Electronic Journal of Statistics
Vol. 18 (2024) 5484–5562
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2292

High-dimensional latent Gaussian count time series: Concentration results for autocovariances and applications

Marie-Christine Düker

Department of Data Science, FAU Erlangen-Nürnberg;
Department of Statistics and Data Science, Cornell University,
e-mail: marie.dueker@fau.de

Robert Lund

Department of Statistics, University of California, Santa Cruz,
e-mail: rolund@ucsc.edu

Vladas Pipiras

Department of Statistics and Operations Research, UNC at Chapel Hill,
e-mail: pipiras@email.unc.edu

Abstract: This work considers stationary vector count time series models defined via deterministic functions of a latent stationary vector Gaussian series. The construction is very general and ensures a pre-specified marginal distribution for the counts in each dimension, depending on unknown parameters that can be marginally estimated. The vector Gaussian series injects flexibility into the model's temporal and cross-dimensional dependencies, perhaps through a parametric model akin to a vector autoregression. We show that the latent Gaussian model can be estimated by relating the covariances of the counts and the latent Gaussian series. In a possibly high-dimensional setting, concentration bounds are established for the differences between the estimated and true latent Gaussian autocovariances, in terms of those for the observed count series and the estimated marginal parameters. The results are applied to the case where the latent Gaussian series is a vector autoregression, and its parameters are estimated sparsely through a LASSO-type procedure.

MSC2020 subject classifications: Primary 62H20, 62H12; secondary 62M10.
Keywords and phrases: High-dimensional time series, count time series, count distributions, autocorrelation matrix, Hermite expansions, vector autoregressions, shrinkage estimation.

Received December 2022.

1. Introduction

The model: This work concerns a strictly stationary multivariate count valued time series model. The d-dimensional count time series at time t is denoted by $X_t = (X_{1,t}, \ldots, X_{d,t})'$, $t \in \mathbb{Z}$, where prime indicates transpose. Count valued

means that $X_{i,t} \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$; in practice, the counts could encode categorical or ordinal observations. By strict stationarity, the ith component series $\{X_{i,t}\}$ has a time-invariant marginal cumulative distribution function (CDF) for each $i = 1, \dots, d$, which is denoted by

$$F_i(x) = \mathrm{P}[X_{i,t} \leq x], \quad x \in \mathbb{R}. \tag{1.1}$$

We are interested in constructing such series through latent standardized Gaussian series. A simple way to ensure that the desired marginal CDF of the ith component is F_i sets $X_{i,t} = F_i^{-1}(\Phi(Z_{i,t}))$, where Φ denotes the standard Gaussian CDF and F_i^{-1} is the inverse of F_i defined below. Thus, we define the functions

$$G_i(z_i) = F_i^{-1}(\Phi(z_i)), \;\; G(z) = (G_1(z_1), \dots, G_d(z_d)), \;\; z \in \mathbb{R}^d,$$

where

$$F_i^{-1}(u) = \inf\{x \mid F_i(x) \geq u\}, \;\; u \in (0, 1),$$

is the generalized inverse (quantile function) of F_i. Our multivariate count model sets

$$X_t = (X_{1,t}, \dots, X_{d,t})' = (G_1(Z_{1,t}), \dots, G_d(Z_{d,t}))' = G(Z_t), \tag{1.2}$$

where $Z_t = (Z_{1,t}, \dots, Z_{d,t})'$ is a d-dimensional stationary Gaussian series with zero mean and a unit variance: $\mathrm{E}[Z_{i,t}] \equiv 0$ and $\mathrm{E}[Z_{i,t}^2] \equiv 1$. We write

$$\Gamma_Z(h) = R_Z(h) = \mathrm{E}[Z_{t+h} Z_t'], \;\; h \in \mathbb{Z}, \tag{1.3}$$

for the lag-h matrix autocovariance and autocorrelation functions (ACVF and ACF) of $\{Z_t\}$. The ACVF and ACF of the count series $\{X_t\}$ will similarly be denoted by $\Gamma_X(h)$ and $R_X(h)$. Since the $X_{i,t}$ s are not standardized, $\Gamma_X(h)$ and $R_X(h)$ are not necessarily equal in contrast to (1.3). Since the means of X_t are not zero either, $\Gamma_X(h) = \mathrm{E}[X_{t+h} X_t'] - \mathrm{E}[X_{t+h}]\,\mathrm{E}[X_t]'$ at lag h. We will often write Γ_X and R_X to refer to the ACVF and ACF of $\{X_t\}$ over some or all lags.

While the construction in (1.2) ensures F_i as the marginal distribution of $X_{i,t}$, temporal and cross-sectional (spatial) dependencies are driven by the latent Gaussian series $\{Z_t\}$ which will allow us to make inference based on second-order properties of the latent process $\{Z_t\}$. We assume that F_i depends on an unknown parameter vector $\theta_i \in \mathbb{R}^{K_i}$. Furthermore, the construction allows to impose a parametric model on the latent $\{Z_t\}$, such as vector autoregressions (VARs) or dynamic factor models (DFMs).

Some advantageous features of the model are worth stating here. Besides being able to accommodate any count marginal F_i whatsoever, negative correlations in the counts are easily achieved. In fact, the model's correlations are the most flexible possible in both a positive and negative sense; see Remark 2.2 in Jia et al. (2023).

Literature review: Modeling discrete time series has been an active research area and is far less developed than the continuous case. There are several classes of different models including those based on thinning operators (e.g. McKenzie (1985), Alzaid and Al-Osh (1993)) and the generalized state-space models (e.g. Davis et al. (2016)), for example, Markov chain and Hidden Markov models, Bayesian dynamic models (e.g. Gamerman et al. (2015)) and integer-valued autoregressive conditional heteroskedasticity modeling (e.g. Ferland, Latour and Oraichi (2006), Fokianos, Rahbek and Tjøstheim (2009)). A recent survey by Davis et al. (2021) discusses several of these classes of count models, including (1.2), and their (dis)advantages. The analysis of multivariate and potentially high-dimensional count series has received considerably less attention. For a review of approaches for multivariate counts, we refer to Karlis (2016).

The model (1.2) for $d = 1$, was popularized by Jia et al. (2023), where several parameter estimation approaches were suggested (Gaussian pseudo-likelihood estimation, a Yule-Walker based approach for latent autoregressive models and particle filtering). For related work by (subsets of) the same authors, see Kong and Lund (2023) for the seasonal case when $d = 1$. Livsey et al. (2018) used the discussed model to predict hurricane counts assuming Poisson marginals in the case $d = 2$.

In the multivariate and potentially high-dimensional setting, Kim et al. (2024) used latent Gaussian dynamic factor series in the model (1.2) for the purpose of gaining insight into and forecasting of count time series. In Chapter 5 of their dissertation, Kim (2023) considered latent Gaussian series parametrized by a possibly sparse vector autoregression model. The author introduced possibly regularized estimation methods for the latent process, including a numerical study to assess estimation performance.

Other work related to (1.2) uses independent and identically distributed (i.i.d.) Z_ts, where t may no longer refer to time but, for example, different individuals. In psychometrics, models of the type (1.2) have been used extensively for discrete data (e.g. Lebo and Nesselroade (1978)) and related models are termed "polychoric correlations" for ordinal data. Developed in structural equation models (e.g. Lee, Poon and Bentler (1992)), they have made their way into various software packages, e.g. Mplus 7.11 Muthén and Muthén (1998–2017), and more recently into the popular R package `lavaan`; see Rosseel (2012). Applications to psychology under consideration of temporal dependence have been considered in more recent works; see Kim et al. (2024), Kim (2023).

In the statistical literature, Liu et al. (2012), Mitra and Zhang (2014), Wegkamp and Zhao (2016), Han and Liu (2017), Fan et al. (2017), Feng and Ning (2019), and Dirksen, Maly and Rauhut (2022) study Gaussian copula models in possibly high-dimensional settings. These authors derive theoretical results guaranteeing consistent estimation of the latent correlation structure. These publications concentrate on Spearman's rho and Kendall's tau matrices, or subsets of these quantities. Under suitable assumptions, the entries of the copula correlation matrix relate to the entries of the Kendall's tau or Spearman's rho matrices through an explicit link function that does not need to be estimated.

Our contributions: As the literature review shows, the considered latent Gaus-

sian count model has found popularity in the more applied literature. In particular, the works mentioned above have pushed forward the development of feasible algorithms to estimate the parameters of the latent Gaussian process under different parametric assumptions including VAR (Chapter 5 in Kim (2023)) and DFM (Kim et al. (2024)). In this work, we aim to give a theoretical justification of a method that has been proven to work well in practice. The applied and count time series literature aside, we also note that our theoretical results generalize existing results significantly and push forward the analysis of latent models. In contrast to the existing statistical literature, our results capture a much more general class of functions and incorporate the potentially necessary estimation of the transformation.

We are broadly interested in making inferences about Γ_Z from the observed counts $X_1, \ldots, X_T$, especially in the high-dimensional setting where d can be much larger than T. We do so by first considering an estimator $\widehat{\Gamma}_Z$ of Γ_Z defined informally as follows. As shown below, there is a deterministic function ℓ, depending only on marginal CDF parameters θ_i, such that

$$\Gamma_X = \ell(\Gamma_Z) \tag{1.4}$$

for all lags h. If $\widehat{\theta}_i$ is an estimator of θ_i used to construct $\widehat{\ell}$, an estimator of ℓ, Γ_Z can be estimated via

$$\widehat{\Gamma}_Z = \widehat{\ell}^{-1}(\widehat{\Gamma}_X),$$

where $\widehat{\Gamma}_X$ is a standard ACVF estimator of Γ_X based on $X_1, \ldots, X_T$. This work consists of providing concentration bounds on $\|\widehat{\Gamma}_Z - \Gamma_Z\|$ in terms of those for $\|\widehat{\Gamma}_X - \Gamma_X\|$ and $\|\widehat{\theta}_i - \theta_i\|$, where the norms $\|\cdot\|$ are suitably chosen. Concentration bounds on $\|\widehat{\Gamma}_X - \Gamma_X\|$ and $\|\widehat{\theta}_i - \theta_i\|$ are extracted from available results in the literature. To the best of our knowledge, our main results are new even in the i.i.d. setting (i.e., with no temporal dependence).

The derived concentration bounds for $\|\widehat{\Gamma}_Z - \Gamma_Z\|$ make inferences possible for the parameters in Γ_Z. We illustrate this with an application to a latent VAR series $\{Z_t\}$ assuming that its coefficient matrices are suitably sparse. Adapting a LASSO-type approach, we show how the VAR coefficient matrices can be estimated sparsely from $\widehat{\Gamma}_Z$, and then use our concentration results to establish consistency of the matrix estimates, including high-dimensional cases. Perhaps somewhat surprisingly, our results here with a latent $\{Z_t\}$ are of the same flavor as those in the seminal work of Basu and Michailidis (2015), who worked with an observable VAR series $\{Z_t\}$.

Organization: The rest of the paper is structured as follows. Section 2 deals with some preliminaries, including issues related to (1.4), quantities of interest, assumptions, and some moment quantities. The main concentration results for $\|\widehat{\Gamma}_Z - \Gamma_Z\|$ are stated in Section 3, The application to sparse latent VAR series is considered in Section 4. Section 5 provides a discussion and conclusions. Technical proofs and additional material is contained in Appendices A–E.

Notation: For the reader's convenience, notations used throughout the paper are collected here. The maximum and minimum eigenvalues of a symmetric

matrix A are denoted by $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively. To indicate that a matrix A is positive (semi)definite, we write $A \succ 0$ ($A \succcurlyeq 0$). The matrix inequality $A \geq B$ means that $A_{ij} \geq B_{ij}$ for $i, j = 1, \ldots, d$. A range of different norms are used here, including the maximum norm and the spectral norm, defined respectively as $\|A\|_{\max} = \max_{1 \leq i,j \leq d} |A_{ij}|$, $\|A\| = \sqrt{\lambda_{\max}(A'A)}$ for a matrix $A \in \mathbb{R}^{d \times d}$. The ℓ_1-norm $\|v\|_1 = \sum_{j=1}^{d} |v_j|$, the Euclidean norm $\|v\|^2 = \sum_{j=1}^{d} |v_j|^2$, and the norm that counts all non-zero elements in $\|v\|_0$ for a vector $v \in \mathbb{R}^d$ are also used. For a $d \times N$ matrix composed of N d-dimensional vectors $v_1, \ldots, v_N$, we write $[v_1 : \cdots : v_N]$. Our proofs use the Hadamard product $A \odot B$ of two matrices $A, B \in \mathbb{R}^{d \times d}$ and the related notations $A \odot A = A^{\odot 2}$, $A^{\odot \frac{1}{2}} = (A_{ij}^{\frac{1}{2}})_{i,j=1,\ldots,d}$ and $A^{\odot -1} = (1/A_{ij})_{i,j=1,\ldots,d}$. The componentwise application of absolute values is denoted $|A| = (|A_{ij}|)_{i,j=1,\ldots,d}$. For two quantities a and b, we use $a \gtrsim b$ if there exists an absolute constant c, independent of the model parameters, such that $a \geq cb$. We write $\nabla_x f$ for the gradient of the function f with respect to a vector $x \in \mathbb{R}^d$. When the gradient is evaluated at a specific value $\widetilde{x}$, we write $\nabla_x f|_{\widetilde{x}}$. The derivative with respect to a scalar is denoted as $\frac{\partial}{\partial x_1} f$ so that $\nabla_x f = (\frac{\partial}{\partial x_1} f, \ldots, \frac{\partial}{\partial x_d} f)'$.

2. Preliminaries

This section first relates the autocovariance matrices of the latent and observed processes in Section 2.1. Our goals are formulated and the estimators are clarified in Section 2.2. We then state our assumptions and main results in Section 2.3 and introduce notation that allows our results and proofs to be compactly presented in Section 2.4.

2.1. Autocovariance matrices and their relationships

Recall that $\Gamma_X(h) = \mathrm{E}[X_{t+h} X_t'] - \mathrm{E}[X_{t+h}]\,\mathrm{E}[X_t]'$ denotes the autocovariance matrix function at lag h of a stationary time series and $R_X(h)$ its corresponding lag h autocorrelation. Individual entries are denoted by $\Gamma_{X,ij}(h)$ and $R_{X,ij}(h)$ for $i, j = 1, \ldots, d$.

The ACVFs of $\{X_t\}$ and $\{Z_t\}$ in (1.2) can be related using Hermite expansions for the components in G:

$$G_i(z) = \sum_{k=0}^{\infty} \frac{c_{i,k}}{k!} H_k(z) \tag{2.1}$$

with the kth Hermite polynomial defined as

$$H_k(z) = (-1)^k e^{z^2/2} \frac{\partial^k}{\partial z^k} e^{-z^2/2}.$$

The Hermite coefficients are

$$c_{i,k} = \mathrm{E}(G_i(Z_{i,0}) H_k(Z_{i,0})); \tag{2.2}$$

see Chapter 5 in Pipiras and Taqqu (2017) for more details on Hermite polynomials.

As stated in Proposition 5.1.4 in Pipiras and Taqqu (2017), the autocovariances of $\{X_t\}$ can be written as

$$\Gamma_X(h) = \left(\sum_{k=1}^{\infty} \frac{c_{i,k} c_{j,k}}{k!} R_{Z,ij}(h)^k \right)_{i,j=1,\ldots,d}; \tag{2.3}$$

the corresponding autocorrelation matrix is

$$R_X(h) = \left(\sum_{k=1}^{\infty} \frac{c_{i,k} c_{j,k}}{k!} \frac{1}{(\Gamma_{X,ii}(0)\Gamma_{X,jj}(0))^{\frac{1}{2}}} R_{Z,ij}(h)^k \right)_{i,j=1,\ldots,d}.$$

Following Jia et al. (2023), we write

$$L(u) = (L_{ij}(u))_{i,j=1,\ldots,d} \quad \text{with} \quad L_{ij}(u) = \sum_{k=1}^{\infty} \frac{c_{i,k} c_{j,k}}{k!} \frac{u^k}{(\Gamma_{X,ii}(0)\Gamma_{X,jj}(0))^{\frac{1}{2}}}$$

and, for what we will refer to as the link function,

$$\ell(u) = (\ell_{ij}(u))_{i,j=1,\ldots,d} \quad \text{with} \quad \ell_{ij}(u) = \sum_{k=1}^{\infty} \frac{c_{i,k} c_{j,k}}{k!} u^k. \tag{2.4}$$

From Section 2.3 in Jia et al. (2023), the Hermite coefficients (2.2) admit the representation

$$c_{i,k} = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} e^{-Q_{i,n}^2/2} H_{k-1}(Q_{i,n}) \tag{2.5}$$

with $Q_{i,n} = \Phi^{-1}(C_{i,n})$ and $C_{i,n} = \mathrm{P}[X_{i,t} \leq n]$. In general, $Q_{i,n}$ depends on θ_i, which contains all CDF parameters for the ith component series. We write $Q_n(\theta_i) := Q_{i,n}$ and $C_n(\theta_i) := C_{i,n}$ to emphasize dependence on θ_i. We use a different notation $C_{i,n} = \mathrm{P}[X_{i,t} \leq n]$ instead of potentially more natural $F_i(n)$ to bring out the dependence on θ_i as the argument (e.g. to be differentiated with respect to).

Proposition 2.1 in Jia et al. (2023) provides an explicit representation for the first derivative of ℓ in (2.4). While that result lies in a univariate setting, the representation here is stated for each individual entry of the covariance matrices and allows different marginal distributions in each dimension.

Proposition 2.1. *Let ℓ be as in (2.4). Then, for $u \in (-1, 1)$,*

$$\ell'_{ij}(u) = \frac{1}{2\pi\sqrt{1-u^2}} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} (Q_{i,n_0}^2 + Q_{j,n_1}^2 - 2uQ_{i,n_0}Q_{j,n_1}) \right). \tag{2.6}$$

As Proposition 2.1 shows, ℓ_{ij} has a positive derivative and is therefore monotonically strictly increasing. The function ℓ_{ij} maps $[-1,1]$ into $[\ell_{ij}(-1), \ell_{ij}(1)]$ with $\ell_{ii}(1) = \Gamma_{ii}(0)$ and crosses zero at $u = 0$ due to (2.4). Note, that since ℓ_{ij} is a strictly increasing function with $\ell_{ij}(0) = 0$, negative or positive correlation of the latent process gets inherited by the observed count series. Figure 2 in the supplementary material of Jia et al. (2023) plots several link functions. Since ℓ_{ij} is strictly increasing, so is its inverse $g_{ij} = \ell_{ij}^{-1}$, which is defined on $[\ell_{ij}(-1), \ell_{ij}(1)]$. Later, we extend the domains of ℓ and g to define an appropriate estimator. To use a plug in estimator for the autocovariance matrices, ℓ and g will need to be evaluated at any point.

Related to (2.6), we introduce, for $x, y \in (-1,1)$, the function

$$h_{ij}(x,y) = \frac{1}{2\pi\sqrt{1-x}} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-x)}(Q_{i,n_0}^2 + Q_{j,n_1}^2 - 2yQ_{i,n_0}Q_{j,n_1})\right).$$

Observe that $\ell'_{ij}(u) = h_{ij}(u^2, u)$. We also define

$$z_{ij}(y) := \frac{1}{2\pi}(h_{ij}(0,y))^{-1}, \quad Z_{ij}(x) := \frac{1}{2\pi\sqrt{1-x}}(h_{ij}(x, \Gamma_{Z,ij}(h)))^{-1}. \tag{2.7}$$

The corresponding matrix-valued quantities can be defined accordingly; for instance, $\ell'(\Gamma_Z(h)) = h(\Gamma_Z(h)^{\odot 2}, \Gamma_Z(h))$.

Example 2.1. Proposition 2.1 explicitly provides the derivative of the link function in a general setting that allows any marginal distribution. This representation simplifies in many cases. For instance, suppose that the marginals in dimension i have a Bernoulli distribution with success probability p_i for $i = 1, \dots, d$. Then G in (1.2) is $G : \mathbb{R}^d \to \{0,1\}^d$ and its components can be written as $G_i(z) = F_i^{-1}(\Phi(z)) = \mathbb{1}_{\{z > \Phi^{-1}(1-p_i)\}}$. In particular, $Q_{i,n} = \Phi^{-1}(1-p_i) =: q_i$. The first derivative of the link function simplifies to

$$\ell'_{ij}(u) = \frac{1}{2\pi(1-u^2)^{\frac{1}{2}}} \exp\left(-\frac{1}{2(1-u^2)}(q_i^2 + q_j^2 - 2uq_iq_j)\right).$$

In particular, for $i = j$, the expression in the exponential reduces to $-q_i^2/(1+u)$.

Example 2.2. Following up on the previous Example 2.1, we simplify even further by assuming that we have a $d = 2$ dimensional count series with Bernoulli marginals and $p_i = \frac{1}{2}$ for $i = 1, 2$. Then, $G_i(z) = F_i^{-1}(\Phi(z)) = \mathbb{1}_{\{z > \Phi^{-1}(1-p_i)\}} = \mathbb{1}_{\{z>0\}}$ for $i = 1, 2$. With $\Gamma_Z(h) = (\Gamma_{Z,ij}(h))_{i,j=1,2}$, we get for the observed count series

$$\Gamma_X(h) = \frac{1}{2\pi}\begin{pmatrix} \arcsin(\Gamma_{Z,11}(h)) & \arcsin(\Gamma_{Z,12}(h)) \\ \arcsin(\Gamma_{Z,21}(h)) & \arcsin(\Gamma_{Z,22}(h)) \end{pmatrix}.$$

In other words, in this particular case, where the CDF parameters are known, it is possible to calculate an explicit link function. This example also illustrates how the count series inherits negative or positive correlation of the latent process. We refer to Lemma 4.1 in Livsey et al. (2018) and also Section III of Van Vleck and Middleton (1966) for this example.

2.2. Quantities of interest and estimation

Our goal here is to extract information about the latent process $\{Z_t\}$, including cases of high-dimensions. Section 4 below estimates coefficient matrices for the latent process, which is assumed to follow a VAR model.

To take advantage of the sparsity assumed on the latent process, it is crucial to derive deviation bounds in a specific norm. Introduce the set $\mathcal{K}(2s) = \{v \in \mathbb{R}^{dL} : \|v\| \leq 1, \|v\|_0 \leq 2s\}$ and define the mapping

$$A \mapsto \|A\|_s := \sup_{v \in \mathcal{K}(2s)} |v' A v|. \tag{2.8}$$

Besides submultiplicativity, the mapping in (2.8) satisfies all properties of a matrix norm. We derive and collect some properties of (2.8) in Section D.1. All results will be expressed in terms of $\|\cdot\|_s$. The mapping allows us to impose sparsity on a matrix A through the vectors v. Since all properties of (2.8) in Section D.1 are also satisfied by the spectral norm, our main result remains true for the spectral norm. This said, results for the spectral norm do not seem to be particularly relevant in our setting.

We aim to derive concentration inequalities for estimates of $\mathbf{\Gamma}_Z = (\Gamma_Z(r-s))_{r,s=1,\ldots,L}$, which collects autocovariances at different low lags. That is, our goal is to bound the probability

$$\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > \delta\right]$$

for $\delta > 0$. A natural estimator of $\mathbf{\Gamma}_Z$ is $\ell^{-1}(\widehat{\mathbf{\Gamma}}_X)$ for a known link function ℓ. We also use the notation $g = \ell^{-1}$, and $g_{ij} = \ell_{ij}^{-1}$. In principle, $\widehat{\mathbf{\Gamma}}_X$ could take on any value, even beyond the domain $[\ell_{ij}(-1), \ell_{ij}(1)]$ of g_{ij}. Therefore, we assume throughout the paper that $g_{ij}(x) = g_{ij}(\ell_{ij}(1))$ for all $x > \ell_{ij}(1)$ (and similarly at the left border).

Assuming that the observations have a zero mean, the autocovariance matrix $\mathbf{\Gamma}_X = (\Gamma_X(r-s))_{r,s=1,\ldots,L}$ can be estimated as $\widehat{\mathbf{\Gamma}}_X = N^{-1}\mathcal{X}_X'\mathcal{X}_X$ with $N = T - L$ and

$$\mathcal{X}_X = \begin{pmatrix} X_L' & \ldots & X_1' \\ \vdots & \ddots & \vdots \\ X_{T-1}' & \ldots & X_{T-L}' \end{pmatrix}. \tag{2.9}$$

With a slight abuse of notation, we write both $\mathbf{\Gamma}_X = \ell(\mathbf{\Gamma}_Z)$ and $\Gamma_X(h) = \ell(\Gamma_Z(h))$.

The link function ℓ_{ij} defined in (2.4) depends on the CDF parameter vectors θ_i and θ_j. We collect these marginal distribution parameters across dimensions into $\theta = (\theta_1', \ldots, \theta_d')'$ and denote their estimators as $\widehat{\theta}$ and $\widehat{\theta}_i$ for $i = 1, \ldots, d$. Each component series is allowed a different marginal distribution, which potentially depends on a different number of parameters. Let K_i denote the number of parameters for the marginal distribution of the ith component series and set $K = \max_{i=1,\ldots,d} K_i$. Explicit dependence on the vectors θ_i is often omitted in

our notations; antipodally, we write $Q_n(\theta_i) = Q_{i,n}$ and $C_n(\theta_i) = C_{i,n}$ when this dependence needs to be emphasized.

Estimates of ℓ and g and other related functions are written as $\widehat{\ell}$ and $\widehat{g}$ and are computed by replacing $C_{i,n}$ with $\widehat{C}_{i,n} = C_n(\widehat{\theta}_i)$ in $Q_{i,n} = \Phi^{-1}(C_{i,n})$. Our estimator for $\mathbf{\Gamma}_Z$ is written as $\widehat{\mathbf{\Gamma}}_Z = \widehat{g}(\widehat{\mathbf{\Gamma}}_X)$ for an unknown link function.

Our main results relate the probability of autocovariance matrix estimator deviations of the latent process to the analogous probability in the observations. To state these, several assumptions are needed.

2.3. Assumptions

We will work with two sets of assumptions. The first assumption set applies to our main result, which relates the probability of how much $\widehat{\mathbf{\Gamma}}_Z$ deviates from $\mathbf{\Gamma}_Z$ to the analogous probabilities in the observed $\{X_t\}$. Assumption M.3 is shown to hold for several common count distributions under Assumption M.2 in Appendix E.

Assumption M.1. There is a constant $\boldsymbol{c}_Z \in (0,1)$ such that $|\Gamma_{Z,ij}(h)| < \boldsymbol{c}_Z$ for $h \neq 0$, $i,j = 1,\dots,d$ and $|\Gamma_{Z,ij}(0)| < \boldsymbol{c}_Z$ for all $i \neq j$.

Assumption M.2. For each $\theta_i = (\theta_{i1},\dots,\theta_{iK_i})'$, there exists an open neighborhood S of θ_i such that the moment $\sup_{\theta_i \in S} \mathrm{E}[|X_{i,t}|^p] = \sup_{\theta_i \in S} \mathrm{E}_{\theta_i}[|X_{i,t}|^p] < \infty$ for some $p > 2$.

Assumption M.3. For each $\theta_i = (\theta_{i1},\dots,\theta_{iK_i})'$, there exists an open neighborhood S of θ_i such that

$$\begin{aligned}&\sup_{\theta_i \in S} \sum_{n=0}^{\infty} (1 - C_n(\theta_i))^{-\frac{1}{2}} \sum_{j=1}^{K_i} \left| \frac{\partial}{\partial \theta_{ij}} C_n(\theta_i) \right| \\ &= \sup_{\theta_i \in S} \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \sum_{j=1}^{K_i} \left| \frac{\partial}{\partial \theta_{ij}} \mathrm{P}[X_{i,t} > n] \right| < \infty.\end{aligned}$$

Assumption M.4. For each $\theta_i = (\theta_{i1},\dots,\theta_{iK_i})'$, there exist an open neighborhood S of θ_i and at least one n such that $\inf_{\theta_i \in S} C_n(\theta_i) > 0$.

Note that we require our moment conditions to hold uniformly in a neighborhood around θ_i. This allows us to infer finiteness on a compact subset of the parameter space of θ_i.

The following two assumptions ensure consistent estimation of $\mathbf{\Gamma}_Z$ with a $\log(Ld^2)/T$ convergence rate. Other bounds may also lead to consistency. Section 4 establishes these assumptions for a causal VAR series $\{Z_t\}$.

Assumption C.1. There exist finite positive constants c_1 and c_2 such that for any $v \in \mathcal{K}(2s)$ and any $\delta > 0$,

$$\mathrm{P}[|v'(\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)v| > c_0(s)\delta] \leq \; c_1 \exp\left(-c_2 N \min\{\delta, \delta^2\}\right)$$

with $N = T - L$ and $c_0(s) \geq 1$.

Assumption C.2. There exist finite positive constants c_1 and c_2 such that for any $\varepsilon > 0$,

$$\mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon] \leq c_1 dK \exp\left(-c_2 T \min\{\varepsilon, \varepsilon^2\}\right).$$

The quantity $c_0(s)$ in Assumption C.1 should be of lower order than q and describe some kind of lower dimensional structure imposed through the vectors $v \in \mathcal{K}(2s)$. We provide a discussion on Assumption C.1 and $c_0(s)$ in Section 4.3 below. The assumption that $c_0(s) \geq 1$ has only aesthetic reasons.

2.4. Moments

We now collect some notation used in the proofs of the main results. Proposition 2.1 shows that ℓ is differentiable on the open interval $(-1, 1)$ and gives an explicit form for this derivative. In our proofs, the on- and off-diagonal elements in the difference $\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z$ are handled separately. In fact, bounds for the off-diagonal elements can be expressed in terms of ℓ' due to (2.6).

Our main results are cast in terms of moment conditions for $\{X_t\}$. The following notation allows us to express our results compactly. As the diagonal terms will be treated separately, quantities used to bound these terms are considered first. Set

$$\Delta_i = \sum_{n=0}^{\infty} n \left\|\nabla_{\theta_i} C_{i,n}\right\|_1 = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2u} Q_{i,n}^2\right) n \|\nabla_{\theta_i} Q_{i,n}\|_1, \tag{2.10}$$

which is uniformly bounded in a neighborhood of θ_i by Lemma D.3 after the assumptions in M.2 and M.3 are invoked. The second representation emphasizes the similarity to the subsequently introduced quantities and is further explained in the proof of Lemma D.3.

Also define

$$m_i^{(k)}(u) = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2u} Q_{i,n}^2\right) |Q_{i,n}|^k. \tag{2.11}$$

Lemma D.4 ensures that (2.11) is uniformly bounded in a neighborhood of θ_i under Assumption M.2. Furthermore, Assumption M.4 ensures that (2.11) is uniformly bounded from below since at least one of the summands in the series expansion is nonzero. We will occasionally write $m_{\theta_i}^{(k)}(u)$ to emphasize that $m_i^{(k)}(u)$ depends on θ_i. Analogously, we introduce the derivative

$$\mu_i^{(k)}(u) = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2u} Q_{i,n}^2\right) |Q_{i,n}|^k \|\nabla_{\theta_i} Q_{i,n}\|_1. \tag{2.12}$$

This expression will be further simplified in Lemma D.5 and is uniformly bounded in a neighborhood of θ_i under Assumption M.3. See Section E for further discussion on Assumption M.3.

Our probability bounds will be expressed in terms of

$$\Delta(\varepsilon) = \max_{i=1,\dots,d} \sup_{\theta_i \in \Theta(\varepsilon)} \Delta_i \tag{2.13}$$

for the diagonal terms and the following four quantities:

$$\begin{aligned} M(\boldsymbol{c}_Z, \varepsilon) &= \max_{\substack{i=1,\dots,d \\ k=0,3}} \sup_{\theta_i \in \Theta(\varepsilon)} m_i^{(k)} \left(1 + \boldsymbol{c}_Z\right), \\ \mu(\boldsymbol{c}_Z, \varepsilon) &= \max_{\substack{i=1,\dots,d \\ k=0,3}} \sup_{\theta_i \in \Theta(\varepsilon)} \mu_i^{(k)} \left(1 + \boldsymbol{c}_Z\right), \end{aligned} \tag{2.14}$$

$$\begin{aligned} M_1(\boldsymbol{c}_Z, \varepsilon) &= \max_{i=1,\dots,d} \sup_{\theta_i \in \Theta(\varepsilon)} \frac{1}{\left(m_i^{(0)} \left(1 - \boldsymbol{c}_Z\right)\right)^2}, \\ M_2(\boldsymbol{c}_Z, \varepsilon) &= \max_{\substack{i=1,\dots,d \\ k=0,2}} \sup_{\theta_i \in \Theta(\varepsilon)} \frac{\left(m_i^{(k)} \left(\frac{1}{1-\boldsymbol{c}_Z}\right)\right)^2}{\left(m_i^{(0)} \left(1 - \boldsymbol{c}_Z\right)\right)^4} \end{aligned} \tag{2.15}$$

with $\Theta(\varepsilon) = \{\theta \in [\Delta_{11}, \Delta_{12}] \times \cdots \times [\Delta_{K_i 1}, \Delta_{K_i 2}] \mid \|\theta - \theta_i\|_{\max} \leq \varepsilon\}$. Here, $[\Delta_{r1}, \Delta_{r2}]$ refers to the interval of admissible estimates for the rth parameter in the vector θ_i.

Note that Lemmas D.3–D.5 in combination with Assumptions M.2–M.4 ensure that the expressions (2.10)–(2.12) are uniformly bounded in a neighborhood of θ_i. Consequently, we can infer that the quantities (2.13)–(2.15) in our probability bounds are finite on the compact set $\Theta(\varepsilon)$.

Example 2.3. Returning to Example 2.1, consider the Bernoulli case where $\mathrm{P}[X_{i,t} = 1] = p_i$ for $i = 1, \dots, d$. Then, (2.11) and (2.12) simplify to

$$m_i^{(k)}(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2u} q_i^2\right) |q_i|^k$$

and

$$\mu_i^{(k)}(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2u} q_i^2\right) |q_i|^k \left|\frac{\partial}{\partial p_i} q_i\right| = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1-u}{2u} q_i^2\right) |q_i|^k$$

since $\frac{\partial}{\partial p_i} q_i = \frac{1}{\phi(q_i)}$, where $\phi(\cdot)$ is the standard normal density function.

3. Concentration inequalities for autocovariance matrix estimates

This section presents our main results. These results allow one to make inferences about $\{Z_t\}$ from $\{X_t\}$. Proofs are delegated to Section A.3 and subsequent appendices.

Proposition 3.1. *Suppose that Assumptions M.1–M.4 hold. Then, for any* $\delta, \widetilde{\delta}, \varepsilon, \widetilde{\varepsilon} > 0$,

$$\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > Q(\mathbf{\Gamma}_Z)\delta\right] \precsim \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon \wedge \widetilde{\varepsilon}] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max}^2 > \delta], \tag{3.1}$$

where $Q(\mathbf{\Gamma}_Z) := Q(\widetilde{\delta}, \varepsilon, \widetilde{\varepsilon}, \mathbf{\Gamma}_Z)$ *is a function of* Δ, M, μ, M_1, *and* M_2 *as defined in* (2.13), (2.14), *and* (2.15). *The explicit representation of* $Q(\mathbf{\Gamma}_Z)$ *can be found in Section A.2.*

We refer the reader to Section A.1 for an outline of the proof and its associated challenges. Section A.1 also provides intuition on how $\delta, \widetilde{\delta}, \varepsilon$ and $\widetilde{\varepsilon}$ arise on the right hand side of (3.1). While the result could be simplified by writing, for instance, $\mathrm{P}[\|\widehat{\theta} - \theta\|_{\max}^2 > \delta] = \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta^{\frac{1}{2}}]$, we purposely bookkept second order terms to emphasize the proof's strategy. For some insight on the quantity $Q(\mathbf{\Gamma}_Z)$, we refer to Example 3.1 below, where we discuss some of its behavior in the Bernoulli case.

