Document de travail
Discussion paper

N°2025-03

https://lem.univ-lille.fr

Efficient parametric estimation of the spatial autoregressive model

Nicolas Debarsy
Univ. Lille, CNRS, IESEG School of Management, UMR 9221 – LEM, Lille Economie Management, F-59000, France | nicolas.debarsy@cnrs.fr

Vincenzo Verardi
Université Catholique de Louvain, LouRIM, Place des doyens 1 /L2.01.01. B-1348 Louvain-la-Neuve

Catherine Vermandele
Université Libre de Bruxelles, LMTD, 50 Av. F. D. Roosevelt, B-1050 Brussels

Efficient parametric estimation of the spatial autoregressive model

Nicolas Debarsy[1], Vincenzo Verardi*[2], and Catherine Vermandele[3]

[1]Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, F-59000 Lille, France
[2]Université catholique de Louvain. LouRIM. Place des Doyens 1/L2.01.01. B-1348 Louvain-la-Neuve.
[3]Université libre de Bruxelles, LMTD, 50 Av. F. D. Roosevelt, B-1050 Brussels.

March 11, 2025

Abstract

This paper introduces a new one-step parametric estimation method for spatial autoregressive (SAR) models, providing an efficient estimator for any error distribution with a defined quantile function. Based on Le Cam's Local Asymptotic Normality (LAN) theory, it extends the maximum likelihood approach to cases like the Laplace distribution, which lacks a globally defined first derivative.

We further develop this estimator for two highly flexible distributions: Tukey's g-and-h and Pewsey and Jones's sinh-arcsinh (SAS), designed to capture skewness and non-normal tail weight. These flexible distributions mitigate the risks of distributional misspecification by approximating a wide range of parametric distributions.

Monte Carlo simulations assess finite-sample performance, showing that our estimator outperforms traditional parametric spatial methods when the error distribution deviates from normality and is well-approximated by these flexible alternatives.

1 Introduction

The linear model with endogenous interaction effects, often referred to as autoregressive model in spatial econometrics literature, cannot generally be estimated by Ordinary Least Squares (OLS) due to the simultaneity arising from these effects. As such, Two-Stage Least

*This research was initiated when Vincenzo Verardi was Associated Researcher of the FNRS (Fonds National de la Recherche Scientifique) at the University of Namur and he gratefully acknowledges their financial support.

Squares (TSLS) (Kelejian & Prucha 1998, 1999, Bramoullé et al. 2009), Generalized Method of Moments (GMM) (Lee 2007, Kelejian & Prucha 2010) and Maximum Likelihood (ML) (Ord 1975) approaches have been derived to obtain consistent estimators. However, all of these estimators are not equally efficient.

When the distribution of the error term is known, the maximum likelihood estimator (MLE) provides the most efficient estimator, as its asymptotic variance attains the Cramer-Rao lower bound. In case the distribution is unknown Quasi-MLE (QLME) for the SAR model under normality of the error term can be considered and Lee (2004) studied its asymptotic properties. An important benefit of the normal distribution is that it belongs to the linear exponential family of distributions (see Gourieroux et al. 1984). As such, even if the true error distribution is not normal, the QMLE will remains consistent provided that the conditional expectation of the outcome is well specified. However, this estimator will no longer be efficient. Hence, alternative approaches, which do not impose distributional assumptions on the error term, have been proposed. Lee (2003) develops the best TSLS estimator, but which has been shown to be less efficient than the MLE under the true distribution; Liu et al. (2010) derive a GMM estimator that incorporates both linear and quadratic moment conditions and demonstrate that, although their estimator is less efficient than the MLE under the true distribution, it outperforms the normality-based QMLE as soon as the underlying distribution deviates from normality. In a different asymptotic setting, similar to the one Lee (2002) used to show the consistency of OLS estimators for the SAR specification, Robinson (2010) proposes an adaptive estimator (ADPE) based on series approximations of the score function. This estimator achieves the same efficiency as the MLE under the true distribution. However, it requires selecting a basis function for the series, which Robinson (2010) restricts to two options, along with determining its power. Finally, Debarsy et al. (2024) develop a semiparametrically efficient estimator (R&S) (i.e. that attains the semiparametric Cramer-Rao bound) that relies on the rank and signs of the residuals of a preliminary consistent estimator.

Relaxing assumptions about the error distribution can eliminate one source of misspecification. However, adopting a parametric distribution offers several advantages, such as improved efficiency and interpretability. When the assumed distribution is correct, parametric methods yield more efficient estimators than nonparametric ones. Additionally, researchers may be interested in specific features of the distribution, such as its symmetry or tail behavior. For example, assessing the asymmetry in the payoffs distribution of a lottery can be crucial for testing differences between loss aversion and acquiring gains. Furthermore, in structural econometrics, the distribution of the error term is often specified.

The classical approach in this setup is to write the (log-)likelihood function derived from the parametric distribution and estimate simultaneously regression coefficients and distributional parameters. However, this approach is not free of problems. Some distributions, such as the Laplace distribution, are not differentiable everywhere, compromising the regularity assumptions of the MLE. In other cases, the density function may not be explicitly defined, making MLE infeasible. Even when all conditions are satisfied, numerical convergence can sometimes be difficult and cumbersome.

In this paper, we first develop a general estimation strategy which *easily* provides efficient estimators for any parametric distribution with an explicitly defined quantile function. Secondly, we also propose to approximate the error distribution by relying on two highly flexible distributions, namely the Tukey g-and-h distribution (Tukey 1977) and the Sinh-Arcsinh distribution (introduced by Jones & Pewsey 2009).

Our methodology is based on one-step estimators derived from the Local Asymptotic Normality (LAN) theory, as developed in the statistical literature. (see Le Cam 1960). The LAN framework is much less demanding in terms of distributional assumptions, allowing it to be viewed as an extension of the Maximum Likelihood approach in the regression setting. Additionally, we choose to express everything in terms of the quantile function, allowing our results to be applicable even in cases where the density function is not explicitly defined.

The structure of this paper is as follows. Following this introduction, we present the Local Asymptotic Normality (LAN) property and the resulting one-step estimator in Section 2. In Section 3, we present a spillover effects model, commonly referred to as the spatial autoregressive model, though the term is somewhat overused, as it has broader applications beyond spatial econometrics. In Section 4, we explain how this model can be estimated by relying on the quantile function and the LAN property. Subsequently, in Section 5, we explore specific cases, such as the Laplace distribution (leading to a spatial L1 estimator), the Tukey g-and-h distribution, and Jones and Pewsey's Sinh-Arcsinh (SAS) distribution. We also explain how a preliminary estimator of the density function can be obtained using quantile least squares. Section 6 focuses on a discussion of flexible distributions. In Section 7, we compare the performance of the proposed estimators against alternative methods while Section 8 concludes.

2 LAN property

Local Asymptotic Normality, first introduced by Le Cam (1960), provides information on the performance of estimators and test procedures when sample sizes go to infinity.

Following Hallin (1996) (from page 129 onwards), let $\mathbf{y}^{(n)} = \left(y_1^{(n)}, \ldots, y_n^{(n)}\right)^{\mathrm{T}}$, $n \in \mathbb{N}_0$, be a sequence of observations described by the sequence of statistical models $\mathcal{E}^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, where $\mathcal{P}^{(n)} = \left\{\mathrm{P}_{\boldsymbol{\theta}}^{(n)} : \boldsymbol{\theta} \in \Theta\right\}$ is a parametric family of probability distributions defined on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and indexed by parameter $\boldsymbol{\theta} \in \Theta$ (with Θ an open set of $\mathbb{R}^L$); observation $\mathbf{y}^{(n)}$ is a random vector, of distribution $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$. Consider the sequences of probability distributions $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$ and $P_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}}^{(n)}$, where $\boldsymbol{\nu}(n)$ is a $(L \times L)$ non singular matrix such that $\|\boldsymbol{\nu}(n)\| \to 0$ for $n \to \infty$, and $\boldsymbol{\tau}^{(n)}$ is a $(L \times 1)$ real vector such that $\sup_n (\boldsymbol{\tau}^{(n)})^{\mathrm{T}} \boldsymbol{\tau}^{(n)} < \infty$; denote by $\Lambda_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta}}^{(n)}$ the logarithm of the likelihood ratio. Hallin (1996) (definition 4.1, page 131) formulates the LAN property of $\mathcal{E}^{(n)}$ as follows:

Definition. The sequence of parametric statistical models $\mathcal{E}^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$ is said to be *locally asymptotically normal* (LAN) if, for all $\boldsymbol{\theta} \in \Theta$, there exists a sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ of L-dimensional and $(\mathbf{y}^{(n)}, \boldsymbol{\theta})$-measurable random vectors, and a $(L \times L)$ symetric positive semi-definite matrix $\mathbf{I}(\boldsymbol{\theta})$, such that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, as $n \to \infty$:

(i) for every sequence $\boldsymbol{\tau}^{(n)}$ such that $\sup_n (\boldsymbol{\tau}^{(n)})^{\mathrm{T}} \boldsymbol{\tau}^{(n)} < \infty$,

$$\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta}} = (\boldsymbol{\tau}^{(n)})^{\mathrm{T}} \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{\tau}^{(n)})^{\mathrm{T}} \mathbf{I}(\boldsymbol{\theta}) \boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1)\,;$$

(ii) $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta})\right).$

Vector $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ is called the *central sequence*. It is only defined up to $o_{\mathrm{P}}(1)$ (under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, as $n \to \infty$).

A sequence of statistical models with the LAN property is therefore a sequence of models whose log-likelihoods $\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}/\boldsymbol{\theta}}$ can be approximated (on a pointwise basis, at each value of $\boldsymbol{\theta}$) by the log-likelihoods of a Gaussian position model of the form $\left(\mathbb{R}^L, \mathcal{B}(\mathbb{R}^L), \mathcal{P}_\theta\right)$, where $\mathcal{P}_\theta = \left\{N(\mathbf{I}(\boldsymbol{\theta})\boldsymbol{\tau}, \mathbf{I}(\boldsymbol{\theta})) : \boldsymbol{\tau} \in \mathbb{R}^L\right\}$. Hence, statistical procedures (tests, estimators) enjoying "good" properties in the limit local Gaussian model also enjoy, asymptotically, these "good" properties in the original sequence of statistical models $\mathcal{E}^{(n)}$.

