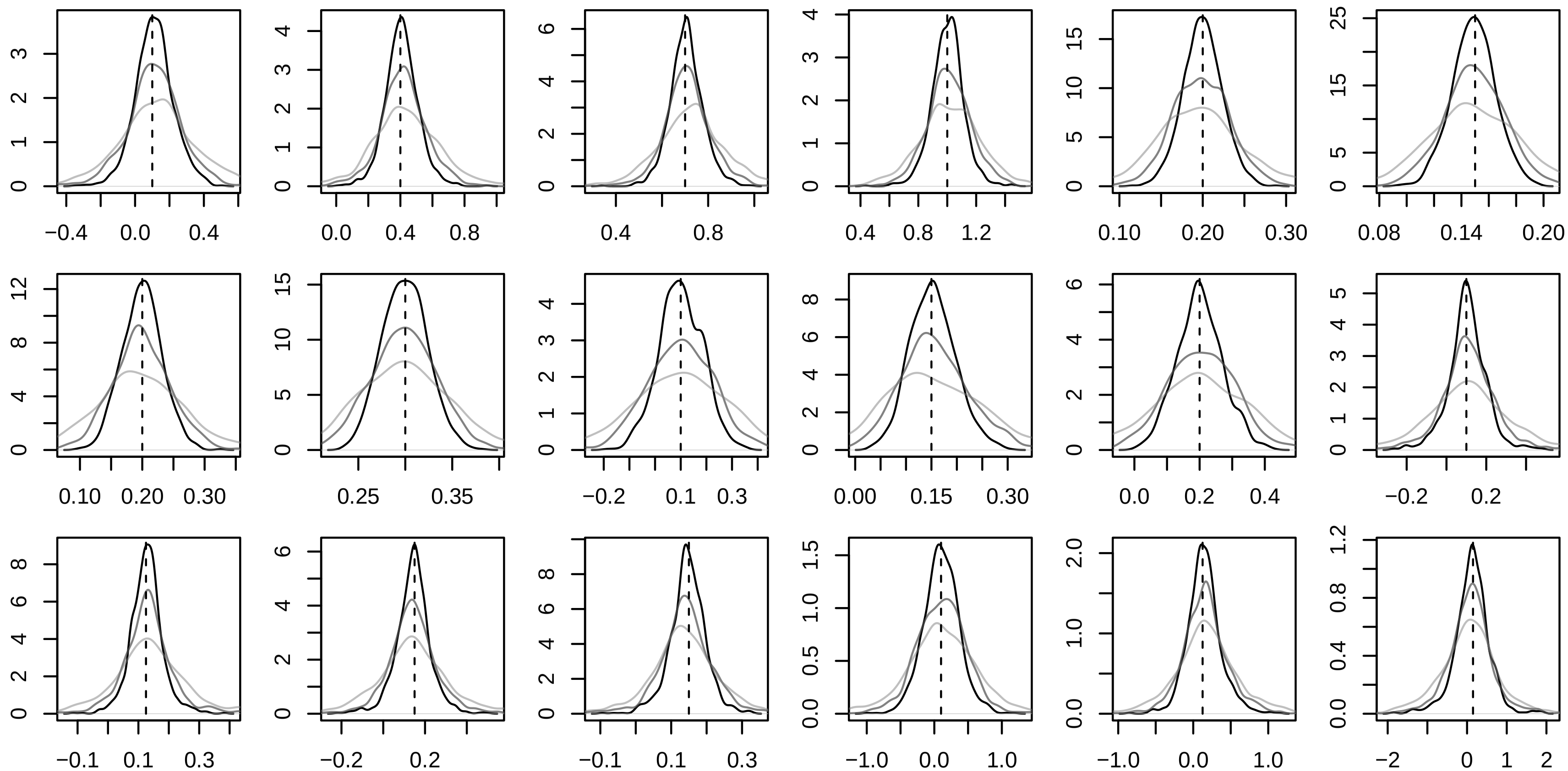

Figure S-4: Kernel density estimator for the distribution of the QML estimates for the copula tensor log-linear INGARCH model with full correlation structure obtained from 1000 Monte Carlo replications for sample sizes 500 (——), 1000 (——) and 2000 (——), true parameter values (- - - -); the bandwidth of the kernel is selected by using Silverman's rule of thumb; the plots are reported rowwise following the order of parameter vector $\theta = (d', \text{vec}(B_2)', \text{vec}(B_1)' \, \text{vec}(A_2)', \text{vec}(A_1)')'$.