Note that once we estimate θ s for specific marginal distributions, the parameters might be restricted to a certain interval. Then, θ does not only have to satisfy $\theta_i - \varepsilon \leq \widehat{\theta}_i \leq \theta_i + \varepsilon$ but also comply with the parameter space imposed by the marginal distribution. In other words, $\widehat{\theta}_i$ lies in the intersection of $[\theta_i - \varepsilon, \theta_i + \varepsilon]$ and the parameter space given by the marginal distribution. The prior knowledge of a set of feasible estimates results in our constants depending on the set $\Theta(\varepsilon) = \{\theta \in [\Delta_{11}, \Delta_{12}] \times \cdots \times [\Delta_{K_i 1}, \Delta_{K_i 2}] \mid \|\theta - \theta_i\|_{\max} \leq \varepsilon\}$. Here, $[\Delta_{r1}, \Delta_{r2}]$ refers to the interval of admissible estimates for the rth parameter in the vector θ_i. Since ε is supposed to go to zero, it will eventually be small enough to ensure for $[\theta_i - \varepsilon, \theta_i + \varepsilon]$ to be shorter than the interval of admissible estimates. Similarly, $\widetilde{\delta}$ will be small enough to ensure $\widehat{\Gamma}_{X,ij}(h) \leq \widetilde{\delta} + \Gamma_{X,ij}(h) < \ell_{ij}(1)$ for $i \neq j$ since $\Gamma_{X,ij}(h) < \ell_{ij}(\boldsymbol{c}_Z) < \ell_{ij}(1)$.

To simplify Proposition 3.1, one can work with a single quantity $\nu = \delta \wedge \widetilde{\delta} \wedge \varepsilon \wedge \widetilde{\varepsilon}$, as stated in Corollary 3.1 below. The corollary is a consequence of Proposition 3.1 and illustrates how convergence rates for the observed process can be used to extract a high probability bound on deviations between $\widehat{\mathbf{\Gamma}}_Z$ and $\mathbf{\Gamma}_Z$.

Corollary 3.1. *Suppose that Assumptions M.1–M.4 and C.1–C.2 hold. Then, for any* $\delta, \widetilde{\delta}, \varepsilon, \widetilde{\varepsilon} > 0$, *there exists finite constants* $c_{i,1}, c_{i,2} > 0$, $i = 1, 2$, *such that*

$$\begin{aligned} &\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > Q(\mathbf{\Gamma}_Z)c_0(s)\delta\right] \\ &\quad \leq c_{1,1}\exp\left(-c_{1,2}N\min\left\{1, \nu^2, \nu/c_0(s)\right\} + 2s\log(Ld)\right) \\ &\qquad + c_{2,1}dK\exp\left(-c_{2,2}T\min\{1, \nu^2\}\right) \end{aligned} \tag{3.2}$$

with $Q(\mathbf{\Gamma}_Z) := Q(\widetilde{\delta}, \varepsilon, \widetilde{\varepsilon}, \mathbf{\Gamma}_Z)$ *defined in Proposition 3.1,* $c_0(s) \geq 1$ *as in Assumption C.1 and* $\nu = \delta \wedge \widetilde{\delta} \wedge \varepsilon \wedge \widetilde{\varepsilon}$.

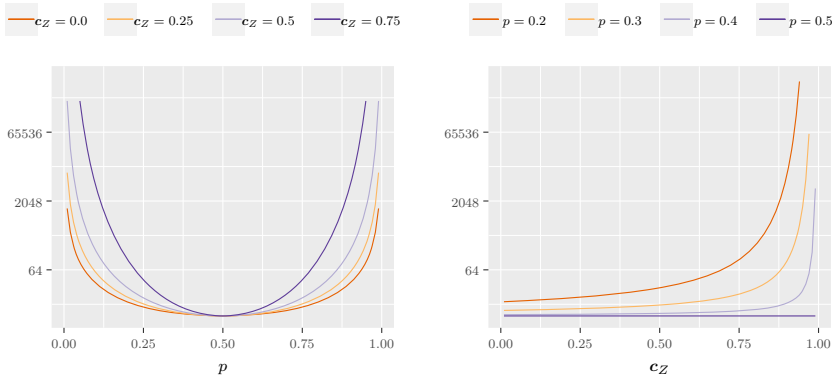

FIG 1. *The function* $M_1(\boldsymbol{c}_Z, 0)$ *in* (2.15) *for* $p := p_1 = \cdots = p_d$ *and as function of* $p \in (0,1)$ *(left panel) and as function of* $\boldsymbol{c}_Z \in (0,1)$ *(right panel).*

From Corollary 3.1, one can further infer the existence of constants $c_1, c_2 > 0$ such that

$$\begin{aligned} &\mathrm{P}\left[\|\widehat{\boldsymbol{\Gamma}}_Z - \boldsymbol{\Gamma}_Z\|_s > Q(\boldsymbol{\Gamma}_Z)c_0(s)\delta\right] \\ &\le c_1 \exp\Big(- c_{1,2}N \min\left\{1, \nu^2, \nu/c_0(s)\right\} \\ &\qquad + 2s\log(Ld) - c_{2,2}T\min\{1,\nu^2\} + \log(dK)\Big) \\ &\le c_1 \exp\left(-c_2 N \min\left\{1, \nu^2, \nu/c_0(s)\right\}\right), \end{aligned} \tag{3.3}$$

where we assume that $N \gtrsim \max\{c_0(s)/\nu, \nu^{-2}, 1\}\max\{s\log(Ld), \log(dK)\}$. Choosing $\delta = \widetilde{\delta} = \varepsilon = \widetilde{\varepsilon} = \sqrt{\frac{\log(Ld^2)}{N}}$ so that $\nu = \sqrt{\frac{\log(Ld^2)}{N}}$, we infer the existence of positive constants c_1 and c_2 such that

$$\begin{aligned} &\mathrm{P}\left[\|\widehat{\boldsymbol{\Gamma}}_Z - \boldsymbol{\Gamma}_Z\|_s > Q(\boldsymbol{\Gamma}_Z)c_0(s)\sqrt{\frac{\log(Ld^2)}{N}}\right] \\ &\le c_1 \exp\left(-c_2 \min\{N, \log(Ld^2), \sqrt{N\log(Ld^2)/c_0^2(s)}\}\right) \\ &\le c_1 \exp\left(-c_2\log(Ld^2)\right) \\ &= c_1 q^{-c_2}, \end{aligned} \tag{3.4}$$

whenever $N \gtrsim c_0^2(s)\log(q)$ with $q = Ld^2$.

For a sense of what parts of $\|\widehat{\boldsymbol{\Gamma}}_Z - \boldsymbol{\Gamma}_Z\|_s$ contribute to the probability bound, we focus on Corollary 3.1 and compare the statement to some existing results. The first part of the bound, $c_{1,1}\exp\left(-c_{1,2}N\min\left\{1,\nu^2,\nu/c_0(s)\right\} + 2s\log(Ld)\right)$, can be separated into two parts. The first summand in the exponential bound is almost the same as the one in Proposition 2.4 in Basu and Michailidis (2015) for expressions of the form $v'(\widehat{\boldsymbol{\Gamma}}_Z - \boldsymbol{\Gamma}_Z)v$. This makes use of Gaussianity for $\{Z_t\}$. The only difference is $\nu/c_0(s)$. However, $\nu/c_0(s)$ results from our second-order

terms and is asymptotically negligible as seen in (3.4). The second summand $2s\log(Ld)$ in the exponential arises after applying Lemma F.2 in the supplementary material of Basu and Michailidis (2015); this effectively extends results uniformly over all sparse vectors $v \in \mathcal{K}(2s)$.

The major difference between our bounds and existing results for autocovariance estimation of Gaussian series is the summand $c_{2,1}dK\exp(-c_{2,2}T\min\{1, \nu^2\})$.
This exponential term comes from estimating the unknown parameters θ in the marginal distributions and link function ℓ. In particular, for a known link function, this summand does not show up in the bound; see also Lemma A.2 below.

Example 3.1. Returning to Example 2.3, consider the Bernoulli case where $\mathrm{P}[X_{i,t} = 1] = p_i$ for $i = 1, \ldots, d$. We aim to shed some light on the constant $Q(\mathbf{\Gamma}_Z)$ in (3.1). The quantity $Q(\mathbf{\Gamma}_Z)$ is a function of (2.14) and (2.15) which are functions of (2.11) and (2.12). In the Bernoulli case, those values depend on the success probabilities p_i, $i = 1, \ldots, d$, and $\boldsymbol{c}_Z$ as in Assumption M.1. As an example, we consider $M_1(\boldsymbol{c}_Z, \varepsilon)$ for known probabilities p_i such that $\varepsilon = 0$. Then,

$$M_1(\boldsymbol{c}_Z, 0) = \max_{i=1,\ldots,d} \frac{1}{\left(m_i^{(0)}(1-\boldsymbol{c}_Z)\right)^2} = \max_{i=1,\ldots,d} \frac{1}{\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2(1-\boldsymbol{c}_Z)}q_i^2\right)\right)^2} \tag{3.5}$$

with $q_i := \Phi^{-1}(1-p_i)$. In Figure 1, we plot $M_1(\boldsymbol{c}_Z, 0)$ for $p := p_1 = \cdots = p_d$ and as function of $\boldsymbol{c}_Z \in (0,1)$ and $p \in (0,1)$. As expected, for strong temporal and cross-sectional correlation, i.e. $\boldsymbol{c}_Z$ close to one as well as for very small and large probabilities, the constants can get quite large. In general, with growing dimension d, and more and more values p_i contributing, we can get values close to the boundary which results in large $Q(\mathbf{\Gamma}_Z)$. One can avoid this phenomenon by restricting the set of possible values to a closed subset of $(0,1)$.

4. Sparse estimation for latent VAR processes

In this section, we suppose that $\{Z_t\}_{t\in\mathbb{Z}}$ follows a causal VAR model of order p (VAR(p)); that is,

$$Z_t = \sum_{u=1}^{p} \Psi_u Z_{t-u} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{4.1}$$

for some $\Psi_u \in \mathbb{R}^{d\times d}$ and white noise series $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ characterized by

$$\mathrm{E}[\varepsilon_t] = 0, \quad \mathrm{E}[\varepsilon_t\varepsilon_t'] = \Sigma_\varepsilon, \quad \mathrm{E}[\varepsilon_s\varepsilon_t'] = 0 \quad \text{for } s \neq t. \tag{4.2}$$

We assume that the VAR(p) process is causal; that is,

$$\det(\Psi(z)) \neq 0, \quad \text{for} \quad |z| \leq 1, z \in \mathbb{C} \quad \text{with} \quad \Psi(z) = I_d - \Psi_1 z - \cdots - \Psi_p z^p. \tag{4.3}$$

Our goal here is to estimate the transition matrices $\Psi_1, \ldots, \Psi_p$ in (4.1) sparsely in a possibly high-dimensional regime.

While our main result in Proposition 3.1 is proven in generality without imposing any assumptions on the observed series, our results here require a couple of more assumptions.

Assumption V.1. The function G in (1.2) satisfies $G : \mathbb{R}^d \to [a, b]^d$.

Assumption V.2. The parameters in θ_i satisfy $\theta_i = \mathrm{E}[X_{i,t}]$ which allows us to estimate θ_i via $\widehat{\theta}_i = \frac{1}{T}\sum_{t=1}^{T} X_{i,t}$.

These assumptions are satisfied for all discrete distributions having a finite support and whose population mean coincides with the unknown parameter characterizing the respective distribution. Examples are Bernoulli, binomial and hypergeometric distributions. In contrast, it excludes discrete distributions with an infinite support set such as Poisson. We refer the reader to the second part of Section 4.3 for a detailed discussion and potential extensions.

In the two subsequent sections, estimation of the coefficient matrices of $\{Z_t\}$ (Section 4.1) and their estimator's theoretical properties (Section 4.2) are presented. We conclude with a discussion of our results and the required assumptions (Section 4.3) and an illustration of the convergence rate (Section 4.4).

4.1. Estimation procedure

To estimate the VAR coefficients sparsely, we adopt a procedure proposed by Basu and Michailidis (2015). However, their procedure needs to be modified so as to base inferences on the observations.

The VAR(p) model can be written in a linear models form as

$$\begin{pmatrix} Z'_{p+1} \\ \vdots \\ Z'_T \end{pmatrix} = \begin{pmatrix} Z'_p & \ldots & Z'_1 \\ \vdots & \ddots & \vdots \\ Z'_{T-1} & \ldots & Z'_{T-p} \end{pmatrix} \begin{pmatrix} \Psi'_1 \\ \vdots \\ \Psi'_p \end{pmatrix} + \begin{pmatrix} \varepsilon'_{p+1} \\ \vdots \\ \varepsilon'_T \end{pmatrix} \quad \text{or} \quad \mathcal{Y}_Z = \mathcal{X}_Z B_0 + \mathcal{E}. \tag{4.4}$$

A vectorized version of (4.4) is then seen to be

$$\begin{aligned} \mathrm{vec}(\mathcal{Y}_Z) &= \mathrm{vec}(\mathcal{X}_Z B_0) + \mathrm{vec}(\mathcal{E}) \\ &= (I_d \otimes \mathcal{X}_Z)\,\mathrm{vec}(B_0) + \mathrm{vec}(\mathcal{E}), \end{aligned} \tag{4.5}$$

or

$$Y = Z\beta_0 + E, \tag{4.6}$$

where $Y \in \mathbb{R}^{Nd}$ with $N = T - p$, $Z \in \mathbb{R}^{Nd \times q}$ with $q = pd^2$, $\beta_0 \in \mathbb{R}^q$, and $E \in \mathbb{R}^{Nd}$. To estimate $\Psi_1, \ldots, \Psi_p$, we simply estimate $\beta_0 \in \mathbb{R}^q$ in (4.6). To impose sparsity on the transition matrices, we state the following assumption on the true vector β_0.

Assumption V.3. Assume that β_0 is a s-sparse vector; that is, $\|\beta_0\|_0 = \sum_{u=1}^{p} \|\,\mathrm{vec}(\Psi_u)\|_0 = s$.

Following Basu and Michailidis (2015), we define a LASSO-type estimator for β_0 by

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^q}{\arg\min} \Big(-2\beta'\widehat{\gamma} + \beta'\widehat{\Gamma}\beta + \lambda_N \|\beta\|_1 \Big) \tag{4.7}$$

with

$$\begin{aligned} \widehat{\gamma} &= \mathrm{vec}(\widehat{\boldsymbol{\gamma}}_Z) = \mathrm{vec}(\widehat{g}(\widehat{\boldsymbol{\gamma}}_X)), \\ \widehat{\Gamma} &= I_d \otimes \widehat{\boldsymbol{\Gamma}}_Z = I_d \otimes \widehat{g}(\widehat{\boldsymbol{\Gamma}}_X), \end{aligned} \tag{4.8}$$

where $\boldsymbol{\gamma}_Z = (\Gamma_Z(1)', \ldots, \Gamma_Z(p)')'$, $\boldsymbol{\Gamma}_Z = (\Gamma_Z(r-s))_{r,s=1,\ldots,p}$, and their estimated counterparts $\widehat{\boldsymbol{\gamma}}_Z$ and $\widehat{\boldsymbol{\Gamma}}_Z$ are defined analogously. Furthermore, for the observed series $\{X_t\}$, we set $\widehat{\boldsymbol{\gamma}}_X = N^{-1}\mathcal{X}_X'\mathcal{Y}_X$ and $\widehat{\boldsymbol{\Gamma}}_X = \frac{1}{N}\mathcal{X}_X'\mathcal{X}_X$, where $\mathcal{Y}_X$ and $\mathcal{X}_X$ are quantities analogous to $\mathcal{Y}_Z$ and $\mathcal{X}_Z$ above formed by replacing $\{Z_t\}_{t=1,\ldots,T}$ with $\{X_t\}_{t=1,\ldots,T}$ in (4.4).

In contrast to Basu and Michailidis (2015), the population quantities γ and Γ in (4.8) cannot be estimated through the VAR series $\{Z_t\}$, which is unobserved. Instead, we need to estimate γ and Γ from $\{X_t\}_{t=1,\ldots,T}$.

4.2. Theoretical properties

The theoretical properties of $\widehat{\beta}$ in (4.7) are derived here. We state consistency results in a possibly high-dimensional regime, allowing d and T to go to infinity.

As suggested in Loh and Wainwright (2012) and Basu and Michailidis (2015), we first establish consistency under a restricted eigenvalue condition and a deviation bound. Subsequently, we verify that these conditions are satisfied by $\{Z_t\}$. Section 4.3 clarifies how our proofs differ from those in Loh and Wainwright (2012) and Basu and Michailidis (2015).

Restricted eigenvalue: A symmetric matrix $\widehat{\Gamma} \in \mathbb{R}^{q\times q}$ satisfies the restricted eigenvalue condition with curvature $\alpha > 0$ and tolerance $\tau > 0$ if

$$x'\widehat{\Gamma}x \geq \alpha\|x\|^2 - \tau\|x\|_1^2 \quad \text{for all} \quad x \in \mathbb{R}^q. \tag{4.9}$$

We write $\widehat{\Gamma} \sim RE(\alpha, \tau)$ for short.

Deviation bound: There exists a deterministic function $\mathcal{Q}(\beta_0)$ such that

$$\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} \leq \mathcal{Q}(\beta_0)\sqrt{\frac{\log(q)}{N}}. \tag{4.10}$$

The following proposition ensures consistent estimation of the coefficients for a latent VAR(p) model obeying the restricted eigenvalue condition and deviation bounds. The statement is effectively the same as Proposition 4.1 in Basu and Michailidis (2015) for observed VAR models but using the estimators (4.8) instead of those for an observed series.

Proposition 4.1. *Suppose that $\widehat{\Gamma}$ in (4.8) satisfies the restricted eigenvalue condition (4.9) with $s\tau \leq \alpha/32$ and that $(\widehat{\Gamma}, \widehat{\gamma})$ in (4.8) satisfies the deviation*

bound in (4.10)*. Then, for any* $\lambda_N \geq 4\mathcal{Q}(\beta_0)\sqrt{\frac{\log(q)}{N}}$,

$$\|\widehat{\beta} - \beta_0\|_1 \leq 64s\frac{\lambda_N}{\alpha}, \quad \|\widehat{\beta} - \beta_0\| \leq 16\sqrt{s}\frac{\lambda_N}{\alpha},$$
$$(\widehat{\beta} - \beta_0)'\widehat{\Gamma}(\widehat{\beta} - \beta_0) \leq 128s\frac{\lambda_N^2}{\alpha}.$$

The following two lemmas provide sufficient conditions for when the restricted eigenvalue condition (4.9) and deviation bound (4.10) hold for a latent VAR series. To state the lemmas, recall the causality assumption in (4.3) on the VAR(p) model and set $\mu_{\max}(\mathcal{A}) := \max_{|z|=1} \lambda_{\max}(\mathcal{A}^*(z)\mathcal{A}(z))$, with $\mathcal{A}(z) = I_d - \sum_{j=1}^p A_j z^j$.

Lemma 4.1 (Verifying restricted eigenvalue)**.** *Suppose that Assumptions M.1–M.4 hold and that the latent process follows a causal VAR model. Then, there are constants* $c_1, c_2 > 0$ *such that for all* $N \gtrsim \max\{c_0(s)\nu^{-1}, \nu^{-2}, 1\}$ $\max\{s\log(dp), \log(dK)\}$*, with probability at least* $1 - c_1\exp(-c_2 N \min\{1, \nu^2, \nu/c_0(s)\})$,

$$\widehat{\Gamma} \sim RE(\alpha, \tau),$$

where

$$\alpha = \frac{\lambda_{\min}(\Sigma_\varepsilon)}{2\mu_{\max}(\mathcal{A})}, \quad \tau = \alpha\max\{c_0(s)\nu^{-1}, \nu^{-2}, 1\}\frac{\log(dp)}{N}, \quad \nu = \frac{\lambda_{\min}(\Sigma_\varepsilon)}{54\mu_{\max}(\mathcal{A})Q(\mathbf{\Gamma}_Z)c_0(s)} \tag{4.11}$$

with $Q(\mathbf{\Gamma}_Z)$ *defined as in Proposition 3.1 and* $c_0(s) = s$.

Lemma 4.2 (Verifying deviation bound)**.** *Suppose that Assumptions M.1–M.4 hold and that the latent process is a causal VAR. Then, there are constants* $c_1, c_2 > 0$ *such that for all* $N \gtrsim \max\{c_0(s)\nu^{-1}, \nu^{-2}, 1\}s\log(dp)$*, with probability at least* $1 - c_1\exp\left(-c_2 N\min\{1, \nu^2, \nu/c_0(s)\}\right)$,

$$\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} \leq \sqrt{\frac{\log(q)}{N}}\mathcal{Q}(\beta_0),$$

where

$$\mathcal{Q}(\beta_0) = Q(\mathbf{\Gamma}_Z)c_0(s)$$

with $Q(\mathbf{\Gamma}_Z)$ *as in Proposition 3.1,* B_0 *in* (4.4)*,* $e_{q,i}$ *denotes the ith basis vector of* $\mathbb{R}^q$ *and* $c_0(s) \leq s$.

4.3. Discussion

A discussion on our results, including a comparison to existing literature and potential relaxations of our assumptions, is now provided.

Comparison to Basu and Michailidis (2015): The statement of Lemma 4.1 is analogous to the one of Proposition 4.2 in Basu and Michailidis (2015). The proof is very similar and leads to almost the same relation between the

sample size $N = T - p$ and the dimension d. More precisely, we get $N \gtrsim \max\{c_0(s)\nu^{-1}, \nu^{-2}, 1\} \max\{s\log(dp), \log(dK)\}$ and Proposition 4.2 in Basu and Michailidis (2015) states that $N \gtrsim \max\{\nu^{-2}, 1\} \max\{s\log(dp)\}$. Since the number of unknown parameters K in the marginal distributions is relatively small, it is expected that $\max\{s\log(dp), \log(dK)\} = s\log(dp)$. Furthermore, $c_0(s)\nu^{-1}$ is due a second-order approximation used in the proofs and will also vanish for certain choices of ν.

The statement of Lemma 4.2 is analogous to the one of Proposition 4.3 in Basu and Michailidis (2015). This said, the deviation bound is significantly different in our setting and results in a different rate between the sample size $N = T - p$ and the dimension d. Basu and Michailidis (2015) require $N \gtrsim \max\{\nu^{-2}, 1\}2\log(dp)$, while our relation $N \gtrsim \max\{c_0(s)\nu^{-1}, \nu^{-2}, 1\}s\log(dp)$ includes the sparsity parameter s. However, this relation only impacts the assumptions on the verification of the deviation bound in Lemma 4.2. Proposition 4.1 is however not impacted and requires the same assumptions on N and d as Proposition 4.1 in Basu and Michailidis (2015). To compare our proof of Lemma 4.2 with that of Proposition 4.3 in Basu and Michailidis (2015), assume that $\{Z_t\}$ is observed. Then, estimation in (4.7) is done through

$$\begin{aligned} \widehat{\gamma} &= \mathrm{vec}(\widehat{\boldsymbol{\gamma}}_Z) = \mathrm{vec}(\mathcal{X}_Z'\mathcal{Y}_Z), \\ \widehat{\Gamma} &= I_d \otimes \widehat{\boldsymbol{\Gamma}}_Z = I_d \otimes \mathcal{X}_Z'\mathcal{X}_Z/N. \end{aligned} \tag{4.12}$$

The estimators in (4.12) simply replace those in (4.8). Basu and Michailidis (2015) use the fact that $\widehat{\gamma} - \widehat{\Gamma}\beta_0 = (I_d \otimes \mathcal{X}_Z')\,\mathrm{vec}(E)/N = \mathrm{vec}(\mathcal{X}_Z'E)/N$, with E in (4.6), to extract a concentration bound on $\|\mathcal{X}_Z'E/N\|_{\max}$. Since $\{Z_t\}$ is unobserved in our setting, we need to reduce the issue to inference of autocovariance matrices to use our main result Proposition 3.1. Concluding, our procedure affects the convergence rate for the deviation bound but not for the main consistency result. As a byproduct, our procedure offers an alternative proof for the deviation bound in Basu and Michailidis (2015).

Discussion of assumptions part I: Our analysis of the latent VAR setup is restricted to bounded G_i in (1.2) as formalized in Assumption V.1. Our proof route will verify Assumptions C.1 and C.2 so that Corollary 3.1 can be applied. While Proposition 3.1 allows us to phrase concentration bounds on the autocovariance matrices of the latent process through those of the observed one, it remains to verify Assumption C.1 for an estimator of the autocovariances of the observed process. Since the observed process is a function G of a Gaussian vector series, bounded G_i s permit use of Hoeffding's inequality for Markov chains (Fan, Jiang and Sun, 2021). Most existing concentration bounds for functionals have been developed under Lipschitz continuity; a general result for our setting requires results for the Hermite polynomials in (2.1). The case for unbounded functions of Gaussian random variables is more challenging and has been studied only in special cases. Here, Adamczak and Bednorz (2015) develop concentration bounds for polynomials of certain degree. However, their bounds are difficult to compute explicitly since they rely on a generalization of the Frobenius norm for multi-indexed matrices. Adamczak and Wolff (2015)

generalized Fan, Jiang and Sun (2021) for unbounded functions, but require much more restrictive conditions than a Markov chain.

Assumption V.2 is chosen for simplicity. The purpose of this section is to illustrate the usefulness of our main result Proposition 3.1. In principle, we should be able to prove concentration results for other estimators for the parameters of the marginal distribution. For example for the variance using the method of moments.

Discussion of assumptions part II: We add here a discussion on Assumption C.1 and $c_0(s)$ therein. If the causal VAR(p) series $\{Z_t\}$ is observed, Assumption C.1 is effectively satisfied by Proposition 2.4 in Basu and Michailidis (2015). In Proposition 2.4 of Basu and Michailidis (2015), $c_0(s) = 2\pi\mathcal{M}(f_Z, s)$ with

$$\mathcal{M}(f_Z, s) := \max_{S\subset\{1,\dots,q\},|S|\le s} \operatorname*{ess\,sup}_{\lambda\in[-\pi,\pi]} \|f_{Z(S)}(\lambda)\|,$$

where $f_{Z(S)}$ describes the spectral density of the subprocess $\{Z(S)\} = \{\widetilde{Z}_{i,t} \mid i \in S\}_{t\in\mathbb{Z}}$ with $\widetilde{Z}_t = (Z_t', Z_{t-1}', \dots, Z_{t-p+1}')'$. In Section 4.4 below, we illustrate our convergence rate and compare it with the one for observed VAR models.

4.4. Illustration of estimation error

In this section, we illustrate our theoretical results on latent VAR estimation in a numerical study. We demonstrate how the estimation error of our estimator for the latent VAR (4.7) scales with the sample size T and dimension d. We simulate d-dimensional count series with marginal Bernoulli distributions and latent Gaussian VAR(1) process. The success probabilities were randomly sampled from $p_i \in (0.4, 0.7)$, $i = 1, \dots, d$. For different values of d ($d = 5, 10, 15, 20, 25, 30$, or in terms of the number of parameters, $d^2 = 25, 100, 225, 400, 625, 900$, resp.), we generated sparse coefficient matrices Ψ ($\beta_0 = \mathrm{vec}(\Psi)$) with sparsity s ($s = (13, 28, 43, 58, 73, 88)$, resp.), that is, the number of non-zero entries; see Assumption V.3. Then, we applied (4.7) with tuning parameter $\lambda_N = \sqrt{\log(d^2)/N}$ on samples of sizes $T \in \{200, 300, 400, 500, 1000, 2000\}$. The ℓ_2-error of estimation $\|\widehat{\beta} - \beta_0\|^2$ is plotted in the first row of Figure 2. The left panel displays the errors for different values of d, plotted against the sample sizes T. As expected, the errors increase with the dimension. The right panel displays the estimation errors against the rescaled sample size $(T-1)/(s\log(d^2))$. Note that this rate is suggested by our theoretical result in Proposition 4.1. That is, the result suggests that $\|\widehat{\beta} - \beta_0\|^2$ is proportional to $s\lambda_N^2 = s\log(d^2)/(T-1) = 1/((T-1)/(s\log(d^2)))$ and hence is proportional to $1/x$ if plotted against $x = (T-1)/(s\log(d^2))$.

As discussed in Section 4.3, it is not expected to get a better rate when the link function is known. The estimation of θ to get estimates of the link function should only impact the rate when K (the number of unknown parameters) is large. To illustrate that, the second row in Figure 2 displays plots for the same setting as described above but for known link function. As one can see, there is no significant difference between the plots of the first and second rows.

As also discussed in Section 4.3, our theoretical results claim that we get the same convergence rate as when we observe a process which follows a VAR model and estimate the transition matrices directly from the data. Those results are proved in Basu and Michailidis (2015). For reference, the third row of Figure 2, displays the results for the observed process following the VAR(1) models in our simulation study. As one can see, the estimation error does scale well with the same rate as for the latent models, the difference being in the scale of the vertical axes where the observed process naturally has smaller estimation errors.

5. Conclusions

This work considered a possibly high-dimensional count time series model whose correlation structure is determined through the correlation of an underlying latent Gaussian process. We derived a relation between consistent estimation of the autocorrelation matrices of the latent process and the autocovariance matrices of the observed process.

Several theoretical challenges needed to be addressed to ensure consistent estimation of the latent model. These include estimation of the link function based on unknown CDF parameters, the non-differentiability of the link function around unity, and the issue of high-dimensionality; see Section A.1. Assuming that the latent process follows a VAR model, our results ensure consistent estimation of the transition matrices of a VAR series at the same rate as for an observable VAR series.

While we illustrated our results on the example of a latent VAR model, our main result can be used for other models. For instance, it is conceivable to assume that the latent process follows a dynamic factor model where the factors follow a stationary VAR structure as in Kim et al. (2024). Consistency results require concentration bounds on functionals of a dynamic factor model. A similar question concerns how to derive consistency for possibly unbounded functions of the latent Gaussian process as discussed in Section 4.3.

Other directions are the modeling of matrix valued count time series or of network data driven by a changing adjacency matrix, as well as extensions to spatial settings. Further questions include extensions to non-stationary models, particularly those involving covariates, and estimation of related model parameters of the latent process like the VAR order.

Appendix A: Proofs of the main results

We give a short roadmap of the proof of Proposition 3.1 (Section A.1). Subsequently, we restate Proposition 3.1 in terms of explicit constants which are omitted in the statement (Section A.2) and continue then with the detailed proofs of our main results (Section A.3).

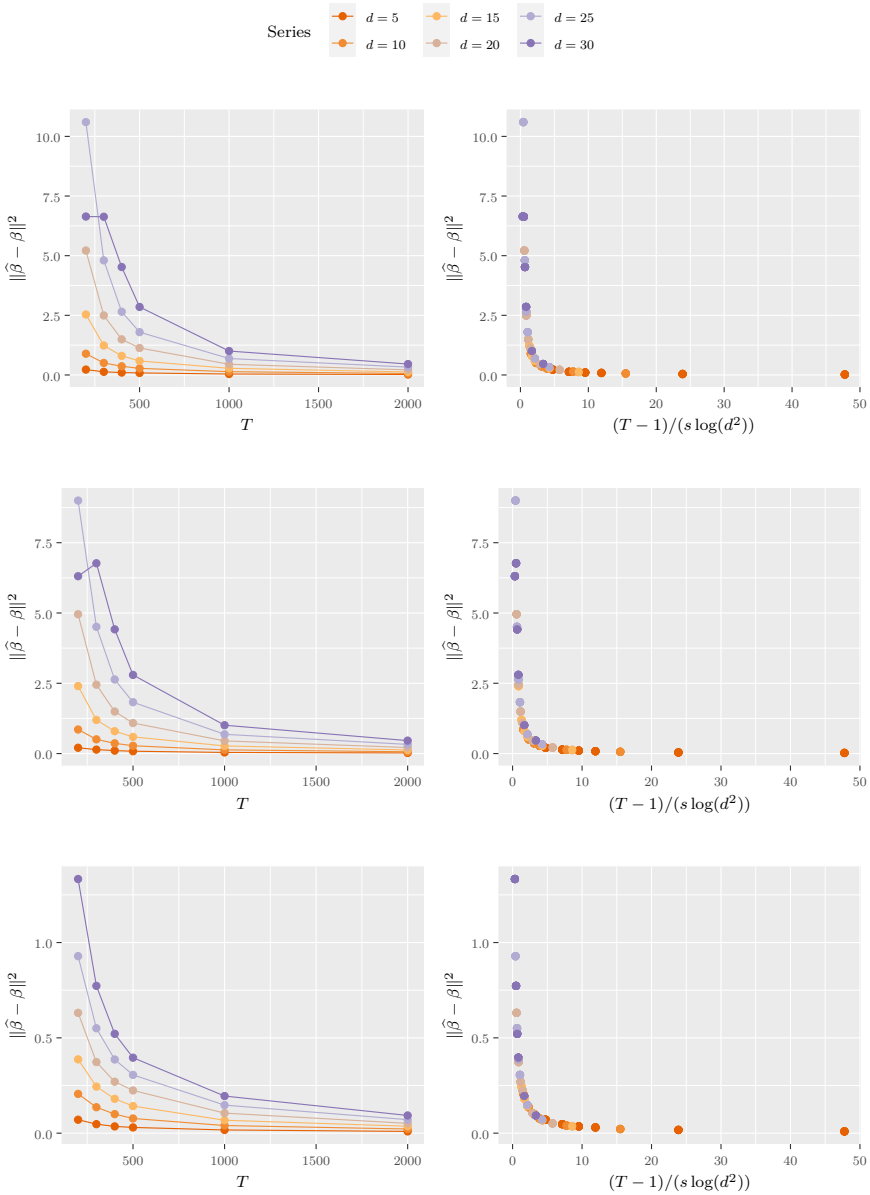

FIG 2. *Estimation error of the LASSO* $\|\widehat{\beta} - \beta\|^2$ *plotted against* T *(left column) and the theoretical rate* $(T-1)/(s\log(d^2))$ *(right column) for three different settings: First row: d-dimensional count series with marginal Bernoulli distributions and latent Gaussian VAR(1) process. Second row: d-dimensional count series with marginal Bernoulli distributions and latent Gaussian VAR(1) process with known link function. Third row: d-dimensional Gaussian VAR(1).*

A.1. Roadmap

We give here a roadmap of the proof of Proposition 3.1 to emphasize the overall structure and some of the main tools. Recall that

$$\mathbf{\Gamma}_Z = g(\mathbf{\Gamma}_X) \quad \text{with} \quad g = \ell^{-1}.$$

In contrast to the link function ℓ, its first derivative ℓ' admits a known explicit representation which is stated in Proposition 2.1. We would like to take advantage of this representation, and identify the following three issues to deal with.

First, the derivative $\ell'(u)$ is defined only on the open interval $u \in (-1, 1)$. Therefore, we have to deal with the cases $u = 1$ and $u = -1$ separately. Assumption M.1 excludes both cases for all off-diagonal elements of $\mathbf{\Gamma}_Z$. However, all diagonal elements of $\mathbf{\Gamma}_Z$ satisfy $\mathbf{\Gamma}_{Z,rr} = \Gamma_{Z,ii}(0) = 1$ since we assume that the latent process is standard Gaussian (has variance one). The general strategy will be to apply a first-order Taylor approximation to the diagonal elements and a second-order Taylor expansion to the off-diagonals. This procedure will be applied to a range of different functions related to ℓ. One of those functions is $g = \ell^{-1}$. In order to exclude all diagonal elements and to treat those separately, we introduce the notation

$$g_{\bullet,rs}(v) = \begin{cases} 1, & \text{for all } r, s \text{ such that } \mathbf{\Gamma}_{Z,rs} = \Gamma_{Z,ii}(0), \\ g_{rs}(v), & \text{else.} \end{cases} \tag{A.1}$$

This notation will also be used for functions other than g but refers to the same edges r, s to be equal to one.