Le Cam (1970) showed that the conventional regularity conditions needed for maximum likelihood theory are excessively stringent.[1] He showed they could be replaced by a simpler assumption called Quadratic Mean Differentiability (QMD) which only requires *single differentiability almost everywhere*. In simple terms, QMD ensures that the log-likelihood behaves smoothly enough so that it is possible to rely on local quadratic approximations of the likelihood to estimate the parameters. In addition, as long as differentiability holds at most points in a neighborhood around the true value, the lack of smoothness or differentiability at a small number of isolated points does not significantly affect the overall behavior of the likelihood. QMD allows to consider distributions of the error term that would have been ruled out in the classical maximum likelihood framework, such as the Laplace distribution.

If the LAN property of the sequence of statistical models $\mathcal{E}^{(n)}$ holds, then for a $\sqrt{n}$-consistent preliminary estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$ of $\boldsymbol{\theta}$, the one-step estimator $\widehat{\boldsymbol{\theta}}^{(n)}$, defined as

$$\widehat{\boldsymbol{\theta}}^{(n)} = \widetilde{\boldsymbol{\theta}}^{(n)} + \frac{1}{\sqrt{n}} \left(\mathbf{I}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)\right)^{-1} \boldsymbol{\Delta}^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right), \tag{1}$$

[1] Explaining the technical aspects of LAN, which involves elements from measure and information theory, goes well beyond the scope of this paper. For a thorough revision of LAN theory, we recommend the works of Van der Vaart (1998) and Le Cam & Yang (2000).

is an *asymptotically efficient* estimator of $\boldsymbol{\theta}$. In other words, $\widehat{\boldsymbol{\theta}}^{(n)}$ is asymptotically equivalent to the maximum likelihood (ML) estimator of $\boldsymbol{\theta}$: under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, (\mathbf{I}(\boldsymbol{\theta}))^{-1}\right).$$

Therefore, an asymptotically efficient one-step estimator can be obtained without having to solve the optimization problem associated with the ML estimation of parameter $\boldsymbol{\theta}$. This result is particularly advantageous when the calculation of the MLE is computationally or analytically complex.

3 The model

3.1 Definition of the model

In this paper, we consider the following linear model with endogenous effects.[2] For $i = 1, \ldots, n$,

$$y_i^{(n)} = (\mathbf{x}_i^{(n)})^{\mathrm{T}}\boldsymbol{\beta} + \lambda \sum_{\substack{j=1 \\ j \neq i}}^{n} w_{ij}^{(n)} y_j^{(n)} + \varepsilon_i^{(n)}, \tag{2}$$

where n is the sample size, $y_1^{(n)}, \ldots, y_n^{(n)}$ are the observations of the dependent variable, $\mathbf{x}_i^{(n)} = \left(x_{i1}^{(n)}, \ldots, x_{iK}^{(n)}\right)^{\mathrm{T}}$ is the vector of covariates for individual i, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^{\mathrm{T}} \in \mathbb{R}^K$ is the associated vector of regression parameters. The term $\sum_{j \neq i}^{n} w_{ij}^{(n)} y_j^{(n)}$ represents endogenous (interaction) effects and consists of a weighted sum of the outcomes for other individuals that affect i. The definition of the relevant weighting scheme is modeled by the elements $w_{ij}^{(n)}$ of the general connectivity matrix $\mathbf{W}^{(n)}$, which depends on the question under study. For instance, in the social-network literature, the peers (people that affect individual i's behavior) may be friends, geographic neighbors, roommates or coworkers to mention a few. The parameter λ measures the intensity of these endogenous effects. Finally, $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ are i.i.d. error terms with (marginal) absolutely continuous distribution characterized by a density function belonging to a family $\mathcal{F}$, defined as

$$\mathcal{F} = \left\{ f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma \subseteq \mathbb{R}^R \right\},$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_R)^{\mathrm{T}}$ is a vector of parameters of location, scale, and shape (typically skewness and tail heaviness).[3]

[2] Contextual effects may be further integrated without any difficulties.

[3] Model (2) does not explicitly include a regression constant. Instead, it is the location parameter of the underlying distribution of the error term that plays the role of the regression constant; the latter is not, strictly speaking, a parameter of interest in the regression model under consideration.

To illustrate the performance of our one-step estimator, we consider three different error density functions. To start with, we assume that the disturbances follow a Laplace distribution, which is not differentiable everywhere and hence does not fall under the regularity conditions of the ML approach. We then study the case of error terms being distributed according to two flexible distributions: a Tukey g-and-h distribution and a sinh-arcsinh (SAS) distribution. The former lacks an explicit density function, which poses challenges when considered in a maximum likelihood setup (see MacGillivray 1992, and Xu & Genton 2015 for numerical solutions) while the latter may pose numerical convergence problems.

Equation (2) actually defines a sequence of parametric models $\mathcal{E}^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, where

$$\mathcal{P}^{(n)} = \left\{ \mathrm{P}_{\boldsymbol{\theta}}^{(n)} : \boldsymbol{\theta} = \left(\boldsymbol{\beta}^{\mathrm{T}}, \lambda, \boldsymbol{\gamma}^{\mathrm{T}}\right)^{\mathrm{T}} \in \Theta \subset \mathbb{R}^K \times \Lambda \times \mathbb{R}^R \right\}.$$

and the parameter space Λ is a compact subset of $\mathbb{R}$ defined in Section 3.3.

3.2 Some notations

Let us introduce some notations that will be used throughout the subsequent text. For a square $(n \times n)$-matrix $\mathbf{A}^{(n)}$:

- $\mathbf{A}_{i\cdot}^{(n)}$ represents the ith row of $\mathbf{A}^{(n)}$;
- $\overline{\mathbf{A}}_{\cdot}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_{i\cdot}^{(n)}$ is the average $(1 \times n)$-vector of the n rows of $\mathbf{A}^{(n)}$;
- $\overline{A}_{\cdot\cdot}^{(n)} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}^{(n)}$ is the average of the n^2 components of matrix $\mathbf{A}^{(n)}$;
- $\mathrm{tr}\left(\mathbf{A}^{(n)}\right)$ is the trace of $\mathbf{A}^{(n)}$.

By writing the model (2) for the entire sample, we find its reduced form:

$$\mathbf{y}^{(n)} = \left(\mathbf{I}_n - \lambda \mathbf{W}^{(n)}\right)^{-1} \left(\mathbf{X}^{(n)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{(n)}\right),$$

where $\mathbf{y}^{(n)} = \left(y_1^{(n)}, \ldots, y_n^{(n)}\right)^{\mathrm{T}}$, $\mathbf{I}_n$ is the $(n \times n)$-identity matrix, $\mathbf{X}^{(n)} = \left(\mathbf{x}_1^{(n)}, \ldots, \mathbf{x}_n^{(n)}\right)^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}^{(n)} = \left(\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}\right)^{\mathrm{T}}$. Further, we have

$$\mathbf{W}^{(n)}\mathbf{y}^{(n)} = \mathbf{G}^{(n)}(\lambda) \left(\mathbf{X}^{(n)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{(n)}\right), \tag{3}$$

with $\mathbf{G}^{(n)}(\lambda) = \mathbf{W}^{(n)} \left(\mathbf{I}_n - \lambda \mathbf{W}^{(n)}\right)^{-1}$.

3.3 Regularity conditions

Lee (2004) lists and discusses the regularity conditions imposed on the interaction weights $w_{ij}^{(n)}$, on the regression parameter λ (see Assumption 1 below), and on the covariate vectors $\mathbf{x}_i^{(n)}$ (Assumption 2), required for the ML estimation of the regression parameters $\boldsymbol{\beta}$ and λ. As Lee (2004) focuses on the quasi-maximum likelihood estimator under normality, he further imposes the distribution of the error term to be Gaussian. In this paper, we relax this assumption using Assumption 3, where we simply say that the density function should be positive, parametrically defined, and differentiable in quadratic mean.

Assumption 1.

(i) The elements $w_{ij}^{(n)}$ of the matrix $\mathbf{W}^{(n)}$ are at most of order $1/h^{(n)}$ — they are $O(1/h^{(n)})$ — uniformly in all i, j, where the rate sequence $\{h^{(n)}\}$ is such that the ratio $h^{(n)}/n \to 0$ as $n \to \infty$.[4] *As a normalization, $w_{ii}^{(n)} = 0$ for all i.*

(ii) Let $\mathbf{I}_n$ be the $(n \times n)$-identity matrix. In model (2), the matrix $\mathbf{I}_n - \lambda\mathbf{W}^{(n)}$ is nonsingular. Moreover, the sequences $\{\mathbf{W}^{(n)}\}$ and $\left\{\left(\mathbf{I}_n - \lambda\mathbf{W}^{(n)}\right)^{-1}\right\}$ are uniformly bounded in both row and column sums (Horn & Johnson 1985).

(iii) In the sequence $\left\{\left(\mathbf{I}_n - \ell\mathbf{W}^{(n)}\right)^{-1}\right\}$, matrices $\left(\mathbf{I}_n - \ell\mathbf{W}^{(n)}\right)^{-1}$ are bounded in either row or column sums, uniformly in ℓ in an open set parameter space Λ. In consequence, the true value of parameter λ in model (2) is assumed to belong to the interior of Λ.