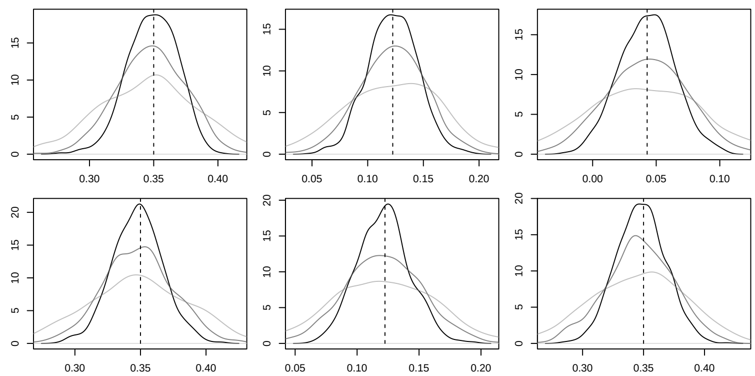

Figure S-5: Kernel density estimator for the distribution of the 2SQML estimates for the copula tensor INGARCH model with full correlation structure obtained from 1000 Monte Carlo replications for sample sizes 500 (——), 1000 (——) and 2000 (——), true parameter value (- - - -); the bandwidth of the kernel is selected by using Silverman's rule of thumb. The plots are reported rowwise following the order of parameter vector $r = (R_{21}, R_{31}, R_{41}, R_{32}, R_{42}, R_{43})'$.

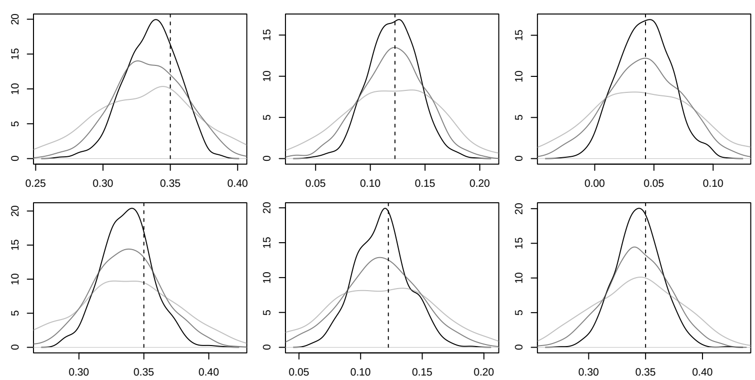

Figure S-6: Kernel density estimator for the distribution of the 2SQML estimates for the copula tensor log-linear INGARCH model with full correlation structure obtained from 1000 Monte Carlo replications for sample sizes 500 (——), 1000 (——) and 2000 (——), true parameter value (- - - -); the bandwidth of the kernel is selected by using Silverman's rule of thumb. The plots are reported rowwise following the order of parameter vector $r = (R_{21}, R_{31}, R_{41}, R_{32}, R_{42}, R_{43})'$.

The first-stage QML estimation results for the log-linear model were presented in the main text. The corresponding first-stage QML estimates for the linear model are $\hat{\alpha} = 0.64^*$ and

$$\hat{B}_1 = \begin{array}{c} \\ \text{Receive} \\ \text{Motor} \\ \text{Retail} \\ \text{Dwell} \\ \text{Person} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Receive} & \text{Motor} & \text{Retail} & \text{Dwell} & \text{Person} \end{array} \\ \begin{pmatrix} 1.000 & 0.000 & 0.003 & 0.043^* & 0.001 \\ 0.002 & 1.138^* & 0.004 & 0.153^* & 0.216^* \\ 0.006 & 0.001 & 1.154^* & 0.006 & 0.002 \\ 0.022 & 0.004 & 0.013 & 1.182^* & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.221^* \end{pmatrix} \end{array}, \quad \hat{B}_2 = \begin{array}{c} \\ \text{Black.} \\ \text{Cant.} \\ \text{Liv.} \\ \text{North.} \\ \text{Parra.} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Black.} & \text{Cant.} & \text{Liv.} & \text{North.} & \text{Par} \end{array} \\ \begin{pmatrix} 0.243^* & 0.002 & 0.032^* & 0.004 & 0.0 \\ 0.000 & 0.277^* & 0.025^* & 0.001 & 0.0 \\ 0.014^* & 0.023^* & 0.221^* & 0.002 & 0.0 \\ 0.001 & 0.026^* & 0.000 & 0.242^* & 0.0 \\ 0.010^* & 0.033^* & 0.002 & 0.000 & 0.23 \end{pmatrix} \end{array}$$

The remaining figures present the autocorrelation functions of the raw data and the Pearson residuals (both for the linear and log-linear specification).