The second and third issues need a little more context. Our goal is to bound the distance $\|\widehat{g}(\widehat{\mathbf{\Gamma}}_X) - g(\mathbf{\Gamma}_X)\|_s$ by $\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s$ involving the autocovariances of the observed process. An application of the mean value theorem is not sufficient. The function g itself is estimated through the unknown parameters θ of the marginal distributions. Second, we address this latter issue by controlling the error we make by estimating θ; the procedure is explained in more detail in Remark C.3.

Third, consider the simpler problem of proving the bound $\|g(\widehat{\mathbf{\Gamma}}_X) - g(\mathbf{\Gamma}_X)\|_s \leq C\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s$ and note that applying the mean value theorem componentwise yields $\|g(\widehat{\mathbf{\Gamma}}_X) - g(\mathbf{\Gamma}_X)\|_s \leq \|g'(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\|_s$ for some Σ. Since $g'(\Sigma)$ is not necessarily positive semidefinite, one can show that the optimal C such that $\|g(\widehat{\mathbf{\Gamma}}_X) - g(\mathbf{\Gamma}_X)\|_s \leq C \max_{i=1,\ldots,d} |g'_{ii}(\Sigma)| \|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s$ is of order $\sqrt{2s}$. This would weaken our bounds. For this reason, we will work with a second-order Taylor expansion; see also Remark C.1 and Wegkamp and Zhao (2016).

The following tree diagram gives an idea of how the subsequent sections contribute to the different proof steps. Each node of the graph below represents a distance which needs to be controlled with high probability. With each layer of the diagram, we reduce the problem further, up to the point where we only rely on the observed process. The edges represent proof steps which are justified

in subsequent sections. Note that we omit the consideration of the diagonal elements, those are studied in Section C.4.

In the fourth and fifth layer of the diagram, the distances have a superscript $r = 1, 2$. This is due to the use of second-order Taylor expansions applied to different functions.

$$\begin{tikzcd}
 & & & \|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s \arrow[d] & & \\
 & & & \|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s \arrow[dl] \arrow[dr] & & \\
 & & \|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\widehat{\mathbf{\Gamma}}_X)\|_s \arrow[dl] \arrow[d] \arrow[dr] & & \|g_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s \arrow[d] & \\
\text{Sections C.1 and C.2} & \|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^r & \|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^r \arrow[d] & \|\widehat{\theta} - \theta\|_{\max} & \|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^r & \\
\text{Sections C.3} & & \|\widehat{\theta} - \theta\|_{\max}^r & & &
\end{tikzcd}$$

Note that the statement of Proposition 3.1 involves a series of parameters $\delta, \widetilde{\delta}, \varepsilon, \widetilde{\varepsilon}$. While δ is supposed to control the difference between $\widehat{\mathbf{\Gamma}}_Z$ and $\mathbf{\Gamma}_Z$, the remaining ones only appear on the right hand side of the relation (3.1). The remaining parameters $\widetilde{\delta}, \varepsilon, \widetilde{\varepsilon}$ ensure respectively that $\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s$, $\|\widehat{\theta} - \theta\|_{\max}$ and $\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s$ are small. See also Remarks A.1 and C.3 below for further discussions.

A.2. Statements of results in Section 3 with explicit constants

We provide here the explicit constants which enter our main result Proposition 3.1. For completeness, we restate the proposition.

Proposition A.1. *Suppose that Assumptions M.1–M.4 hold. Then, for any* $\delta, \widetilde{\delta}, \varepsilon, \widetilde{\varepsilon} > 0$,

$$\begin{aligned} \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > Q(\mathbf{\Gamma}_Z)\delta\right] \precsim\ & \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] \\ & + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon \wedge \widetilde{\varepsilon}] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max}^2 > \delta] \end{aligned} \tag{A.2}$$

with

$$Q(\mathbf{\Gamma}_Z) := Q(\widetilde{\delta}, \varepsilon, \widetilde{\varepsilon}, \mathbf{\Gamma}_Z) = 4\max\{D(\mathbf{\Gamma}_Z), 4R(\mathbf{\Gamma}_Z), 2U(\mathbf{\Gamma}_Z), T(\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z), 1\}$$

and the quantities D, R, S, T, U *defined below.*

The quantities entering (A.2) are:

$$
\begin{aligned}
D(\mathbf{\Gamma}_Z) &:= D(\varepsilon, \mathbf{\Gamma}_Z) = M_1^{\frac{1}{2}}(1/2, \varepsilon) 2 \max\{3\Delta(\varepsilon), 1\}, \\
R(\mathbf{\Gamma}_Z) &:= R(\varepsilon, \mathbf{\Gamma}_Z) = \Big(8\pi M_1(0, \varepsilon) + 24\pi \frac{1}{(1-\boldsymbol{c}_Z^2)^2} M_2(\boldsymbol{c}_Z, \varepsilon)\Big) \|\mathbf{\Gamma}_Z\|_s, \\
S(\mathbf{\Gamma}_Z) &:= S(\varepsilon, \mathbf{\Gamma}_Z) = \frac{12}{(1-\boldsymbol{c}_Z^2)^{\frac{7}{2}}} M(\boldsymbol{c}_Z, \varepsilon) \mu(\boldsymbol{c}_Z, \varepsilon) \|\mathbf{\Gamma}_Z\|_s, \\
T(\mathbf{\Gamma}_Z) &:= T(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z) = \frac{6}{1-\boldsymbol{c}(\widetilde{\delta})^2} M_1(\boldsymbol{c}(\widetilde{\delta}), \varepsilon) M_2(\boldsymbol{c}(\widetilde{\delta}), \varepsilon), \\
U(\mathbf{\Gamma}_Z) &:= U(\varepsilon, \widetilde{\varepsilon}, \mathbf{\Gamma}_Z) = T(\varepsilon, \max\{S^2(\mathbf{\Gamma}_Z), 1\}\widetilde{\varepsilon}, \mathbf{\Gamma}_Z),
\end{aligned}
\tag{A.3}
$$

for some constant $\boldsymbol{c}(\widetilde{\delta}) \in (0,1)$, which depends on $\widetilde{\delta}$; here, Δ, M, μ and M_1, M_2 are defined in (2.13), (2.14), and (2.15).

Proposition 3.1/ A.1 is a consequence of Lemmas A.1 and A.2 below, which respectively cover the cases when g is estimated or known.

Lemma A.1. *Suppose that Assumptions M.1–M.4 hold. Then, for any* $\delta, \widetilde{\delta}, \varepsilon, \widetilde{\varepsilon} > 0$,

$$
\begin{aligned}
&\mathrm{P}\left[\|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\widehat{\mathbf{\Gamma}}_X)\|_s > Q_1(\mathbf{\Gamma}_Z)\delta\right] \\
&\quad \precsim \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}\right] + \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta\right] \\
&\qquad + \mathrm{P}\left[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon \wedge \widetilde{\varepsilon}\right] + \mathrm{P}\left[\|\widehat{\theta} - \theta\|_{\max}^2 > \delta\right],
\end{aligned}
$$

where

$$
Q_1(\mathbf{\Gamma}_Z) := Q_1(\widetilde{\delta}, \varepsilon, \widetilde{\varepsilon}, \mathbf{\Gamma}_Z) = 4\max\{4R(\mathbf{\Gamma}_Z), 2U(\mathbf{\Gamma}_Z), T(\mathbf{\Gamma}_Z)\} \max\{S^2(\mathbf{\Gamma}_Z), 1\}
$$

with R, S, T, *and* U *as in* (A.3).

Lemma A.2. *Suppose that Assumptions M.1–M.4 hold. Then, for any* $\delta, \widetilde{\delta} > 0$,

$$
\begin{aligned}
\mathrm{P}\left[\|g_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > Q_2(\mathbf{\Gamma}_Z)\delta\right] &\precsim \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}\right] \\
&\quad + \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta\right],
\end{aligned}
$$

where $Q_2(\mathbf{\Gamma}_Z) := Q_2(\widetilde{\delta}, \mathbf{\Gamma}_Z) = \max\{2R(0, \mathbf{\Gamma}_Z), T(0, \widetilde{\delta}, \mathbf{\Gamma}_Z)\}$ *with* R, T *as in* (A.3).

In the upcoming proof sections, we refer to the statements in Section 3 as opposed to those in the current Section A.2 but we certainly use the constants in (A.3).

A.3. Proofs

Proof of Proposition 3.1/ A.1. By treating the diagonal and off-diagonal elements of $\mathbf{\Gamma}_Z$ separately, we get

$$
\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > Q(\mathbf{\Gamma}_Z)\delta\right]
$$

$$
\begin{aligned}
&= \mathrm{P}\left[\|\widehat{g}(\widehat{\mathbf{\Gamma}}_X) - g(\mathbf{\Gamma}_X)\|_s > Q(\mathbf{\Gamma}_Z)\delta\right] \\
&\le \mathrm{P}\left[\|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > Q(\mathbf{\Gamma}_Z)\frac{\delta}{2}\right] \\
&\qquad + \mathrm{P}\left[\max_{i=1,\dots,d} |\widehat{g}_{ii}(\widehat{\Gamma}_{X,ii}(0)) - g_{ii}(\Gamma_{X,ii}(0))| > D(\mathbf{\Gamma}_Z)\delta\right] \qquad \text{(A.4)} \\
&\lesssim \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \widetilde{\varepsilon} \wedge \varepsilon] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max}^2 > \delta]. \qquad \text{(A.5)}
\end{aligned}
$$

To obtain (A.4), we used the triangle inequality and the fact that for the diagonal elements of any matrix $A \in \mathbb{R}^{d\times d}$, $\|I_L \otimes \mathrm{diag}(a_{11}, \dots, a_{dd})\|_s \le \max_{i=1,\dots,d} |a_{ii}|$. To obtain (A.5), the second summand is bounded by Lemma C.15. The rest of the proof bounds the first summand of (A.4). We have

$$
\begin{aligned}
&\mathrm{P}\left[\|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > Q(\mathbf{\Gamma}_Z)\frac{\delta}{2}\right] \\
&\le \mathrm{P}\left[\|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\widehat{\mathbf{\Gamma}}_X)\|_s > Q(\mathbf{\Gamma}_Z)\frac{\delta}{4}\right] \\
&\qquad + \mathrm{P}\left[\|g_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > Q(\mathbf{\Gamma}_Z)\frac{\delta}{4}\right] \\
&\le \mathrm{P}\left[\|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\widehat{\mathbf{\Gamma}}_X)\|_s > Q_1(\mathbf{\Gamma}_Z)\delta\right] \\
&\qquad + \mathrm{P}\left[\|g_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > Q_2(\mathbf{\Gamma}_Z)\delta\right] \qquad \text{(A.6)} \\
&\lesssim \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon \wedge \widetilde{\varepsilon}] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max}^2 > \delta],
\end{aligned}
$$

where (A.6) follows from

$$
\begin{aligned}
&Q(\mathbf{\Gamma}_Z) \\
&= 4\max\{D(\mathbf{\Gamma}_Z), 4R(\mathbf{\Gamma}_Z), 2U(\mathbf{\Gamma}_Z), T(\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z), 1\} \\
&\ge 4\max\{4R(\varepsilon, \mathbf{\Gamma}_Z), 2U(\varepsilon, \widetilde{\varepsilon}, \mathbf{\Gamma}_Z), T(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z), 1\} \\
&= 4\max\{4R(\varepsilon, \mathbf{\Gamma}_Z), 2U(\varepsilon, \widetilde{\varepsilon}, \mathbf{\Gamma}_Z), T(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z), 2R(0, \mathbf{\Gamma}_Z), T(0, \widetilde{\delta}, \mathbf{\Gamma}_Z)\} \\
&\qquad\qquad \times \max\{S^2(\mathbf{\Gamma}_Z), 1\} \\
&= 4\max\{Q_1(\mathbf{\Gamma}_Z), Q_2(\mathbf{\Gamma}_Z)\}
\end{aligned}
$$

with $Q(\mathbf{\Gamma}_Z) \ge 4\max\{Q_1(\mathbf{\Gamma}_Z), Q_2(\mathbf{\Gamma}_Z)\}$ and $Q_1(\mathbf{\Gamma}_Z), Q_2(\mathbf{\Gamma}_Z)$ defined in Lemmas A.1 and A.2. The two summands in (A.6) can then be bounded by the results in Lemmas A.1 and A.2, respectively. □

Proof of Corollary 3.1. The corollary is a consequence of Proposition 3.1. Indeed, set $\nu = \delta \wedge \widetilde{\delta} \wedge \varepsilon \wedge \widetilde{\varepsilon}$. Then,

$$
\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > Q(\mathbf{\Gamma}_Z)c_0(s)\delta\right]
$$

$$
\begin{aligned}
&\precsim \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > c_0(s)\nu] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > c_0(s)\nu] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \nu] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max}^2 > \nu] \qquad \text{(A.7)} \\
&\le c_{1,1} \exp\Big(- c_{1,2} N \min\Big\{1, \nu^2, \sqrt{\nu/c_0(s)}, \nu/c_0(s)\Big\} \\
&\qquad + 2s \min\{\log(Ld), \log(21e\ Ld/2s)\}\Big) \\
&\qquad\qquad + c_{2,1} dK \exp\left(-c_{2,2} T \min\{\nu, \nu^2\}\right) \qquad \text{(A.8)} \\
&\le c_{1,1} \exp\Big(-c_{1,2} N \min\Big\{1, \nu^2, \sqrt{\nu/c_0(s)}\Big\} + 2s\log(Ld)\Big) \\
&\qquad + c_{2,1} dK \exp\left(-c_{2,2} T \min\{1, \nu^2\}\right), \qquad \text{(A.9)}
\end{aligned}
$$

where (A.7) follows from Proposition 3.1 and since $c_0(s) \ge 1$ in Assumption C.1, (A.8) is discussed in more detail below, and (A.9) is due to

$$
\min\left\{\nu, \nu^2, \sqrt{\frac{\nu}{c_0(s)}}, \frac{\nu}{c_0(s)}\right\} \ge \min\left\{1, \nu^2, \frac{\nu}{c_0(s)}\right\}.
$$

Turning back to relation (A.8), the last two probabilities in (A.7) are bounded by Assumption C.2. The first two probabilities in (A.7) can be bounded from the following observation. By Assumption C.1 and Lemma F.2 in the supplementary material of Basu and Michailidis (2015),

$$
\begin{aligned}
\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > c_0(s)\nu\right] &\le c_1 \exp\big(- c_2 N \min\{\nu, \nu^2\} \\
&\qquad + 2s \min\{\log(Ld), \log(21e\ Ld/2s)\}\big), \\
\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > c_0(s)\nu\right] &\le c_1 \exp\Bigg(- c_2 N \min\left\{\sqrt{\frac{\nu}{c_0(s)}}, \frac{\nu}{c_0(s)}\right\} \\
&\qquad + 2s \min\{\log(Ld), \log(21e\ Ld/2s)\}\Bigg).
\end{aligned}
$$

Lemma F.2 in the supplementary material of Basu and Michailidis (2015) requires Assumption C.1 to be true for any vector v with $\|v\| \le 1$. A close look into the proof reveals that it is sufficient to have a result for any vector $v \in \mathcal{K}(2s)$. The proof uses a discretization argument to approximate the set $\mathcal{K}(2s)$. More precisely, the authors construct an ε-net to approximate $\mathcal{K}(2s)$ following Definition 3.1 in Vershynin (2009) and the proofs therein utilizing the concept of ε-nets. Following Lemma 3.5 in Vershynin (2009), there is an ε-net which is a subset of $\mathcal{K}(2s)$. The proof of Lemma F.2 needs Assumption C.1 to be satisfied for any vector in the ε-net and therefore elements of $\mathcal{K}(2s)$. □

Proof of Lemma A.1. For shortness, we set $\delta^* = \max\{S^2(\mathbf{\Gamma}_Z), 1\}\delta$ and $\varepsilon^* = \max\{S^2(\mathbf{\Gamma}_Z), 1\}\widetilde{\varepsilon}$ for some $\widetilde{\varepsilon} > 0$. With explanations given below, we have

$$
\mathrm{P}\left[\|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\widehat{\mathbf{\Gamma}}_X)\|_s > Q_1(\mathbf{\Gamma}_Z)\delta\right]
$$

$$
\begin{aligned}
&\le \mathrm{P}\left[\|\widehat{g}_\bullet(\mathbf{\Gamma}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > Q_1(\mathbf{\Gamma}_Z)\frac{\delta}{4}\right] \\
&\qquad + \mathrm{P}\left[\left\|(\widehat{g}'_\bullet(\mathbf{\Gamma}_X) - g'_\bullet(\mathbf{\Gamma}_X)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s > Q_1(\mathbf{\Gamma}_Z)\frac{\delta}{4}\right] \\
&\qquad\qquad + \mathrm{P}\left[\frac{1}{2}\left\|g''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > Q_1(\mathbf{\Gamma}_Z)\frac{\delta}{4}\right] \\
&\qquad\qquad\qquad + \mathrm{P}\left[\frac{1}{2}\left\|\widehat{g}''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > Q_1(\mathbf{\Gamma}_Z)\frac{\delta}{4}\right] \qquad \text{(A.10)} \\
&\le \mathrm{P}\left[\|\widehat{g}_\bullet(\mathbf{\Gamma}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > q_1(\varepsilon, \varepsilon^*, \mathbf{\Gamma}_Z)\delta^*\right] \\
&\qquad + \mathrm{P}\left[\left\|(\widehat{g}'_\bullet(\mathbf{\Gamma}_X) - g'_\bullet(\mathbf{\Gamma}_X)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s > q_2(\varepsilon, \varepsilon^*, \mathbf{\Gamma}_Z)\delta^*\right] \\
&\qquad\qquad + \mathrm{P}\left[\left\|g''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_3(\widetilde{\delta}, \mathbf{\Gamma}_Z)\delta\right] \\
&\qquad\qquad\qquad + \mathrm{P}\left[\left\|\widehat{g}''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_4(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z)\delta\right] \qquad \text{(A.11)} \\
&\precsim \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > \delta^* \wedge \varepsilon^*] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > \delta^*] \\
&\qquad + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > \varepsilon^*] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > \delta^*] \\
&\qquad\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon] \\
&\qquad + \mathrm{P}\left[\left\|g''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_3(\widetilde{\delta}, \mathbf{\Gamma}_Z)\delta\right] \\
&\qquad\qquad + \mathrm{P}\left[\left\|\widehat{g}''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_4(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z)\delta\right] \qquad \text{(A.12)} \\
&\precsim \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > \delta^* \wedge \varepsilon^*] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > \delta^*] \\
&\qquad + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > \varepsilon^*] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > \delta^*] \\
&\qquad\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon] \qquad \text{(A.13)} \\
&\precsim \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > \delta^* \wedge \varepsilon^*] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > \delta^*] \\
&\qquad + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] \\
&\qquad\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon]. \qquad \text{(A.14)}
\end{aligned}
$$

The bound (A.10) follows by applying the second-order Taylor expansion to the function $x \mapsto \widehat{g}(x) - g(x)$ around the true covariance matrix, where Σ is such that $|\Sigma - \mathbf{\Gamma}_X| < \left|\Sigma - \widehat{\mathbf{\Gamma}}_X\right|$ (entry-wise); see also Remark A.1. To bound the probabilities in (A.10) further, we aim to use Lemmas C.1, C.2, C.3 and C.4. The four lemmas respectively introduce the constants q_1, q_2, q_3 and q_4. In the following, we argue why $Q_1(\mathbf{\Gamma}_Z)$ is always larger than either one of them. This will give us (A.11). Indeed, note that

$$
\begin{aligned}
&Q_1(\mathbf{\Gamma}_Z) \\
&= 4\max\{4R(\mathbf{\Gamma}_Z), 2U(\mathbf{\Gamma}_Z), T(\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z), 1\}
\end{aligned}
$$

$$
\begin{aligned}
&= 4\max\{4R(\varepsilon,\mathbf{\Gamma}_Z), 2U(\varepsilon,\widetilde{\varepsilon},\mathbf{\Gamma}_Z), T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z),1\}\\
&= 4\max\{4R(\varepsilon,\mathbf{\Gamma}_Z), 2T(\varepsilon,\max\{S^2(\mathbf{\Gamma}_Z),1\}\widetilde{\varepsilon},\mathbf{\Gamma}_Z), T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z),1\}\\
&= 4\max\{4R(\varepsilon,\mathbf{\Gamma}_Z), 2T(\varepsilon,\varepsilon^*,\mathbf{\Gamma}_Z), T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z),1\}\\
&= 4\max\{q_2(\varepsilon,\varepsilon^*,\mathbf{\Gamma}_Z), T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z),1\} \qquad \text{(A.15)}\\
&= 4\max\{q_1(\varepsilon,\varepsilon^*,\mathbf{\Gamma}_Z), q_2(\varepsilon,\varepsilon^*,\mathbf{\Gamma}_Z), T(0,\widetilde{\delta},\mathbf{\Gamma}_Z), T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z),1\} \qquad \text{(A.16)}\\
&= 4\max\{q_1(\varepsilon,\varepsilon^*,\mathbf{\Gamma}_Z), q_2(\varepsilon,\varepsilon^*,\mathbf{\Gamma}_Z), q_3(\widetilde{\delta},\mathbf{\Gamma}_Z), q_4(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\}\max\{S^2(\mathbf{\Gamma}_Z),1\}, \qquad \text{(A.17)}
\end{aligned}
$$

where (A.15) follows since $q_2(\varepsilon,\varepsilon^*,\mathbf{\Gamma}_Z) := \max\{4R(\varepsilon,\mathbf{\Gamma}_Z), 2T(\varepsilon,\varepsilon^*,\mathbf{\Gamma}_Z)\}$ as defined in Lemma C.2. The inequality (A.16) is due to the relation $q_2 = 2q_1$ and $T(0,\widetilde{\delta},\mathbf{\Gamma}_Z) \le T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)$. Finally, (A.17) follows since $q_3(\widetilde{\delta},\mathbf{\Gamma}_Z) = T(0,\widetilde{\delta},\mathbf{\Gamma}_Z)$ and $q_4(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z) = T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)$ as stated in Lemmas C.3 and C.4. Then, (A.12) can be inferred by applying Lemmas C.1 and C.2 with $\delta = \delta^*$ and $\widetilde{\delta} = \varepsilon^*$. The inequality (A.13) follows by Lemmas C.3 and C.4. Finally, note that by Lemma C.12,

$$
\begin{aligned}
&\mathrm{P}\left[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > S(\mathbf{\Gamma}_Z)\delta\right] \le \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \delta\wedge\varepsilon],\\
&\mathrm{P}\left[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > S^2(\mathbf{\Gamma}_Z)\delta\right] \le \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}^2 > \delta] + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \varepsilon].
\end{aligned}
$$

Hence, using (A.14), we can infer

$$
\begin{aligned}
&\mathrm{P}\left[\|\widehat{g}_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\widehat{\mathbf{\Gamma}}_X)\|_s > Q_1(\mathbf{\Gamma}_Z)\delta\right]\\
&\precsim \mathrm{P}\left[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > S(\mathbf{\Gamma}_Z)(\delta\wedge\widetilde{\varepsilon})\right] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > S^2(\mathbf{\Gamma}_Z)\delta]\\
&\qquad + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta\wedge\widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \varepsilon]\\
&\precsim \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta\wedge\widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta]\\
&\qquad + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \delta\wedge\widetilde{\varepsilon}\wedge\varepsilon] + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}^2 > \delta].
\end{aligned}
$$

□

Remark A.1. In light of our extensive use of second-order Taylor approximations applied to the link function ℓ and its inverse g, we pause here to discuss some differentiability issues. Note first that as a power series with absolutely summable coefficients, the function $\ell_{ij}(u)$ is differentiable infinitely many times for $u \in (-1,1)$. An expression for its first derivative is given in Proposition 2.1. The inverse function g_{ij} of ℓ_{ij} is defined on $(\ell_{ij}(-1), \ell_{ij}(1))$ and is differentiable infinitely many times on this interval, since the same holds for ℓ_{ij} on $(-1,1)$. As an example for differentiability requirements in the proof of our results, we discuss (A.10), where we applied a second-order Taylor approximation as

$$
g_\bullet(\widehat{\mathbf{\Gamma}}_X) = g_\bullet(\mathbf{\Gamma}_X) + g'_\bullet(\mathbf{\Gamma}_X)\odot(\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X) + \frac{1}{2}g''_\bullet(\Sigma)\odot(\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}
$$

for some Σ such that $|\Sigma - \mathbf{\Gamma}_X| < \left|\Sigma - \widehat{\mathbf{\Gamma}}_X\right|$. Strictly speaking, this requires twice differentiability of g on $|\Sigma - \mathbf{\Gamma}_X| < \left|\Sigma - \widehat{\mathbf{\Gamma}}_X\right|$. However, g_{ij} is twice differentiable only on $(\ell_{ij}(-1), \ell_{ij}(1))$ but the estimator $\widehat{\mathbf{\Gamma}}_X$ can certainly take values outside of this interval. We assume implicitly for now that $\widehat{\mathbf{\Gamma}}_X$ is close enough to the true $\mathbf{\Gamma}_X$ to ensure differentiability. In Lemmas C.3 and C.4, we will address this issue by intersecting with the event $\{\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s \le \widetilde{\delta}\}$, whenever we aim to bound the second-order terms.

Proof of Lemma A.2. By the second-order Taylor approximation of each component of g around $\mathbf{\Gamma}_X$ and subsequent application of the triangle inequality, we get, for some Σ such that $|\Sigma - \mathbf{\Gamma}_X| < \left|\Sigma - \widehat{\mathbf{\Gamma}}_X\right|$,

$$\begin{aligned}
&\left\|g_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\right\|_s \\
&\le \left\|g'_\bullet(\mathbf{\Gamma}_X) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s + \frac{1}{2}\left\|g''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s \\
&= \left\|(\ell')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s + \frac{1}{2}\left\|g''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s \qquad \text{(A.18)} \\
&\le R(0, \mathbf{\Gamma}_Z)\left\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\right\|_s + \frac{1}{2}\left\|g''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s . \qquad \text{(A.19)}
\end{aligned}$$

The equality (A.18) follows since $g'(\mathbf{\Gamma}_X) = (\ell^{-1})'(\ell(\mathbf{\Gamma}_Z)) = (\ell')^{\odot(-1)}(\mathbf{\Gamma}_Z)$; see Proposition 2.1 for an explicit representation of ℓ'. Furthermore, relation (A.19) follows by Lemma C.6. While the first term in (A.19) is already what appears in the bound of Lemma A.2, the second term needs to be considered further. We thus have

$$\begin{aligned}
&\mathrm{P}\left[\left\|g_\bullet(\widehat{\mathbf{\Gamma}}_X) - g_\bullet(\mathbf{\Gamma}_X)\right\|_s > Q_2(\mathbf{\Gamma}_Z)\delta\right] \\
&\le \mathrm{P}\left[R(0, \mathbf{\Gamma}_Z)\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > Q_2(\mathbf{\Gamma}_Z)\frac{\delta}{2}\right] \\
&\qquad + \mathrm{P}\left[\left\|g''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > Q_2(\mathbf{\Gamma}_Z)\delta\right] \\
&\le \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta\right] + \mathrm{P}\left[\left\|g''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s^2 > q_3(\widetilde{\delta}, \mathbf{\Gamma}_Z)\delta\right] \qquad \text{(A.20)} \\
&\le \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta\right] + \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta\right] + \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \widetilde{\delta}\right] \qquad \text{(A.21)} \\
&\precsim \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}\right] + \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta\right],
\end{aligned}$$

where (A.20) follows since

$$Q_2(\mathbf{\Gamma}_Z) = \max\{2R(0, \mathbf{\Gamma}_Z), T(0, \widetilde{\delta}, \mathbf{\Gamma}_Z)\} = \max\{2R(0, \mathbf{\Gamma}_Z), q_3(\widetilde{\delta}, \mathbf{\Gamma}_Z)\}$$

and (A.21) is a consequence of applying Lemma C.3. □

Appendix B: Case of latent VAR processes

B.1. Proofs of results in Section 4

In this section, we provide all proofs concerning the transition matrix estimation of the latent VAR(p) process.

The following proof of Proposition 4.1 is exactly the same as the proof of Proposition 4.1 in Basu and Michailidis (2015) and is only included for completeness. The actual contributions below are the proofs of Lemmas 4.1 and 4.2. Those lemmas show that the restricted eigenvalue condition and the deviation bound can be verified for the latent process based on the observed count series.

Proof of Proposition 4.1. Recall $\widehat{\Gamma} = I_d \otimes \widehat{\mathbf{\Gamma}}_Z$ from (4.8). Since $\widehat{\beta}$ minimizes the objective function, we get that, for all β,

$$-2\widehat{\beta}'\widehat{\gamma} + \widehat{\beta}'\widehat{\Gamma}\widehat{\beta} + \lambda_N\|\widehat{\beta}\|_1 \leq -2\beta_0'\widehat{\gamma} + \beta_0'\widehat{\Gamma}\beta_0 + \lambda_N\|\beta_0\|_1.$$

The above inequality reduces to

$$v'\widehat{\Gamma}v \leq 2v'(\widehat{\gamma} - \widehat{\Gamma}\beta_0) + \lambda_N(\|\beta_0\|_1 - \|\beta_0 + v\|_1),$$

where $v = \widehat{\beta} - \beta_0$. By the restricted eigenvalue condition

$$v'\widehat{\Gamma}v \geq \alpha\|v\|^2 - \tau(N,q)\|v\|_1^2 \geq (\alpha - 16s\tau(N,q))\|v\|^2 \geq \frac{\alpha}{2}\|v\|^2, \tag{B.1}$$

where the second last inequality is due to $\|v\|_1 \leq 4\|v_S\|_1 \leq 4\sqrt{s}\|v\|$ which follows by Cauchy-Schwarz inequality. The last inequality in (B.1) follows by the assumed relationship between $\tau(N,q)$ and α.

Set $S = \text{supp}\{\beta_0\}$, such that $\beta_{0,j} = 0$ for all $j \in S^c$, where S^c denotes the complement of S. We write $\beta_{0,S}$ for the corresponding non-zero entries of β_0. By (B.1), $v'\widehat{\Gamma}v \geq 0$ for all $v \in \mathbb{R}^q$. Then,

$$\begin{aligned}
0 \leq v'\widehat{\Gamma}v &\leq 2v'(\widehat{\gamma} - \widehat{\Gamma}\beta_0) + \lambda_N(\|\beta_0\|_1 - \|\beta_0 + v\|_1) \\
&\leq 2v'(\widehat{\gamma} - \widehat{\Gamma}\beta_0) + \lambda_N(\|\beta_{0,S}\|_1 - \|\beta_{0,S} + v_S\|_1 - \|\beta_{0,S^c} + v_{S^c}\|_1) \\
&\leq 2\|v\|_1\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} + \lambda_N(\|v_S\|_1 - \|v_{S^c}\|_1) \\
&\leq \frac{\lambda_N}{2}\|v\|_1 + \lambda_N(\|v_S\|_1 - \|v_{S^c}\|_1) \\
&\leq \frac{3\lambda_N}{2}\|v_S\|_1 - \frac{\lambda_N}{2}\|v_{S^c}\|_1
\end{aligned} \tag{B.2}$$

since $\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} \leq \mathcal{Q}(\beta_0)\sqrt{\frac{\log(q)}{N}}$ by the deviation bound and since we suppose that $\lambda_N \geq 4\mathcal{Q}(\beta_0)\sqrt{\frac{\log(q)}{N}}$. Then, (B.2) ensures that $\|v_{S^c}\|_1 \leq 3\|v_S\|_1$ and hence $\|v\|_1 \leq 4\|v_S\|_1 \leq 4\sqrt{s}\|v\|$.

Due to (B.1), the upper and lower bounds on (B.2) and (B.1) lead

$$\|v\| \leq 16\sqrt{s}\frac{\lambda_N}{\alpha}, \quad \|v\|_1 \leq 4\sqrt{s}\lambda_N\|v\| \leq 64s\frac{\lambda_N}{\alpha}, \quad v'\widehat{\Gamma}v \leq 2\lambda_N\|v\|_1 \leq 128s\frac{\lambda_N^2}{\alpha}.$$

□

Proof of Lemma 4.1. (*Restricted Eigenvalue condition*) By Lemma B.1 in the supplementary material of Basu and Michailidis (2015), we have

$$\widehat{\Gamma} \sim RE(\alpha, \tau) \quad \text{if} \quad \widehat{\mathbf{\Gamma}}_Z \sim RE(\alpha, \tau).$$

For this reason, it is sufficient to prove the restricted eigenvalue condition for $\widehat{\mathbf{\Gamma}}_Z$.