The parameter space Λ depends on the specification of $\mathbf{W}^{(n)}$. When its eigenvalues are real, Λ may be defined as the open subset $\left(1/\omega_{\max}^{(n)}, 1/\omega_{\min}^{(n)}\right)$, where $\omega_{\min}^{(n)}$ and $\omega_{\max}^{(n)}$ are respectively the minimal and maximal eigenvalues of $\mathbf{W}^{(n)}$. However, to ensure comparable values of λ for different connectivity matrices, $\mathbf{W}^{(n)}$ is most of the time normalized. Kelejian & Prucha (2010) mention the spectral radius and the minimum between the absolute row and column sum norms as normalizations, which restrict Λ to be the open subset $(-1, 1)$.[5]

Assumption 2. *The elements of $\mathbf{x}_i^{(n)}$ are uniformly bounded constants for all i and all n. Besides, the $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{x}_i^{(n)}(\mathbf{x}_i^{(n)})^{\mathrm{T}}/n$ exists and is non-singular.*

The LAN property of the sequence of the parametric SAR models $\mathcal{E}^{(n)}$ considered in this paper holds under the following regularity conditions for the marginal density $f_{\boldsymbol{\gamma}}$ of the i.i.d. error terms $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ (see for instance, Hallin 1996, for the first two conditions):

[4] That is, for some real constant c, there exists a finite integer N such that, for all $n \geq N$, $|h^{(n)}w_{ij}^{(n)}| < c$ for all i, j (see, e.g. White 1984, p.14).

[5] Another normalization, namely the row-normalization, is widely used in applied work. However, unless it is theoretically grounded (see, for instance, Patacchini & Zenou 2012), or for special cases, such as assigning the same number of neighbors to each observation, this normalization should not be used as it introduces misspecification in the model (see Neumayer & Plümper 2016).

Assumption 3.

(i) $f_{\boldsymbol{\gamma}}(e) > 0$ *for* $e \in \mathbb{R}$*;*

(ii) $f_{\boldsymbol{\gamma}}$ *is absolutely continuous with (almost everywhere) derivative* $f'_{\boldsymbol{\gamma}}$ *and finite Fisher information for location* $\mathcal{I}_{f_{\boldsymbol{\gamma}}} = \int_{-\infty}^{\infty} \phi^2_{f_{\boldsymbol{\gamma}}}(e) f_{\boldsymbol{\gamma}}(e) de$*, where* $\phi_{f_{\boldsymbol{\gamma}}}(e) = -\frac{f'_{\boldsymbol{\gamma}}(e)}{f_{\boldsymbol{\gamma}}(e)}$*.*

(iii) $\int_{-\infty}^{\infty} |e|^{4+\nu} f_{\boldsymbol{\gamma}}(e) de < \infty$ *for some* $\nu > 0$*.*

4 Efficient one-step estimator of $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^{\mathrm{T}}, \lambda, \boldsymbol{\gamma}^{\mathrm{T}}\right)^{\mathrm{T}}$

4.1 Efficient estimation of $\boldsymbol{\theta}$: general principles

Let $e_i^{(n)}(\boldsymbol{\beta}, \lambda)$ $(i = 1, \ldots, n)$ be the regression residuals associated with the values $\boldsymbol{\beta}$ and λ of the regression coefficients:

$$e_i^{(n)}(\boldsymbol{\beta}, \lambda) = y_i^{(n)} - (\mathbf{x}_i^{(n)})^{\mathrm{T}}\boldsymbol{\beta} - \lambda \mathbf{W}_{i\boldsymbol{.}}^{(n)} \mathbf{y}^{(n)}. \tag{4}$$

Then, the log-likelihood function associated to the parametric SAR model $\mathcal{E}^{(n)}$ is

$$\ln L\left(\boldsymbol{\theta} \middle| \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) = \ln \left|\det\left(\mathbf{I}_n - \lambda \mathbf{W}^{(n)}\right)\right| + \sum_{i=1}^{n} \ln f_{\boldsymbol{\gamma}}\left(e_i^{(n)}(\boldsymbol{\beta}, \lambda)\right). \tag{5}$$

Note that we may also characterize the distribution of the error term by its distribution function $F_{\boldsymbol{\gamma}}$ or, equivalently, its quantile function

$$Q_{\boldsymbol{\gamma}} : (0, 1) \to \mathbb{R} : u \mapsto Q_{\boldsymbol{\gamma}}(u) = F_{\boldsymbol{\gamma}}^{-1}(u).$$

Consequently, since

$$f_{\boldsymbol{\gamma}}(e) = \frac{dF_{\boldsymbol{\gamma}}(e)}{de} = \frac{d}{de}\left\{Q_{\boldsymbol{\gamma}}^{-1}(e)\right\} = \frac{1}{Q'_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}(e)\right)}$$

with $Q'_{\boldsymbol{\gamma}}(u) = \frac{dQ_{\boldsymbol{\gamma}}(u)}{du}$, the log-likelihood function for $\mathcal{E}^{(n)}$ may be written as follows:

$$\ln L\left(\boldsymbol{\theta} \middle| \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) = \ln \left|\det\left(\mathbf{I}_n - \lambda \mathbf{W}^{(n)}\right)\right| - \sum_{i=1}^{n} \ln Q'_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}\left(e_i^{(n)}(\boldsymbol{\beta}, \lambda)\right)\right). \tag{6}$$

Note also that, under $P_{\boldsymbol{\theta}}^{(n)}$, the random variables $u_i^{(n)}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} Q_{\boldsymbol{\gamma}}^{-1}\left(e_i^{(n)}(\boldsymbol{\beta}, \lambda)\right)$, $i = 1, \ldots, n$, are i.i.d. $\mathcal{U}(0, 1)$.

If $f_{\boldsymbol{\gamma}}$ is the density function of a centered Laplace distribution, the assumptions needed for the ML estimation are not satisfied as $f_{\boldsymbol{\gamma}}$ is not differentiable at zero. If $f_{\boldsymbol{\gamma}}$ is the density function of a SAS distribution (see Jones & Pewsey 2009), the optimization program for the MLE could pose technical difficulties given the complicated interdependence of the parameters determining the shape of the pdf. If on the other hand we consider that the error term distribution is a Tukey g-and-h, the ML estimation of $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^{\mathrm{T}}, \lambda, \boldsymbol{\gamma}^{\mathrm{T}}\right)^{\mathrm{T}}$ is even more difficult. Since the density function of the Tukey g-and-h distribution has no explicit form, the optimization program for the MLE needs a numeric inversion of the quantile function, which is a computationally demanding task (see Rayner & MacGillivray (2002) or Xu & Genton (2015) for a simpler numerical solution). These three examples show why we may benefit from avoiding the obstacles related to fitting the ML estimator of $\boldsymbol{\theta}$. As a solution, we propose to rely on the LAN property of the sequence $\mathcal{E}^{(n)}$ of SAR models and to estimate $\boldsymbol{\theta}$ using a one-step estimator of the form (1).

4.2 General expressions for $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$

The central sequence is, up to $o_{\mathrm{P}}(1)$, given by

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ \Delta^{(n)}_{\lambda}(\boldsymbol{\theta}) \\ \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}}\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\ln L\left(\boldsymbol{\theta}\left|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right.\right)\right\} \\ \frac{1}{\sqrt{n}}\frac{\partial}{\partial\lambda}\left\{\ln L\left(\boldsymbol{\theta}\left|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right.\right)\right\} \\ \frac{1}{\sqrt{n}}\frac{\partial}{\partial\boldsymbol{\gamma}}\left\{\ln L\left(\boldsymbol{\theta}\left|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right.\right)\right\} \end{pmatrix}.$$

Considering the log-likelihood function shown in (5) and the regression residuals presented in (4), we have, as in Debarsy et al. (2024), that:

$$\begin{aligned} \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f'_{\boldsymbol{\gamma}}\left(e_i^{(n)}(\boldsymbol{\beta},\lambda)\right)}{f_{\boldsymbol{\gamma}}\left(e_i^{(n)}(\boldsymbol{\beta},\lambda)\right)}\frac{\partial}{\partial\boldsymbol{\beta}}\left\{e_i^{(n)}(\boldsymbol{\beta},\lambda)\right\} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{f_{\boldsymbol{\gamma}}}\left(e_i^{(n)}(\boldsymbol{\beta},\lambda)\right)\mathbf{x}_i^{(n)}, \end{aligned} \tag{7}$$

$$\begin{aligned} \Delta_{\lambda}(\boldsymbol{\theta}) &= \frac{1}{\sqrt{n}}\frac{\partial}{\partial\lambda}\left\{\ln\left|\det\left(\mathbf{I}_n - \lambda\mathbf{W}^{(n)}\right)\right|\right\} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\lambda}\left\{\ln f_{\boldsymbol{\gamma}}\left(e_i^{(n)}(\boldsymbol{\beta},\lambda)\right)\right\} \\ &= -\frac{1}{\sqrt{n}}\mathrm{tr}\left(G^{(n)}(\lambda)\right) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{f_{\boldsymbol{\gamma}}}\left(e_i^{(n)}(\boldsymbol{\beta},\lambda)\right)\mathbf{W}_{i\boldsymbol{.}}^{(n)}\mathbf{y}^{(n)}, \end{aligned} \tag{8}$$

and

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) = \left(\Delta^{(n)}_{\gamma_1}(\boldsymbol{\theta}), \ldots, \Delta^{(n)}_{\gamma_R}(\boldsymbol{\theta})\right)^{\mathrm{T}},$$

where, for $r = 1, \ldots, R$,

$$\Delta^{(n)}_{\gamma_r}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{f_{\boldsymbol{\gamma}}\left(e_i^{(n)}(\boldsymbol{\beta},\lambda)\right)}\frac{\partial}{\partial\gamma_r}\left\{f_{\boldsymbol{\gamma}}\left(e_i^{(n)}(\boldsymbol{\beta},\lambda)\right)\right\}. \tag{9}$$

Remember that $f_{\boldsymbol{\gamma}}(e) = \frac{1}{Q'_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}(e)\right)}$, such that $f'_{\boldsymbol{\gamma}}(e) = -\frac{Q''_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}(e)\right)}{\left[Q'_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}(e)\right)\right]^3}$ and

$$\phi_{f_{\boldsymbol{\gamma}}}(e) = -\frac{f'_{\boldsymbol{\gamma}}(e)}{f_{\boldsymbol{\gamma}}(e)} = \frac{Q''_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}(e)\right)}{\left[Q'_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}(e)\right)\right]^2} \stackrel{\text{def}}{=} \widetilde{Q}_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}(e)\right), \qquad e \in \mathbb{R}.$$