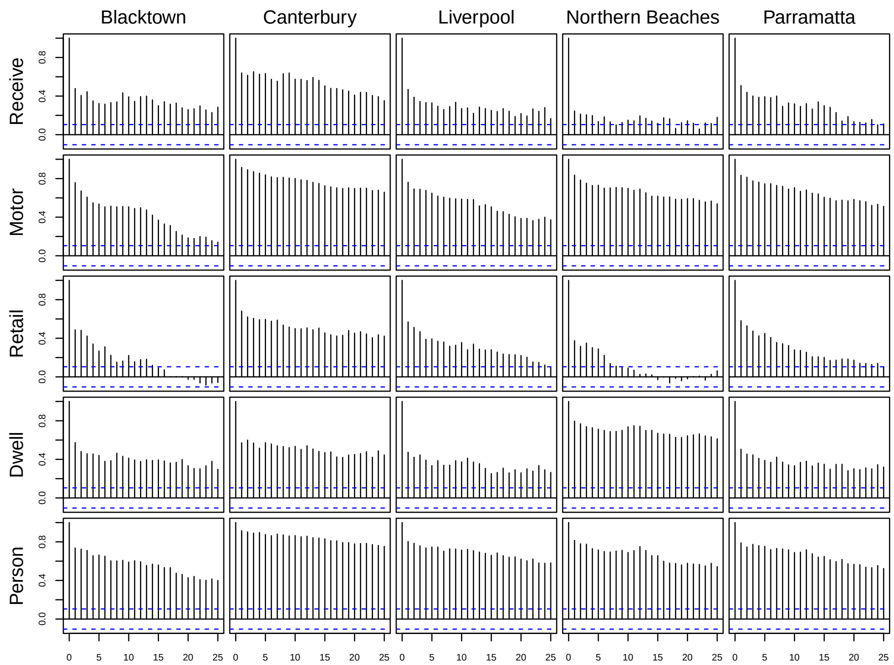

Figure S-7: Sample auto-correlation function of the monthly number of police reports for five theft-related crimes (Receive: Receiving or handling stolen goods; Motor: Steal from motor vehicle; Retail: Steal from retail store; Dwell: Steal from dwelling; Person: Steal from person) in five cities of Australia (Blacktown; Canterbury-Bankstown; Liverpool; Northern Beaches; Parramatta).

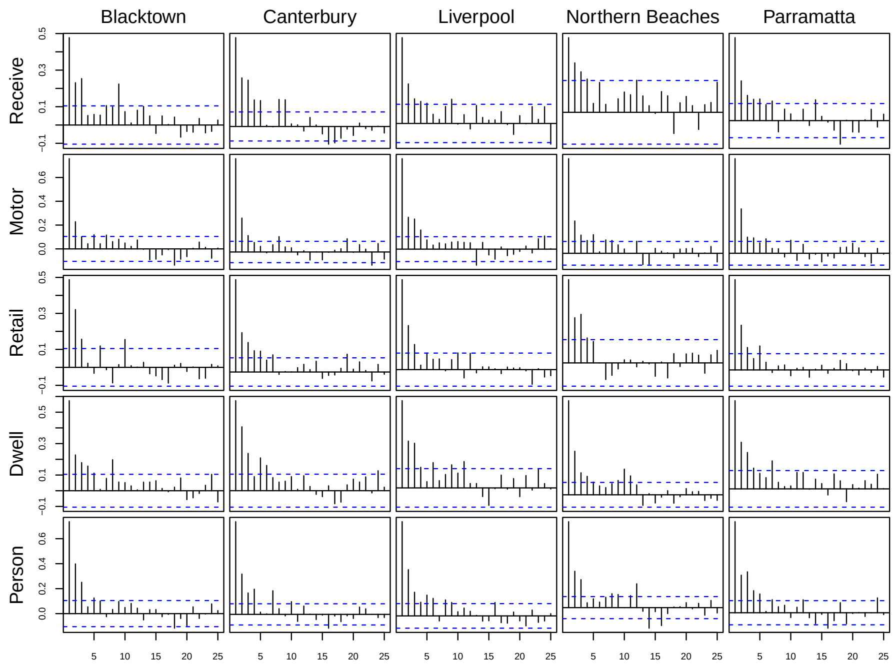

Figure S-8: Sample partial auto-correlation function of the monthly number of police reports for five theft-related crimes (Receive: Receiving or handling stolen goods; Motor: Steal from motor vehicle; Retail: Steal from retail store; Dwell: Steal from dwelling; Person: Steal from person) in five cities of Australia (Blacktown; Canterbury-Bankstown; Liverpool; Northern Beaches; Parramatta).