Under Assumptions C.1 and C.2, we apply Corollary 3.1 with $\delta = \widetilde{\delta} = \varepsilon = \widetilde{\varepsilon} = \nu$ so that

$$\begin{aligned}
&\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > Q(\mathbf{\Gamma}_Z)c_0(s)\nu\right] \\
&\leq c_{1,1}\exp\left(-c_{1,2}N\min\{1,\nu^2\} + 2s\log(dp)\right) + c_{2,1}dK\exp\left(-c_{2,2}T\min\{1,\nu^2\}\right) \\
&\leq c_1\exp\left(-c_{1,2}N\min\{1,\nu^2\} + 2s\log(dp) - c_{2,2}T\min\{1,\nu^2\} + \log(dK)\right) \\
&\leq c_1\exp\left(-c_2N\min\{1,\nu^2\}\right)
\end{aligned}$$

since $N = T - p \leq T$ and $N \gtrsim \max\{c_0(s)/\nu, \nu^{-2}, 1\}\max\{s\log(dp), \log(dK)\}$. Then,

$$\left|v'(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)v\right| \leq Q(\mathbf{\Gamma}_Z)c_0(s)\nu = \frac{\lambda_{\min}(\Sigma_\varepsilon)}{54\mu_{\max}(\mathcal{A})} \quad \text{for} \quad v \in \mathcal{K}(2s) \tag{B.3}$$

with probability at least $1 - c_1\exp\left(-c_2N\min\{1,\nu^2,\nu/c_0(s)\}\right)$ and choosing

$$\nu = \frac{\lambda_{\min}(\Sigma_\varepsilon)}{54\mu_{\max}(\mathcal{A})Q(\mathbf{\Gamma}_Z)c_0(s)}.$$

From (B.3) and by applying Lemma 12 in the supplement of Loh and Wainwright (2012) we infer

$$v'\widehat{\mathbf{\Gamma}}_Z v \geq v'\mathbf{\Gamma}_Z v - \frac{\lambda_{\min}(\Sigma_\varepsilon)}{2\mu_{\max}(\mathcal{A})}(\|v\|_2^2 + \frac{1}{s}\|v\|_1^2).$$

Then,

$$v'\widehat{\mathbf{\Gamma}}_Z v \geq \frac{\lambda_{\min}(\Sigma_\varepsilon)}{2\mu_{\max}(\mathcal{A})}\|v\|_2^2 - \frac{\lambda_{\min}(\Sigma_\varepsilon)}{2\mu_{\max}(\mathcal{A})}\frac{1}{s}\|v\|_1^2 \tag{B.4}$$

$$\geq \alpha\|v\|_2^2 - \alpha\frac{1}{N}\max\{\nu^{-2},1\}4\log(dp)\|v\|_1^2 = \alpha\|v\|_2^2 - \tau\|v\|_1^2. \tag{B.5}$$

The bound (B.4) follows since $v'\mathbf{\Gamma}_Z v \geq \frac{\lambda_{\min}(\Sigma_\varepsilon)}{\mu_{\max}(\mathcal{A})}\|v\|_2^2$ by Proposition 2.3 and relation (4.1) in Basu and Michailidis (2015). With $s = \lceil N\min\{1,\nu^2,\nu/c_0(s)\}/4\log(dp)\rceil$ and α and τ in (4.11), we infer (B.5). □

Proof of Lemma 4.2. (Deviation bound) Note that

$$\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} \leq \|(\widehat{\Gamma} - \Gamma)\beta_0\|_{\max} + \|\widehat{\gamma} - \gamma\|_{\max}, \tag{B.6}$$

since $\gamma - \Gamma\beta_0 = 0$, where γ and Γ are the population quantities of $\widehat{\gamma}$ and $\widehat{\Gamma}$. Let $e_{q,i}$ denote the ith basis vector of $\mathbb{R}^q$. Then, considering both summands in (B.6) separately, we get for the first summand,

$$\begin{aligned}
\|(\widehat{\Gamma} - \Gamma)\beta_0\|_{\max} &= \max_{i=1,\ldots,q} |e'_{q,i}(I_d \otimes \widehat{\mathbf{\Gamma}}_Z - \Gamma)\beta_0| \\
&= \max_{i=1,\ldots,q} |e'_{q,i}(I_d \otimes (\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z))\beta_0| \\
&= \max_{i=1,\ldots,q} |e'_{q,i} \operatorname{vec}((\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)B_0)| \qquad \text{(B.7)} \\
&= \max_{i=1,\ldots,pd;j=1,\ldots,d} |e'_{pd,i}(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)B_0 e_{d,j}|, \qquad \text{(B.8)}
\end{aligned}$$

where $\beta_0 = \operatorname{vec}(B_0)$ with B_0 in (4.4) and (B.7) is due to Theorem 2, Section 4 in Magnus and Neudecker (2007). Note that $\|B_0 e_{d,j}\| < 1$ since $\|B_0 e_{d,j}\| = \|B_0 e_{d,j}\|/\|e_{d,j}\| \le \sup_v \|B_0 v\|/\|v\| < 1$. By using (B.8) and with further explanations given below, for $\mu_1 = 3Q(\mathbf{\Gamma}_Z)c_0(s)\delta$ with $v_j = B_0 e_{d,j}$,

$$\begin{aligned}
&\mathrm{P}\left[\|(\widehat{\Gamma} - \Gamma)\beta_0\|_{\max} > \mu_1\right] \\
&= \mathrm{P}\left[\max_{i=1,\ldots,pd;j=1,\ldots,d} |e'_{pd,i}(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)B_0 e_{d,j}| > \mu_1\right] \\
&\le \sum_{i,j=1}^{pd} \mathrm{P}\left[|e'_{pd,i}(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)v_j| > \mu_1\right] \qquad \text{(B.9)} \\
&\le \sum_{i,j=1}^{pd} \Bigg(\mathrm{P}\left[|e'_{pd,i}(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)e_{pd,i}| > \frac{2}{3}\mu_1\right] + \mathrm{P}\left[|v'_j(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)v_j| > \frac{2}{3}\mu_1\right] \\
&\qquad + \mathrm{P}\left[|(e_{pd,i} + v_j)'(\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z)(e_{pd,i} + v_j)| > \frac{2}{3}\mu_1\right]\Bigg) \qquad \text{(B.10)} \\
&\le (pd)^2 3\,\mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z - \mathbf{\Gamma}_Z\|_s > \frac{1}{3}\mu_1\right] \\
&\le c_{1,1} \exp\left(-c_{1,2} N \min\left\{1, \nu^2, \nu/c_0(s)\right\} + 2\log(pd)\right). \qquad \text{(B.11)}
\end{aligned}$$

We used a union bound to infer (B.9). The relation $2|v'Aw| \le |v'Av| + |w'Aw| + |(v+w)'A(v+w)|$ implies (B.10). Under Assumptions C.1 and C.2, Corollary 3.1 and (3.3) with $N \gtrsim \max\{c_0(s)\nu^{-1}, \nu^{-2}, 1\}\max\{s\log(Ld), \log(dK)\}$ and $N = T - p \le T$ give (B.11).

For the second summand in (B.6), recall from (4.8) that $\gamma = \operatorname{vec}(\boldsymbol{\gamma}_Z)$ with $\boldsymbol{\gamma}_Z = (\Gamma_Z(1)', \ldots, \Gamma_Z(p)')'$. We further introduce $\mathbf{\Gamma}_Z^{p+1} = (\Gamma_Z(r-s))_{r,s=1,\ldots,(p+1)}$ and $\widehat{\mathbf{\Gamma}}_Z^{p+1} = (\widehat{\Gamma}_Z(r-s))_{r,s=1,\ldots,(p+1)}$.

$$\begin{aligned}
&\|\widehat{\gamma} - \gamma\|_{\max} \\
&= \|\operatorname{vec}(\widehat{\boldsymbol{\gamma}}_Z) - \operatorname{vec}(\boldsymbol{\gamma}_Z)\|_{\max} \\
&= \|(e'_{(p+1),1} \otimes I_d)(\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1})((e_{(p+1)d,2}, \ldots, e_{(p+1),(p+1)}) \otimes I_d)\|_{\max}
\end{aligned}$$

$$
\begin{aligned}
&= \max_{i=1,\dots,d;j=1,\dots,dp} |e'_{d,i}(e'_{(p+1),1} \otimes I_d)(\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1}) \\
&\qquad\qquad \times ((e_{(p+1)d,2}, \dots, e_{(p+1),(p+1)}) \otimes I_d) e_{dp,j}| \\
&= \max_{i=1,\dots,d;j=1,\dots,dp} |w'_{1,i}(\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1}) w_{2,j}|, \qquad \text{(B.12)}
\end{aligned}
$$

where $w'_{1,i} = e'_{d,i}(e'_{(p+1),1} \otimes I_d)$ and $w_{2,j} = ((e_{(p+1)d,2}, \dots, e_{(p+1),(p+1)}) \otimes I_d) e_{dp,j}$. In particular, using (B.12) and with further explanations given below, for $\mu_2 = 3Q(\mathbf{\Gamma}_Z)c_0(s)\delta$,

$$
\begin{aligned}
&\mathrm{P}\left[\|\widehat{\gamma} - \gamma\|_{\max} > \mu_2\right] \\
&= \mathrm{P}\left[\max_{i=1,\dots,d;j=1,\dots,dp} |w'_{1,i}(\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1}) w_{2,j}| > \mu_2\right] \\
&\le \sum_{i,j=1}^{dp} \mathrm{P}\left[|w'_{1,i}(\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1}) w_{2,j}| > \mu_2\right] \qquad \text{(B.13)} \\
&\le \sum_{i,j=1}^{dp} \Bigg(\mathrm{P}\left[|w'_{1,i}(\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1}) w_{1,i}| > \frac{2}{3}\mu_2\right] \\
&\qquad + \mathrm{P}\left[|w'_{2,j}(\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1}) w_{2,j}| > \frac{2}{3}\mu_2\right] \\
&\qquad + \mathrm{P}\left[|(w_{1,i} + w_{2,j})'(\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1})(w_{1,i} + w_{2,j})| > \frac{2}{3}\mu_2\right]\Bigg) \qquad \text{(B.14)} \\
&\le 3(dp)^2\, \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_Z^{p+1} - \mathbf{\Gamma}_Z^{p+1}\|_s > \frac{1}{3}\mu_2\right] \\
&\le c_1 \exp\left(-c_2 N \min\{1, \nu^2, \nu/c_0(s)\} + 2\log(pd)\right). \qquad \text{(B.15)}
\end{aligned}
$$

We used a union bound to infer (B.13). The relation $2|v'Aw| \le |v'Av| + |w'Aw| + |(v+w)'A(v+w)|$ implies (B.14). Under Assumptions C.1 and C.2, Corollary 3.1 and (3.3) with $N \gtrsim \max\{c_0(s)\nu^{-1}, \nu^{-2}, 1\} \max\{s\log(Ld), \log(dK)\}$ gives (B.15). Note that $N = T - p \le T$ and $\log((p+1)d) = \log(1 + 1/p) + \log(pd) \le 2\log(pd)$.

Combining (B.11) and (B.15), there are constants $c_1, c_2 > 0$ such that

$$
\mathrm{P}\left[\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} > 6Q(\mathbf{\Gamma}_Z)c_0(s)\nu\right] \le c_1 \exp\left(-c_2 N \min\{1, \nu^2, \nu/c_0(s)\}\right).
$$

Choosing $\nu = \sqrt{\frac{\log(q)}{N}}$, we get

$$
\|\widehat{\gamma} - \widehat{\Gamma}\beta_0\|_{\max} \le 6Q(\mathbf{\Gamma}_Z)c_0(s)\sqrt{\frac{\log(q)}{N}}
$$

with high probability. □

B.2. Verification of Assumptions C.1 and C.2

In this section, we verify Assumptions C.1 and C.2 in certain cases. More specifically, we show that Assumption C.1 is satisfied with $c_0(s) = s$ and for VAR(p) models whenever the function G is bounded. We refer to Section 4.3 for discussions on $c_0(s) = s$ and boundedness of G. We prove that Assumption C.2 is satisfied whenever the unknown CDF parameters θ_i can be estimated through the mean of the observed process. Examples include Bernoulli, binomial and negative hypergeometric distributions.

Write a VAR(p) model as a pd-dimensional VAR(1) model, that is,

$$\begin{pmatrix} Z_t \\ Z_{t-1} \\ \vdots \\ Z_{t-p+1} \end{pmatrix} = \begin{pmatrix} \Psi_1 & \cdots & \Psi_{p-1} & \Psi_p \\ I_d & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_d & 0 \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ \vdots \\ Z_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{or} \quad Y_t = AY_{t-1} + \widetilde{\varepsilon}_t. \tag{B.16}$$

A VAR(1) model in (B.16) is known to satisfy the Markov property and is also geometrically ergodic under assumption (4.3); see p. 944 in An and Huang (1996). Under geometric ergodicity, Theorem 2.1 in Roberts and Rosenthal (1997) implies that there is a spectral gap λ with $1 - \lambda > 0$.

For the verification of both Assumptions C.1 and C.2, we will use the following concentration inequality for bounded functions of general-state-space Markov chains derived in Fan, Jiang and Sun (2021). The result is expressed in terms of λ_r, the rightmost value of the spectrum $[-\lambda, \lambda]$. We refer to $1 - \lambda_r$ as the right spectral gap of the Markov chain.

Theorem B.1 (Theorem 3 in Fan, Jiang and Sun (2021))**.** *Let $\{Y_t\}_{t\geq 1}$ be a Markov chain on $\mathcal{X}$ with right spectral gap $1 - \lambda_r > 0$. For any $\varepsilon > 0$ and bounded function $f : \mathcal{X} \to [a, b]$,*

$$\mathrm{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T} f(Y_t) - \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}[f(Y_t)]\right| > \varepsilon\right] \leq 2\exp\left(-\frac{1 - \max\{0, \lambda_r\}}{1 + \max\{0, \lambda_r\}} \frac{T\varepsilon^2}{(b-a)^2/2}\right).$$

We start with the verification of Assumption C.2 since it is slightly simpler.

Verification of Assumption C.2: First, consider the expected value and note that

$$\mathrm{E}[\widehat{\theta}_i] = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} X_{i,t}\right] = \theta_i.$$

Then,

$$\begin{aligned} \mathrm{P}[\max_{i=1,\ldots,d} |\widehat{\theta}_i - \theta_i| > \varepsilon] &= \mathrm{P}[\max_{i=1,\ldots,d} |\widehat{\theta}_i - \mathrm{E}[\widehat{\theta}_i]| > \varepsilon] \\ &\leq d\,\mathrm{P}[|\widehat{\theta}_i - \mathrm{E}[\widehat{\theta}_i]| > \varepsilon] \\ &\leq d\,\mathrm{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T} X_{i,t} - \mathrm{E}[X_{i,t}]\right| > \varepsilon\right] \end{aligned}$$

$$= d\,\mathrm{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T} f(Y_t) - \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}[f(Y_t)]\right| > \varepsilon\right] \tag{B.17}$$

$$\leq 2d\exp\left(-\frac{1-\max\{0,\lambda_r\}}{1+\max\{0,\lambda_r\}}\frac{T\varepsilon^2}{2b^2}\right), \tag{B.18}$$

where (B.17) follows by choosing the function f as $f : y \mapsto (e_{p,i} \otimes e_{d,i})'G(y)$. Then, it remains to verify that f is bounded which follows since $f(y) = (e_{p,i} \otimes e_{d,i})'G(y) \leq b(e_{p,i} \otimes e_{d,i})'j_{dp} = b$, where j_d denotes a dp-dimensional column vector with all entries equal to one. Finally, (B.18) is a consequence of applying Theorem B.1.

Verification of Assumption C.1: We consider here the centered random variables $\widetilde{X}_t = X_t - \mathrm{E}\,X_t$ to estimate $\mathbf{\Gamma}_X$. First, note that the expected value of $\widehat{\mathbf{\Gamma}}_X$ can be calculated as

$$\begin{aligned}
&\mathrm{E}\,\widehat{\mathbf{\Gamma}}_X \\
&= \frac{1}{N}\,\mathrm{E}\,\mathcal{X}_X'\mathcal{X}_X = \left(\frac{1}{N}\,\mathrm{E}\sum_{t=p}^{T-1}\widetilde{X}_{t-r+1}\widetilde{X}_{t-s+1}'\right)_{r,s=1,\ldots,p} \\
&= \left(\frac{1}{N}\sum_{t=p}^{T-1}(\mathrm{E}(G(Z_{t-r+1})G(Z_{t-s+1})') - \mathrm{E}\,G(Z_{t-r+1})\,\mathrm{E}\,G(Z_{t-s+1})')\right)_{r,s=1,\ldots,p} \\
&= \mathbf{\Gamma}_X,
\end{aligned} \tag{B.19}$$

We now aim to apply Theorem B.1. Write

$$\begin{aligned}
\mathrm{P}[|v'(\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)v| > s\delta] &= \mathrm{P}[|v'(\widehat{\mathbf{\Gamma}}_X - \mathrm{E}\,\widehat{\mathbf{\Gamma}}_X)v| > s\delta] \\
&= \mathrm{P}\left[\left|v'\Big(\frac{1}{N}\mathcal{X}_X'\mathcal{X}_X - \mathrm{E}\,\frac{1}{N}\mathcal{X}_X'\mathcal{X}_X\Big)v\right| > s\delta\right] \\
&= \mathrm{P}\left[\left|\frac{1}{N}\sum_{t=p}^{T-1} f(Y_t) - \frac{1}{N}\sum_{t=p}^{T-1}\mathrm{E}[f(Y_t)]\right| > s\delta\right] \qquad \text{(B.20)} \\
&\leq 2\exp\left(-\frac{1-\max\{0,\lambda_r\}}{1+\max\{0,\lambda_r\}}\frac{N\delta^2}{8b^4}\right). \qquad \text{(B.21)}
\end{aligned}$$

The function f in (B.20) is characterized below and satisfies $|f(y)| \leq b^2 4s$. Finally, (B.21) is a consequence of applying Theorem B.1.

To find the function f in (B.20), set $v = \mathrm{vec}([v_1 : \cdots : v_p])$ with $v_r \in \mathbb{R}^d$ and $\widetilde{G}(Y_t) = (G(Z_t)' - \mathrm{E}\,G(Z_t)', \ldots, G(Z_{t-p+1})' - \mathrm{E}\,G(Z_{t-p+1})')'$. Then, the function f in (B.20) can be determined through the following calculations:

$$\begin{aligned}
&v'\mathcal{X}_X'\mathcal{X}_X v \\
&= v'[\widetilde{G}(Y_p) : \widetilde{G}(Y_{p+1}) : \cdots : \widetilde{G}(Y_{T-1})][\widetilde{G}(Y_p) : \widetilde{G}(Y_{p+1}) : \cdots : \widetilde{G}(Y_{T-1})]'v \\
&= v'\left(\sum_{t=p}^{T-1}(e_{p,r}' \otimes I_d)\widetilde{G}(Y_t)\widetilde{G}(Y_t)'(e_{p,s}' \otimes I_d)'\right)_{r,s=1,\ldots,p} v
\end{aligned}$$

$$= \sum_{t=p}^{T-1} \sum_{r,s=1}^{p} v_r'(e_{p,r}' \otimes I_d)\widetilde{G}(Y_t)\widetilde{G}(Y_t)'(e_{p,s}' \otimes I_d)'v_s = \sum_{t=p}^{T-1} f(Y_t)$$

with $f : y \mapsto \sum_{r,s=1}^{p} v_r'(e_{p,r}' \otimes I_d)\widetilde{G}(y)\widetilde{G}(y)'(e_{p,s}' \otimes I_d)'v_s$. Then, it remains to verify that f is bounded. Denote J_d as a $d \times d$-matrix with all entries equal to one and j_d as a d-dimensional column vector with all entries equal to one. Then, with explanations given below,

$$\begin{aligned}
|f(y)| &= \left| \sum_{r,s=1}^{p} v_r'(e_{p,r}' \otimes I_d)\widetilde{G}(y)\widetilde{G}(y)'(e_{p,s}' \otimes I_d)'v_s \right| \\
&\leq 4b^2 \sum_{r,s=1}^{p} |v_r'|(e_{p,r}' \otimes I_d)J_{dp}(e_{p,s}' \otimes I_d)'|v_s| \qquad \text{(B.22)} \\
&= 4b^2 \sum_{r,s=1}^{p} |v_r'|(e_{p,r}' \otimes I_d)(J_p \otimes J_d)(e_{p,s} \otimes I_d)|v_s| \\
&= 4b^2 \sum_{r,s=1}^{p} |v_r'|(e_{p,r}' \otimes I_d)(j_p \otimes J_d)|v_s| \qquad \text{(B.23)} \\
&= 4b^2 \sum_{r,s=1}^{p} |v_r'|(1 \otimes J_d)|v_s| \qquad \text{(B.24)} \\
&= 4b^2 \sum_{r,s=1}^{p} \sum_{i=1}^{d} |v_{r,i}| \sum_{j=1}^{d} |v_{s,j}| \\
&\leq 2b^2 \sum_{r,s=1}^{p} \sum_{i,j=1}^{d} (v_{r,i}^2 + v_{s,j}^2) = b^2 4s, \qquad \text{(B.25)}
\end{aligned}$$

where (B.22) follows since we assume that $|G_i(y_i) - \mathrm{E}\, G_i(y_i)| \leq 2b$, (B.23) and (B.24) use a Kronecker product property in equation (4) on page 32 in Magnus and Neudecker (2007); and for (B.25) note that $\|v\| = 1$.

Appendix C: Technical lemmas and their proofs

Our technical lemmas required to prove our main results are separated into results on the inverse link function and its derivatives (Section C.1), the reciprocal of the first derivative of the link function (Section C.2) and bounds on the link function itself and its derivatives (Section C.3). Finally, we consider the diagonal elements separately (Section C.4).

C.1. Inverse link function and its derivatives

This section provides high probability bounds for expressions of the form

$$\left\| (\widehat{g}_{\bullet}^{(a)}(\mathbf{\Sigma}) - g_{\bullet}^{(a)}(\mathbf{\Sigma})) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^a \right\|_s, \quad a \in \{0, 1, 2\}, \qquad \text{(C.1)}$$

where the function g is the inverse of the link function ℓ, $g_\bullet$ is defined in (A.1) and $\mathbf{\Sigma}$ is such that $\mathbf{\Sigma} = \mathbf{\Gamma}_X$ or $|\mathbf{\Sigma} - \mathbf{\Gamma}_X| < \left|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\right|$. Note that $g^{(a)}$ diverges from our previous notation and denotes the function g itself ($a = 0$) and its first and second derivatives ($a = 1, 2$). The following lemma covers the case $a = 0$ in (C.1).

Lemma C.1. *Suppose Assumptions M.1–M.4. Then, for any $\delta, \widetilde{\delta}, \varepsilon > 0$,*

$$\begin{aligned}
&\mathrm{P}[\|\widehat{g}_\bullet(\mathbf{\Gamma}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > q_1(\mathbf{\Gamma}_Z)\delta] \\
&\lesssim \mathrm{P}\left[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > \delta \wedge \widetilde{\delta}\right] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon]
\end{aligned}$$

with $q_1(\mathbf{\Gamma}_Z) := q_1(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z) = \max\{2R(\varepsilon, \mathbf{\Gamma}_Z), T(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z)\}$.

Proof. Noting that $g(\mathbf{\Gamma}_X) = \ell^{-1}(\ell(\mathbf{\Gamma}_Z)) = \mathbf{\Gamma}_Z = \widehat{\ell}^{-1}(\widehat{\ell}(\mathbf{\Gamma}_Z))$, we get

$$\begin{aligned}
&\|\widehat{g}_\bullet(\mathbf{\Gamma}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s \\
&= \left\|\widehat{\ell}_\bullet^{-1}(\ell(\mathbf{\Gamma}_Z)) - \ell_\bullet^{-1}(\ell(\mathbf{\Gamma}_Z))\right\|_s = \left\|\widehat{\ell}_\bullet^{-1}(\ell(\mathbf{\Gamma}_Z)) - \widehat{\ell}_\bullet^{-1}(\widehat{\ell}(\mathbf{\Gamma}_Z))\right\|_s \\
&\leq \left\|(\widehat{\ell}_\bullet^{-1})'(\widehat{\ell}(\mathbf{\Gamma}_Z)) \odot (\ell_\bullet(\mathbf{\Gamma}_Z) - \widehat{\ell}_\bullet(\mathbf{\Gamma}_Z))\right\|_s \\
&\qquad + \frac{1}{2}\left\|(\widehat{\ell}_\bullet^{-1})''(\Sigma) \odot (\ell_\bullet(\mathbf{\Gamma}_Z) - \widehat{\ell}_\bullet(\mathbf{\Gamma}_Z))^{\odot 2}\right\|_s \qquad \text{(C.2)} \\
&= \left\|(\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\ell_\bullet(\mathbf{\Gamma}_Z) - \widehat{\ell}_\bullet(\mathbf{\Gamma}_Z))\right\|_s \\
&\qquad + \frac{1}{2}\left\|(\widehat{\ell}_\bullet^{-1})''(\Sigma) \odot (\ell_\bullet(\mathbf{\Gamma}_Z) - \widehat{\ell}_\bullet(\mathbf{\Gamma}_Z))^{\odot 2}\right\|_s, \qquad \text{(C.3)}
\end{aligned}$$

where (C.2) follows by the second-order Taylor expansion of $\widehat{\ell}^{-1}$ around $\widehat{\ell}(\mathbf{\Gamma}_Z)$ for some Σ such that $|\Sigma - \ell(\mathbf{\Gamma}_Z)| < \left|\widehat{\ell}(\mathbf{\Gamma}_Z) - \ell(\mathbf{\Gamma}_Z)\right|$. For the equality (C.3), we use (D.46). We further bound the probabilities of the two terms in (C.3) separately.

$$\begin{aligned}
&\mathrm{P}\left[\|\widehat{g}_\bullet(\mathbf{\Gamma}_X) - g_\bullet(\mathbf{\Gamma}_X)\|_s > q_1(\mathbf{\Gamma}_Z)\delta\right] \\
&\leq \mathrm{P}\left[\left\|(\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z))\right\|_s > q_1(\mathbf{\Gamma}_Z)\frac{\delta}{2}\right] \\
&\qquad + \mathrm{P}\left[\frac{1}{2}\left\|\widehat{g}_\bullet''(\Sigma) \odot (\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z))^{\odot 2}\right\|_s > q_1(\mathbf{\Gamma}_Z)\frac{\delta}{2}\right] \\
&\leq \mathrm{P}\left[\left\|(\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z))\right\|_s > R(\varepsilon, \mathbf{\Gamma}_Z)\delta\right] \\
&\qquad + \mathrm{P}\left[\left\|\widehat{g}_\bullet''(\Sigma) \odot (\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z))^{\odot 2}\right\|_s > T(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z)\delta\right] \qquad \text{(C.4)} \\
&\lesssim \mathrm{P}\left[\left\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\right\|_s > \delta \wedge \widetilde{\delta}\right] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon], \qquad \text{(C.5)}
\end{aligned}$$

where (C.4) follows since $q_1(\mathbf{\Gamma}_Z) = \max\{2R(\varepsilon, \mathbf{\Gamma}_Z), T(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z)\}$, and (C.5) follows from Lemmas C.11 and C.5 since $R(\mathbf{\Gamma}_Z) := R(\varepsilon, \mathbf{\Gamma}_Z)$ and $q_4(\mathbf{\Gamma}_Z) = T(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z)$. □

The following lemma concerns the case $a = 1$ in (C.1).

Lemma C.2. *Suppose Assumptions M.1–M.4. Then, for any* $\delta, \widetilde{\delta}, \varepsilon > 0$,

$$\begin{aligned}
&\mathrm{P}\left[\left\|(\widehat{g}'_\bullet(\mathbf{\Gamma}_X) - g'_\bullet(\mathbf{\Gamma}_X)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s > q_2(\mathbf{\Gamma}_Z)\delta\right] \\
&\lesssim \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta \wedge \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s > \widetilde{\delta}] + \mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon]
\end{aligned}$$

with $q_2(\mathbf{\Gamma}_Z) := q_2(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z) = \max\{4R(\varepsilon, \mathbf{\Gamma}_Z), 2T(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z)\}$.

Proof. With explanations given below, we bound the quantity of interest as follows:

$$\begin{aligned}
&\left\|(\widehat{g}'_\bullet(\mathbf{\Gamma}_X) - g'_\bullet(\mathbf{\Gamma}_X)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \\
&= \left\|(\widehat{g}'_\bullet(\ell(\mathbf{\Gamma}_Z)) - g'_\bullet(\ell(\mathbf{\Gamma}_Z))) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \\
&\le \left\|(\widehat{g}'_\bullet(\ell(\mathbf{\Gamma}_Z)) - \widehat{g}'_\bullet(\widehat{\ell}(\mathbf{\Gamma}_Z))) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \\
&\qquad + \left\|(\widehat{g}'_\bullet(\widehat{\ell}(\mathbf{\Gamma}_Z)) - g'_\bullet(\ell(\mathbf{\Gamma}_Z))) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \\
&= \left\|\widehat{g}''_\bullet(\Sigma) \odot (\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \\
&\qquad + \left\|\left((\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) - (\ell')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z)\right) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \qquad \text{(C.6)} \\
&\le \frac{1}{2}\left\||\widehat{g}''_\bullet(\Sigma)| \odot (\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z))^{\odot 2}\right\|_s + \frac{1}{2}\left\||\widehat{g}''_\bullet(\Sigma)| \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s \\
&\qquad + \left\|(\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s + \left\|(\ell')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s, \qquad \text{(C.7)}
\end{aligned}$$

where (C.6) follows by the mean value theorem for some Σ such that $|\Sigma - \ell(\mathbf{\Gamma}_Z)| < \left|\widehat{\ell}(\mathbf{\Gamma}_Z) - \ell(\mathbf{\Gamma}_Z)\right|$ and the last line (C.7) is a consequence of $|A \odot B| \le \frac{1}{2}(A^{\odot 2} + B^{\odot 2})$ and (D.18) in Lemma D.2. The four summands in (C.7) can be handled through subsequent Lemmas as follows

$$\begin{aligned}
&\mathrm{P}\left[\left\|(\widehat{g}'_\bullet(\mathbf{\Gamma}_X) - g'_\bullet(\mathbf{\Gamma}_X)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s > q_2(\mathbf{\Gamma}_Z)\delta\right] \\
&\le \mathrm{P}\left[\left\||\widehat{g}''_\bullet(\Sigma)| \odot (\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z))^{\odot 2}\right\|_s > q_2(\mathbf{\Gamma}_Z)\delta/2\right] \\
&\qquad + \mathrm{P}\left[\left\||\widehat{g}''_\bullet(\Sigma)| \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_2(\mathbf{\Gamma}_Z)\delta/2\right] \\
&\qquad\qquad + \mathrm{P}\left[\left\|(\ell')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s > q_2(\mathbf{\Gamma}_Z)\delta/4\right]
\end{aligned}$$

$$
\begin{aligned}
&\qquad + \mathrm{P}\left[\left\|(\widehat{\ell}')_{\bullet}^{\odot(-1)}(\mathbf{\Gamma}_Z)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s > q_2(\mathbf{\Gamma}_Z)\delta/4\right]\\
&\le \mathrm{P}\left[\left\||\widehat{g}''_{\bullet}(\Sigma)|\odot(\widehat{\ell}_{\bullet}(\mathbf{\Gamma}_Z)-\ell_{\bullet}(\mathbf{\Gamma}_Z))^{\odot 2}\right\|_s > T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\delta\right]\\
&\quad + \mathrm{P}\left[\left\||\widehat{g}''_{\bullet}(\Sigma)|\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)^{\odot 2}\right\|_s > T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\delta\right]\\
&\quad + \mathrm{P}\left[\left\|(\widehat{\ell}')_{\bullet}^{\odot(-1)}(\mathbf{\Gamma}_Z)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s > R(\varepsilon,\mathbf{\Gamma}_Z)\delta\right]\\
&\quad + \mathrm{P}\left[\left\|(\ell')_{\bullet}^{\odot(-1)}(\mathbf{\Gamma}_Z)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s > R(0,\mathbf{\Gamma}_Z)\delta\right] \qquad \text{(C.8)}\\
&\lesssim \mathrm{P}[\|\widehat{\ell}_{\bullet}(\mathbf{\Gamma}_Z)-\ell_{\bullet}(\mathbf{\Gamma}_Z)\|_s^2>\delta] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>\delta]\\
&\quad + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s>\widetilde{\delta}] + \mathrm{P}[\|\widehat{\ell}_{\bullet}(\mathbf{\Gamma}_Z)-\ell_{\bullet}(\mathbf{\Gamma}_Z)\|_s>\widetilde{\delta}]\\
&\quad + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s>\delta] + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon] \qquad \text{(C.9)}\\
&\lesssim \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s>\delta\wedge\widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>\delta)\\
&\quad + \mathrm{P}[\|\widehat{\ell}_{\bullet}(\mathbf{\Gamma}_Z)-\ell_{\bullet}(\mathbf{\Gamma}_Z)\|_s>\widetilde{\delta}] + \mathrm{P}[\|\widehat{\ell}_{\bullet}(\mathbf{\Gamma}_Z)-\ell_{\bullet}(\mathbf{\Gamma}_Z)\|_s^2>\delta]\\
&\quad + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon]
\end{aligned}
$$

where (C.8) follows since

$$
\begin{aligned}
q_2(\mathbf{\Gamma}_Z) &= \max\{4R(\varepsilon,\mathbf{\Gamma}_Z), 2T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\}\\
&= \max\{4R(0,\mathbf{\Gamma}_Z), 4R(\varepsilon,\mathbf{\Gamma}_Z), 2T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)\}
\end{aligned}
$$

with R, T as in (A.3). The relation (C.9) follows from Lemmas C.5 C.4, C.10 and C.6 since $R(\mathbf{\Gamma}_Z) := R(\varepsilon,\mathbf{\Gamma}_Z)$ and $q_4(\mathbf{\Gamma}_Z) = T(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)$. □

We proceed with finding bounds on (C.1) with $a = 2$ and reduce the problem to expressions of the form

$$
\left\|g''_{\bullet}(\Sigma)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)^{\odot 2}\right\|_s.
$$

In particular, we distinguish the cases when g is known (Lemma C.3) and when g is estimated (Lemma C.4).

Lemma C.3. *Suppose Assumptions M.1–M.4 and let Σ be such that $|\Sigma-\mathbf{\Gamma}_X| < \left|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\right|$. Then, for any $\delta,\widetilde{\delta}>0$,*

$$
\begin{aligned}
\mathrm{P}\left[\left\|g''_{\bullet}(\Sigma)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_3(\mathbf{\Gamma}_Z)\delta\right] &\le \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>\delta\right]\\
&\quad + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s>\widetilde{\delta}]
\end{aligned}
$$

with

$$
q_3(\mathbf{\Gamma}_Z) := q_3(\widetilde{\delta},\mathbf{\Gamma}_Z) = \frac{6}{1-\boldsymbol{c}(\widetilde{\delta})^2}M_1(\boldsymbol{c}(\widetilde{\delta}),0)M_2(\boldsymbol{c}(\widetilde{\delta}),0),
$$

where M_1, M_2 are as in (2.15).

Proof. With explanations given below, we get

$$
\begin{aligned}
&\mathrm{P}\left[\left\|g''_{\bullet}(\Sigma)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_3(\mathbf{\Gamma}_Z)\delta\right] \\
&= \mathrm{P}\Bigg[\left\{\left\|g''_{\bullet}(\Sigma)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_3(\mathbf{\Gamma}_Z)\delta\right\} \\
&\qquad \cap\left(\{\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s\le\widetilde{\delta}\}\cup\{\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s>\widetilde{\delta}\}\right)\Bigg] \\
&\le \mathrm{P}\left[\left\{\left\|g''_{\bullet}(\Sigma)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_3(\mathbf{\Gamma}_Z)\delta\right\}\cap\{\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_{\max}\le 4\widetilde{\delta}\}\right] \\
&\qquad + \mathrm{P}\left[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s>\widetilde{\delta}\right] \qquad \text{(C.10)} \\
&\le \mathrm{P}\left[\sup_{\sigma\in\Omega(\widetilde{\delta})}|g''_{\bullet}(\sigma)|\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>q_3(\mathbf{\Gamma}_Z)\delta\right]+\mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s>\widetilde{\delta}] \qquad \text{(C.11)} \\
&\le \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>\delta]+\mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s>\widetilde{\delta}]. \qquad \text{(C.12)}
\end{aligned}
$$

The bound (C.10) follows since

$$
\begin{aligned}
2\|A\|_{\max} &= \max_{i,j=1,\dots,d} 2|e'_{d,i}Ae_{d,j}| \\
&\le \max_{i=1,\dots,d}|e'_{d,i}Ae_{d,i}|+\max_{j=1,\dots,d}|e'_{d,j}Ae_{d,j}| \\
&\qquad + 2\max_{i,j=1,\dots,d}\left|\left(\frac{e_{d,i}+e_{d,j}}{\sqrt{2}}\right)'A\left(\frac{e_{d,i}+e_{d,j}}{\sqrt{2}}\right)\right| \\
&\le 4\sup_{v\in\mathcal{K}(2s)}|v'Av| = 4\|A\|_s
\end{aligned}
$$

for a matrix $A\in\mathbb{R}^{d\times d}$. For (C.11), set $\Omega(\widetilde{\delta})=\{\widetilde{\sigma}_{X,ij}\mid \max_{i,j=1,\dots,d}|\widetilde{\sigma}_{X,ij}-\sigma_{X,ij}|\le 2\widetilde{\delta}\}$.

Note that $g''(x)=f(\ell^{-1}(x))$ as stated in (D.47) in the appendix. In order to apply Lemma D.6, we need to find a bound on $\ell^{-1}(\sigma)$ uniformly over all $\sigma\in\Omega(\widetilde{\delta})$.

The function ℓ is strictly increasing as a consequence of Proposition 2.1. For strictly increasing functions, its inverse ℓ^{-1} exists and is also strictly increasing. Recall that by Assumption M.1 there is a a constant $\boldsymbol{c}_Z\in(0,1)$ such that $|\Gamma_{Z,ij}(h)|<\boldsymbol{c}_Z<1$. Then, $-\boldsymbol{c}_Z<\Gamma_{Z,ij}(h)<\boldsymbol{c}_Z$ and $\ell_{ij}(-\boldsymbol{c}_Z)<\ell_{ij}(\Gamma_{Z,ij}(h))=\Gamma_{X,ij}(h)<\ell_{ij}(\boldsymbol{c}_Z)<\ell_{ij}(1)$. Since $\max_{i,j=1,\dots,d}|\widetilde{\sigma}_{X,ij}-\sigma_{X,ij}|\le 2\widetilde{\delta}$ we further argue that for $\sigma\in\Omega(\widetilde{\delta})$,

$$
\ell_{ij}^{-1}(\sigma)\le\ell_{ij}^{-1}(\Gamma_{X,ij}(h)+2\widetilde{\delta})<\ell_{ij}^{-1}(\ell_{ij}(1))=1.
$$

Therefore, for $\widetilde{\delta}$ small, there is a constant $\boldsymbol{c}(\widetilde{\delta})<1$ such that the assumptions in Lemma D.6 are satisfied and therefore $\sup_{\sigma\in\Omega(\widetilde{\delta})}|g''_{\bullet}(\sigma)|\le q_3(\mathbf{\Gamma}_Z)$. □

The following lemma is the analogue of Lemma C.3 for the estimated counterpart $\widehat{g}''$.