Consequently, if we rather consider expression (6) of the log-likelihood function and denote, as previously, $u_i^{(n)}(\boldsymbol{\theta}) = Q_{\boldsymbol{\gamma}}^{-1}\left(e_i^{(n)}(\boldsymbol{\beta}, \lambda)\right) = F_{\boldsymbol{\gamma}}\left(e_i^{(n)}(\boldsymbol{\beta}, \lambda)\right)$, we get:

$$\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}(\boldsymbol{\theta})\right) \mathbf{x}_i^{(n)}, \tag{10}$$

$$\Delta_{\lambda}^{(n)}(\boldsymbol{\theta}) = -\frac{1}{\sqrt{n}} \operatorname{tr}\left(G^{(n)}(\lambda)\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}(\boldsymbol{\theta})\right) \mathbf{W}_{i\boldsymbol{\cdot}}^{(n)} \mathbf{y}^{(n)}, \tag{11}$$

and

$$\boldsymbol{\Delta}_{\boldsymbol{\gamma}}^{(n)}(\boldsymbol{\theta}) = \left(\Delta_{\gamma_1}^{(n)}(\boldsymbol{\theta}), \ldots, \Delta_{\gamma_R}^{(n)}(\boldsymbol{\theta})\right)^{\mathrm{T}},$$

where, for $r = 1, \ldots, R$,

$$\begin{aligned}
\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta}) &= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \gamma_r} \left\{\ln Q'_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}\left(e_i^{(n)}(\boldsymbol{\beta}, \lambda)\right)\right)\right\} \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{1}{Q'_{\boldsymbol{\gamma}}\left(u_i^{(n)}(\boldsymbol{\theta})\right)} \frac{\partial}{\partial \gamma_r} \left\{Q'_{\boldsymbol{\gamma}}\left(Q_{\boldsymbol{\gamma}}^{-1}\left(e_i^{(n)}(\boldsymbol{\beta}, \lambda)\right)\right)\right\}\right] \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{Q'_{\boldsymbol{\gamma}}\left(u_i^{(n)}(\boldsymbol{\theta})\right)} \\
&\qquad \times \left[\frac{\partial}{\partial \gamma_r} \left\{Q'_{\boldsymbol{\gamma}}\left(u_i^{(n)}(\boldsymbol{\theta})\right)\right\} - \frac{Q''_{\boldsymbol{\gamma}}\left(u_i^{(n)}(\boldsymbol{\theta})\right)}{Q'_{\boldsymbol{\gamma}}\left(u_i^{(n)}(\boldsymbol{\theta})\right)} \frac{\partial}{\partial \gamma_r} \left\{Q_{\boldsymbol{\gamma}}\left(u_i^{(n)}(\boldsymbol{\theta})\right)\right\}\right]
\end{aligned}$$

(see Rayner & MacGillivray 2002, p.63). Defining

$$\begin{aligned}
H_{\boldsymbol{\gamma};r}(u) &= \frac{(-1)}{Q'_{\boldsymbol{\gamma}}(u)} \left[\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \gamma_r} - \frac{Q''_{\boldsymbol{\gamma}}(u)}{Q'_{\boldsymbol{\gamma}}(u)} \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \gamma_r}\right] \\
&= \widetilde{Q}_{\boldsymbol{\gamma}}(u) \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \gamma_r} - \frac{1}{Q'_{\boldsymbol{\gamma}}(u)} \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \gamma_r}, \qquad u \in (0, 1),
\end{aligned}$$

we have, for $r = 1, \ldots, R$,

$$\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_{\boldsymbol{\gamma};r}\left(u_i^{(n)}(\boldsymbol{\theta})\right). \tag{12}$$

The information matrix

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathbf{I}_{\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) & \mathbf{I}_{\boldsymbol{\beta},\boldsymbol{\gamma}}(\boldsymbol{\theta}) \\ (\mathbf{I}_{\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}))^{\mathrm{T}} & I_{\lambda}(\boldsymbol{\theta}) & \mathbf{I}_{\lambda,\boldsymbol{\gamma}}(\boldsymbol{\theta}) \\ (\mathbf{I}_{\boldsymbol{\beta},\boldsymbol{\gamma}}(\boldsymbol{\theta}))^{\mathrm{T}} & (\mathbf{I}_{\lambda,\boldsymbol{\gamma}}(\boldsymbol{\theta}))^{\mathrm{T}} & \mathbf{I}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) \end{pmatrix}$$

can be obtained as

$$\lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})\left(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})\right)^{\mathrm{T}}\right], \tag{13}$$

the expectation being determined under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, i.e. assuming that the regression residuals $e_i^{(n)}(\boldsymbol{\beta}, \lambda)$ $(i = 1, \ldots, n)$ are i.i.d. with marginal density function $f_{\boldsymbol{\gamma}}$ and quantile function $Q_{\boldsymbol{\gamma}}$.

Let us define:

$$\mu_{\boldsymbol{\gamma}} = \int_0^1 Q_{\boldsymbol{\gamma}}(u)du, \quad \nu_{\boldsymbol{\gamma}} = \int_0^1 Q_{\boldsymbol{\gamma}}^2(u)du, \quad \mathcal{I}_{\boldsymbol{\gamma}} = \int_0^1 \widetilde{Q}_{\boldsymbol{\gamma}}^2(u)du,$$

$$\mathcal{K}_{\boldsymbol{\gamma}} = \int_0^1 \widetilde{Q}_{\boldsymbol{\gamma}}^2(u)Q_{\boldsymbol{\gamma}}(u)du, \quad \mathcal{L}_{\boldsymbol{\gamma}} = \int_0^1 \widetilde{Q}_{\boldsymbol{\gamma}}^2(u)Q_{\boldsymbol{\gamma}}^2(u)du,$$

and, for $r, s \in \{1, \ldots, R\}$, $r \neq s$:

$$\mathcal{H}_{\boldsymbol{\gamma};r} = \int_0^1 \widetilde{Q}_{\boldsymbol{\gamma}}(u)H_{\boldsymbol{\gamma};r}(u)du, \quad \mathcal{M}_{\boldsymbol{\gamma};r} = \int_0^1 \widetilde{Q}_{\boldsymbol{\gamma}}(u)Q_{\boldsymbol{\gamma}}(u)H_{\boldsymbol{\gamma};r}(u)du,$$

$$\mathcal{J}_{\boldsymbol{\gamma};r} = \int_0^1 H_{\boldsymbol{\gamma};r}^2(u)du, \quad \mathcal{J}_{\boldsymbol{\gamma};r,s} = \int_0^1 H_{\boldsymbol{\gamma};r}(u)H_{\boldsymbol{\gamma};s}(u)du.$$

We have[6] ($r, s \in \{1, \ldots, R\}$, $r \neq s$):

$$\mathbb{E}\left[\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})\left(\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})\right)^{\mathrm{T}}\right] = \mathcal{I}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(n)}(\mathbf{x}_i^{(n)})^{\mathrm{T}}\right\}, \tag{14}$$

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})\Delta_{\lambda}^{(n)}(\boldsymbol{\theta})\right] &= \mathcal{I}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(n)}\left(\mathbf{G}_{i\cdot}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right)\right\} + \mathcal{K}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(n)}G_{ii}^{(n)}(\lambda)\right\} \\ &\quad + \mathcal{I}_{\boldsymbol{\gamma}}\mu_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbf{x}_i^{(n)}G_{ij}^{(n)}(\lambda)\right\},\end{aligned} \tag{15}$$

$$\mathbb{E}\left[\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{H}_{\boldsymbol{\gamma};r}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(n)}\right\}, \tag{16}$$

$$\begin{aligned}\mathbb{E}\left[\left(\Delta_{\lambda}^{(n)}(\boldsymbol{\theta})\right)^2\right] &= \mathcal{I}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{G}_{i\cdot}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right)^2\right\} + (\mathcal{L}_{\boldsymbol{\gamma}} - 1)\left\{\frac{1}{n}\sum_{i=1}^{n}\left(G_{ii}^{(n)}(\lambda)\right)^2\right\} \\ &\quad + \mathcal{I}_{\boldsymbol{\gamma}}\nu_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(G_{ij}^{(n)}(\lambda)\right)^2\right\} + \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}G_{ij}^{(n)}(\lambda)G_{ji}^{(n)}(\lambda)\right\} \\ &\quad + \mathcal{I}_{\boldsymbol{\gamma}}\mu_{\boldsymbol{\gamma}}^2\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i,j}}^{n}G_{ij}^{(n)}(\lambda)G_{ik}^{(n)}(\lambda)\right\} \\ &\quad + \mathcal{K}_{\boldsymbol{\gamma}}\mu_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}G_{ii}^{(n)}(\lambda)G_{ij}^{(n)}(\lambda)\right\} \\ &\quad + \mathcal{K}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{G}_{i\cdot}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right)G_{ii}^{(n)}(\lambda)\right\} \\ &\quad + \mathcal{I}_{\boldsymbol{\gamma}}\mu_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\mathbf{G}_{i\cdot}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right)G_{ij}^{(n)}(\lambda)\right\},\end{aligned} \tag{17}$$

[6] Appendix A contains some details on the derivation of the different expectations.

$$\mathbb{E}\left[\Delta_{\lambda}^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{H}_{\boldsymbol{\gamma};r}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right\} + \mathcal{M}_{\boldsymbol{\gamma};r}\frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n} + \mathcal{H}_{\boldsymbol{\gamma};r}\mu_{\boldsymbol{\gamma}}\left\{n\overline{G}_{\boldsymbol{\cdot\cdot}}^{(n)}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n}\right\}, \quad (18)$$

$$\mathbb{E}\left[\left(\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right)^2\right] = \mathcal{J}_{\boldsymbol{\gamma};r}, \quad (19)$$

$$\mathbb{E}\left[\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_s}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{J}_{\boldsymbol{\gamma};r,s}. \quad (20)$$

Remark 1. All the results obtained in this paper can be readily applied to the classical linear model (i.e., one without endogenous effects). To do so, one requires to impose $\lambda = 0$ in the regression residuals (see 4) and in the log-likelihood function (see 5) and also simplify the central sequence vector and the information matrix by removing all terms involving λ. As such, the vector of parameters to estimate becomes $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}}\right)^{\mathrm{T}}$. Finally, the consistent preliminary estimator required for the estimation (as explained later) could be a standard ordinary least squares fit.