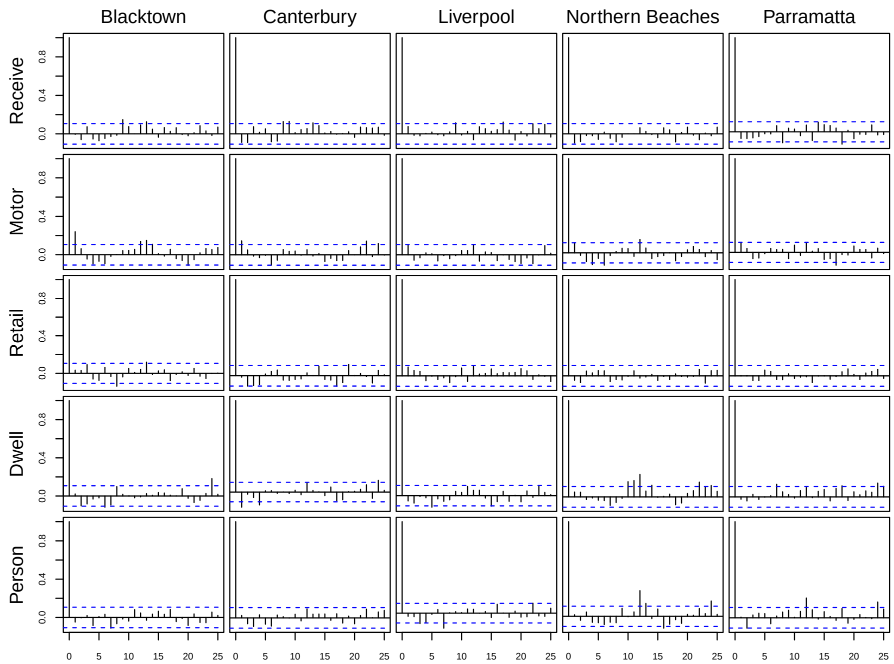

Figure S-9: Sample auto-correlation function of the residuals of the tensor INGARCH model (21) for the New South Wales dataset.

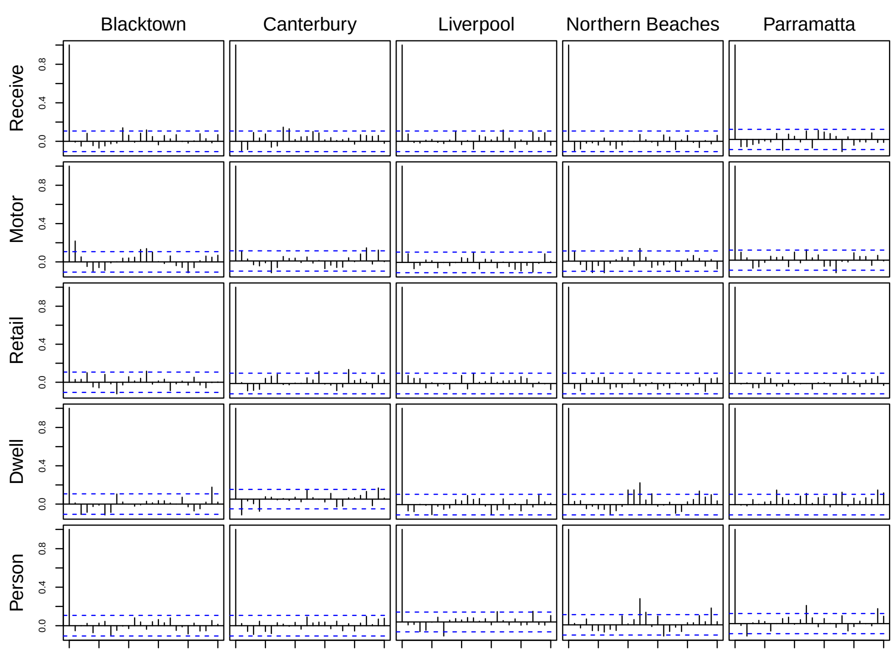

Figure S-10: Sample auto-correlation function of the residuals of the tensor log-linear INGARCH model (22) for the New South Wales dataset.

References

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Curnow, R. N. and C. W. Dunnett (1962). The numerical evaluation of certain multivariate normal integrals. *The Annals of Mathematical Statistics 33*(2), 571–579.

Debaly, Z.-M. and L. Truquet (2023). Multivariate time series models for mixed data. *Bernoulli 29*, 669–695.

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

Magnus, J. R. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Seber, G. A. F. (2008). *A Matrix Handbook for Statisticians*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.