Lemma C.4. *Suppose Assumptions M.1–M.4 and let Σ be such that $|\Sigma - \mathbf{\Gamma}_X| < \left|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\right|$. Then, for any $\delta, \widetilde{\delta}, \varepsilon > 0$,*

$$\begin{aligned}
&\mathrm{P}\left[\left\|\widehat{g}''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_4(\mathbf{\Gamma}_Z)\delta\right] \\
&\le \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \widetilde{\delta}] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > \delta] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon]
\end{aligned}$$

with

$$q_4(\mathbf{\Gamma}_Z) := q_4(\varepsilon, \widetilde{\delta}, \mathbf{\Gamma}_Z) = \frac{6}{1 - \boldsymbol{c}(\widetilde{\delta})^2} M_1(\boldsymbol{c}(\widetilde{\delta}), \varepsilon) M_2(\boldsymbol{c}(\widetilde{\delta}), \varepsilon), \tag{C.13}$$

where M_1, M_2 are as in (2.15).

Proof. The proof consists of two parts in order to bound $\widehat{g}''(\Sigma)$ across all elements. The matrix Σ satisfies $|\Sigma - \mathbf{\Gamma}_X| < \left|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\right|$ but g'' is generally not bounded on the whole interval $(-1, 1)$. Therefore, we need to control how much Σ deviates from the true $\mathbf{\Gamma}_X$ (Step 1). Furthermore, $\widehat{g}''$ depends on $\widehat{\theta}$ and needs to be bounded across all possible values of θ (Step 2).

Step 1: With explanations given below,

$$\begin{aligned}
&\mathrm{P}\left[\left\|\widehat{g}''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_4(\mathbf{\Gamma}_Z)\delta\right] \\
&= \mathrm{P}\Bigg[\left\{\left\|\widehat{g}''_\bullet(\Sigma) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_4(\mathbf{\Gamma}_Z)\delta\right\} \\
&\qquad \cap \left(\{\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s \le \widetilde{\delta}\} \cup \{\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \widetilde{\delta}\}\right)\Bigg] \\
&\le \mathrm{P}\Bigg[\left\{\left\|\sup_{\sigma \in \Omega(\widetilde{\delta})} |\widehat{g}''_\bullet(\Sigma)| \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)^{\odot 2}\right\|_s > q_4(\mathbf{\Gamma}_Z)\delta\right\} \\
&\qquad \cap \{\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_{\max} \le 4\widetilde{\delta}\}\Bigg] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \widetilde{\delta}] \qquad \text{(C.14)} \\
&\le \mathrm{P}\left[\frac{6}{1 - \boldsymbol{c}(\widetilde{\delta})^2} \widehat{M}_1(\boldsymbol{c}(\widetilde{\delta}), 0) \widehat{M}_2(\boldsymbol{c}(\widetilde{\delta}), 0) \|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > q_4(\mathbf{\Gamma}_Z)\delta\right] \\
&\qquad + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \widetilde{\delta}] \qquad \text{(C.15)} \\
&\le \mathrm{P}\left[\widehat{q}_4(0, \widetilde{\delta}, \mathbf{\Gamma}_Z) \|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s^2 > q_4(\mathbf{\Gamma}_Z)\delta\right] + \mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \widetilde{\delta}],
\end{aligned}$$

where $\widehat{q}_4(0, \widetilde{\delta}, \mathbf{\Gamma}_Z)$ denotes the estimated counterpart of $q_4(0, \widetilde{\delta}, \mathbf{\Gamma}_Z)$ in (C.13). The bound (C.14) follows from (C.13). Then, applying the same strategy as to get from (C.11) to (C.12) but for an estimator $\widehat{g}$ of g, we get (C.15).

Step 2: For the second step, we argue similarly,

$$
\begin{aligned}
&\mathrm{P}\left[\widehat{q}_4(0,\widetilde{\delta},\mathbf{\Gamma}_Z)\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>q_4(\mathbf{\Gamma}_Z)\delta\right]\\
&=\mathrm{P}\left[\left\{\widehat{q}_4(0,\widetilde{\delta},\mathbf{\Gamma}_Z)\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>q_4(\mathbf{\Gamma}_Z)\delta\right\}\right.\\
&\qquad\left.\cap\left(\{\|\widehat{\theta}-\theta\|_{\max}\le\varepsilon\}\cup\{\|\widehat{\theta}-\theta\|_{\max}>\varepsilon\}\right)\right]\\
&\le\mathrm{P}\left[\left\{\widehat{q}_4(0,\widetilde{\delta},\mathbf{\Gamma}_Z)\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>q_4(\mathbf{\Gamma}_Z)\delta\right\}\cap\{\|\widehat{\theta}-\theta\|_{\max}\le\varepsilon\}\right]\\
&\qquad+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon]\\
&\le\mathrm{P}\left[\sup_{\theta_i\in\Theta(\varepsilon)}q_4(0,\widetilde{\delta},\mathbf{\Gamma}_Z)\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>q_4(\mathbf{\Gamma}_Z)\delta\right]\\
&\qquad+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon]\\
&\le\mathrm{P}[\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\|_s^2>\delta]+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon],
\end{aligned}
$$

since $\sup_{\theta_i\in\Theta(\varepsilon)}q_4(0,\widetilde{\delta},\mathbf{\Gamma}_Z)=q_4(\varepsilon,\widetilde{\delta},\mathbf{\Gamma}_Z)$. □

We also state the result in terms of $\widehat{\ell}(\mathbf{\Gamma}_Z)-\ell(\mathbf{\Gamma}_Z)$ instead of $\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X$ and omit its proof.

Lemma C.5. *Suppose Assumptions M.1–M.4 and let Σ be such that $|\Sigma-\ell(\mathbf{\Gamma}_Z)|<\left|\widehat{\ell}(\mathbf{\Gamma}_Z)-\ell(\mathbf{\Gamma}_Z)\right|$. Then, for any $\delta,\widetilde{\delta},\varepsilon>0$,*

$$
\begin{aligned}
&\mathrm{P}\left[\left\|\widehat{g}''_\bullet(\Sigma)\odot(\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z)-\ell_\bullet(\mathbf{\Gamma}_Z))^{\odot 2}\right\|_s>q_4(\mathbf{\Gamma}_Z)\delta\right]\\
&\le\mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z)-\ell_\bullet(\mathbf{\Gamma}_Z)\|_s>\widetilde{\delta}]+\mathrm{P}[\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z)-\ell_\bullet(\mathbf{\Gamma}_Z)\|_s^2>\delta]\\
&\qquad+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon]
\end{aligned}
$$

with $q_4(\mathbf{\Gamma}_Z)$ as in (C.13).

C.2. Reciprocal of the first derivative of link function

This section concerns high probability bounds for expressions of the form

$$
\left\|(\ell')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s. \tag{C.16}
$$

We study the cases when ℓ is known (Section C.2.1) and when ℓ is estimated (Section C.2.2).

C.2.1. Case of known link function

In this section, we focus on (C.16) given that the link function ℓ is known.

Lemma C.6. *Suppose Assumptions M.1–M.4. Then,*

$$\left\|(\ell')_{\bullet}^{\odot(-1)}(\mathbf{\Gamma}_Z)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s \le R(0,\mathbf{\Gamma}_Z)\left\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\right\|_s \tag{C.17}$$

with

$$R(0,\mathbf{\Gamma}_Z)=\left(24\pi\frac{1}{(1-\boldsymbol{c}_Z^2)^2}M_2(\boldsymbol{c}_Z,0)+8\pi M_1(0,0)\right)\|\mathbf{\Gamma}_Z\|_s$$

and M_1, M_2 are as in (2.15).

Proof. With explanations given below,

$$\begin{aligned}
&\left\|(\ell')_{\bullet}^{\odot(-1)}(\mathbf{\Gamma}_Z)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s\\
&=2\pi\left\|(J_{dL}-\mathbf{\Gamma}_Z^{\odot 2})^{\odot\frac{1}{2}}\odot Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s && \text{(C.18)}\\
&=2\pi\left\|(J_{dL}-\mathbf{\Gamma}_Z^{\odot 2})^{\odot\frac{1}{2}}\odot M\odot Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s && \text{(C.19)}\\
&\le 2\pi\sum_{k=0}^{\infty}\left|\binom{1/2}{k}\right|\left\|\mathbf{\Gamma}_Z^{\odot 2k}\odot M\odot Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s && \text{(C.20)}\\
&\le 4\pi\left\|M\odot Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s. && \text{(C.21)}
\end{aligned}$$

The relation (C.18) is rewritten in terms of the function Z defined in (2.7) and the matrix J_d denotes a $d\times d$-matrix with all entries equal to one. For the equality (C.19), note that $Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})$ is zero on the diagonals. Then, we may include a matrix M being a $0-1$ matrix with the diagonal entries equal to zero. In the next step, we will take advantage of this construction by replacing $Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})$ by a sum of different functions having either zero or non-zero diagonals. For (C.20), we applied the Taylor series expansion $\sqrt{1-y}=\sum_{k=0}^{\infty}\binom{1/2}{k}y^k$. For (C.21), note that $\sum_{k=0}^{\infty}\left|\binom{1/2}{k}\right|=2$; see p. 1206 in Wegkamp and Zhao (2016). Furthermore, since $\mathbf{\Gamma}_Z$ is positive semidefinite, so is $\mathbf{\Gamma}_Z^{\odot 2k}$ due to (D.3). An application of (D.9) proves (C.21) since $\Gamma_{Z,ii}(0)=1$ for all $i=1,\dots,d$.

We continue bounding the expression in (C.21). Note that due to the definition of M all diagonal elements are zero. Therefore, with z defined in (2.7),

$$\begin{aligned}
&M\odot Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\\
&=M\odot\Big((Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})-Z_{\bullet}(0))+(Z_{\bullet}(0)-z_{\bullet}(0))+z_{\bullet}(0)\Big)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\\
&=M\odot\Big((Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})-Z_{\bullet}(0))+(Z_{\bullet}(0)-z_{\bullet}(0))+z(0)\Big)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\\
&=\Big((Z_{\bullet}(\mathbf{\Gamma}_Z^{\odot 2})-Z_{\bullet}(0))+M\odot(z(\mathbf{\Gamma}_Z)-z(0))+M\odot z(0)\Big)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X),
\end{aligned}\tag{C.22}$$

where (C.22) follows since $z(\mathbf{\Gamma}_Z)=Z(0)$. Combining (C.21) and (C.22), with explanations given below,

$$\left\|(\ell')_{\bullet}^{\odot(-1)}(\mathbf{\Gamma}_Z)\odot(\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X)\right\|_s$$

$$
\begin{aligned}
&\leq 4\pi\Big(\left\| (Z_\bullet(\boldsymbol{\Gamma}_Z^{\odot 2}) - Z_\bullet(0)) \odot (\widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X) \right\|_s \\
&\qquad + \left\| M \odot (z(\boldsymbol{\Gamma}_Z) - z(0)) \odot (\widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X) \right\|_s \\
&\qquad\qquad + \left\| M \odot z(0) \odot (\widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X) \right\|_s \Big) \qquad \text{(C.23)} \\
&\leq 4\pi \Bigg(\frac{1}{(1-\boldsymbol{c}_Z^2)^2} M_2(\boldsymbol{c}_Z, 0) + 2M_2(0,0) \\
&\qquad\qquad + 2M_2(\boldsymbol{c}_Z, 0) + 2M_1(0,0) \Bigg) \|\boldsymbol{\Gamma}_Z\|_s \left\| \widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X \right\|_s \qquad \text{(C.24)} \\
&\leq \left(24\pi \frac{1}{(1-\boldsymbol{c}_Z^2)^2} M_2(\boldsymbol{c}_Z, 0) + 8\pi M_1(0,0) \right) \|\boldsymbol{\Gamma}_Z\|_s \left\| \widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X \right\|_s ,
\end{aligned}
$$

where (C.24) follows by Lemmas C.7, C.8 and C.9. □

Remark C.1. We pause here to note that (C.18)–(C.21) are borrowed from Wegkamp and Zhao (2016) and have served as an inspiration for the rest of the proofs. Wegkamp and Zhao (2016) consider a semi-parametric elliptical copula model. Similarly to the link function ℓ, the entries of the copula correlation matrix (Σ) relate to the entries of the Kendall's tau matrix (T) via the formula $\Sigma = \sin\left(\frac{\pi}{2}T\right)$. Then, $\sin'\left(\frac{\pi}{2}T\right) = \cos\left(\frac{\pi}{2}T\right) = (J_d - \sin^{\odot 2}\left(\frac{\pi}{2}T\right))^{\odot\frac{1}{2}} = (J_d - \Sigma^{\odot 2})^{\odot\frac{1}{2}}$ and Lemma 4.3. in Wegkamp and Zhao (2016) provides a result similar to Lemma C.6, relating consistent estimation of the copula correlation matrix to that of Kendall's tau matrix. In their scenario, the cosine function, however, does not depend on any unknown parameters which need to be estimated. Furthermore, the function and its derivative are bounded on the whole interval $(-1, 1)$.

Remark C.2. One can use a different approach than the one pursued in the proof of Lemma C.6 to deal with $(\ell')^{\odot -1}(\boldsymbol{\Gamma}_Z) \odot (\widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X)$. An alternative is to apply a second-order Taylor expansion around zero componentwise so that $(\ell')^{\odot -1}(\boldsymbol{\Gamma}_Z) = (\ell')^{\odot -1}(0) + ((\ell')^{\odot -1})'(0) \odot \boldsymbol{\Gamma}_Z + \frac{1}{2}((\ell')^{\odot -1})''(\Sigma) \odot \boldsymbol{\Gamma}_Z^{\odot 2}$ for some Σ. The challenges here are to show that $((\ell')^{\odot -1})'(0)$ is positive semidefinite and to bound $((\ell')^{\odot -1})''(\Sigma)$ which involves the third derivative of the link function. It seems like the quantities arising in our approach in the proof of Lemma C.6 are slightly simpler to handle.

Lemma C.7. *Suppose Assumptions M.1–M.4. For z in (2.7) and M_1 as in (2.15),*

$$
\left\| M \odot z(0) \odot (\widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X) \right\|_s \leq 2M_1(0,0) \left\| \widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X \right\|_s . \qquad \text{(C.25)}
$$

Proof. With further explanations given below,

$$
\begin{aligned}
&\left\| M \odot z(0) \odot (\widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X) \right\|_s \\
&\leq \max_{i=1,\ldots,dL} (\sqrt{MM'})_{ii} \left\| z(0) \odot (\widehat{\boldsymbol{\Gamma}}_X - \boldsymbol{\Gamma}_X) \right\|_s \qquad \text{(C.26)}
\end{aligned}
$$

$$\leq 2 \left\| z(0) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X) \right\|_s \tag{C.27}$$

$$\leq 2M_1(0,0) \left\| \widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X \right\|_s, \tag{C.28}$$

where (C.26) follows by (D.8). For (C.27), note that with $n = dL$, $\sqrt{MM'} = \sqrt{M_1} = M_2$ with $M_{1,ij} = (n-1)\mathbb{1}_{\{i=j\}} + (n-2)\mathbb{1}_{\{i\neq j\}}$ and $M_{2,ij} = 2\frac{n-1}{n}\mathbb{1}_{\{i=j\}} + \frac{n-2}{n}\mathbb{1}_{\{i\neq j\}}$ (which we leave as an exercise), so that $\max_{i=1,\dots,n}(\sqrt{MM'})_{ii} = 2\frac{n-1}{n} \leq 2$. The relation (C.28) follows by positive semidefiniteness of $z(0)$ (proven below) and application of (D.9).

We prove that $z(0)$ is positive semidefnite by expressing it as a vector product. The matrix $z(0)$ is a $dL \times dL$-dimensional block matrix consisting of blocks of the form

$$\begin{aligned} &\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}(Q_{i,n_0}^2 + Q_{j,n_1}^2) \right) \right)_{i,j=1,\dots,d}^{\odot -1} \\ &= \left(\sum_{n_0=0}^{\infty} \exp\left(-\frac{1}{2}Q_{i,n_0}^2 \right) \sum_{n_1=0}^{\infty} \exp\left(-\frac{1}{2}Q_{j,n_1}^2 \right) \right)_{i,j=1,\dots,d}^{\odot -1} = QQ' \succcurlyeq 0 \end{aligned} \tag{C.29}$$

with

$$Q' = \left(\sum_{n=0}^{\infty} \exp\left(-\frac{1}{2}Q_{1,n}^2 \right), \dots, \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2}Q_{d,n}^2 \right) \right)^{\odot -1}. \tag{C.30}$$

Due to the block structure $z(0) = (QQ')_{r,s=1,\dots,L}$ is also positive semidefinite. □

Lemma C.8. *Suppose Assumptions M.1–M.4. For z in (2.7) and M_2 as in (2.15),*

$$\begin{aligned} &\left\| M \odot (z(\mathbf{\Gamma}_Z) - z(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X) \right\|_s \\ &\leq 2(M_2(0,0) + M_2(\boldsymbol{c}_Z, 0)) \|\mathbf{\Gamma}_Z\|_s \left\| \widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X \right\|_s. \end{aligned} \tag{C.31}$$

Proof. The second-order Taylor approximation of $z(y)$ around zero, applied componentwise in (C.32) below, gives, for some $|\Sigma| < |\mathbf{\Gamma}_Z|$,

$$\begin{aligned} &\left\| M \odot (z(\mathbf{\Gamma}_Z) - z(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X) \right\|_s \\ &\leq \left\| M \odot z'(0) \odot \mathbf{\Gamma}_Z \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X) \right\|_s + \frac{1}{2} \left\| z''_\bullet(\Sigma) \odot \mathbf{\Gamma}_Z^{\odot 2} \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X) \right\|_s \end{aligned} \tag{C.32}$$

$$\leq 2 \|z'(0) \odot \mathbf{\Gamma}_Z\|_s \left\| \widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X \right\|_s + \frac{1}{2} \|z''_\bullet(\Sigma) \odot \mathbf{\Gamma}_Z^{\odot 2}\|_s \left\| \widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X \right\|_s \tag{C.33}$$

$$\leq 2M_2(0,0)\|\mathbf{\Gamma}_Z\|_s \left\| \widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X \right\|_s + 2M_2(\boldsymbol{c}_Z, 0)\|\mathbf{\Gamma}_Z\|_s \left\| \widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X \right\|_s, \tag{C.34}$$

where (C.33) follows by (D.7) and by the same arguments as (C.26) and (C.27) in the proof of Lemma C.7 to handle the matrix M. The first summand in (C.34)

is a consequence of (D.9) since $(-z'(0))$ is positive semidefinite and its diagonals are bounded by $M_2(0,0)$; see Lemma D.7. The second summand in (C.34) follows by Lemma D.8 and $\|\mathbf{\Gamma}_Z^{\odot 2}\|_s \leq \|\mathbf{\Gamma}_Z\|_s$ which is satisfied due to (D.9), positive semidefiniteness of $\mathbf{\Gamma}_Z$ and $\Gamma_{Z,ii}(0) = 1$ for all $i = 1, \ldots, d$. □

Lemma C.9. *Suppose Assumptions M.1–M.4. For Z in (2.7) and M_2 as in (2.15),*

$$\begin{aligned} &\left\|(Z_\bullet(\mathbf{\Gamma}_Z^{\odot 2}) - Z_\bullet(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \\ &\leq \frac{1}{(1-\boldsymbol{c}_Z^2)^2} M_2(\boldsymbol{c}_Z, 0)\, \|\mathbf{\Gamma}_Z\|_s \left\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\right\|_s . \end{aligned} \tag{C.35}$$

Proof. By the mean value theorem there is a Σ such that $|\Sigma| < \mathbf{\Gamma}_Z^{\odot 2}$ and

$$\begin{aligned} &\left\|(Z_\bullet(\mathbf{\Gamma}_Z^{\odot 2}) - Z_\bullet(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \\ &= \left\|Z'_\bullet(\Sigma) \odot \mathbf{\Gamma}_Z^{\odot 2} \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \\ &\leq \left\|Z'_\bullet(\Sigma) \odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s \left\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\right\|_s \qquad \text{(C.36)} \\ &\leq \frac{1}{(1-\boldsymbol{c}_Z^2)^2} M_2(\boldsymbol{c}_Z, 0)\, \|\mathbf{\Gamma}_Z\|_s \left\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\right\|_s , \qquad \text{(C.37)} \end{aligned}$$

where (C.36) follows by (D.7) and (C.37) is a consequence of Lemma D.9 and $\|\mathbf{\Gamma}_Z^{\odot 2}\|_s \leq \|\mathbf{\Gamma}_Z\|_s$ which is satisfied due to (D.9), positive semidefiniteness of $\mathbf{\Gamma}_Z$ and $\Gamma_{Z,ii}(0) = 1$ for all $i = 1, \ldots, d$. □

C.2.2. Case of estimated link function

The following lemma is the analogue of Lemma C.6 for an estimated link function.

Lemma C.10. *Suppose Assumptions M.1–M.4 and let M_2 be as in (2.15). For any $\delta, \varepsilon > 0$,*

$$\begin{aligned} \mathrm{P}\left[\left\|(\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s > \delta\right] \precsim \mathrm{P}\left[R(\mathbf{\Gamma}_Z)\|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\|_s > \delta\right] \\ + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon] \end{aligned} \tag{C.38}$$

with

$$R(\mathbf{\Gamma}_Z) := R(\varepsilon, \mathbf{\Gamma}_Z) = \left(24\pi \frac{1}{(1-\boldsymbol{c}_Z^2)^2} M_2(\boldsymbol{c}_Z, \varepsilon) + 8\pi M_1(0, \varepsilon)\right) \|\mathbf{\Gamma}_Z\|_s \tag{C.39}$$

and M_1, M_2 are as in (2.15).

Proof. The proof is similar to that of Lemma C.6 by replacing all functions with their estimated counterparts. In particular, we can follow the proof of Lemma C.6 up to (C.23), that is,

$$\left\|(\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\right\|_s \tag{C.40}$$

$$\leq 4\pi\Big(\Big\|(\widehat{Z}_\bullet(\mathbf{\Gamma}_Z^{\odot 2}) - \widehat{Z}_\bullet(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s + \Big\|(\widehat{z}_\bullet(\mathbf{\Gamma}_Z) - \widehat{z}_\bullet(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s$$
$$+ \Big\|M \odot \widehat{z}(0) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s \Big). \tag{C.41}$$

In contrast to Lemma C.6, the functions in (C.41) are random and one needs to control the error made by estimating the CDF parameters θ_i. Continuing with (C.41), and with further explanations given below,

$$\begin{aligned}
&\mathrm{P}\left[\Big\|(\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s > \delta\right] \\
&\leq \mathrm{P}\Bigg[\Bigg\{4\pi\Big(\Big\|(\widehat{Z}_\bullet(\mathbf{\Gamma}_Z^{\odot 2}) - \widehat{Z}_\bullet(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s \\
&\qquad + \Big\|(\widehat{z}_\bullet(\mathbf{\Gamma}_Z) - \widehat{z}_\bullet(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s \\
&\qquad\qquad + \Big\|M \odot \widehat{z}(0) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s \Big) > \delta\Bigg\} \\
&\qquad\qquad\qquad \cap \Big(\{\|\widehat{\theta} - \theta\|_{\max} \leq \varepsilon\} \cup \{\|\widehat{\theta} - \theta\|_{\max} > \varepsilon\}\Big)\Bigg] \qquad \text{(C.42)} \\
&\leq \mathrm{P}\Bigg[\Bigg\{4\pi\Big(\Big\|(\widehat{Z}_\bullet(\mathbf{\Gamma}_Z^{\odot 2}) - \widehat{Z}_\bullet(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s \\
&\qquad + \Big\|M \odot (\widehat{z}_\bullet(\mathbf{\Gamma}_Z) - \widehat{z}_\bullet(0)) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s \\
&\qquad\qquad + \Big\|M \odot \widehat{z}(0) \odot (\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X)\Big\|_s \Big) > \delta\Bigg\} \cap \{\|\widehat{\theta} - \theta\|_{\max} \leq \varepsilon\}\Bigg] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon] \qquad \text{(C.43)} \\
&\precsim \mathrm{P}\left[\Big(24\pi \frac{1}{(1-\boldsymbol{c}_Z^2)^2} M_2(\boldsymbol{c}_Z, \varepsilon) + 8\pi M_1(0,\varepsilon)\Big)\|\mathbf{\Gamma}_Z\|_s \|\widehat{\mathbf{\Gamma}}_X - \mathbf{\Gamma}_X\| > \delta\right] \\
&\qquad + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon]. \qquad \text{(C.44)}
\end{aligned}$$

The inequality (C.42) follows from (C.41) and we intersect with the event $\{\|\widehat{\theta} - \theta\|_{\max} \leq \varepsilon\}$ in (C.42) to control the estimation error we made by using $\widehat{\theta}_i$; see also Remark C.3. The bound (C.43) follows as in (C.23). The three summands in (C.43) can then be handled as in the proof of Lemme C.6 but through probabilistic versions of Lemmas C.7, C.8 and C.9, and using the fact that $\widehat{\theta}_i$ is in an ε-region of the true θ_i. □

Remark C.3. The step of intersecting with the event $\{\|\widehat{\theta}-\theta\|_{\max} \leq \varepsilon\}$ in (C.42) is crucial in order to control how much the estimated CDF parameters deviate from the true model parameters. The function Z in (2.7) is not necessarily bounded for all possible values of θ. For this reason, we can only ensure that $M_1(\boldsymbol{c}_Z, \varepsilon), M_2(\boldsymbol{c}_Z, \varepsilon)$ are finite for small ε. A similar approach was pursued in Baek, Düker and Pipiras (2023). Baek, Düker and Pipiras (2023) deal with high-dimensional spectral density estimation under long-range dependence. In

order to consistently estimate the spectral density matrix under long-range dependence, one needed to control the memory parameter matrix as well; see Proposition 3.5 in Baek, Düker and Pipiras (2023) and its proof.

For completeness, we also state the result analogous to Lemma C.10 in terms of $\ell(\mathbf{\Gamma}_Z)$ instead of $\mathbf{\Gamma}_X$. We omit the proof since it is similar to the proof of Lemma C.10.

Lemma C.11. *Suppose Assumptions M.1–M.4 and let M_2 be as in (2.15). For any $\delta, \varepsilon > 0$,*

$$\begin{aligned} &\mathrm{P}\left[\left\|(\widehat{\ell}')_\bullet^{\odot(-1)}(\mathbf{\Gamma}_Z) \odot (\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z))\right\|_s > \delta\right] \\ &\qquad \precsim \mathrm{P}\left[R(\mathbf{\Gamma}_Z)\left\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\right\|_s > \delta\right] + \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \varepsilon] \end{aligned} \tag{C.45}$$

with $R(\mathbf{\Gamma}_Z)$ as in (C.39).

C.3. Link function and its derivatives

In this section, we provide results for the link function ℓ and its derivatives.

Lemma C.12. *Suppose Assumptions M.1–M.4. For any $\delta, \varepsilon > 0$,*

$$\mathrm{P}\left[\left\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\right\|_s > S(\mathbf{\Gamma}_Z)\delta\right] \precsim \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon]$$

with

$$S(\mathbf{\Gamma}_Z) := S(\varepsilon, \mathbf{\Gamma}_Z) = 4\frac{18}{(1 - \boldsymbol{c}_Z^2)^{\frac{7}{2}}} M(\boldsymbol{c}_Z, \varepsilon)\mu(\boldsymbol{c}_Z, \varepsilon)\|\mathbf{\Gamma}_Z\|_s \tag{C.46}$$

and M, μ are as in (2.14).

Proof. We have

$$\begin{aligned} &\left\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\right\|_s \\ &\le \left\|\widehat{\ell}(0) - \ell(0) + (\widehat{\ell}'_\bullet(\mathbf{\Gamma}_Z) - \ell'_\bullet(\mathbf{\Gamma}_Z)) \odot \mathbf{\Gamma}_Z\right\|_s + \left\|(\widehat{\ell}''_\bullet(\Sigma_1) - \ell''_\bullet(\Sigma_1)) \odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s \end{aligned} \tag{C.47}$$

$$= \left\|(\widehat{\ell}'_\bullet(\mathbf{\Gamma}_Z) - \ell'_\bullet(\mathbf{\Gamma}_Z)) \odot \mathbf{\Gamma}_Z\right\|_s + \left\|(\widehat{\ell}''_\bullet(\Sigma_1) - \ell''_\bullet(\Sigma_1)) \odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s \tag{C.48}$$

$$\begin{aligned} &\le \left\|(\widehat{\ell}'_\bullet(\mathbf{\Gamma}_Z) - \ell'_\bullet(\mathbf{\Gamma}_Z) - (\widehat{\ell}'_\bullet(0) - \ell'_\bullet(0))) \odot \mathbf{\Gamma}_Z\right\|_s + 2\left\|(\widehat{\ell}'(0) - \ell'(0)) \odot \mathbf{\Gamma}_Z\right\|_s \\ &\qquad + \left\|(\widehat{\ell}''_\bullet(\Sigma_1) - \ell''_\bullet(\Sigma_1)) \odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s \end{aligned} \tag{C.49}$$

$$\begin{aligned} &\le \left\|(\widehat{\ell}''_\bullet(\Sigma_2) - \ell''_\bullet(\Sigma_2)) \odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s + 2\left\|(\widehat{\ell}'(0) - \ell'(0)) \odot \mathbf{\Gamma}_Z\right\|_s \\ &\qquad + \left\|(\widehat{\ell}''_\bullet(\Sigma_1) - \ell''_\bullet(\Sigma_1)) \odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s . \end{aligned} \tag{C.50}$$

The bound (C.47) follows for Σ_1, such that $|\Sigma_1| < |\mathbf{\Gamma}_Z|$ by applying the second-order Taylor expansion to the function $x \mapsto \widehat{\ell}(x) - \ell(x)$ around $x = 0$. The equality (C.48) follows since $\widehat{\ell}(0) = \ell(0) = 0$ due to (2.4). By subtracting and adding the function $(\widehat{\ell}'_\bullet(0) - \ell'_\bullet(0)) \odot \mathbf{\Gamma}_Z$ and subsequent application of triangle inequality, we get (C.49). The second summand of (C.49) is explained at the end of this proof. The bound (C.50) results from the mean value theorem for some Σ_2 such that $|\Sigma_2| < |\mathbf{\Gamma}_Z|$. It follows from (C.50) that

$$\begin{aligned}
&\mathrm{P}\left[\left\|\widehat{\ell}_\bullet(\mathbf{\Gamma}_Z) - \ell_\bullet(\mathbf{\Gamma}_Z)\right\|_s > S(\mathbf{\Gamma}_Z)\delta\right] \\
&\le \mathrm{P}\left[2\left\|(\widehat{\ell}'(0) - \ell'(0)) \odot \mathbf{\Gamma}_Z\right\|_s > S(\mathbf{\Gamma}_Z)\frac{\delta}{2}\right] \\
&\qquad + \mathrm{P}\left[2\left\|(\widehat{\ell}''_\bullet(\Sigma) - \ell''_\bullet(\Sigma)) \odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s > S(\mathbf{\Gamma}_Z)\frac{\delta}{2}\right] \\
&\le \mathrm{P}\left[\left\|(\widehat{\ell}'(0) - \ell'(0)) \odot \mathbf{\Gamma}_Z\right\|_s > s(\mathbf{\Gamma}_Z)\delta\right] \\
&\qquad + \mathrm{P}\left[\left\|(\widehat{\ell}''_\bullet(\Sigma) - \ell''_\bullet(\Sigma)) \odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s > S(\mathbf{\Gamma}_Z)\frac{\delta}{4}\right] \qquad \text{(C.51)} \\
&\le 2\,\mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon], \qquad \text{(C.52)}
\end{aligned}$$

where (C.51) follows since $S(\mathbf{\Gamma}_Z) \ge s(\mathbf{\Gamma}_Z)$ with $s(\mathbf{\Gamma}_Z)$ in (C.55) and the summands in (C.51) are respectively bounded through Lemmas C.13 and C.14 to get (C.52).