5 Estimation of models assuming a parametric distribution of the errors

The formulas provided in the previous section can be used to fit estimators that are asymptotically equivalent to the MLE. However, the density function of the error is only required to be differentiable in quadratic mean and not to be twice or three times differentiable everywhere, as it is usually the case for maximum likelihood. Moreover, since all the formulas have been derived using the quantile function, there is no necessity to explicitly define the density function. In the three sections below, we give the explicit formulas for $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$ in the case of three specific distributions of the error term: (i) the Laplace distribution that is well known not to be differentiable everywhere (but satisfying the QMD condition), (ii) the flexible Tukey g-and-h distribution that approximates well a vast variety of distributions but that does not have an explicitly defined density, and (iii) the SAS distribution, which can also approximate many parametric distributions, but despite being characterized by a well-defined density function, might be associated with computational difficulties when using standard ML algorithm, due to complicated interdependence of the parameters determining the shape of the density.

5.1 For a Laplace distributed error term

Let us assume that, in model (2), the error terms $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ are i.i.d. with Laplace(μ, b) marginal distribution (i.e., the Laplace distribution with mean equal to μ and scale parameter $b > 0$). In this particular case, $R = 2$ and $\boldsymbol{\gamma} = (\mu, b)^{\mathrm{T}}$. When E is Laplace(μ, b)-distributed, its density function is equal to

$$f_{\boldsymbol{\gamma}}(e) = \frac{1}{2b} \exp\left(-\frac{|e-\mu|}{b}\right) = \begin{cases} \frac{1}{2b} \exp\left(\frac{e-\mu}{b}\right) & \text{if } e \leq \mu \\ \frac{1}{2b} \exp\left(-\frac{e-\mu}{b}\right) & \text{if } e > \mu. \end{cases}$$

Its distribution function is given by

$$F_{\boldsymbol{\gamma}}(e) = \begin{cases} \frac{1}{2} \exp\left(\frac{e-\mu}{b}\right) & \text{if } e \leq \mu \\ 1 - \frac{1}{2} \exp\left(-\frac{e-\mu}{b}\right) & \text{if } e > \mu, \end{cases}$$

such that the quantile function that characterizes the Laplace(μ, b)-distribution is

$$Q_{\boldsymbol{\gamma}}(u) = \begin{cases} \mu + b \ln(2u) & \text{if } u \in (0, 1/2] \\ \mu - b \ln(2(1-u)) & \text{if } u \in (1/2, 1). \end{cases}$$

It follows that

$$Q'_{\boldsymbol{\gamma}}(u) = \begin{cases} \frac{b}{u} & \text{if } u \in (0, 1/2] \\ \frac{b}{1-u} & \text{if } u \in (1/2, 1) \end{cases} \quad \text{and} \quad Q''_{\boldsymbol{\gamma}}(u) = \begin{cases} -\frac{b}{u^2} & \text{if } u \in (0, 1/2] \\ \frac{b}{(1-u)^2} & \text{if } u \in (1/2, 1). \end{cases}$$

Moreover,

$$\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \mu} = 1, \quad \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \mu} = 0,$$

$$\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial b} = \begin{cases} \ln(2u) & \text{if } u \in (0, 1/2] \\ -\ln(2(1-u)) & \text{if } u \in (1/2, 1) \end{cases} \quad \text{and} \quad \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial b} = \begin{cases} \frac{1}{u} & \text{if } u \in (0, 1/2] \\ \frac{1}{1-u} & \text{if } u \in (1/2, 1). \end{cases}$$

Consequently,

$$\widetilde{Q}_{\boldsymbol{\gamma}}(u) = \frac{Q''_{\boldsymbol{\gamma}}(u)}{\left[Q'_{\boldsymbol{\gamma}}(u)\right]^2} = \begin{cases} -\frac{1}{b} & \text{if } u \in (0, 1/2] \\ \frac{1}{b} & \text{if } u \in (1/2, 1), \end{cases}$$

such that

$$H_{\boldsymbol{\gamma};1}(u) = \widetilde{Q}_{\boldsymbol{\gamma}}(u) \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \mu} - \frac{1}{Q'_{\boldsymbol{\gamma}}(u)} \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \mu} = \widetilde{Q}_{\boldsymbol{\gamma}}(u)$$

and

$$\begin{aligned} H_{\boldsymbol{\gamma};2}(u) &= \widetilde{Q}_{\boldsymbol{\gamma}}(u) \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial b} - \frac{1}{Q'_{\boldsymbol{\gamma}}(u)} \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial b} \\ &= \begin{cases} -\frac{1}{b}\left[1 + \ln(2u)\right] & \text{if } u \in (0, 1/2] \\ -\frac{1}{b}\left[1 + \ln(2(1-u))\right] & \text{if } u \in (1/2, 1). \end{cases} \end{aligned}$$

To obtain the spatial L1 estimator, we only have to plug the formulas here above in expressions for the central sequence (10 to 12) and the Information matrix (14 to 20), and then compute the efficient one-step estimator shown in (1).

5.2 For a Tukey g-and-h distributed error term

Tukey (1977) develop a flexible family of distributions known as Tukey g-and-h distributions. These distributions are derived from transformations of the standard normal distribution and are designed to model random variables characterized by heavy tails and/or skewness.

5.2.1 Definition

Let Z be a random variable with standard normal distribution $N(0,1)$.[7] Define the random variable E through the transformation

$$E = \xi + \sigma \tau_{g,h}(Z),$$

where $\xi \in \mathbb{R}$, $\sigma \in \mathbb{R}_0^+$, and

$$\tau_{g,h}(z) = \frac{1}{g}\left(\exp(gz) - 1\right)\exp(hz^2/2),$$

with $g \in \mathbb{R}$ (for $g = 0$, we have $\tau_{0,h}(z) = \lim_{g \to 0} \tau_{g,h}(z) = z\exp(hz^2/2)$) and $h \geq 0$.[8] Variable E is said to have a Tukey g-and-h distribution with location parameter (median) ξ and scale parameter σ: $E \sim T_{g,h}(\xi, \sigma)$. Parameter g controls the skewness ($g = 0$ corresponds to a symmetric distribution; $g > 0$ yields a right-skewed distribution and $g < 0$ gives a left-skewed distribution), while h controls the tail heaviness (also called elongation) of the distribution ($h > 0$ leads to heavy-tailed distributions). When $g = h = 0$, the $T_{g,h}(\xi, \sigma)$-distribution reduces to the normal distribution with mean ξ and standard deviation σ. Finally, to guarantee the existence of 4th order moment, Martinez & Iglewicz (1984) have shown that the elongation parameter h should be smaller than 0.25.

The $T_{g,h}(\xi, \sigma)$ distribution is very flexible and approximates well many commonly used unimodal distributions (see Martinez & Iglewicz 1984, MacGillivray 1992, Jiménez Moscoso & Arunachalam 2011).

5.2.2 Density and quantile functions of the $T_{g,h}(\xi, \sigma)$-distribution

Let $\boldsymbol{\gamma} = (\xi, \sigma, g, h)^{\mathrm{T}}$. If E is $T_{g,h}(\xi, \sigma)$-distributed, its density is written as

$$f_{\boldsymbol{\gamma}}(e) = \frac{\phi\left(\tau_{g,h}^{-1}\left(\frac{e-\xi}{\sigma}\right)\right)}{\sigma \tau'_{g,h}\left(\tau_{g,h}^{-1}\left(\frac{e-\xi}{\sigma}\right)\right)}, \qquad e \in \mathbb{R},$$

[7] Jones (2015, p.179) provides a general formula which extends the Normal assumption of the variable to be transformed.

[8] As explained by Xu & Genton (2015), the restriction of non-negative h ensures that the function $\tau_{g,h}(\cdot)$ is strictly monotone, regardless of the value of g.

where $\phi(\cdot)$ is the standard normal density function, $\tau_{g,h}^{-1}(\cdot)$ is the inverse function of $\tau_{g,h}(\cdot)$, and $\tau'_{g,h}(z) = \frac{d\tau_{g,h}(z)}{dz}$. As no explicit form for $\tau_{g,h}^{-1}(\cdot)$ exists, it is not possible to explicitly write the density function $f_{\boldsymbol{\gamma}}(\cdot)$. As such, it is more convenient to work with the quantile function $Q_{\boldsymbol{\gamma}}(\cdot)$ that characterizes the $T_{g,h}(\xi, \sigma)$-distribution:

$$Q_{\boldsymbol{\gamma}} : (0,1) \to \mathbb{R} : u \mapsto Q_{\boldsymbol{\gamma}}(u),$$

where $Q_{\boldsymbol{\gamma}}(u)$ is the quantile of order u of the $T_{g,h}(\xi, \sigma)$-distribution. Denoting by $\Phi(\cdot)$ the distribution function of the $N(0,1)$-distribution and by $z_u = \Phi^{-1}(u)$ the $N(0,1)$-quantile of order u, we have:

$$Q_{\boldsymbol{\gamma}}(u) = \xi + \sigma\tau_{g,h}(z_u), \qquad u \in (0,1).$$

This quantile function, as the function $\tau_{g,h}(\cdot)$, is strictly monotone and smooth, for all values of ξ, σ, g, and h.