Regarding the second summand in (C.49), we can always write $\|\widehat{\ell}'_\bullet(0) - \ell'_\bullet(0)\|_s$ with diagonals set to zero as $\|M \odot (\widehat{\ell}'(0) - \ell'(0))\|_s$ with M being a $0-1$ matrix with the diagonal entries equal to zero. Then,

$$\begin{aligned}
\left\|(\widehat{\ell}'_\bullet(0) - \ell'_\bullet(0)) \odot \mathbf{\Gamma}_Z\right\|_s &= \left\|M \odot (\widehat{\ell}'(0) - \ell'(0)) \odot \mathbf{\Gamma}_Z\right\|_s \\
&\le \max_{i=1,\dots,dL} (\sqrt{MM^*})_{ii} \left\|(\widehat{\ell}'(0) - \ell'(0)) \odot \mathbf{\Gamma}_Z\right\|_s \qquad \text{(C.53)} \\
&\le 2\left\|(\widehat{\ell}'(0) - \ell'(0)) \odot \mathbf{\Gamma}_Z\right\|_s, \qquad \text{(C.54)}
\end{aligned}$$

where (C.53) follows by (D.8). For (C.54), note that with $n = dL$, $\sqrt{MM^*} = \sqrt{M_1} = M_2$ with $M_{1,ij} = (n-1)\mathbb{1}_{\{i=j\}} + (n-2)\mathbb{1}_{\{i\neq j\}}$ and $M_{2,ij} = 2\frac{n-1}{n}\mathbb{1}_{\{i=j\}} + \frac{n-2}{n}\mathbb{1}_{\{i\neq j\}}$, so that $\max_{i=1,\dots,n}(\sqrt{MM^*})_{ii} = 2\frac{n-1}{n} \le 2$. □

Lemma C.13. *Suppose Assumptions M.1–M.4. For $\delta, \varepsilon > 0$,*

$$\mathrm{P}\left[\left\|(\widehat{\ell}'(0) - \ell'(0)) \odot \mathbf{\Gamma}_Z\right\|_s > s(\mathbf{\Gamma}_Z)\delta\right] \precsim \mathrm{P}[\|\widehat{\theta} - \theta\|_{\max} > \delta \wedge \varepsilon]$$

with

$$s(\mathbf{\Gamma}_Z) := s(\varepsilon, \mathbf{\Gamma}_Z) = 4 \max_{i=1,\dots,d} \sup_{\theta_i \in \Theta(\varepsilon)} m_i^{(0)}(1) \max_{j=1,\dots,d} \sup_{\theta_j \in \Theta(\varepsilon)} \mu_j^{(1)}(1) \|\mathbf{\Gamma}_Z\|_s. \qquad \text{(C.55)}$$

Proof. Note that

$$\ell'(0) = \left(\frac{1}{2\pi}\sum_{n_0,n_1=0}^{\infty}\exp\left(-\frac{1}{2}(Q_{i,n_0}^2+Q_{j,n_1}^2)\right)\right)_{i,j=1,\dots,d} = QQ' \succcurlyeq 0,$$

where $Q' = (m_1^{(0)}(1),\dots,m_d^{(0)}(1))$. We write $(QQ')_{r,s=1,\dots,L}$ for a $dL\times dL$ block matrix, where each $d\times d$ block is the same matrix QQ'. Similar quantities can be defined for $\widehat{\ell}'(\cdot)$ in terms of $\widehat{Q}$. For some $\widetilde{\theta}_j$ with $|\widetilde{\theta}_j-\theta_j| < |\widehat{\theta}_j-\theta_j|$, we further introduce $R' = (R_1,\dots,R_d)$ with

$$R_j = \frac{1}{\sqrt{2\pi}}\sum_{n=0}^{\infty}\exp\left(-\frac{1}{2}\widetilde{Q}_{j,n}^2\right)(-\widetilde{Q}_{j,n})\langle\nabla Q_n(\theta_j)|_{\widetilde{\theta}_j},(\widehat{\theta}_j-\theta_j)\rangle,$$

where $\widetilde{Q}_{j,n} = Q_n(\widetilde{\theta}_j)$ and $Q_{i,n} = Q_n(\theta_i)$. Using Lemma D.5, R_j can be bounded as

$$\begin{aligned}|R_j| &\le \frac{1}{\sqrt{2\pi}}\sum_{n=0}^{\infty}\exp\left(-\frac{1}{2}\widetilde{Q}_{j,n}^2\right)|\widetilde{Q}_{j,n}|\|\nabla Q_n(\theta_j)|_{\widetilde{\theta}_j}\|\|\widehat{\theta}_j-\theta_j\|_{\max}\\ &\le \widetilde{\mu}_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max},\end{aligned} \tag{C.56}$$

where $\widetilde{\mu}_j^{(1)}(1)$ is defined as $\mu_j^{(1)}(1)$ in (2.12) but $Q_{i,n}$ replaced with $\widetilde{Q}_{j,n}$. Then, with explanations given below,

$$\begin{aligned}&\left\|(\widehat{\ell}'(0)-\ell'(0))\odot\mathbf{\Gamma}_Z\right\|_s\\ &= \left\|(\widehat{Q}\widehat{Q}'-QQ')_{r,s=1,\dots,L}\odot\mathbf{\Gamma}_Z\right\|_s\\ &\le \left\|(\widehat{Q}\widehat{Q}'-\widehat{Q}Q')_{r,s=1,\dots,L}\odot\mathbf{\Gamma}_Z\right\|_s + \left\|(\widehat{Q}Q'-QQ')_{r,s=1,\dots,L}\odot\mathbf{\Gamma}_Z\right\|_s\\ &= \left\|(\widehat{Q}R')_{r,s=1,\dots,L}\odot\mathbf{\Gamma}_Z\right\|_s + \left\|(RQ')_{r,s=1,\dots,L}\odot\mathbf{\Gamma}_Z\right\|_s \qquad \text{(C.57)}\\ &= \left\|([\widehat{Q}:\cdots:\widehat{Q}]\odot[R:\cdots:R]')_{r,s=1,\dots,L}\odot\mathbf{\Gamma}_Z\right\|_s\\ &\qquad + \left\|([R:\cdots:R]\odot[Q:\cdots:Q]')_{r,s=1,\dots,L}\odot\mathbf{\Gamma}_Z\right\|_s \qquad \text{(C.58)}\\ &\le \max_{i=1,\dots,d}|\widehat{Q}_i|\max_{i=1,\dots,d}|R_i|\|\mathbf{\Gamma}_Z\|_s + \max_{i=1,\dots,d}|Q_i|\max_{i=1,\dots,d}|R_i|\|\mathbf{\Gamma}_Z\|_s \qquad \text{(C.59)}\\ &\le \max_{i=1,\dots,d}\widehat{m}_i^{(0)}(1)\max_{j=1,\dots,d}\widetilde{\mu}_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s\\ &\qquad + \max_{i=1,\dots,d}m_i^{(0)}(1)\max_{j=1,\dots,d}\widetilde{\mu}_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s. \qquad \text{(C.60)}\end{aligned}$$

The equality (C.57) results from componentwise application of the mean value theorem. The relation (C.58) follows by noting that for two vectors $a,b\in\mathbb{R}^d$ one can write $ab' = [a:\cdots:a]\odot[b:\cdots:b]'$. The inequality (C.59) follows by (D.19) in Lemma D.2. The last inequality (C.60) is due to (C.56).

It remains to get high probability bounds on the two summands in (C.60). We have

$$
\begin{aligned}
&\mathrm{P}[\max_{i=1,\dots,d} \widehat{m}_i^{(0)}(1) \max_{j=1,\dots,d} \widetilde{\mu}_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s > s(\mathbf{\Gamma}_Z)\delta/2] \\
&= \mathrm{P}\Bigg[\Bigg\{\max_{i=1,\dots,d} \widehat{m}_i^{(0)}(1) \max_{j=1,\dots,d} \widetilde{\mu}_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s > s(\mathbf{\Gamma}_Z)\delta/2\Bigg\} \\
&\qquad \cap \Big(\{\|\widehat{\theta}-\theta\|_{\max} \le \varepsilon\} \cup \{\|\widehat{\theta}-\theta\|_{\max} > \varepsilon\}\Big)\Bigg] \\
&\le \mathrm{P}\Bigg[\Bigg\{\max_{i=1,\dots,d} \widehat{m}_i^{(0)}(1) \max_{j=1,\dots,d} \widetilde{\mu}_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s > s(\mathbf{\Gamma}_Z)\delta/2\Bigg\} \\
&\qquad \cap \{\|\widehat{\theta}-\theta\|_{\max} \le \varepsilon\}\Bigg] + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \varepsilon] \\
&\le \mathrm{P}\Bigg[\max_{i=1,\dots,d} \sup_{\theta_i\in\Theta(\varepsilon)} m_i^{(0)}(1) \max_{j=1,\dots,d} \sup_{\theta_j\in\Theta(\varepsilon)} \mu_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s > s(\mathbf{\Gamma}_Z)\delta/2\Bigg] \\
&\qquad + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \varepsilon] \\
&\le 2\,\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \delta\wedge\varepsilon],
\end{aligned}
$$

due to the definition of $s(\mathbf{\Gamma}_Z)$ in (C.55). Analogously, we can infer that

$$
\begin{aligned}
&\mathrm{P}[\max_{i=1,\dots,d} m_i^{(0)}(1) \max_{j=1,\dots,d} \widetilde{\mu}_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s > s(\mathbf{\Gamma}_Z)\delta/2] \\
&= \mathrm{P}\Bigg[\Bigg\{\max_{i=1,\dots,d} m_i^{(0)}(1) \max_{j=1,\dots,d} \widetilde{\mu}_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s > s(\mathbf{\Gamma}_Z)\delta/2\Bigg\} \\
&\qquad \cap \{\|\widehat{\theta}-\theta\|_{\max} \le \varepsilon\}\Bigg] + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \varepsilon] \\
&\le \mathrm{P}\Bigg[\max_{i=1,\dots,d} m_i^{(0)}(1) \max_{j=1,\dots,d} \sup_{\theta_j\in\Theta(\varepsilon)} \mu_j^{(1)}(1)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z\|_s > s(\mathbf{\Gamma}_Z)\delta/2\Bigg] \\
&\qquad + \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \varepsilon] \\
&\le \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \delta\wedge\varepsilon],
\end{aligned}
$$

since $s(\mathbf{\Gamma}_Z) \ge 2\max_{i=1,\dots,d} m_i^{(0)}(1)\max_{j=1,\dots,d}\sup_{\theta_j\in\Theta(\varepsilon)} \mu_j^{(1)}(1)\|\mathbf{\Gamma}_Z\|_s$. □

The following lemma provides a high probability bound on the difference between the estimated and true second derivatives of the link function.

Lemma C.14. *Suppose Assumptions M.1–M.4 and let Σ be such that $|\Sigma| < |\mathbf{\Gamma}_Z|$. Then,*

$$
\mathrm{P}\left[\left\|\left(\widehat{\ell}''_\bullet(\Sigma) - \ell''_\bullet(\Sigma)\right)\odot \mathbf{\Gamma}_Z^{\odot 2}\right\|_s > S(\mathbf{\Gamma}_Z)\delta\right] \le \mathrm{P}[\|\widehat{\theta}-\theta\|_{\max} > \delta\wedge\varepsilon]
$$

with $S(\mathbf{\Gamma}_Z)$ as in (C.46).

Proof. With further explanations given below,

$$
\begin{aligned}
&\mathrm{P}\left[\left\|\left(\widehat{\ell}''(\Sigma)-\ell''(\Sigma)\right)\odot\mathbf{\Gamma}_Z^{\odot 2}\right\|_s>S(\mathbf{\Gamma}_Z)\delta\right]\\
&\le \mathrm{P}\left[\left\{\frac{18}{(1-\boldsymbol{c}_Z^2)^{\frac{7}{2}}}\widetilde{M}(\boldsymbol{c}_Z,0)\widetilde{\mu}(\boldsymbol{c}_Z,0)\|\widehat{\theta}-\theta\|_{\max}\|\mathbf{\Gamma}_Z^{\odot 2}\|_s>S(\mathbf{\Gamma}_Z)\delta\right\}\right.\\
&\qquad\left.\cap\{\|\widehat{\theta}-\theta\|_{\max}\le\varepsilon\}\right]+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon) \qquad \text{(C.61)}\\
&\le \mathrm{P}\left[\frac{18}{(1-\boldsymbol{c}_Z^2)^{\frac{7}{2}}}M(\boldsymbol{c}_Z,\varepsilon)\mu(\boldsymbol{c}_Z,\varepsilon)\|\mathbf{\Gamma}_Z\|_s\|\widehat{\theta}-\theta\|_{\max}>S(\mathbf{\Gamma}_Z)\delta\right]\\
&\qquad+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon] \qquad \text{(C.62)}\\
&\le 2\,\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\delta\wedge\varepsilon],
\end{aligned}
$$

where (C.61) follows by Lemma D.10 and (C.62) is a consequence of incorporating $\{\|\widehat{\theta}-\theta\|_{\max}\le\varepsilon\}$ and $\|\mathbf{\Gamma}_Z\|_s^{\odot 2}\le\|\mathbf{\Gamma}_Z\|_s$ due to (D.9) and $\mathbf{\Gamma}_{Z,rr}=1$ for all $r=1,\ldots,dL$. □

C.4. Diagonal elements

The lemma stated in this section concerns the diagonal elements of $\mathbf{\Gamma}_Z$ and its estimators. The arguments follow a strategy very close to the ones used in the proofs of Lemmas A.1 and A.2. However, we work only with the mean value theorem rather than a second-order Taylor approximation.

Lemma C.15. *Suppose Assumptions M.2–M.4. Then, for any $\delta,\varepsilon>0$,*

$$
\begin{aligned}
&\mathrm{P}\left[\max_{i=1,\ldots,d}|\widehat{g}_{ii}(\widehat{\Gamma}_{X,ii}(0))-g_{ii}(\Gamma_{X,ii}(0))|>D(\mathbf{\Gamma}_Z)\delta\right]\\
&\precsim \mathrm{P}\left[\left\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\right\|_s>\delta\right]+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\delta\wedge\varepsilon]
\end{aligned}
\qquad \text{(C.63)}
$$

with

$$D(\mathbf{\Gamma}_Z)=M_1^{\frac{1}{2}}(1/2,\varepsilon)2\max\{3\Delta(\varepsilon),1\},$$

where M_1 and Δ are as in (2.15) and (2.13).

Proof. As discussed in Section 2.2, we assume that the domain of g (and $\widehat{g}$) is naturally extended such that $\widehat{g}_{ii}(x)=\widehat{g}_{ii}(\widehat{\ell}_{ii}(1))=1$ for all $x>\widehat{\ell}_{ii}(1)$. We then have $0\le\widehat{g}_{ii}(\widehat{\Gamma}_{X,ii}(0))=\widehat{\Gamma}_{Z,ii}(0)\le 1$.

Therefore, it is sufficient to consider the probability

$$
\begin{aligned}
&\mathrm{P}\left[\max_{i=1,\ldots,d}|\widehat{g}_{ii}(\widehat{\Gamma}_{X,ii}(0))-g_{ii}(\Gamma_{X,ii}(0))|>D(\mathbf{\Gamma}_Z)\delta\right]\\
&=\mathrm{P}\left[\max_{i=1,\ldots,d}(1-\widehat{g}_{ii}(\widehat{\Gamma}_{X,ii}(0)))>D(\mathbf{\Gamma}_Z)\delta\right].
\end{aligned}
\qquad \text{(C.64)}
$$

Set $\delta^* = D(\mathbf{\Gamma}_Z)\delta$, with explanations given below, the probability on the right-hand side of (C.64) can be bounded as

$$
\begin{aligned}
&\mathrm{P}\left[\max_{i=1,\dots,d}(1-\widehat{g}_{ii}(\widehat{\Gamma}_{X,ii}(0)))>\delta^*\right]\\
&=\mathrm{P}\left[\bigcup_{i=1,\dots,d}\left\{\widehat{\ell}_{ii}(1-\delta^*)>\widehat{\Gamma}_{X,ii}(0))\right\}\right]\\
&=\mathrm{P}\left[\bigcup_{i=1,\dots,d}\left\{\Gamma_{X,ii}(0)-\widehat{\Gamma}_{X,ii}(0)+\widehat{\ell}_{ii}(1)-\ell_{ii}(1)>\widehat{\ell}_{ii}(1)-\widehat{\ell}_{ii}(1-\delta^*)\right\}\right]\\
&=\mathrm{P}\left[\bigcup_{i=1,\dots,d}\left\{\Gamma_{X,ii}(0)-\widehat{\Gamma}_{X,ii}(0)+\widehat{\ell}_{ii}(1)-\ell_{ii}(1)>\widehat{\ell}'_{ii}(c_{\delta^*})\delta^*\right\}\right] \qquad \text{(C.65)}\\
&\le \mathrm{P}\left[\left\{\max_{i=1,\dots,d}\frac{1}{\widehat{\ell}'_{ii}(c_{\delta^*})}|\Gamma_{X,ii}(0)-\widehat{\Gamma}_{X,ii}(0)+\widehat{\ell}_{ii}(1)-\ell_{ii}(1)|>\delta^*\right\}\right.\\
&\qquad\left.\cap\left(\{\|\widehat{\theta}-\theta\|_{\max}\le\varepsilon\}\cup\{\|\widehat{\theta}-\theta\|_{\max}>\varepsilon\}\right)\right] \qquad \text{(C.66)}\\
&\le \mathrm{P}\left[M_1^{\frac{1}{2}}(1/2,\varepsilon)\max_{i=1,\dots,d}|\Gamma_{X,ii}(0)-\widehat{\Gamma}_{X,ii}(0)\right.\\
&\qquad\left.+3\sup_{\theta_i\in\Theta(\varepsilon)}\sum_{n=0}^{\infty}n\left\|\nabla_{\theta_i}C_n(\theta_i)\right\|_1\|\widehat{\theta}-\theta\|_{\max}|>\delta^*\right]\\
&\qquad\qquad+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\varepsilon]. \qquad \text{(C.67)}\\
&\precsim \mathrm{P}\left[\left\|\widehat{\mathbf{\Gamma}}_X-\mathbf{\Gamma}_X\right\|_s>\delta\right]+\mathrm{P}[\|\widehat{\theta}-\theta\|_{\max}>\delta\wedge\varepsilon]. \qquad \text{(C.68)}
\end{aligned}
$$

Application of the mean value theorem gives (C.65) for some $c_{\delta^*}\in(1-\delta^*,1)$. The relation (C.66) is a consequence of intersecting with the event $\{\|\widehat{\theta}-\theta\|_{\max}\le\varepsilon\}$ and its complement. For (C.67), we derive a lower bound on $\ell'_{ii}(u)$ in Lemma D.13. Lemma D.12 provides a bound on $|\widehat{\ell}_{ii}(1)-\ell_{ii}(1)|$ which is finite under Assumptions M.2–M.4. Finally, with $\delta^*=D(\mathbf{\Gamma}_Z)\delta$, we can bound the first probability in (C.67) further to get (C.68). □

Appendix D: Additional results and their proofs

Section D.1 provides results for the mapping (2.8) and its interplay with the Hadamard product. Section D.2 states some results and their proofs to ensure that the constants in our main results are finite. Finally, Section D.3 collects the majority of derivatives of the link function used throughout this work.

D.1. Hadamard product

Our proofs make extensive use of multiple properties of the Hadamard product. Chapter 5 in Horn and Johnson (1991) provides a survey on the Hadamard product. For the reader's convenience, we collect the properties used in our proofs here and will refer to those instead of the respective statements in Horn and Johnson (1991). Let $A, B \in \mathbb{R}^{d\times d}$ with $A = (a_{ij})_{i,j=1,\dots,d}$ and $B = (b_{ij})_{i,j=1,\dots,d}$ be symmetric matrices. Then, the following statements are true.

1. Theorem 5.5.18 in Horn and Johnson (1991): If $A \succcurlyeq 0$, then
$$\|A \odot B\| \leq \max_{i=1,\dots,d} |a_{ii}| \|B\|. \tag{D.1}$$
2. Theorem 5.5.4 in Horn and Johnson (1991):
$$\|A \odot B\| \leq \|A\| \|B\|. \tag{D.2}$$
3. Chapter 5.2, Problem 3 in Horn and Johnson (1991): If $A, B \succcurlyeq 0$, then
$$A \odot B \succcurlyeq 0. \tag{D.3}$$
4. Theorem 5.5.19 in Horn and Johnson (1991):
$$\|A \odot B\| \leq \max_{i=1,\dots,p} (\sqrt{AA'})_{ii} \|B\|. \tag{D.4}$$
5. Problem 5.6.P42 in Horn and Johnson (2012): If $|a_{ij}| \leq b_{ij}$ for all $i, j = 1, \dots, d$, then
$$\|A\| \leq \|B\|. \tag{D.5}$$

Note that Problem 5.6.P42 in Horn and Johnson (2012) actually states that if $0 \leq a_{ij} \leq b_{ij}$ for all $i, j = 1, \dots, d$, then $\|A\| \leq \|B\|$. However, the proof given in Problem 5.6.P42 in Horn and Johnson (2012) can be easily adapted to our milder assumptions. A proof of the statement as written in (D.5) can also be found in the proof of Lemma 4.4. in Wegkamp and Zhao (2016).

The goal here is to prove (D.1), (D.2), (D.4), (D.5) and one additional property for the norm (but not matrix norm) defined in (2.8), that is,
$$A \mapsto \sup_{v \in \mathcal{K}(2s)} |v'Av|, \tag{D.6}$$
where $\mathcal{K}(2s) = \{v \in \mathbb{R}^d : \|v\| \leq 1, \|v\|_0 \leq 2s\}$. The following Lemmas D.1 and D.2 state the same properties for (D.6) as those known for the spectral norm.

Lemma D.1. *Let $A = (a_{ij})_{i,j=1,\dots,d} \in \mathbb{R}^{d\times d}$ and $B = (b_{ij})_{i,j=1,\dots,d} \in \mathbb{R}^{d\times d}$ be symmetric. Then,*
$$\sup_{v \in \mathcal{K}(2s)} |v'(A \odot B)v| \leq \sup_{v \in \mathcal{K}(2s)} |v'Av| \sup_{v \in \mathcal{K}(2s)} |v'Bv|, \tag{D.7}$$

$$\sup_{v\in\mathcal{K}(2s)} |v'(A\odot B)v| \le \max_{i=1,\dots,d} (\sqrt{AA'})_{ii} \sup_{v\in\mathcal{K}(2s)} |v'Bv|. \tag{D.8}$$

If A is also positive semidefinite, then

$$\sup_{v\in\mathcal{K}(2s)} |v'(A\odot B)v| \le \max_{i=1,\dots,d} |a_{ii}| \sup_{v\in\mathcal{K}(2s)} |v'Bv|. \tag{D.9}$$

Proof of Lemma D.1. Write D_v for the diagonal matrix which corresponds to a vector $v\in\mathbb{R}^d$ such that $D_v = \operatorname{diag}(v_1,\dots,v_d)$. We further introduce the matrix $D_s(v)$, $\|v\|_0 \le 2s$, which is a $0-1$ matrix with the same sparsity pattern as D_v. We prove (D.9) and (D.7) separately and omit the proof of (D.8) since it follows analogously by using (D.4).

Proof of (D.9)*:* With explanations given below,

$$\sup_{v\in\mathcal{K}(2s)} |v'(A\odot B)v| = \sup_{v\in\mathcal{K}(2s)} |\operatorname{tr}(AD_vBD_v)| \tag{D.10}$$

$$= \sup_{v\in\mathcal{K}(2s)} |\operatorname{tr}(AD_vD_s(v)BD_s(v)D_v)|$$

$$= \sup_{v\in\mathcal{K}(2s)} |v'(A\odot D_s(v)BD_s(v))v|$$

$$\le \sup_{v\in\mathcal{K}(2s)} \|A\odot D_s(v)BD_s(v)\|$$

$$\le \max_{i=1,\dots,d} |a_{ii}| \sup_{v\in\mathcal{K}(2s)} \|D_s(v)BD_s(v)\| \tag{D.11}$$

$$= \max_{i=1,\dots,d} |a_{ii}| \sup_{v\in\mathcal{K}(2s)} \max\{\lambda_{\max}(D_s(v)BD_s(v)), -\lambda_{\min}(D_s(v)BD_s(v))\} \tag{D.12}$$

$$= \max_{i=1,\dots,d} |a_{ii}| \sup_{v\in\mathcal{K}(2s)} \max\{\lambda_{\max}(D_s(v)BD_s(v)), \lambda_{\max}(D_s(v)(-B)D_s(v))\} \tag{D.13}$$

$$\le \max_{i=1,\dots,d} |a_{ii}| \sup_{v\in\mathcal{K}(2s)} |v'Bv|, \tag{D.14}$$

where (D.10) is due to Lemma 5.1.5 in Horn and Johnson (1991) and (D.11) follows by (D.1). The representations (D.12) and (D.13) can be used since B is symmetric. Finally, (D.14) follows by the min-max theorem for eigenvalues

$$\sup_{v\in\mathcal{K}(2s)} \lambda_{\max}(D_s(v)BD_s(v)) = \sup_{v\in\mathcal{K}(2s)} \sup_{x:\|x\|=1} x'D_s(v)BD_s(v)x \le \sup_{v\in\mathcal{K}(2s)} v'Bv, \tag{D.15}$$

since $\|D_s(v)x\| \le \|x\| = 1$ and $\|D_s(v)x\|_0 \le \|D_s(v)\|_0 \le 2s$.

Proof of (D.7)*:* Proceeding as for (D.9).

$$\sup_{v\in\mathcal{K}(2s)} |v'(A\odot B)v| = \sup_{v\in\mathcal{K}(2s)} |\operatorname{tr}(AD_vBD_v)|$$
$$= \sup_{v\in\mathcal{K}(2s)} |\operatorname{tr}(AD_vD_s(v)BD_s(v)D_v)|$$

$$
\begin{aligned}
&= \sup_{v\in\mathcal{K}(2s)} |\operatorname{tr}(D_v A D_v D_s(v) B D_s(v))| \\
&= \sup_{v\in\mathcal{K}(2s)} |\operatorname{tr}(D_v D_s(v) A D_s(v) D_v D_s(v) B D_s(v))| \\
&= \sup_{v\in\mathcal{K}(2s)} |v'(D_s(v) A D_s(v) \odot D_s(v) B D_s(v))v| \\
&\leq \sup_{v\in\mathcal{K}(2s)} \|D_s(v) A D_s(v) \odot D_s(v) B D_s(v)\| \\
&\leq \sup_{v\in\mathcal{K}(2s)} \|D_s(v) A D_s(v)\| \|D_s(v) B D_s(v)\| \qquad \text{(D.16)} \\
&\leq \sup_{v\in\mathcal{K}(2s)} |v'Av| \sup_{v\in\mathcal{K}(2s)} |v'Bv|, \qquad \text{(D.17)}
\end{aligned}
$$

where (D.16) follows by (D.2) and (D.17) by the same arguments to go from (D.11) to (D.14). □

Lemma D.2. *Let $A, \widetilde{A}, B \in \mathbb{R}^{d\times d}$ with $A = (a_{ij})_{i,j=1,\ldots,d}$, $\widetilde{A} = (\widetilde{a}_{ij})_{i,j=1,\ldots,d}$ and $B = (b_{ij})_{i,j=1,\ldots,d}$.*

(i) If $|a_{ij}| \leq b_{ij}$ for all $i, j = 1, \ldots, d$, then

$$\sup_{v\in\mathcal{K}(2s)} |v'Av| \leq \sup_{v\in\mathcal{K}(2s)} |v'Bv|. \qquad \text{(D.18)}$$

(ii) If $a_{ij} = a_j$, $\widetilde{a}_{ij} = \widetilde{a}_i$ for $i, j = 1, \ldots, d$ and B symmetric, then

$$\sup_{v\in\mathcal{K}(2s)} |v'(A \odot \widetilde{A} \odot B)v| \leq \max_{j=1,\ldots,d} |a_j| \max_{i=1,\ldots,d} |\widetilde{a}_i| \sup_{v\in\mathcal{K}(2s)} |v'Bv|. \qquad \text{(D.19)}$$

Proof of Lemma D.2. We prove the two inequalities (D.18) and (D.19) separately.

Proof of (D.19)*:* The ideas follow the proof of Lemma D.1. With further explanations given below, we have

$$
\begin{aligned}
&\sup_{v\in\mathcal{K}(2s)} |v'(A \odot \widetilde{A} \odot B)v| \\
&\leq \sup_{v\in\mathcal{K}(2s)} \|A \odot \widetilde{A} \odot D_s(v) B D_s(v)\| \\
&= \sup_{v\in\mathcal{K}(2s)} \|\operatorname{diag}(\widetilde{a}_1, \ldots, \widetilde{a}_p) D_s(v) B D_s(v) \operatorname{diag}(a_1, \ldots, a_p)\| \qquad \text{(D.20)} \\
&\leq \max_{j=1,\ldots,d} |a_j| \max_{i=1,\ldots,d} |\widetilde{a}_i| \sup_{v\in\mathcal{K}(2s)} \|D_s(v) B D_s(v)\| \qquad \text{(D.21)} \\
&\leq \max_{j=1,\ldots,d} |a_j| \max_{i=1,\ldots,d} |\widetilde{a}_i| \sup_{v\in\mathcal{K}(2s)} |v'Bv|. \qquad \text{(D.22)}
\end{aligned}
$$

Given that $A = (a_{ij})_{i,j=1,\ldots,d}$ with $a_{ij} = a_j$ for $j = 1, \ldots, d$, one has $A \odot C = C \operatorname{diag}(a_1, \ldots, a_d)$ so that $\|A \odot C\| = \|C \operatorname{diag}(a_1, \ldots, a_d)\| \leq \max_{j=1,\ldots,d} |a_j| \|C\|$ due to the submultiplicativity of the spectral norm (and similarly for $\widetilde{A}$), which explains (D.20) and (D.21). The inequality (D.22) follows from (D.11) to (D.14).

Proof of (D.18)*:* If $|a_{ij}| \leq b_{ij}$ for all $i, j = 1, \ldots, d$, then

$$\sup_{v \in \mathcal{K}(2s)} |v'Av| \leq \sup_{v \in \mathcal{K}(2s)} \sum_{i,j=1}^{d} |v_i||a_{ij}||v_j| \leq \sup_{v \in \mathcal{K}(2s)} \sum_{i,j=1}^{d} |v_i|b_{ij}|v_j| \leq \sup_{v \in \mathcal{K}(2s)} |v'Bv|.$$

□

D.2. Moment conditions

The results in this section ensure that the constants in our main results are finite. Lemmas D.3, D.4 and D.5 consider respectively the quantities in (2.10), (2.11) and (2.12).

Lemma D.3. *Suppose Assumptions M.2 and M.3. Then, for an open set S,*

$$\begin{aligned} \sup_{\theta_i \in S} \Delta_i &:= \sup_{\theta_i \in S} \sum_{n=0}^{\infty} n \left\| \nabla_{\theta_i} C_{i,n} \right\|_1 \\ &= \sup_{\theta_i \in S} \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2u} Q_{i,n}^2\right) n \|\nabla_{\theta_i} Q_{i,n}\|_1 < \infty. \end{aligned} \tag{D.23}$$

Proof. We have

$$\begin{aligned} &\sum_{n=0}^{\infty} n \left\| \nabla_{\theta_i} C_n(\theta_i) \right\|_1 \\ &= \sum_{n=0}^{\infty} n (\mathrm{P}[X_{i,t} > n])^{\frac{1}{2}} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \left\| \nabla_{\theta_i} C_n(\theta_i) \right\|_1 \\ &\leq (\mathrm{E}\,|X_{i,t}|^2)^{\frac{1}{2}} \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \left\| \nabla_{\theta_i} C_n(\theta_i) \right\|_1 \qquad \text{(D.24)} \\ &= (\mathrm{E}\,|X_{i,t}|^2)^{\frac{1}{2}} \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \sum_{j=1}^{K_i} \left| \frac{\partial}{\partial \theta_{ij}} \mathrm{P}[X_{i,t} > n] \right| < \infty, \qquad \text{(D.25)} \end{aligned}$$

where we applied Markov's inequality in (D.24). After taking the supremum over all θ_i in an open set S on both sides of (D.25), the expression (D.25) is uniformly bounded due to Assumption M.3. For the equality in (D.23), let ϕ denote the Gaussian density and note that

$$\nabla_{\theta_i} Q_{i,n} = \nabla_{\theta_i} \Phi^{-1}(C_{i,n}) = \frac{1}{\phi(\Phi^{-1}(C_{i,n}))} \nabla_{\theta_i} C_n(\theta_i) = \frac{1}{\phi(Q_{i,n})} \nabla_{\theta_i} C_n(\theta_i), \tag{D.26}$$

since $Q_{i,n} = \Phi^{-1}(C_{i,n})$ with $C_{i,n} = \mathrm{P}[X_{i,t} \leq n] = \sum_{j=0}^{n} \mathrm{P}[X_{i,t} = j] = \sum_{j=0}^{n} p_{\theta_i,t}(j) = C_n(\theta_i)$. Then, the relation (D.26) allows us to write

$$\sum_{n=0}^{\infty} n \left\|\nabla_{\theta_i} C_{i,n}\right\|_1 = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2u} Q_{i,n}^2\right) n \|\nabla_{\theta_i} Q_{i,n}\|_1.$$

□

The following lemma is similar to Lemma 2.1 in Jia et al. (2023) and coincides with it for $u = 1$.

Lemma D.4. *Suppose $u > 0$ and, Assumption M.2 is satisfied for some $p > u$. Then, for an open set S and any $k \in \mathbb{N}_0$,*

$$\sup_{\theta_i \in S} m_i^{(k)}(u) := \sup_{\theta_i \in S} \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2u} Q_{i,n}^2\right) |Q_{i,n}|^k < \infty. \tag{D.27}$$

Proof. We follow the proof of Lemma 2.1 in Jia et al. (2023). Note that by Mill's ratio, we have

$$1 - \Phi(x) \sim e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi} x}, \quad \text{as } x \to \infty. \tag{D.28}$$

Then, substituting $x = \Phi^{-1}(y)$ in (D.28) leads $1 - y \sim e^{-\frac{\Phi^{-1}(y)^2}{2}} \frac{1}{\sqrt{2\pi}\Phi^{-1}(y)}$, as $y \uparrow 1$. Taking the logarithm on both sides, we get

$$\log(1-y) \sim -\frac{\Phi^{-1}(y)^2}{2} - \log(\sqrt{2\pi}\Phi^{-1}(y)), \quad \text{as } y \uparrow 1, \tag{D.29}$$

and can infer

$$\sqrt{2} |\log(1-y)|^{\frac{1}{2}} \sim \Phi^{-1}(y), \quad \text{as } y \uparrow 1. \tag{D.30}$$

Finally, applying (D.29) and (D.30) for $y = C_{i,n}$ and substitution into (D.27) with $Q_{i,n} = \Phi^{-1}(C_{i,n})$ show that $m_i^{(k)}(u)$ can be bounded (up to a constant) by

$$\begin{aligned}
&\sum_{n=0}^{\infty} \exp\left(\frac{1}{2u} 2\log(1 - C_{i,n})\right) (2|\log(1 - C_{i,n})|)^{\frac{k}{2}} \\
&\leq c \sum_{n=0}^{\infty} (1 - C_{i,n})^{\frac{1}{u}} |\log(1 - C_{i,n})|^{\frac{k}{2}} \\
&\leq c \sum_{n=0}^{\infty} (1 - C_{i,n})^{\frac{1}{u} - \frac{k}{2}\delta} \left(\frac{1}{\delta}\right)^{\frac{k}{2}} && \text{(D.31)} \\
&= c \sum_{n=0}^{\infty} \mathrm{P}[X_{i,t} > n]^{\frac{1}{u} - \frac{k}{2}\delta} \left(\frac{1}{\delta}\right)^{\frac{k}{2}}, && \text{(D.32)}
\end{aligned}$$

where (D.31) follows since for any $\delta > 0$ and $x \in (0, 1)$, $-\log(x) \leq \frac{x^{-\delta}}{\delta}$, and (D.32) follows since $C_{i,n} = 1 - \mathrm{P}[X_{i,t} > n]$. Using Markov's inequality with $\mathrm{P}[X_{i,t} > n] = \mathrm{P}[X_{i,t}^p > n^p] \leq \mathrm{E}\,|X_{i,t}|^p / n^p$, we get

$$\sum_{n=0}^{\infty} \mathrm{P}[X_{i,t} > n]^{\frac{1}{u} - \frac{k}{2}\delta} \leq (\mathrm{E}\,|X_{i,t}|^p)^{(\frac{1}{u} - \frac{k}{2}\delta)} \sum_{n=0}^{\infty} \frac{1}{n^{p(\frac{1}{u} - \frac{k}{2}\delta)}}, \tag{D.33}$$

which converges as long as $p(\frac{1}{u}-\frac{k}{2}\delta)>1$ is satisfied. After taking the supremum of $m_i^{(k)}(u)$ over all θ_i in an open set S, the expression is uniformly bounded due to (D.33) and Assumption M.3. □

Lemma D.5. *Suppose $u\in(0,2)$ and Assumption M.3. Then, for an open set S and any $k\in\mathbb{N}_0$,*

$$\sup_{\theta_i\in S}\mu_i^{(k)}(u):=\sup_{\theta_i\in S}\frac{1}{\sqrt{2\pi}}\sum_{n=0}^{\infty}\exp\left(-\frac{1}{2u}Q_{i,n}^2\right)|Q_{i,n}|^k\|\nabla_{\theta_i}Q_{i,n}\|_1<\infty.$$

Proof. Recall $\nabla_{\theta_i}Q_{i,n}=\frac{1}{\phi(Q_{i,n})}\nabla_{\theta_i}C_n(\theta_i)$ in (D.26). Then,

$$\begin{aligned}
&\sum_{n=0}^{\infty}\exp\left(-\frac{1}{2u}Q_{i,n}^2\right)|Q_{i,n}|^k\|\nabla_{\theta_i}Q_{i,n}\|_1\\
&=\sum_{n=0}^{\infty}\exp\left(-\frac{1}{2u}Q_{i,n}^2\right)|Q_{i,n}|^k\frac{1}{\phi(Q_{i,n})}\|\nabla_{\theta_i}C_n(\theta_i)\|_1 &\text{(D.34)}\\
&=\sqrt{2\pi}\sum_{n=0}^{\infty}\exp\left(-\frac{1-u}{2u}Q_{i,n}^2\right)|Q_{i,n}|^k\|\nabla_{\theta_i}C_n(\theta_i)\|_1, &\text{(D.35)}
\end{aligned}$$

where (D.34) results from substituting (D.26). In order to continue bounding (D.35), we take advantage of several relations derived in the proof of Lemma D.4.

$$\begin{aligned}
&\sum_{n=0}^{\infty}\exp\left(-\frac{1-u}{2u}Q_{i,n}^2\right)|Q_{i,n}|^k\|\nabla_{\theta_i}C_n(\theta_i)\|_1\\
&\le c\sum_{n=0}^{\infty}(1-C_n(\theta_i))^{\frac{1-u}{u}}(2|\log(1-C_n(\theta_i))|)^{\frac{k}{2}}\|\nabla_{\theta_i}C_n(\theta_i)\|_1 &\text{(D.36)}\\
&\le c\sum_{n=0}^{\infty}(1-C_n(\theta_i))^{\frac{1-u}{u}-\frac{k}{2}\delta}\left(\frac{1}{\delta}\right)^{\frac{k}{2}}\|\nabla_{\theta_i}C_n(\theta_i)\|_1 &\text{(D.37)}\\
&\le c\sum_{n=0}^{\infty}(\mathrm{P}[X_{i,t}>n])^{-\frac{1}{2}}\|\nabla_{\theta_i}C_n(\theta_i)\|_1 &\text{(D.38)}\\
&\le c\sum_{n=0}^{\infty}(\mathrm{P}[X_{i,t}>n])^{-\frac{1}{2}}\sum_{j=1}^{K_i}\left|\frac{\partial}{\partial\theta_{ij}}\mathrm{P}[X_{i,t}>n]\right|<\infty, &\text{(D.39)}
\end{aligned}$$

where (D.36) follows by (D.29) and (D.30). Since $-\log(x)\le\frac{x^{-\delta}}{\delta}$ for any $\delta>0$ and $x\in(0,1)$ the inequality (D.37) follows. In (D.38), we choose δ such that $\frac{1-u}{u}-\frac{k}{2}\delta\ge-\frac{1}{2}$ for $u\in(0,2)$. After taking the supremum over all θ_i in an open set S on both sides of (D.39), the expression in (D.39) is uniformly bounded due to Assumption M.3. □

D.3. Properties of higher order derivatives of the link function

This section collects and derives all bounds and derivatives of the link function as needed throughout this paper.