5.2.3 Expressions for $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$

To calculate the central sequence (10 to 12) and the Information matrix (14 to 20) elements, one needs to compute the expressions below, which are all explicit.

$$
\begin{aligned}
Q_{\boldsymbol{\gamma}}(u) = &\ \xi + \sigma\tau_{g,h}(z_u) = \xi + \sigma\tau_{g,h}\left(\Phi^{-1}(u)\right), \qquad (21)\\
Q'_{\boldsymbol{\gamma}}(u) = &\ \frac{\sigma}{\phi(z_u)}\left[\exp\left(\frac{hz_u^2}{2} + gz_u\right) + hz_u\tau_{g,h}(z_u)\right],\\
Q''_{\boldsymbol{\gamma}}(u) = &\ \frac{\sigma}{\phi^2(z_u)}\left[(2hz_u + z_u + g)\exp\left(\frac{hz_u^2}{2} + gz_u\right) + h\left(1 + z_u^2 + hz_u^2\right)\tau_{g,h}(z_u)\right],\\
\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial\gamma_1} = &\ \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial\xi} = 1,\\
\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial\gamma_1} = &\ \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial\xi} = 0,\\
\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial\gamma_2} = &\ \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial\sigma} = \tau_{g,h}(z_u),\\
\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial\gamma_2} = &\ \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial\sigma} = \frac{1}{\sigma}Q'_{\boldsymbol{\gamma}}(u),\\
\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial\gamma_3} = &\ \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial g} = \frac{\sigma}{g}\left[z_u\exp\left(\frac{hz_u^2}{2} + gz_u\right) - \tau_{g,h}(z_u)\right],\\
\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial\gamma_3} = &\ \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial g} = \frac{\sigma z_u}{\phi(z_u)}\left[\left(1 + \frac{hz_u}{g}\right)\exp\left(\frac{hz_u^2}{2} + gz_u\right) - \frac{h}{g}\tau_{g,h}(z_u)\right],\\
\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial\gamma_4} = &\ \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial h} = \frac{\sigma\, z_u^2}{2}\tau_{g,h}(z_u),\\
\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial\gamma_4} = &\ \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial h} = \frac{\sigma z_u}{\phi(z_u)}\left[\frac{z_u}{2}\exp\left(\frac{hz_u^2}{2} + gz_u\right) + \left(1 + \frac{hz_u^2}{2}\right)\tau_{g,h}(z_u)\right].
\end{aligned}
$$

5.3 For a SAS-distributed error term

Another relatively simple four-parameter family of distributions on $\mathbb{R}$ has been proposed by Jones & Pewsey (2009). It is called the SAS (for sinh-arcsinh) distribution. In contrast to Tukey's g-and-h distribution, the SAS also permits tails to be lighter than those of the Normal distribution. We briefly summarize here below the definition and properties of the SAS and show how the one-step efficient estimator can be computed.

5.3.1 Definition

Let $\sinh(\cdot)$ be the sinus hyperbolic function:

$$
\sinh(t) = \frac{\exp(t) - \exp(-t)}{2}, \qquad t \in \mathbb{R},
$$

and $S_{\epsilon,\delta}(\cdot)$ be the sinh-arcsinh transformation characterized by the parameters $\epsilon \in \mathbb{R}$ and $\delta > 0$:

$$S_{\epsilon,\delta}(t) = \sinh\left(\delta \sinh^{-1}(t) - \epsilon\right), \qquad t \in \mathbb{R},$$

where $\sinh^{-1}(\cdot)$ is the inverse function of the sinus hyperbolic function:

$$\sinh^{-1}(t) = \ln\left(t + \sqrt{1+t^2}\right), \qquad t \in \mathbb{R}.$$

Note that, if $S_{\epsilon,\delta}(t) = v$, then

$$t = S^{-1}_{\epsilon,\delta}(v) = \sinh\left(\frac{1}{\delta}\sinh^{-1}(v) + \frac{\epsilon}{\delta}\right) = S_{-\epsilon/\delta,1/\delta}(v)\,;$$

hence, the inverse of the sinh-arcsinh transformation with parameters ϵ and δ is the sinh-arcsinh transformation with parameters $-\epsilon/\delta$ and $1/\delta$.

Let Z be a random variable with standard normal distribution $N(0,1)$. Define the random variable E through the transformation

$$E = \xi + \sigma S^{-1}_{\epsilon,\delta}(Z),$$

where $\xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+_0$. Variable E is said to follow a sinh-arcsinh normal distribution with location parameter ξ, scale parameter σ, and shape parameters ϵ and δ: $E \sim \mathrm{SAS}_{\epsilon,\delta}(\xi,\sigma)$. Parameter $\epsilon \in \mathbb{R}$ controls the skewness ($\epsilon = 0$ corresponds to a symmetric distribution; $\epsilon > 0$ (resp. $\epsilon < 0$) yields a right-skewed (resp. left-skewed) distribution), while $\delta > 0$ controls the heaviness of the tails (tailweight decreases when δ increases; When $\delta < 1$, the tails of the distribution are heavier than those of the normal distribution; On the contrary, when $\delta > 1$, the tails are lighter). The $\mathrm{SAS}_{0,1}(\xi,\sigma)$-distribution coincides with the $N(\xi,\sigma^2)$-distribution.

5.3.2 Density and quantile functions of the $\mathrm{SAS}_{\epsilon,\delta}(\xi,\sigma)$-distribution

Note first that, for $t \in \mathbb{R}$,

$$S'_{\epsilon,\delta}(t) = \frac{dS_{\epsilon,\delta}(t)}{dt} = C_{\epsilon,\delta}(t)\delta(1+t^2)^{-1/2},$$

where

$$C_{\epsilon,\delta}(t) = \cosh\left(\delta \sinh^{-1}(t) - \epsilon\right) = \left(1 + S^2_{\epsilon,\delta}(t)\right)^{1/2},$$

with

$$\cosh(t) = \frac{\exp(t) + \exp(-t)}{2}.$$

Moreover,

$$C'_{\epsilon,\delta}(t) = \frac{dC_{\epsilon,\delta}(t)}{dt} = S_{\epsilon,\delta}(t)\delta(1+t^2)^{-1/2}.$$

Note also that

$$\frac{d\sinh(t)}{dt} = \cosh(t), \quad \frac{d\cosh(t)}{dt} = \sinh(t), \quad \frac{d}{dt}\left\{\sinh^{-1}(t)\right\} = \frac{1}{\sqrt{1+t^2}}.$$

Let $\boldsymbol{\gamma} = (\xi, \sigma, \epsilon, \delta)^{\mathrm{T}}$. When E is $\mathrm{SAS}_{\epsilon,\delta}(\xi, \sigma)$-distributed, its density is written as

$$f_{\boldsymbol{\gamma}}(e) = \phi\left(S_{\epsilon,\delta}\left(\frac{e-\xi}{\sigma}\right)\right) C_{\epsilon,\delta}\left(\frac{e-\xi}{\sigma}\right) \frac{\delta}{\sigma}\left(1+\left(\frac{e-\xi}{\sigma}\right)^2\right)^{-1/2}, \qquad e \in \mathbb{R},$$

The associated quantile function $Q_{\boldsymbol{\gamma}}(\cdot)$ is

$$\begin{aligned} Q_{\boldsymbol{\gamma}}(u) &= \xi + \sigma S^{-1}_{\epsilon,\delta}(z_u) \\ &= \xi + \sigma S_{-\epsilon/\delta, 1/\delta}(z_u), \end{aligned} \tag{22}$$

where $z_u = \Phi^{-1}(u)$ is the $N(0,1)$-quantile of order u.

Finally note that the position parameter ξ does not coincide with the median of the $\mathrm{SAS}_{\epsilon,\delta}(\xi, \sigma)$-distribution.

$$Q_{\boldsymbol{\gamma}}(1/2) = \xi + \sigma S_{-\epsilon/\delta, 1/\delta}(0) = \xi + \sigma \sinh\left(\frac{\epsilon}{\delta}\right).$$

5.3.3 Expressions for $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$

To calculate the central sequence (10 to 12) and the Information matrix (14 to 20) elements, one needs to compute the expressions below, which are again all explicit:

$$
\begin{aligned}
Q_{\boldsymbol{\gamma}}(u) &= \xi + \sigma S_{-\epsilon/\delta,1/\delta}(z_u) = \xi + \sigma S_{-\epsilon/\delta,1/\delta}\left(\Phi^{-1}(u)\right),\\
Q'_{\boldsymbol{\gamma}}(u) &= \frac{\sigma}{\delta\phi(z_u)\sqrt{1+z_u^2}} C_{-\epsilon/\delta,1/\delta}(z_u),\\
Q''_{\boldsymbol{\gamma}}(u) &= \frac{\sigma}{\delta^2\phi^2(z_u)\left(1+z_u^2\right)}\left[S_{-\epsilon/\delta,1/\delta}(z_u) + \delta C_{-\epsilon/\delta,1/\delta}(z_u)\frac{z_u^3}{\sqrt{1+z_u^2}}\right],\\
\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \gamma_1} &= \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \xi} = 1,\\
\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \gamma_1} &= \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \xi} = 0,\\
\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \gamma_2} &= \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \sigma} = S_{-\epsilon/\delta,1/\delta}(z_u),\\
\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \gamma_2} &= \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \sigma} = \frac{1}{\sigma} Q'_{\boldsymbol{\gamma}}(u),\\
\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \gamma_3} &= \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \epsilon} = \frac{\sigma}{\delta} C_{-\epsilon/\delta,1/\delta}(z_u),\\
\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \gamma_3} &= \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \epsilon} = \frac{\sigma}{\delta^2\phi(z_u)\sqrt{1+z_u^2}} S_{-\epsilon/\delta,1/\delta}(z_u),\\
\frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \gamma_4} &= \frac{\partial Q_{\boldsymbol{\gamma}}(u)}{\partial \delta} = \frac{(-\sigma)}{\delta} C_{-\epsilon/\delta,1/\delta}(z_u)\left(\frac{1}{\delta}\sinh^{-1}(z_u) + \frac{\epsilon}{\delta}\right),\\
\frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \gamma_4} &= \frac{\partial Q'_{\boldsymbol{\gamma}}(u)}{\partial \delta} = \frac{(-\sigma)}{\delta^2\phi(z_u)\sqrt{1+z_u^2}}\left[C_{-\epsilon/\delta,1/\delta}(z_u) + S_{-\epsilon/\delta,1/\delta}(z_u)\left(\frac{1}{\delta}\sinh^{-1}(z_u) + \frac{\epsilon}{\delta}\right)\right].
\end{aligned}
$$

5.4 Practical computation of the efficient one-step estimator of $\boldsymbol{\theta}$

As explained in Section 2, an efficient one-step estimator of $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^{\mathrm{T}}, \lambda, \boldsymbol{\gamma}^{\mathrm{T}}\right)^{\mathrm{T}}$ is defined by (1), i.e.,

$$
\widehat{\boldsymbol{\theta}}^{(n)} = \widetilde{\boldsymbol{\theta}}^{(n)} + \frac{1}{\sqrt{n}}\left(\mathbf{I}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)\right)^{-1}\boldsymbol{\Delta}^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right), \tag{23}
$$

where $\widetilde{\boldsymbol{\theta}}^{(n)}$ is a $\sqrt{n}$-consistent preliminary estimator of $\boldsymbol{\theta}$.