We start with the first two derivatives of ℓ_{ij} and its inverse. For shortness sake, we introduce the following notation:

$$G_{ij}^{n_0,n_1}(u) := Q_{i,n_0}^2 + Q_{j,n_1}^2 - 2uQ_{i,n_0}Q_{j,n_1}, \quad g_{ij}^{n_0,n_1} := Q_{i,n_0}Q_{j,n_1}$$

and

$$\begin{aligned} S_1(u) &:= \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right), \\ S_2^G(u) &:= \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) G_{ij}^{n_0,n_1}(u), \\ S_3^g(u) &:= \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) g_{ij}^{n_0,n_1}. \end{aligned} \tag{D.40}$$

Then, the first derivative of the link function in Proposition 2.1 can be written as

$$\begin{aligned} \ell_{ij}'(u) &= \frac{1}{2\pi\sqrt{1-u^2}} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)}(Q_{i,n_0}^2 + Q_{j,n_1}^2 - 2uQ_{i,n_0}Q_{j,n_1})\right) \\ &= \frac{1}{2\pi\sqrt{1-u^2}} S_1(u). \end{aligned} \tag{D.41}$$

In order to derive the higher order derivatives of ℓ_{ij}, we first derive the first derivatives of the quantities in (D.40):

$$\frac{\partial}{\partial u} S_1(u) = \frac{-u}{(1-u^2)^2} S_2^G(u) + \frac{1}{1-u^2} S_3^g(u), \tag{D.42}$$

$$\begin{aligned} \frac{\partial}{\partial u} S_2^G(u) = &-2S_3^g(u) \\ &- \frac{u}{(1-u^2)^2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) (G_{ij}^{n_0,n_1}(u))^2 \\ &+ \frac{1}{1-u^2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) G_{ij}^{n_0,n_1}(u) g_{ij}^{n_0,n_1}, \end{aligned} \tag{D.43}$$

$$\begin{aligned} \frac{\partial}{\partial u} S_3^g(u) = &-\frac{u}{(1-u^2)^2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) G_{ij}^{n_0,n_1}(u) g_{ij}^{n_0,n_1} \\ &+ \frac{1}{1-u^2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) (g_{ij}^{n_0,n_1})^2. \end{aligned} \tag{D.44}$$

Using the introduced notation in (D.40) and the corresponding derivative (D.42), we can then write the second derivative of ℓ_{ij} as

$$\begin{aligned}\ell''_{ij}(u) &= \frac{\partial}{\partial u}\frac{1}{2\pi\sqrt{1-u^2}}S_1(u)\\ &= \frac{u}{2\pi(1-u^2)^{\frac{3}{2}}}S_1(u) - \frac{u}{2\pi(1-u^2)^{\frac{5}{2}}}S_2^G(u) + \frac{1}{2\pi(1-u^2)^{\frac{3}{2}}}S_3^g(u). \quad \text{(D.45)}\end{aligned}$$

The derived quantities will help expressing the first and second derivatives of the inverse link function $g = \ell^{-1}$. For this, note first that

$$(\ell^{-1})'(x) = \frac{1}{\ell'(\ell^{-1}(x))}, \quad \text{(D.46)}$$

$$(\ell^{-1})''(x) = \frac{\partial}{\partial x}\frac{1}{\ell'(\ell^{-1}(x))} = -\frac{\ell''(\ell^{-1}(x))}{(\ell'(\ell^{-1}(x)))^3} =: f(\ell^{-1}(x)) \quad \text{with} \quad f(u) = -\frac{\ell''(u)}{(\ell'(u))^3}. \quad \text{(D.47)}$$

The following Sections D.3.1, D.3.2 and D.3.3 consider respectively the derivatives used in Sections C.1, C.2 and C.3. We aim to express all bounds in terms of $M(\boldsymbol{c},\varepsilon)$, $\mu(\boldsymbol{c},\varepsilon)$, $M_1(\boldsymbol{c},\varepsilon)$ and $M_2(\boldsymbol{c},\varepsilon)$ in (2.14) and (2.15). All bounds are simple consequences of finding upper and lower bounds on $G_{ij}^{n_0,n_1}(u)$. Those bounds can be derived easily through $a^2+b^2-2uab \le a^2+b^2+2|ab||u| \le (1+\boldsymbol{c})(a^2+b^2)$ and $a^2+b^2-2uab \ge a^2+b^2-2|ab||u| \ge (1-\boldsymbol{c})(a^2+b^2)$ for $0<\boldsymbol{c}<1$ and $|u|<\boldsymbol{c}$.

D.3.1. Derivatives I

In this section, we consider the derivatives and their bounds used in Section C.1.

Lemma D.6. *For $f_{ij}(u)$ in (D.47) and $|u|<\boldsymbol{c}$ for $\boldsymbol{c}\in(0,1)$, we have*

$$|f_{ij}(u)| \le \frac{6}{1-\boldsymbol{c}^2}\max_{k=0,2}\max_{i=1,\dots,d}\frac{\left(m_i^{(k)}(1+\boldsymbol{c})\right)^2}{\left(m_i^{(0)}(1-\boldsymbol{c})\right)^6} \le \frac{6}{1-\boldsymbol{c}^2}M_1(\boldsymbol{c},0)M_2(\boldsymbol{c},0). \quad \text{(D.48)}$$

Proof. Using (D.47), (D.45) and (D.41), we can write

$$\begin{aligned}&\left|-\frac{\ell''_{ij}(u)}{(\ell'_{ij}(u))^3}\right|\\ &= \frac{\left|\frac{u}{2\pi(1-u^2)^{\frac{3}{2}}}S_1(u) - \frac{u}{2\pi(1-u^2)^{\frac{5}{2}}}S_2^G(u) + \frac{1}{2\pi(1-u^2)^{\frac{3}{2}}}S_3^g(u)\right|}{\left(\frac{1}{2\pi\sqrt{1-u^2}}S_1(u)\right)^3}\\ &= (2\pi)^2\left|u\frac{1}{S_1^2(u)} - \frac{u}{1-u^2}\frac{S_2^G(u)}{S_1^3(u)} + \frac{S_3^g(u)}{S_1^3(u)}\right| \quad \text{(D.49)}\end{aligned}$$

$$\leq \frac{3}{1-\boldsymbol{c}^2} \frac{m_j^{(0)}\,(1+\boldsymbol{c})\,m_i^{(0)}\,(1+\boldsymbol{c}) + m_i^{(2)}\,(1+\boldsymbol{c})\,m_j^{(0)}\,(1+\boldsymbol{c}) + m_j^{(0)}\,(1+\boldsymbol{c})\,m_i^{(2)}\,(1+\boldsymbol{c})}{\left(m_i^{(0)}\,(1-\boldsymbol{c})\,m_j^{(0)}\,(1-\boldsymbol{c})\right)^3} \tag{D.50}$$

$$\leq \frac{6}{1-\boldsymbol{c}^2} \max_{k=0,2} \max_{i=1,\ldots,d} \frac{\left(m_i^{(k)}\,(1+\boldsymbol{c})\right)^2}{\left(m_i^{(0)}\,(1-\boldsymbol{c})\right)^6},$$

where the quantities in (D.49) are bounded separately below to get (D.50).

Bound on $\frac{1}{S_1(u)}$*:*

$$\begin{aligned}
\frac{1}{S_1(u)} &= \left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right)\right)^{-1} \\
&\leq \left(\sum_{n=0}^{\infty} \exp\left(-\frac{1}{2(1-\boldsymbol{c})} Q_{i,n}^2\right) \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2(1-\boldsymbol{c})} Q_{j,n}^2\right)\right)^{-1} && \text{(D.51)} \\
&= \frac{1}{2\pi}\left(m_i^{(0)}\,(1-\boldsymbol{c})\,m_j^{(0)}\,(1-\boldsymbol{c})\right)^{-1}. && \text{(D.52)}
\end{aligned}$$

where (D.51) follows since $a^2+b^2-2uab \leq a^2+b^2+2|ab||u| \leq (1+|u|)(a^2+b^2)$ for all $a,b \in \mathbb{R}$.

Bound on $S_2^G(u)$*:*

$$\begin{aligned}
\frac{1}{2\pi}\left|S_2^G(u)\right| &\leq \frac{1}{2\pi} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) |G_{ij}^{n_0,n_1}(u)| \\
&\leq \frac{1+\boldsymbol{c}}{2\pi} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1-|u|}{2(1-u^2)} (Q_{i,n_0}^2 + Q_{j,n_1}^2)\right) (Q_{i,n_0}^2 + Q_{j,n_1}^2) && \text{(D.53)} \\
&\leq 2\left(m_i^{(2)}\,(1+\boldsymbol{c})\,m_j^{(0)}\,(1+\boldsymbol{c}) + m_i^{(0)}\,(1+\boldsymbol{c})\,m_j^{(2)}\,(1+\boldsymbol{c})\right),
\end{aligned}$$

where (D.53) follows since $a^2+b^2-2uab \geq a^2+b^2-2|ab||u| \geq (1-|u|)(a^2+b^2)$ for all $a,b \in \mathbb{R}$.

Bound on $S_3^g(u)$*:* Similarly as above,

$$\begin{aligned}
\frac{1}{2\pi}\left|S_3^g(u)\right| &\leq \frac{1}{2\pi} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) |g_{ij}^{n_0,n_1}| \\
&\leq \frac{1}{2\pi} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1-|u|}{2(1-u^2)} (Q_{i,n_0}^2 + Q_{j,n_1}^2)\right) |Q_{i,n_0} Q_{j,n_1}| && \text{(D.54)} \\
&\leq m_i^{(1)}\,(1+\boldsymbol{c})\,m_j^{(1)}\,(1+\boldsymbol{c})
\end{aligned}$$

$$\leq m_i^{(2)}\,(1+\boldsymbol{c})\,m_j^{(0)}\,(1+\boldsymbol{c}) + m_i^{(0)}\,(1+\boldsymbol{c})\,m_j^{(2)}\,(1+\boldsymbol{c})\,,$$

where the last inequality follows from $|Q_{i,n_0}Q_{j,n_1}| \leq (Q_{i,n_0}^2 + Q_{j,n_1}^2)/2 \leq Q_{i,n_0}^2 + Q_{j,n_1}^2$.

Finally, for the last relation in (D.48), note that $1+\boldsymbol{c} \leq \frac{1}{1-\boldsymbol{c}}$. □

D.3.2. Derivatives II

In this section, we consider the derivatives and their bounds used in Section C.2.

Lemma D.7. *For $z_{ij}(y)$ in (2.7), $-z'(0) = (-z'_{ij}(0))_{i,j=1,\dots,d}$ is positive semidefinite and*

$$|z'_{ii}(0)| \leq \max_{i=1,\dots,d} \frac{\left(m_i^{(1)}(1)\right)^2}{\left(m_i^{(0)}(1)\right)^4} \leq M_2(0,0). \tag{D.55}$$

Proof. The componentwise derivative of z can be derived as

$$\begin{aligned} z'_{ij}(y) &= \frac{\partial}{\partial y}\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right)\right)^{-1} \\ &= -\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right)\right)^{-2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right) g_{ij}^{n_0,n_1}. \end{aligned} \tag{D.56}$$

In particular, z' evaluated at zero, gives

$$\begin{aligned} -z'(0) &= \left(\frac{\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}(Q_{i,n_0}^2 + Q_{j,n_1}^2)\right) Q_{i,n_0}Q_{j,n_1}}{\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}(Q_{i,n_0}^2 + Q_{j,n_1}^2)\right)\right)^2}\right)_{i,j=1,\dots,d} \\ &= \left(\frac{m_1^{(1)}(1)}{\left(m_1^{(0)}(1)\right)^2}, \dots, \frac{m_d^{(1)}(1)}{\left(m_d^{(0)}(1)\right)^2}\right)' \left(\frac{m_1^{(1)}(1)}{\left(m_1^{(0)}(1)\right)^2}, \dots, \frac{m_d^{(1)}(1)}{\left(m_d^{(0)}(1)\right)^2}\right) \succcurlyeq 0. \end{aligned}$$

Furthermore,

$$|z'_{ii}(0)| = \frac{\left(\sum_{n=0}^{\infty} \exp\left(-\frac{1}{2}Q_{i,n}^2\right) Q_{i,n}\right)^2}{\left(\sum_{n=0}^{\infty} \exp\left(-\frac{1}{2}Q_{i,n}^2\right)\right)^4} = \frac{\left(m_i^{(1)}(1)\right)^2}{\left(m_i^{(0)}(1)\right)^4}.$$

□

Lemma D.8. *For $z_{ij}(y)$ in (2.7) and $|y| < \boldsymbol{c}$ for $\boldsymbol{c} \in (0,1)$,*

$$|z''_{ij}(y)| \leq 3 \max_{i=1,\ldots,d} \frac{\left(m_i^{(2)}\left(\frac{1}{1-\boldsymbol{c}}\right)\right)^2}{\left(m_i^{(0)}(1-\boldsymbol{c})\right)^4} \leq 3M_2(\boldsymbol{c},0). \tag{D.57}$$

Proof. Continuing from (D.56),

$$\begin{aligned}
&z''_{ij}(y)\\
&= -\frac{\partial}{\partial y}\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right)\right)^{-2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right) g_{ij}^{n_0,n_1}\\
&= 2\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right)\right)^{-3} \left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right) g_{ij}^{n_0,n_1}\right)^2\\
&\quad -\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right)\right)^{-2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right) \left(g_{ij}^{n_0,n_1}\right)^2.
\end{aligned} \tag{D.58}$$

Then,

$$\begin{aligned}
|z''_{ij}(y)| &\leq 3\frac{\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right)\left(Q_{i,n_0}Q_{j,n_1}\right)^2}{\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2}G_{ij}^{n_0,n_1}(y)\right)\right)^2} && \text{(D.59)}\\
&\leq 3\frac{\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1-\boldsymbol{c}}{2}(Q_{i,n_0}^2+Q_{j,n_1}^2)\right)\left(Q_{i,n_0}Q_{j,n_1}\right)^2}{\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1+\boldsymbol{c}}{2}(Q_{i,n_0}^2+Q_{j,n_1}^2)\right)\right)^2} && \text{(D.60)}\\
&\leq 3\frac{\sum_{n=0}^{\infty} \exp\left(-\frac{1-\boldsymbol{c}}{2}Q_{i,n}^2\right)Q_{i,n}^2 \sum_{n=0}^{\infty} \exp\left(-\frac{1-\boldsymbol{c}}{2}Q_{j,n}^2\right)Q_{j,n}^2}{\left(\sum_{n=0}^{\infty} \exp\left(-\frac{1+\boldsymbol{c}}{2}Q_{i,n}^2\right)\sum_{n=0}^{\infty} \exp\left(-\frac{1+\boldsymbol{c}}{2}Q_{j,n}^2\right)\right)^2}\\
&\leq 3\frac{m_i^{(2)}\left(\frac{1}{1-\boldsymbol{c}}\right)m_j^{(2)}\left(\frac{1}{1-\boldsymbol{c}}\right)}{\left(m_i^{(0)}(1-\boldsymbol{c})m_j^{(0)}(1-\boldsymbol{c})\right)^2} && \text{(D.61)}\\
&\leq 3\max_{i=1,\ldots,d}\frac{\left(m_i^{(2)}\left(\frac{1}{1-\boldsymbol{c}}\right)\right)^2}{\left(m_i^{(0)}(1-\boldsymbol{c})\right)^4},
\end{aligned}$$

where (D.59) follows from (D.58) and the Cauchy-Schwarz inequality. Finally, (D.60) is a consequence of $a^2+b^2-2uab \leq a^2+b^2+2|ab||u| \leq (1+\boldsymbol{c})(a^2+b^2)$ and $a^2+b^2-2uab \geq a^2+b^2-2|ab||u| \leq (1-\boldsymbol{c})(a^2+b^2)$. □

Lemma D.9. *Suppose there is a constant $\boldsymbol{c} \in (0,1)$ such that $|\Gamma_{Z,ij}(h)| < \boldsymbol{c}$ for all $i \neq j$ and $h \neq 0$. Then, for $Z_{ij}(x)$ in (2.7) and $|x| < \boldsymbol{c}^2$ for $\boldsymbol{c} \in (0,1)$,*

$$|Z'_{ij}(x)| \leq \frac{1}{(1-\boldsymbol{c}^2)^2}M_2(\boldsymbol{c},0). \tag{D.62}$$

Proof. Set $\sigma_{ij} := \Gamma_{Z,ij}(h)$. Then, for $i \neq j$ and $h \neq 0$,

$$\begin{aligned} Z'_{ij}(x) &= \frac{\partial}{\partial x}\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-x)} G_{ij}^{n_0,n_1}(\sigma_{ij})\right)\right)^{-1} \\ &= -\frac{\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-x)} G_{ij}^{n_0,n_1}(\sigma_{ij})\right) \frac{-1}{2(1-x)^2} G_{ij}^{n_0,n_1}(\sigma_{ij})}{\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-x)} G_{ij}^{n_0,n_1}(\sigma_{ij})\right)\right)^2}. \end{aligned}$$

Note that $G_{ij}^{n_0,n_1}(u) = Q_{i,n_0}^2 + Q_{j,n_1}^2 - 2uQ_{i,n_0}Q_{j,n_1} \geq (Q_{i,n_0}u + Q_{j,n_1})^2 \geq 0$. Then, with explanations given below,

$$\begin{aligned} &|Z'_{ij}(x)| \\ &\leq \frac{\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2} G_{ij}^{n_0,n_1}(\sigma_{ij})\right) \frac{1}{2(1-x)^2} |G_{ij}^{n_0,n_1}(\sigma_{ij})|}{\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-x)} G_{ij}^{n_0,n_1}(\sigma_{ij})\right)\right)^2} \\ &\leq \frac{1}{2(1-\boldsymbol{c}^2)^2} \frac{\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1-\boldsymbol{c}}{2}(Q_{i,n_0}^2 + Q_{j,n_1}^2)\right)(1+\boldsymbol{c})(Q_{i,n_0}^2 + Q_{j,n_1}^2)}{\left(\sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1+\boldsymbol{c}}{2(1-\boldsymbol{c}^2)}(Q_{i,n_0}^2 + Q_{j,n_1}^2)\right)\right)^2} \qquad \text{(D.63)} \\ &\leq \frac{1}{(1-\boldsymbol{c}^2)^2}\left(\sum_{n=0}^{\infty} \exp\left(-\frac{1+\boldsymbol{c}}{2(1-\boldsymbol{c}^2)} Q_{i,n}^2\right) \sum_{n=0}^{\infty} \exp\left(-\frac{1+\boldsymbol{c}}{2(1-\boldsymbol{c}^2)} Q_{j,n}^2\right)\right)^{-2} \\ &\qquad \times \left(\sum_{n=0}^{\infty} \exp\left(-\frac{1-\boldsymbol{c}}{2} Q_{i,n}^2\right) \sum_{n=0}^{\infty} \exp\left(-\frac{1-\boldsymbol{c}}{2} Q_{j,n}^2\right) Q_{j,n}^2 \right. \\ &\qquad\qquad \left. + \sum_{n=0}^{\infty} \exp\left(-\frac{1-\boldsymbol{c}}{2} Q_{j,n}^2\right) \sum_{n=0}^{\infty} \exp\left(-\frac{1-\boldsymbol{c}}{2} Q_{i,n}^2\right) Q_{i,n}^2\right) \\ &= \frac{1}{(1-\boldsymbol{c}^2)^2} \frac{m_i^{(0)}\left(\frac{1}{1-\boldsymbol{c}}\right) m_j^{(2)}\left(\frac{1}{1-\boldsymbol{c}}\right) + m_i^{(2)}\left(\frac{1}{1-\boldsymbol{c}}\right) m_j^{(0)}\left(\frac{1}{1-\boldsymbol{c}}\right)}{\left(m_i^{(0)}(1-\boldsymbol{c})\, m_j^{(0)}(1-\boldsymbol{c})\right)^2} \qquad \text{(D.64)} \\ &\leq \frac{1}{(1-\boldsymbol{c}^2)^2} \max_{r=0,2} \max_{i=1,\ldots,d} \frac{\left(m_i^{(r)}\left(\frac{1}{1-\boldsymbol{c}}\right)\right)^2}{\left(m_i^{(0)}(1-\boldsymbol{c})\right)^4}, \end{aligned}$$

where (D.63) is a consequence of $a^2 + b^2 - 2uab \leq a^2 + b^2 + 2|ab||u| \leq (1+\boldsymbol{c})(a^2+b^2)$ and $a^2 + b^2 - 2uab \geq a^2 + b^2 - 2|ab||u| \leq (1-\boldsymbol{c})(a^2+b^2)$. □

D.3.3. Derivatives III

In this section, we consider the derivatives and their bounds used in Section C.3. We define $\widetilde{\Delta}(0), \widetilde{M}(\boldsymbol{c},0)$ and $\widetilde{\mu}(\boldsymbol{c},0)$ as $M(\boldsymbol{c},0)$ and $\mu(\boldsymbol{c},0)$ in (2.13)–(2.15) with

the true θ replaced by some $\widetilde{\theta}$, further specified in the proofs below. Similarly, we define $\widehat{S}_1(u), \widehat{S}_2^G(u)$ and $\widehat{S}_3^g(u)$ as the estimated counterparts of $S_1(u), S_2^G(u)$ and $S_3^g(u)$ by replacing θ_i with $\widehat{\theta}_i$.

Lemma D.10. *Suppose $|u| < \boldsymbol{c}$ for $\boldsymbol{c} \in (0,1)$. Then, the second derivative of the link function given in (D.45) satisfies, for some $\widetilde{\theta}_i, \widetilde{\theta}_j$ such that $|\widetilde{\theta}_i - \theta_i| < |\widehat{\theta}_i - \theta_i|$, $|\widetilde{\theta}_j - \theta_j| < |\widehat{\theta}_j - \theta_j|$ and $|u| < \boldsymbol{c}$ for $\boldsymbol{c} \in (0,1)$,*

$$|\widehat{\ell}''_{ij}(u) - \ell''_{ij}(u)| \leq \frac{18}{(1-\boldsymbol{c}^2)^{\frac{7}{2}}} \widetilde{M}(\boldsymbol{c},0)\widetilde{\mu}(\boldsymbol{c},0)\|\widehat{\theta} - \theta\|_{\max}. \tag{D.65}$$

Proof. From (D.45), we have

$$\begin{aligned}
&|\widehat{\ell}''_{ij}(u) - \ell''_{ij}(u)| \\
&= \left| \frac{u}{2\pi(1-u^2)^{\frac{3}{2}}}(\widehat{S}_1(u) - S_1(u)) - \frac{u}{2\pi(1-u^2)^{\frac{5}{2}}}(\widehat{S}_2^G(u) - S_2^G(u)) \right. \\
&\qquad \left. + \frac{1}{2\pi(1-u^2)^{\frac{3}{2}}}(\widehat{S}_3^g(u) - S_3^g(u)) \right| \\
&\leq \frac{1}{2\pi(1-\boldsymbol{c}^2)^{\frac{5}{2}}}(|\widehat{S}_1(u) - S_1(u)| + |\widehat{S}_2^G(u) - S_2^G(u)| + |\widehat{S}_3^g(u) - S_3^g(u)|) \qquad \text{(D.66)} \\
&\leq \frac{1}{2\pi(1-\boldsymbol{c}^2)^{\frac{5}{2}}} \max_{i,j=1,\ldots,d} \Big(\|\nabla_{\theta_i} S_1(u)|_{(\widetilde{\theta}_i,\widetilde{\theta}_j)}\|_1 + \|\nabla_{\theta_i} S_2^G(u)|_{(\widetilde{\theta}_i,\widetilde{\theta}_j)}\|_1 \\
&\qquad + \|\nabla_{\theta_i} S_3^g(u)|_{(\widetilde{\theta}_i,\widetilde{\theta}_j)}\|_1 \Big) \|\widehat{\theta} - \theta\|_{\max} \qquad \text{(D.67)} \\
&\leq \frac{18}{(1-\boldsymbol{c}^2)^{\frac{7}{2}}} \max_{k=0,\ldots,3} \max_{i=1,\ldots,d} \widetilde{m}_i^{(k)}(1+\boldsymbol{c}) \max_{k=0,\ldots,3} \max_{i=1,\ldots,d} \widetilde{\mu}_i^{(k)}(1+\boldsymbol{c}) \|\widehat{\theta} - \theta\|_{\max} \\
&= \frac{18}{(1-\boldsymbol{c}^2)^{\frac{7}{2}}} \widetilde{M}(\boldsymbol{c},0)\widetilde{\mu}(\boldsymbol{c},0)\|\widehat{\theta} - \theta\|_{\max}.
\end{aligned}$$

We consider the first summand in (D.66) in detail. The remaining ones can be handled analogously.

$$\begin{aligned}
&|\widehat{S}_1(u) - S_1(u)| \\
&= |\langle \nabla_{\theta_i} S_1(u)|_{(\widetilde{\theta}_i,\widetilde{\theta}_j)}, \widehat{\theta}_j - \theta_j\rangle + \langle \nabla_{\theta_j} S_1(u)|_{(\widetilde{\theta}_i,\widetilde{\theta}_j)}, \widehat{\theta}_i - \theta_i\rangle| \\
&\leq \|\nabla_{\theta_i} S_1(u)|_{(\widetilde{\theta}_i,\widetilde{\theta}_j)}\|_1 \|\widehat{\theta}_j - \theta_j\|_{\max} + \|\nabla_{\theta_j} S_1(u)|_{(\widetilde{\theta}_i,\widetilde{\theta}_j)}\|_1 \|\widehat{\theta}_i - \theta_i\|_{\max} \\
&\leq \max_{i,j=1,\ldots,d} \|\nabla_{\theta_i} S_1(u)|_{(\widetilde{\theta}_i,\widetilde{\theta}_j)}\|_1 \max_{j=1,\ldots,d} \|\widehat{\theta}_j - \theta_j\|_{\max}.
\end{aligned}$$

The subsequent relation (D.66) can then be bounded through Lemma D.11 below. □

The following Lemma D.11 provides bounds on the derivatives of (D.40) with respect to the model parameters θ.

Lemma D.11. *Suppose $|u| < \boldsymbol{c}$ for $\boldsymbol{c} \in (0,1)$. Then, the derivatives of the quantities (D.40) with respect to the model parameter θ_i can be bounded as*

$$\begin{aligned}&\|\nabla_{\theta_i} S_1(u)\|_1 + \|\nabla_{\theta_i} S_2^G(u)\|_1 + \|\nabla_{\theta_i} S_3^g(u)\|_1\\ &\leq \frac{18}{1-\boldsymbol{c}^2} 2\pi \max_{k=0,\dots,3} \max_{i=1,\dots,d} m_i^{(k)}(1+\boldsymbol{c}) \max_{k=0,\dots,3} \max_{i=1,\dots,d} \mu_i^{(k)}(1+\boldsymbol{c})\end{aligned}$$

with $m_i^{(k)}, \mu_i^{(k)}$ in (2.11) and (2.12).

Proof. We consider the three quantities separately.

Bound on $\|\nabla_{\theta_i} S_1(u)\|_1$:

$$\begin{aligned}&\|\nabla_{\theta_i} S_1(u)\|_1\\ =&\|\nabla_{\theta_i} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right)\|_1\\ \leq& \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right)\left|-\frac{1}{1-u^2}(Q_{i,n_0} - uQ_{j,n_1})\right| \|\nabla_{\theta_i} Q_{i,n_0}\|_1\\ \leq& \frac{1}{1-\boldsymbol{c}^2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1-|u|}{2(1-u^2)}(Q_{i,n_0}^2 + Q_{j,n_1}^2)\right)|Q_{i,n_0} - uQ_{j,n_1}| \|\nabla_{\theta_i} Q_{i,n_0}\|_1\\ \leq& \frac{1}{1-\boldsymbol{c}^2}\left(\sum_{n=0}^{\infty} \exp\left(-\frac{1}{2(1+\boldsymbol{c})} Q_{j,n}^2\right) \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2(1+\boldsymbol{c})} Q_{i,n}^2\right)|Q_{i,n}| \|\nabla_{\theta_i} Q_{i,n_0}\|_1\right.\\ &\left.+ \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2(1+\boldsymbol{c})} Q_{j,n}^2\right)|Q_{j,n}| \sum_{n=0}^{\infty} \exp\left(-\frac{1}{2(1+\boldsymbol{c})} Q_{i,n}^2\right)\|\nabla_{\theta_i} Q_{i,n_0}\|_1\right)\\ \leq& \frac{1}{1-\boldsymbol{c}^2} 2\pi\left(m_j^{(0)}(1+\boldsymbol{c})\,\mu_i^{(1)}(1+\boldsymbol{c}) + m_j^{(1)}(1+\boldsymbol{c})\,\mu_i^{(0)}(1+\boldsymbol{c})\right)\\ \leq& \frac{2}{1-\boldsymbol{c}^2} 2\pi \max_{k=0,1} \max_{i=1,\dots,d} m_i^{(k)}(1+\boldsymbol{c}) \max_{k=0,1} \max_{i=1,\dots,d} \mu_i^{(k)}(1+\boldsymbol{c}).\end{aligned} \tag{D.68}$$

Bound on $\|\nabla_{\theta_i} S_2^G(u)\|_1$:

$$\begin{aligned}&\|\nabla_{\theta_i} S_2^G(u)\|_1\\ &= \|\nabla_{\theta_i} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right) G_{ij}^{n_0,n_1}(u)\|_1\\ &\leq \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right)\\ &\qquad \times \left|G_{ij}^{n_0,n_1}(u)\left(-\frac{1}{1-u^2}(Q_{i,n} - uQ_{j,n})\right)\right| \|\nabla_{\theta_i} Q_{i,n_0}\|_1\\ &\qquad + \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u)\right)\end{aligned}$$

$$
\begin{aligned}
&\times \left| -\frac{1}{1-u^2}(Q_{i,n} - uQ_{j,n}) \right| \|\nabla_{\theta_i} Q_{i,n_0}\|_1 \\
\leq\ & \frac{1}{1-\boldsymbol{c}^2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1-|u|}{2(1-u^2)} (Q_{i,n_0}^2 + Q_{j,n_1}^2) \right) \\
&\quad \times 2(Q_{i,n_0}^2 + Q_{j,n_1}^2)(|Q_{i,n_0}| + |Q_{j,n_1}|) \|\nabla_{\theta_i} Q_{i,n_0}\|_1 \\
&\quad + \frac{1}{1-\boldsymbol{c}^2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1-|u|}{2(1-u^2)} (Q_{i,n_0}^2 + Q_{j,n_1}^2) \right) \\
&\quad \times |Q_{i,n_0} - uQ_{j,n_1}| \|\nabla_{\theta_i} Q_{i,n_0}\|_1 \\
\leq\ & \frac{2}{1-\boldsymbol{c}^2} 2\pi \Bigg(m_j^{(3)} (1+\boldsymbol{c}) \mu_i^{(0)} (1+\boldsymbol{c}) + m_j^{(0)} (1+\boldsymbol{c}) \mu_i^{(3)} (1+\boldsymbol{c}) \\
&\quad + m_j^{(2)} (1+\boldsymbol{c}) \mu_i^{(1)} (1+\boldsymbol{c}) + m_j^{(1)} (1+\boldsymbol{c}) \mu_i^{(2)} (1+\boldsymbol{c}) \\
&\quad + m_j^{(1)} (1+\boldsymbol{c}) \mu_i^{(0)} (1+\boldsymbol{c}) + m_j^{(0)} (1+\boldsymbol{c}) \mu_i^{(1)} (1+\boldsymbol{c}) \Bigg) \\
\leq\ & \frac{12}{1-\boldsymbol{c}^2} 2\pi \max_{k=0,\ldots,3} \max_{i=1,\ldots,d} m_i^{(k)} (1+\boldsymbol{c}) \max_{k=0,\ldots,3} \max_{i=1,\ldots,d} \mu_i^{(k)} (1+\boldsymbol{c}) . \qquad \text{(D.69)}
\end{aligned}
$$

Bound on $\|\nabla_{\theta_i} S_3^g(u)\|_1$*:*

$$
\begin{aligned}
&\|\nabla_{\theta_i} S_3^g(u)\|_1 \\
&= \|\nabla_{\theta_i} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u) \right) g_{ij}^{n_0,n_1} \|_1 \\
&\leq \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u) \right) \\
&\qquad \times \left| -\frac{1}{1-u^2} (Q_{i,n_0} - uQ_{j,n_1}) g_{ij}^{n_0,n_1} \right| \|\nabla_{\theta_i} Q_{i,n_0}\|_1 \\
&\qquad\quad + \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)} G_{ij}^{n_0,n_1}(u) \right) |Q_{j,n_1}| \|\nabla_{\theta_i} Q_{i,n_0}\|_1 \\
&\leq \frac{1}{1-\boldsymbol{c}^2} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1-|u|}{2(1-u^2)} (Q_{i,n_0}^2 + Q_{j,n_1}^2) \right) \\
&\qquad \times (|Q_{i,n_0} - uQ_{j,n_1}| |Q_{i,n_0} Q_{j,n_1}| + |Q_{j,n_1}|) \|\nabla_{\theta_i} Q_{i,n_0}\|_1 \\
&\leq \frac{2}{1-\boldsymbol{c}^2} 2\pi \Big(m_j^{(1)} (1+\boldsymbol{c}) \mu_i^{(1)} (1+\boldsymbol{c}) + m_j^{(2)} (1+\boldsymbol{c}) \mu_i^{(1)} (1+\boldsymbol{c}) \\
&\qquad + m_j^{(2)} (1+\boldsymbol{c}) \mu_i^{(0)} (1+\boldsymbol{c}) + m_j^{(1)} (1+\boldsymbol{c}) \mu_i^{(2)} (1+\boldsymbol{c}) \Big) \\
&\leq \frac{4}{1-\boldsymbol{c}^2} 2\pi \max_{k=1,2} \max_{i=1,\ldots,d} m_i^{(k)} (1+\boldsymbol{c}) \max_{k=0,1,2} \max_{i=1,\ldots,d} \mu_i^{(k)} (1+\boldsymbol{c}) . \qquad \text{(D.70)}
\end{aligned}
$$

Finally, combining (D.68), (D.69) and (D.70), we infer that

$$\begin{aligned}&\|\nabla_{\theta_i} S_1(u)\|_1 + \|\nabla_{\theta_i} S_2^G(u)\|_1 + \|\nabla_{\theta_i} S_3^g(u)\|_1 \\ &\leq \frac{18}{1-\boldsymbol{c}^2} 2\pi \max_{k=0,\dots,3} \max_{i=1,\dots,d} m_i^{(k)} (1+\boldsymbol{c}) \max_{k=0,\dots,3} \max_{i=1,\dots,d} \mu_i^{(k)} (1+\boldsymbol{c}) .\end{aligned}$$

□

D.3.4. Derivatives IV

In this section, we consider the derivatives and their bounds used in Section C.4.