5.4.1 The preliminary estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$

The preliminary estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$ of $\boldsymbol{\theta}$ contains both regression coefficients and the parameters of the error distribution. As such, we proceed in two steps to obtain a preliminary value of all its elements.

Step 1 We estimate regression parameters $\boldsymbol{\beta}$ and λ using either the TSLS estimator[9] (Kelejian & Prucha 1998, Bramoullé et al. 2009, Lee 2003), the GMM estimator (Lee 2007, Liu et al. 2010), or the QML estimator of Lee (2004). This gives us our preliminary estimates $\widetilde{\boldsymbol{\beta}}^{(n)} = \left(\widetilde{\beta}_1^{(n)}, \ldots, \widetilde{\beta}_K^{(n)}\right)^{\mathrm{T}}$ of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^{\mathrm{T}}$, and $\widetilde{\lambda}^{(n)}$ of λ.

Step 2 Once we have a preliminary $\sqrt{n}$-consistent point estimate for the regression coefficients, we may compute the residuals as follows

$$e_i^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right) = y_i^{(n)} - \sum_{k=1}^{K} \widetilde{\beta}_k^{(n)} x_{ik}^{(n)} - \widetilde{\lambda}^{(n)} \mathbf{W}_{i\boldsymbol{.}}^{(n)} \mathbf{y}^{(n)}, \qquad i = 1, \ldots, n. \tag{24}$$

Then, we may search for a preliminary estimate $\widetilde{\boldsymbol{\gamma}}^{(n)}$ of $\boldsymbol{\gamma}$ by minimizing the (squared) distance between some empirical quantiles and the corresponding theoretical quantiles of the distribution characterized by the parameter $\boldsymbol{\gamma}$. Let us consider m fixed probabilities $0 < p_1 < \ldots < p_m < 1$:

$$\widetilde{\boldsymbol{\gamma}}^{(n)} = \arg\min_{\boldsymbol{\gamma}\in\Gamma} \sum_{\ell=1}^{m} \left[e_{p_\ell}^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right) - Q_{\boldsymbol{\gamma}}(p_\ell)\right]^2, \tag{25}$$

where $e_{p_\ell}^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right)$ is the empirical quantile of order p_ℓ among the residuals computed in (24) and $Q_{\boldsymbol{\gamma}}(p_\ell)$ is the corresponding theoretical quantile of the distribution associated with parameter $\boldsymbol{\gamma}$. This nonlinear least squares procedure is borrowed from Xu et al. (2014) and ensures the $\sqrt{n}$-consistency of the preliminary estimator of $\boldsymbol{\gamma}$.[10] This approach has also been used in Ricci et al. (2018).

Remark 2. In the case of the $T_{g,h}(\xi, \sigma)$-distribution, the location parameter ξ is the median of the distribution. Then, it is quite natural to simply take $\widetilde{\xi}^{(n)}$ as the empirical median of the regression residuals $e_i^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right)$ $(i = 1, \ldots, n)$.

[9] The existence of the TSLS relies upon the presence of at least a significant exogenous variable, as instruments are constructed from spatial lags of determinants. Otherwise, a GMM estimator might be used.

[10] Here, we consider the ventiles.

5.4.2 The practical computation of $\mathbf{\Delta}^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)$ and $\mathbf{I}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)$

Computation of $\mathbf{\Delta}^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)$ requires the knowledge of $u_i^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right) = F_{\widetilde{\boldsymbol{\gamma}}^{(n)}}\left(e_i^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right)\right)$, for $i = 1, \ldots, n$. When the error terms are $\text{SAS}_{\epsilon,\delta}(\xi, \sigma)$-distributed, we have

$$F_{\boldsymbol{\gamma}}(e) = Q_{\boldsymbol{\gamma}}^{-1}(e) = \Phi\left(S_{\epsilon,\delta}\left(\frac{e-\xi}{\sigma}\right)\right), \qquad e \in \mathbb{R},$$

and calculating $u_i^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)$ presents no particular difficulties. On the other hand, in the case of the $T_{g,h}(\xi, \sigma)$-distribution,

$$F_{\boldsymbol{\gamma}}(e) = Q_{\boldsymbol{\gamma}}^{-1}(e) = \Phi\left(\tau_{g,h}^{-1}\left(\frac{e-\xi}{\sigma}\right)\right), \qquad e \in \mathbb{R}.$$

As no explicit form for the function $\tau_{g,h}^{-1}(\cdot)$ exists, calculating the terms $u_i^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)$ poses some technical difficulties. To avoid the computationally expensive solution of numeric inversion of the quantile function, we suggest to replace $u_i^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)$ by

$$\breve{u}_i^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right) = \frac{R_i^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right)}{n+1}, \qquad i = 1, \ldots, n,$$

where $R_i^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right)$ is the rank of the regression residual $e_i^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right)$. As such, the ratio $\frac{R_i^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right)}{n+1}$ estimates $F^{(n)}\left(e_i^{(n)}\left(\widetilde{\boldsymbol{\beta}}^{(n)}, \widetilde{\lambda}^{(n)}\right)\right)$, where $F^{(n)}(\cdot)$ is the empirical cumulative distribution function of the regression residuals.[11]

To finally obtain a consistent estimator of the (asymptotic) information matrix $\mathbf{I}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)$, we simply plug in the values of the estimated parameters $\widetilde{\boldsymbol{\beta}}^{(n)}$, $\widetilde{\lambda}^{(n)}$, and $\widetilde{\boldsymbol{\gamma}}^{(n)}$ in expressions (13)-(20).

[11] This strategy could also be used for the SAS distribution, as an alternative to the computation of the exact $u_i^{(n)}\left(\widetilde{\boldsymbol{\theta}}^{(n)}\right)$.

Remark 3. In the case of the $T_{g,h}(\xi,\sigma)$-distribution, h is required to be non-negative. As such, a correction needs to be applied to $\widehat{\boldsymbol{\theta}}^{(n)}$ when $\widehat{h}^{(n)}$ resulting from plugging expressions in the one-step estimator (1) takes a negative value. As proposed in Xu & Genton (2015), this correction consists in imposing the point estimate $\widehat{h}^{(n)}$ to zero and replacing $\widehat{\boldsymbol{\theta}}^{(n)}$ by $\widehat{\widehat{\boldsymbol{\theta}}}^{(n)}$ with

$$\widehat{\widehat{\boldsymbol{\theta}}}^{(n)} = \begin{pmatrix} \widehat{\widehat{\boldsymbol{\beta}}}^{(n)} \\ \widehat{\widehat{\lambda}}^{(n)} \\ \widehat{\widehat{\sigma}}^{(n)} \\ \widehat{\widehat{g}}^{(n)} \\ \widehat{\widehat{h}}^{(n)} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}^{(n)} - \frac{\mathbf{I}^{\boldsymbol{\beta},h}}{I^{h,h}}\widehat{h}^{(n)} \\ \widehat{\lambda}^{(n)} - \frac{\mathbf{I}^{\lambda,h}}{I^{h,h}}\widehat{h}^{(n)} \\ \widehat{\sigma}^{(n)} - \frac{\mathbf{I}^{\sigma,h}}{I^{h,h}}\widehat{h}^{(n)} \\ \widehat{g}^{(n)} - \frac{\mathbf{I}^{g,h}}{I^{h,h}}\widehat{h}^{(n)} \\ 0 \end{pmatrix},$$

where $I^{p,q}$ is the (p,q)-th element of $\left(\mathbf{I}\left(\widehat{\boldsymbol{\theta}}^{(n)}\right)\right)^{-1}$.

6 Benefits of considering flexible distributions

In econometrics, there is an increasing body of research focused on developing estimators for SAR models that do not require specific assumptions on the distribution of the error term. For example Lee (2004) develops a QML estimator, based on the normality assumption of the error, that remains consistent even if the distribution is not well specified. Liu et al. (2010) propose a GMM estimator that does not need any distributional assumption and relies on a set of linear and quadratic moment conditions. Finally, Robinson (2010), Lee & Robinson (2020), and Debarsy et al. (2024) develop semiparametric models. The former two rely on a series approximation of the score function while the latter is based on a one-step improvement (based on the ranks and the signs of the residuals) of a preliminary consistent estimator.

In this literature, the estimators developed in 5.2 and 5.3 could serve as alternatives since both the Tukey g-and-h and the SAS distributions approximate well a wide variety of unimodal distributions. Estimating the skewness and elongation of the error term distribution together with the regression parameters could be an empirically attractive strategy. Indeed, by remaining in the purely parametric world, models can have better predictive accuracy than their semiparametric counterparts. Furthermore, when the parametric assumptions about the data distribution are right (or at least reasonably accurate), there could be a gain in efficiency with respect to estimators with a less structured approach to modeling data.

To illustrate the quality of approximation of the flexible distributions considered in this paper, we generate synthetic data from several commonly used distributions. The process begins by constructing 1000 relative ranks, uniformly distributed between 0 and 1. These values serve as cumulative probabilities and are transformed using the inverse cumulative distribution function (quantile function) of the target distribution. This ensures that the generated data follows the desired statistical properties.[12]

We consider ten distributions: the normal distribution, which serves as a benchmark, the Laplace distribution, which has a sharper peak and heavier tails, and the Tukey g-and-h and SAS distributions, which introduce parameters to control skewness and tail behavior. The beta distribution is bounded and light-tailed, making it useful for modeling proportions, while the Student's t-distribution exhibits heavier tails than the normal, making it suitable for capturing extreme values. The lognormal and Weibull distributions, frequently used in reliability studies, introduce skewness and take only positive values. The gamma and chi-squared distributions are also positively skewed, with various applications in statistical modeling.