Lemma D.12. *Suppose Assumptions M.2 and M.3. Then, for some $\widetilde{\theta}_i$ such that $|\widetilde{\theta}_i - \theta_i| < |\widehat{\theta}_i - \theta_i|$,*

$$|\widehat{\ell}_{ii}(1) - \ell_{ii}(1)| \leq 3\widetilde{\Delta}(0)\|\widehat{\theta} - \theta\|_{\max}. \tag{D.71}$$

Proof. Note that $\ell_{ii}(1)$ can be written as

$$\begin{aligned}\ell_{ii}(1) &= \sum_{k=1}^{\infty} \frac{c_{i,k}^2}{k!} = \mathrm{Var}[X_{i,t}] \\ &= \sum_{n=0}^{\infty} (2n+1)\,\mathrm{P}[X_{i,t} > n] - \left(\sum_{n=0}^{\infty} \mathrm{P}[X_{i,t} > n]\right)^2 \\ &= \sum_{n=0}^{\infty} (2n+1)(1 - C_n(\theta_i)) - \left(\sum_{n=0}^{\infty} (1 - C_n(\theta_i))\right)^2 \end{aligned} \tag{D.72}$$

and is a function in θ_i. Using (D.72), its partial derivative with respect to θ_i can be derived as

$$\begin{aligned}\nabla_{\theta_i} \ell_{ii}(1) &= \sum_{n=0}^{\infty} (2n+1) \nabla_{\theta_i} C_n(\theta_i) - 2 \sum_{n=0}^{\infty} (1 - C_n(\theta_i)) \sum_{n=0}^{\infty} \nabla_{\theta_i} C_n(\theta_i) \\ &= \sum_{n=0}^{\infty} (2n + 1 - 2\,\mathrm{E}[X_{i,t}]) \nabla_{\theta_i} C_n(\theta_i). \end{aligned} \tag{D.73}$$

The estimated counterpart $\widehat{\ell}_{ii}(1)$ of $\ell_{ii}(1)$ is then given by replacing θ_i with its estimator $\widehat{\theta}_i$ in (D.72). Using the representation (D.72) and the derivative (D.73), we can write

$$\begin{aligned}|\widehat{\ell}_{ii}(1) - \ell_{ii}(1)| &= \left| \left\langle \sum_{n=0}^{\infty} (2n + 1 - \mathrm{E}[X_{i,t}]) \nabla_{\theta_i} C_n(\theta_i) \Big|_{\widetilde{\theta}_i}, \widehat{\theta}_i - \theta_i \right\rangle \right| \\ &\leq 3 \sum_{n=0}^{\infty} n \|\nabla_{\theta_i} C_n(\theta_i)\|_1 \Big|_{\widetilde{\theta}_i} \|\widehat{\theta} - \theta\|_{\max}, \end{aligned} \tag{D.74}$$

where we applied the mean value theorem in (D.74). □

Lemma D.13. *On the diagonal, the reciprocal of ℓ' satisfies $\frac{1}{\ell'_{ii}(u)} \leq M_1^{\frac{1}{2}}(1/2,0)$ with M_1 as in (2.15).*

Proof. For the diagonal elements of the first derivative ℓ'_{ii}, we can find a lower bound across all $u \in (-1,1)$. That is,

$$\begin{aligned}\ell'_{ii}(u) &= \frac{1}{2\pi\sqrt{1-u^2}} \sum_{n_0,n_1=0}^{\infty} \exp\left(-\frac{1}{2(1-u^2)}(Q_{i,n_0}^2 + Q_{i,n_1}^2 - 2uQ_{i,n_0}Q_{i,n_1})\right) \\ &\geq \sum_{n=0}^{\infty} \exp\left(-\frac{1}{1+u}Q_{i,n_0}^2\right) \\ &\geq \sum_{n=0}^{\infty} \exp\left(-Q_{i,n_0}^2\right) = m_i^{(0)}(1/2)\end{aligned}$$

such that $\frac{1}{\ell'_{ii}(u)} \leq M_1^{\frac{1}{2}}(1/2,0)$ with M_1 as in (2.15). □

Appendix E: Discussion on Assumption M.3

In this section, we verify Assumption M.3 for mixture Poisson, Conway-Maxwell-Poisson, binomial and negative binomial distributions.

For shortness' sake, we state an inequality for discrete random variables used in the subsequent examples in a small lemma at the end of this section.

Example E.1 (Mixture Poisson)**.** The mixture Poisson distribution is given by

$$\mathrm{P}[X_{i,t} = k] = \sum_{m=1}^{M} p_m e^{-\lambda_m} \frac{\lambda_m^k}{k!}, \quad k = 0, 1, \ldots,$$

with mixture probabilities $\boldsymbol{p} = (p_1, \ldots, p_M)$ such that $\sum_{m=1}^{M} p_m = 1$, $p_m > 0$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)$, $\lambda_m > 0$.

In order to verify Assumption M.3, set $\theta_i = (\theta_{i1}, \ldots, \theta_{i2M}) = (\boldsymbol{p}, \boldsymbol{\lambda})$. Then, with explanations given below, there is a constant $c > 0$ not depending on any model parameters such that,

$$\begin{aligned}&\sup_{\theta_i \in S} \sum_{n=0}^{\infty} (1 - C_n(\theta_i))^{-\frac{1}{2}} \sum_{j=1}^{2M} \left| \frac{\partial}{\partial \theta_{ij}} C_n(\theta_i) \right| \\ &\leq c \sup_{(\boldsymbol{p},\boldsymbol{\lambda}) \in S} \max_{m=1,\ldots,M} \frac{1}{p_m} \left((\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}} \left(1 + \lambda_1^{-\frac{1}{2}}\right) + 1 \right) < \infty. \qquad \text{(E.1)}\end{aligned}$$

Boundedness on S follows by Assumption M.2 and since the functions $x \mapsto \frac{1}{x}$ and $x \mapsto \frac{1}{\sqrt{x}}$ are both locally bounded on $(0, \infty)$. We turn to explaining the inequality in (E.1). With more details given below and with constant $c > 0$ not depending on any model parameters and possibly changing from line to line,

$$\sum_{n=0}^{\infty} (1 - C_n(\theta_i))^{-\frac{1}{2}} \sum_{j=1}^{2M} \left| \frac{\partial}{\partial \theta_{ij}} C_n(\theta_i) \right|$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \sum_{j=1}^{2M} \left| \frac{\partial}{\partial \theta_{ij}} \mathrm{P}[X_{i,t} > n] \right| \\
&\leq \max_{m=1,\dots,M} \frac{1}{p_m} \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} (\mathrm{P}[X_{i,t} > n] + \mathrm{P}[X_{i,t} = n]) \qquad \text{(E.2)} \\
&= \max_{m=1,\dots,M} \frac{1}{p_m} \sum_{n=0}^{\infty} \left(\mathrm{P}[X_{i,t} > n]^{\frac{1}{2}} + \frac{\mathrm{P}[X_{i,t} = n]}{\mathrm{P}[X_{i,t} > n]^{\frac{1}{2}}} \right) \\
&= \max_{m=1,\dots,M} \frac{1}{p_m} \left(\sum_{n=0}^{\infty} \mathrm{P}[X_{i,t} > n]^{\frac{1}{2}} + \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} = n])^{\frac{1}{2}} \left(\frac{\mathrm{P}[X_{i,t} = n]}{\mathrm{P}[X_{i,t} > n]} \right)^{\frac{1}{2}} \right) \\
&\leq \max_{m=1,\dots,M} \frac{1}{p_m} \left(\sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{\frac{1}{2}} + c \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} = n])^{\frac{1}{2}} \left(\frac{n}{\lambda_1} \right)^{\frac{1}{2}} \right) \qquad \text{(E.3)} \\
&\leq \max_{m=1,\dots,M} \frac{1}{p_m} \left(c(\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}} + 1 + c \sum_{n=1}^{\infty} (\mathrm{P}[X_{i,t} = n])^{\frac{1}{2}} n^{\frac{3}{2}} n^{-1} \right) \qquad \text{(E.4)} \\
&\leq \max_{m=1,\dots,M} \frac{1}{p_m} \left(c(\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}} + 1 + c \left(\sum_{n=0}^{\infty} n^3 \,\mathrm{P}[X_{i,t} = n] \sum_{n=1}^{\infty} n^{-2} \right)^{\frac{1}{2}} \lambda_1^{-\frac{1}{2}} \right) \qquad \text{(E.5)} \\
&= c \max_{m=1,\dots,M} \frac{1}{p_m} \left((\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}} \left(1 + \lambda_1^{-\frac{1}{2}} \right) + 1 \right),
\end{aligned}
$$

where (E.4) follows by Lemma E.1 and (E.5) by Hölder's inequality. We consider (E.2) and (E.3) separately. The bound (E.2) follows since

$$
\begin{aligned}
&\sum_{j=1}^{2M} \left| \frac{\partial}{\partial \theta_{ij}} \mathrm{P}[X_{i,t} > n] \right| = \sum_{j=1}^{2M} \left| \frac{\partial}{\partial \theta_{ij}} \sum_{k=n+1}^{\infty} \mathrm{P}[X_{i,t} = k] \right| \\
&= \sum_{j=1}^{M} \left| \frac{\partial}{\partial p_j} \sum_{k=n+1}^{\infty} \sum_{m=1}^{M} p_m e^{-\lambda_m} \frac{\lambda_m^k}{k!} \right| + \sum_{j=1}^{M} \left| \frac{\partial}{\partial \lambda_j} \sum_{k=n+1}^{\infty} \sum_{m=1}^{M} p_m e^{-\lambda_m} \frac{\lambda_m^k}{k!} \right| \\
&= \sum_{j=1}^{M} \left| \sum_{k=n+1}^{\infty} e^{-\lambda_j} \frac{\lambda_j^k}{k!} \right| + \sum_{j=1}^{M} \left| \sum_{k=n+1}^{\infty} p_j \left(e^{-\lambda_j} \frac{\lambda_j^{k-1}}{(k-1)!} - e^{-\lambda_j} \frac{\lambda_j^k}{k!} \right) \right| \\
&= \sum_{j=1}^{M} \left| \sum_{k=n+1}^{\infty} e^{-\lambda_j} \frac{\lambda_j^k}{k!} \right| + \sum_{j=1}^{M} p_j \left| e^{-\lambda_j} \frac{\lambda_j^n}{n!} \right| \\
&\leq \max_{m=1,\dots,M} \frac{1}{p_m} \sum_{k=n+1}^{\infty} \sum_{j=1}^{M} p_j e^{-\lambda_j} \frac{\lambda_j^n}{n!} + \mathrm{P}[X_{i,t} = n] \\
&= \max_{m=1,\dots,M} \frac{1}{p_m} \mathrm{P}[X_{i,t} > n] + \mathrm{P}[X_{i,t} = n].
\end{aligned}
$$

For the bound (E.3), assume without loss of generality that $\lambda_1 > \cdots > \lambda_M$. Note that the ratio between probability function and tail distribution behaves

asymptotically as

$$
\begin{aligned}
&\frac{\mathrm{P}[X_{i,t}=n]}{\mathrm{P}[X_{i,t}>n]} \\
&= \frac{\sum_{m=1}^{M} p_m e^{-\lambda_m} \frac{\lambda_m^n}{n!}}{\sum_{k=n+1}^{\infty} \sum_{m=1}^{M} p_m e^{-\lambda_m} \frac{\lambda_m^k}{k!}} \\
&= \frac{p_1 e^{-\lambda_1} \frac{\lambda_1^n}{n!} + \sum_{m=2}^{M} p_m e^{-\lambda_m} \frac{\lambda_m^n}{n!}}{p_1 e^{-\lambda_1} \frac{\lambda_1^{n+1}}{(n+1)!} + \sum_{m=2}^{M} p_m e^{-\lambda_m} \frac{\lambda_m^{n+1}}{(n+1)!} + \sum_{k=n+2}^{\infty} \sum_{m=1}^{M} p_m e^{-\lambda_m} \frac{\lambda_m^k}{k!}} \\
&= \frac{p_1 e^{-\lambda_1} \frac{\lambda_1^n}{n!} \left(1 + \sum_{m=2}^{M} \frac{p_m}{p_1} e^{-\lambda_m+\lambda_1} \left(\frac{\lambda_m}{\lambda_1}\right)^n\right)}{p_1 e^{-\lambda_1} \frac{\lambda_1^{n+1}}{(n+1)!} M_n(\lambda_1,\ldots,\lambda_M)} \qquad \text{(E.6)} \\
&\sim \frac{n}{\lambda_1} \frac{1 + \sum_{m=2}^{M} \frac{p_m}{p_1} e^{-\lambda_m+\lambda_1} \left(\frac{\lambda_m}{\lambda_1}\right)^n}{M_n(\lambda_1,\ldots,\lambda_M)} \sim \frac{n}{\lambda_1},
\end{aligned}
$$

where $M_n(\lambda_1,\ldots,\lambda_M)$ in (E.6) is defined in (E.7) below. Furthermore, the last relation follows since $\frac{(n+1)!}{k!} \to 0$ for $k \geq n+2$ and $\left(\frac{\lambda_m}{\lambda_1}\right)^k \to 0$ as $n \to \infty$. We conclude with the definition of $M_n(\lambda_1,\ldots,\lambda_M)$,

$$
\begin{aligned}
M_n(\lambda_1,\ldots,\lambda_M) = 1 &+ \sum_{m=2}^{M} \frac{p_m}{p_1} e^{-\lambda_m+\lambda_1} \left(\frac{\lambda_m}{\lambda_1}\right)^{n+1} \\
&+ \sum_{k=n+2}^{\infty} \lambda_1^{k-(n+1)} \sum_{m=1}^{M} \frac{p_m}{p_1} e^{-\lambda_m+\lambda_1} \left(\frac{\lambda_m}{\lambda_1}\right)^k \frac{(n+1)!}{k!}.
\end{aligned} \qquad \text{(E.7)}
$$

Example E.2 (Conway-Maxwell-Poisson)**.** The Conway-Maxwell-Poisson distribution is

$$
\mathrm{P}[X_{i,t}=k] = \frac{\lambda^k}{(k!)^\nu Z(\lambda,\nu)}, \quad k=0,1,\ldots, \quad \text{with} \quad Z(\lambda,\nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu},
$$

for $\lambda,\nu>0$. Let $\theta_i = (\theta_{i1},\theta_{i2}) = (\lambda,\nu)$. Then, with explanations given below, there is a constant $c>0$ not depending on any model parameters such that,

$$
\begin{aligned}
&\sup_{\theta_i \in S} \sum_{n=0}^{\infty} (1-C_n(\theta_i))^{-\frac{1}{2}} \sum_{j=1}^{2} \left| \frac{\partial}{\partial \theta_{ij}} C_n(\theta_i) \right| \\
&\leq \sup_{(\lambda,\nu)\in S} 2c \left((\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}} + 1 \right) \left(\frac{1}{\lambda} \mathrm{E}\,|X_{i,t}| + \mathrm{E}\,|\log(X_{i,t}!)| \right) < \infty. \qquad \text{(E.8)}
\end{aligned}
$$

Boundedness on S follows by Assumption M.2 and since the function $x \mapsto \frac{1}{x}$ is locally bounded on $(0,\infty)$. We turn to explaining the inequality in (E.8). With

more details given below,

$$
\begin{aligned}
&\sum_{n=0}^{\infty}(1-C_n(\theta_i))^{-\frac{1}{2}}\sum_{j=1}^{2}\left|\frac{\partial}{\partial\theta_{ij}}C_n(\theta_i)\right| \\
&=\sum_{n=0}^{\infty}(\mathrm{P}[X_{i,t}>n])^{-\frac{1}{2}}\sum_{j=1}^{2}\left|\frac{\partial}{\partial\theta_{ij}}\mathrm{P}[X_{i,t}>n]\right| \\
&=\sum_{n=0}^{\infty}(\mathrm{P}[X_{i,t}>n])^{-\frac{1}{2}}\left(\left|\frac{\partial}{\partial\lambda}\mathrm{P}[X_{i,t}>n]\right|+\left|\frac{\partial}{\partial\nu}\mathrm{P}[X_{i,t}>n]\right|\right) \\
&\le 2\sum_{n=0}^{\infty}(\mathrm{P}[X_{i,t}>n])^{-\frac{1}{2}}\left(\frac{1}{\lambda}\mathrm{P}[X_{i,t}>n]\,\mathrm{E}\,|X_{i,t}|+\mathrm{P}[X_{i,t}>n]\,\mathrm{E}\,|\log(X_{i,t}!)|\right) \qquad \text{(E.9)} \\
&=2\sum_{n=0}^{\infty}(\mathrm{P}[X_{i,t}>n])^{\frac{1}{2}}\left(\frac{1}{\lambda}\mathrm{E}\,|X_{i,t}|+\mathrm{E}\,|\log(X_{i,t}!)|\right) \\
&\le 2c\left((\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}}+1\right)\left(\frac{1}{\lambda}\mathrm{E}\,|X_{i,t}|+\mathrm{E}\,|\log(X_{i,t}!)|\right),
\end{aligned}
$$

where the last step follows by Lemma E.1. The bound (E.9) is obtained as follows. The derivative with respect to λ can be bounded as

$$
\begin{aligned}
\frac{\partial}{\partial\lambda}\mathrm{P}[X_{i,t}>n] &= \frac{\partial}{\partial\lambda}\sum_{k=n+1}^{\infty}\frac{\lambda^k}{(k!)^\nu Z(\lambda,\nu)} \\
&=\sum_{k=n+1}^{\infty}k\frac{\lambda^{k-1}}{(k!)^\nu Z(\lambda,\nu)}-\sum_{k=n+1}^{\infty}\frac{\lambda^k}{(k!)^\nu Z(\lambda,\nu)^2}\sum_{j=0}^{\infty}j\frac{\lambda^{j-1}}{(j!)^\nu} \\
&\le\frac{1}{\lambda Z^2(\lambda,\nu)}\left(\sum_{k=n+1}^{\infty}k\frac{\lambda^k}{(k!)^\nu}\sum_{j=0}^{\infty}\frac{\lambda^j}{(j!)^\nu}+\sum_{k=n+1}^{\infty}\frac{\lambda^k}{(k!)^\nu}\sum_{j=0}^{\infty}j\frac{\lambda^j}{(j!)^\nu}\right) \\
&\le\frac{2}{\lambda}\sum_{k=n+1}^{\infty}\frac{\lambda^k}{(k!)^\nu Z(\lambda,\nu)}\sum_{j=0}^{\infty}j\frac{\lambda^j}{(j!)^\nu Z(\lambda,\nu)} \\
&\le\frac{2}{\lambda}\mathrm{P}[X_{i,t}>n]\,\mathrm{E}\,|X_{i,t}|.
\end{aligned}
$$

The derivative with respect to ν can be bounded as

$$
\begin{aligned}
&\frac{\partial}{\partial\nu}\mathrm{P}[X_{i,t}>n] \\
&=\frac{\partial}{\partial\nu}\sum_{k=n+1}^{\infty}\frac{\lambda^k}{(k!)^\nu Z(\lambda,\nu)} \\
&=\frac{1}{Z(\lambda,\nu)^2}\left(\sum_{k=n+1}^{\infty}(-1)\log(k!)\frac{\lambda^k}{(k!)^\nu}Z(\lambda,\nu)+\sum_{k=n+1}^{\infty}\frac{\lambda^k}{(k!)^\nu}\sum_{j=0}^{\infty}\log(j!)\frac{\lambda^j}{(j!)^\nu}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{Z(\lambda,\nu)^2}\left(\sum_{k=n+1}^{\infty}\log(k!)\frac{\lambda^k}{(k!)^\nu}\sum_{j=0}^{\infty}\frac{\lambda^j}{(j!)^\nu}+\sum_{k=n+1}^{\infty}\frac{\lambda^k}{(k!)^\nu}\sum_{j=0}^{\infty}\log(j!)\frac{\lambda^j}{(j!)^\nu}\right)\\
&\leq 2\sum_{k=n+1}^{\infty}\frac{\lambda^k}{(k!)^\nu Z(\lambda,\nu)}\sum_{j=0}^{\infty}\log(j!)\frac{\lambda^j}{(j!)^\nu Z(\lambda,\nu)}\\
&= 2\,\mathrm{P}[X_{i,t}>n]\,\mathrm{E}\,|\log(X_{i,t}!)|.
\end{aligned}
$$

Note that $\mathrm{E}\,|\log(X_{i,t}!)|$ maximizes the Conway-Maxwell-Poisson likelihood function and does not have an analytic solution.

Example E.3 (Binomial)**.** The binomial distribution is

$$
\mathrm{P}[X_{i,t}=k]=\binom{N}{k}p^k(1-p)^{N-k},\quad k=0,1,\ldots,N.
$$

In order to verify Assumption M.3, note that we assume that the number of trials N is known and $\theta_i=p$ for the unknown probability of success in each trial. Then, with explanations given below, there is a constant $c>0$ not depending on any model parameters such that,

$$
\sup_{\theta_i\in S}\sum_{n=0}^{\infty}(1-C_n(\theta_i))^{-\frac{1}{2}}\left|\frac{\partial}{\partial\theta_i}C_n(\theta_i)\right|\leq c\sup_{p\in S}\frac{N}{p}\left((\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}}+1\right)<\infty. \tag{E.10}
$$

Boundedness on S follows by Assumption M.2 and since the function $x\mapsto\frac{1}{x}$ and is locally bounded on $(0,1)$. We turn to explaining the inequality in (E.10). With more details given below and with constant $c>0$ not depending on any model parameters,

$$
\sum_{n=0}^{\infty}(\mathrm{P}[X_{i,t}>n])^{-\frac{1}{2}}\left|\frac{\partial}{\partial p}\mathrm{P}[X_{i,t}>n]\right|\leq\frac{N}{p}\sum_{n=0}^{\infty}(\mathrm{P}[X_{i,t}>n])^{\frac{1}{2}} \tag{E.11}
$$

$$
\leq c\frac{N}{p}\left((\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}}+1\right), \tag{E.12}
$$

where (E.12) follows by Lemma E.1 and (E.11) is true since

$$
\begin{aligned}
\frac{\partial}{\partial p}\mathrm{P}[X_{i,t}>n]&=\frac{\partial}{\partial p}\sum_{k=n+1}^{N}\binom{N}{k}p^k(1-p)^{N-k}\\
&=\sum_{k=n+1}^{N}\binom{N}{k}(kp^{k-1}(1-p)^{N-k}-(N-k)p^k(1-p)^{N-k-1})\\
&=\sum_{k=n+1}^{N}\binom{N}{k}p^k(1-p)^{N-k}\frac{k(1-p)-Np}{p(1-p)}\\
&\leq\frac{N}{p}\mathrm{P}[X_{i,t}>n].
\end{aligned}
$$

Example E.4 (Negative binomial)**.** The negative binomial distribution is

$$\mathrm{P}[X_{i,t} = k] = \binom{k+r-1}{r-1}(1-p)^k p^r, \quad k = 0, 1, \ldots,$$

for $r \in \mathbb{N}$. Let $\theta_i = (\theta_{i1}, \theta_{i2}) = (p, r)$. Note that in contrast to Example E.3, both the number of failures until the experiment is stopped and the success probability in each experiment are assumed to be unknown. Then, with explanations given below, there is a constant $c > 0$ not depending on any model parameters such that,

$$\begin{aligned}
&\sup_{\theta_i \in S} \sum_{n=0}^{\infty} (1 - C_n(\theta_i))^{-\frac{1}{2}} \sum_{j=1}^{K_i} \left| \frac{\partial}{\partial \theta_{ij}} C_n(\theta_i) \right| \\
&\leq c \sup_{(p,r) \in S} \left((\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}} + 1 \right) \left(\frac{r}{p} + |\log(p)| \right) \qquad \text{(E.13)}
\end{aligned}$$

Boundedness on S follows by Assumption M.2 and since the functions $x \mapsto \frac{1}{x}$, $x \mapsto x$ and $x \mapsto |\log(x)|$ are locally bounded on $(0, \infty)$. We turn to explaining the inequality in (E.13). With more details given below and with constant $c > 0$ not depending on any model parameters,

$$\begin{aligned}
&\sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \left(\left| \frac{\partial}{\partial p} \mathrm{P}[X_{i,t} > n] \right| + \left| \frac{\partial}{\partial r} \mathrm{P}[X_{i,t} > n] \right| \right) \\
&\leq \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \sum_{k=n+1}^{\infty} \binom{k+r-1}{r-1} (1-p)^k p^r \left(\frac{r}{p} - \frac{k}{1-p} + \frac{k}{r} + \log(p) \right) \qquad \text{(E.14)} \\
&\leq \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \sum_{k=n+1}^{\infty} \binom{k+r-1}{r-1} (1-p)^k p^r \left(\frac{r}{p} + |\log(p)| \right) \qquad \text{(E.15)} \\
&= \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{-\frac{1}{2}} \mathrm{P}[X_{i,t} > n] \left(\frac{r}{p} + |\log(p)| \right) \\
&= \sum_{n=0}^{\infty} (\mathrm{P}[X_{i,t} > n])^{\frac{1}{2}} \left(\frac{r}{p} + |\log(p)| \right) \leq c \left((\mathrm{E}\,|X_{i,t}|^3)^{\frac{1}{2}} + 1 \right) \left(\frac{r}{p} + |\log(p)| \right). \qquad \text{(E.16)}
\end{aligned}$$

The inequality (E.14) follows by calculating the two derivatives with respect to p and r. The results are given in (E.17) and (E.19) below, respectively. Then, (E.15) is due to $r \geq 1 - p$ since $r \in \mathbb{N}$. The last line (E.16) follows by Lemma E.1.

The derivative of $\mathrm{P}[X_{i,t} > n]$ with respect to p is calculated as follows

$$\frac{\partial}{\partial p} \mathrm{P}[X_{i,t} > n] = \frac{\partial}{\partial p} \sum_{k=n+1}^{\infty} \binom{k+r-1}{r-1} (1-p)^k p^r$$

$$= \sum_{k=n+1}^{\infty} \binom{k+r-1}{r-1} (-k(1-p)^{k-1}p^r + (1-p)^k r p^{r-1})$$

$$= \sum_{k=n+1}^{\infty} \binom{k+r-1}{r-1} (1-p)^k p^r \left(\frac{r}{p} - \frac{k}{1-p}\right). \tag{E.17}$$

For the derivative with respect to n, we get

$$\begin{aligned}
\frac{\partial}{\partial r} \mathrm{P}[X_{i,t} > n] &= \frac{\partial}{\partial r} \sum_{k=n+1}^{\infty} \binom{k+r-1}{r-1} (1-p)^k p^r \\
&= \sum_{k=n+1}^{\infty} (1-p)^k \left(p^r \frac{\partial}{\partial r} \binom{k+r-1}{r-1} + \log(p) p^r \binom{k+r-1}{r-1} \right) \\
&\leq \sum_{k=n+1}^{\infty} (1-p)^k \left(p^r \frac{k}{r} \binom{k+r-1}{r-1} + \log(p) p^r \binom{k+r-1}{r-1} \right) \qquad \text{(E.18)} \\
&= \sum_{k=n+1}^{\infty} \binom{k+r-1}{r-1} (1-p)^k p^r \left(\frac{k}{r} + \log(p) \right), \qquad \text{(E.19)}
\end{aligned}$$

where the derivative of the binomial coefficient in (E.18) can be written and bounded as

$$\begin{aligned}
\frac{\partial}{\partial r} \binom{k+r-1}{r-1} &= \sum_{j=1}^{k} \frac{1}{k!} \prod_{i \in \{1,\ldots,k\}/\{j\}} (k+r-i) \\
&= \sum_{j=1}^{k} \frac{1}{k!} \frac{k+r-j}{k+r-j} \prod_{i \in \{1,\ldots,k\}/\{j\}} (k+r-i) \\
&= \sum_{j=1}^{k} \frac{1}{k+r-j} \binom{k+r-1}{r-1} \leq \frac{k}{r} \binom{k+r-1}{r-1}.
\end{aligned}$$

We state here a generic result which has found extensive use in showing that Assumption M.3 is satisfied.

Lemma E.1. *For a nonnegative discrete random variable X, there is a constant $c > 0$ such that*

$$\sum_{n=0}^{\infty} (\mathrm{P}[X > n])^{\frac{1}{2}} \leq c(\mathrm{E}\,|X|^3)^{\frac{1}{2}} + 1.$$

Proof. We make use of the following formula for moments of discrete random variables

$$\mathrm{E}\, X^p = \sum_{n=0}^{\infty} ((n+1)^p - n^p)\, \mathrm{P}[X > n], \quad \text{for } p = 1, 2, \ldots. \tag{E.20}$$

Then,

$$\begin{aligned}\sum_{n=0}^{\infty}(\mathrm{P}[X>n])^{\frac{1}{2}} &= \sum_{n=1}^{\infty}(n^2\,\mathrm{P}[X>n])^{\frac{1}{2}}n^{-1} + (\mathrm{P}[X>0])^{\frac{1}{2}} \\ &\leq \left(\sum_{n=1}^{\infty} n^2\,\mathrm{P}[X>n]\right)^{\frac{1}{2}}\left(\sum_{n=1}^{\infty} n^{-2}\right)^{\frac{1}{2}} + 1 \qquad \text{(E.21)} \\ &\leq \left(\sum_{n=0}^{\infty}((n+1)^3-n^3)\,\mathrm{P}[X>n]\right)^{\frac{1}{2}}\left(\frac{\pi^2}{6}\right)^{\frac{1}{2}} + 1 \qquad \text{(E.22)} \\ &= c(\mathrm{E}\,|X|^3)^{\frac{1}{2}} + 1,\end{aligned}$$

where (E.21) is a consequence of applying Hölder's inequality and (E.22) follows since $n^2 \leq 3n^2 + 3n + 1 = (n+1)^3 - n^3$ which allows us to use (E.20). □

Funding

The authors would like to thank Younghoon Kim for sharing his code for estimation of latent Gaussian time series. We also thank the two anonymous reviewers and the associate editor for their helpful comments and suggestions. Marie Düker's reseach was supported by NSF grant DMS-1934985, Robert Lund thanks NSF grant DMS-1407480, and Vladas Pipiras acknowledges NSF grants DMS-2113662 and DMS-2134107.

References

Adamczak, R. and Bednorz, W. (2015). Exponential concentration inequalities for additive functionals of Markov chains. *ESAIM: Probability and Statistics* **19** 440–481. MR3423302

Adamczak, R. and Wolff, P. (2015). Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probability Theory and Related Fields* **162** 531–586. MR3383337

Alzaid, A. A. and Al-Osh, M. A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics* **45** 223–232. MR1232490

An, H. Z. and Huang, F. C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* **6** 943–956. MR1422412

Baek, C., Düker, M.-C. and Pipiras, V. (2023). Local Whittle estimation of high-dimensional long-run variance and precision matrices. *The Annals of Statistics* **51** 2386–2414. MR4682702

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* **43** 1535–1567. MR3357870

Davis, R. A., Holan, S. H., Lund, R. and Ravishanker, N. (2016). *Handbook of discrete-valued time series*. CRC Press. MR3642975

Davis, R. A., Fokianos, K., Holan, S. H., Joe, H., Livsey, J., Lund, R., Pipiras, V. and Ravishanker, N. (2021). Count time series: A methodological review. *Journal of the American Statistical Association* **116** 1–15. MR4309291

Dirksen, S., Maly, J. and Rauhut, H. (2022). Covariance estimation under one-bit quantization. *The Annals of Statistics* **50** 3538–3562. MR4524507

Fan, J., Jiang, B. and Sun, Q. (2021). Hoeffding's Inequality for General Markov Chains and Its Applications to Statistical Learning. *Journal of Machine Learning Research* **22** 1–35. MR4318495

Fan, J., Liu, H., Ning, Y. and Zou, H. (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 405–421. MR3611752

Feng, H. and Ning, Y. (2019). High-dimensional mixed graphical model with ordinal data: Parameter estimation and statistical inference. In *The 22nd International Conference on Artificial Intelligence and Statistics* 654–663.

Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis* **27** 923–942. MR2328548

Fokianos, K., Rahbek, A. and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* **104** 1430–1439. MR2596998

Gamerman, D., Abanto-Valle, C. A., Silva, R. S., Martins, T. G., Davis, R., Holan, S., Lund, R. and Ravishanker, N. (2015). Dynamic Bayesian models for discrete-valued time series. *Handbook of discrete-valued time series* 165–186. MR3699405

Han, F. and Liu, H. (2017). Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli* **23** 23–57. MR3556765

Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press. MR1091716

Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press. MR2978290

Jia, Y., Kechagias, S., Livsey, J., Lund, R. and Pipiras, V. (2023). Latent Gaussian count time series. *Journal of the American Statistical Association* **118** 596–606. MR4571144

Karlis, D. (2016). Models for multivariate count time series. *Handbook of discrete-valued time series* **407** 424. MR3699415

Kim, Y. (2023). Modeling Multiple-Subject and Discrete-Valued High-Dimensional Time Series. *https://doi.org/10.17615/pdjz-yg02*.

Kim, Y., Düker, M. C., Fisher, Z. F. and Pipiras, V. (2024). Latent Gaussian dynamic factor modeling and forecasting for high-dimensional count time series. *arXiv preprint arXiv:2307.10454*.

Kong, J. and Lund, R. (2023). Seasonal count time series. *Journal of Time Series Analysis* **44** 93–124. MR4562413

Lebo, M. A. and Nesselroade, J. R. (1978). Intraindividual differences dimensions of mood change during pregnancy identified in five P-technique factor analyses. *Journal of Research in Personality* **12** 205–224.

Lee, S.-Y., Poon, W.-Y. and Bentler, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika* **57** 89–105.

Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics* **40** 2293–2326. MR3059084

Livsey, J., Lund, R., Kechagias, S. and Pipiras, V. (2018). Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *The Annals of Applied Statistics* **12** 408–431. MR3773399

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics* **40** 1637–1664. MR3015038

Magnus, J. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley. MR1698873

McKenzie, E. (1985). Some simple models for discrete variate time series. *JAWRA Journal of the American Water Resources Association* **21** 645–650. MR1973555

Mitra, R. and Zhang, C.-H. (2014). Multivariate analysis of nonparametric estimates of large correlation matrices. *arXiv preprint arXiv:1403.6195*.

Muthén, L. K. and Muthén, B. O. (1998–2017). Mplus 7.11. *Los Angeles, CA: Muthén & Muthén*.

Pipiras, V. and Taqqu, M. S. (2017). *Long-range Dependence and Self-Similarity* **45**. Cambridge University Press. MR3729426

Roberts, G. and Rosenthal, J. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability* **2** 13–25. MR1448322

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software* **48** 1–36.

Van Vleck, J. H. and Middleton, D. (1966). The spectrum of clipped noise. *Proceedings of the IEEE* **54** 2–19.

Vershynin, R. (2009). *Lectures in Geometric Functional Analysis*. Available at `https://www.math.uci.edu/~rvershyn/papers/GFA-book.pdf`.

Wegkamp, M. and Zhao, Y. (2016). Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli* **22** 1184–1226. MR3449812