Once the data are generated, we fit them using two flexible distributions. To do so, we rely on a nonlinear least squares procedure, shown in (25) but rather use all the quantiles of the generated distribution as the dependent variable. The theoretical quantiles used in (25) either come from the $T_{g,h}$ distribution (see 21) or from the SAS distribution (see 22). The standard R^2 statistic is used to quantify how well each model explains variability in the data.

To graphically represent the estimated densities, a nontrivial task for the Tukey g-and-h model since its density is not explicitly defined, we first compute the cumulative distribution function of the predicted values from the nonlinear least squares fit. We then take its numerical derivative to approximate the probability density function. The use of a regular grid ensures a smooth and accurate density representation, effectively capturing the key features of the distribution.

Figure 1 below represents the histogram for each considered distribution, together with estimated densities that are obtained using either a $T_{g,h}(\xi, \sigma)$ (in red) or a $\text{SAS}_{\epsilon,\delta}(\xi, \sigma)$ (in blue) distribution. The first general remark is that both distributions constitute very good approximations of the true distributions, with R^2 values no lower than 0.89, which occurs when the $T_{g,h}$ distribution is used to approximate the SAS. Except for this case, we observe a quality of fit of at least 97% for all other distributions. In addition, as expected, for distributions with lighter tails than the normal, for instance, the Beta(2,2), the SAS approximation performs better than the $T_{g,h}$. For skewed distributions, the elongation parameter of the $T_{g,h}$ will be positive only for those with a very long tail. Otherwise, it is the skewness parameter (g) which determines the length of the tail. This issue does not arise with the SAS distribution, as its elongation parameter (δ) has no positivity constraint. Finally, we see that both flexible distributions capture well tail heaviness.

[12]Essentially, a random variable is obtained by applying a nonlinear transformation to uniform values regularly spaced over the interval $[0, 1]$, where the quantile function maps probabilities to corresponding quantiles.

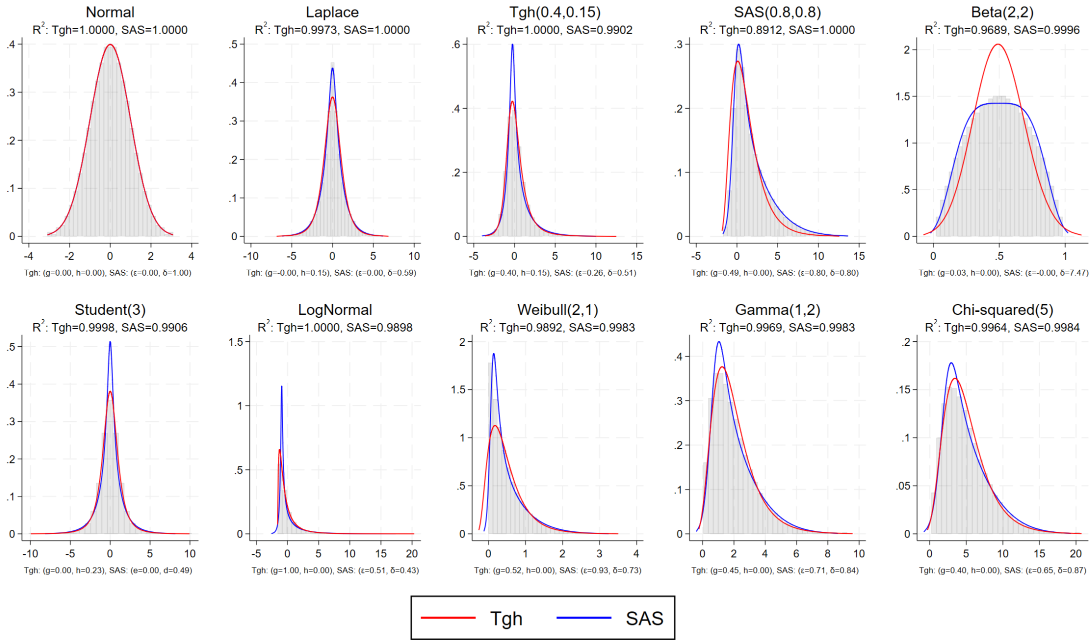

Figure 1: Adjustment of $T_{g,h}$ and $\text{SAS}_{\epsilon,\delta}$ distributions

7 Simulations

The experimental design considered is

$$y_i = \beta_1 x_i + \lambda \mathbf{W}_{i\cdot}\mathbf{y}^{(n)} + \varepsilon_i^{(n)}, \quad i = 1, \dots, n \tag{26}$$

where the x_i's are generated once (and kept constant over all the simulations) from a standard normal. We also have $\beta_1 = 1$, and λ spans values from -2.1 to 0.7, increasing in steps of 0.2, and also includes the value 0.[13] We consider as a connectivity matrix the binary 10 nearest neighbors constructed from Euclidian distance, based on random coordinates using two $U(0, 10)$ distributions (also kept constant across the simulations). This matrix has been normalized using the spectral radius norm of Kelejian & Prucha (2010). The distributions considered for the simulations are those considered in Figure 1. In total, 144 alternative scenarios are considered, and each of them has been replicated 1000 times. The simulation setup is run for 3 sample sizes: $n = 300, 500$, and 900.[14]

[13] The constant term, β_0, is viewed as a nuisance parameter and estimated as the location parameter of the error distribution.

[14] All simulations have been run with Matlab R2019a on the calculation center of the Université de Lille (Mésocentre de Calcul Scientifique Intensif de l'Université de Lille). Our proposed estimator has also been programmed in Stata software.

For each setup, we assess the performance of $\widehat{\lambda}$ and $\widehat{\beta}_1$ for five alternative estimators: the QML estimator of Lee (2004), the efficient GMM estimator of Liu et al. (2010), the R&S semiparametric estimator proposed by Debarsy et al. (2024), and estimators based on the two flexible distributions considered here, namely the $T_{g,h}(\xi, \sigma)$ (TGH estimator) and the $\text{SAS}_{\epsilon,\delta}(\xi, \sigma)$ (SAS estimator).[15] All LAN-based estimators use the TSLS as preliminary estimator for λ and β_1 while the preliminary estimator for the distribution parameters of the TGH and SAS are computed from equation (25) using ventiles of the residuals. Finally, to improve finite sample performance of LAN estimators, 100 possible refinements of the one-step estimator is allowed.[16]

The summary measures considered to assess the performances of estimators are the median difference of the estimated coefficients to the true values as a measure of their bias and the interquartile range (divided by 1.349 to guarantee Gaussian consistency towards the standard deviation) of the point estimates as a measure of dispersion. Finally, we compare the median standard error to the dispersion of $\hat{\lambda}$ over repeated samples to assess the bias of the estimated standard errors.

7.1 Bias

Figures 2 to 7 report the bias for λ and the β_1 over the three considered sample sizes. We observe a small and constant bias for all classical estimators. We also note that the flexible distributions and the semiparametric estimator (R&S) perform as well as the usual estimators, regardless of the true distribution of the error term.

7.2 Efficiency

Figures 8 to 10 compare the dispersion for all estimators of λ across sample sizes. We note that, except for normally distributed error terms, the flexible distributions proposed here, along with the semiparametric estimator introduced in Debarsy et al. (2024), are significantly more efficient than classical estimators. As expected, we observe an inverted U-shaped relation between dispersion and the value of λ, driven by the fact that when the parameter gets close to its boundary values, its variance naturally shrinks, similar to the behavior of the variance of a sample correlation coefficient when it approaches -1 or 1.[17] Regarding the efficiency of the estimators for β_1, shown in Figures 11 to 13, the flexible distributions proposed here demonstrate comparable or superior performance compared to the classical estimators for all the distributions considered. In addition, their performance is comparable to the R&S estimator. This result holds consistently across all sample sizes.

[15] For the normal distribution setup, we use the ML estimator of Lee (2004).

[16] These refinements are optional and a single application of the one-step estimator is theoretically sufficient.

[17] Let us nevertheless remember that λ is not a correlation coefficient as boundaries of its parameter space may exceed 1 in absolute value.

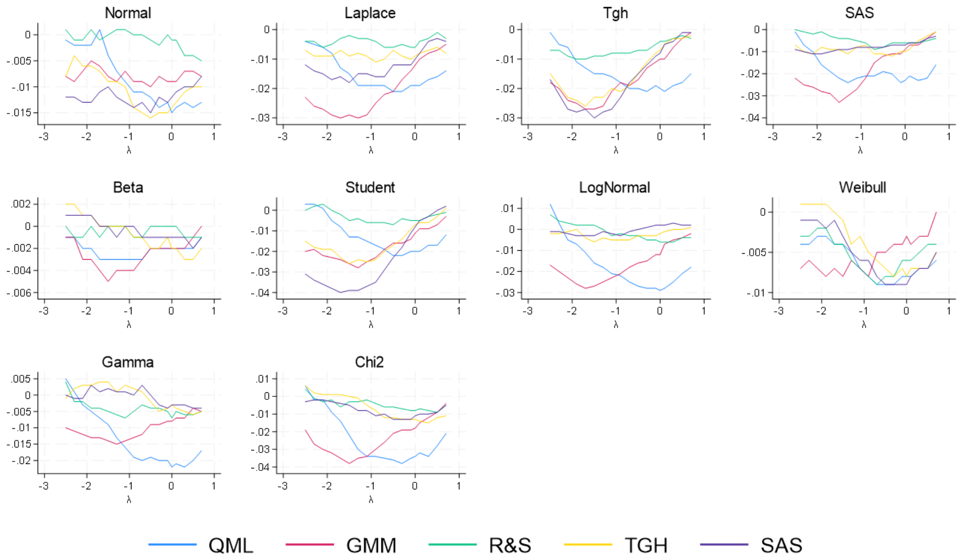

Figure 2: Bias of $\widehat{\lambda}$, $n = 300$

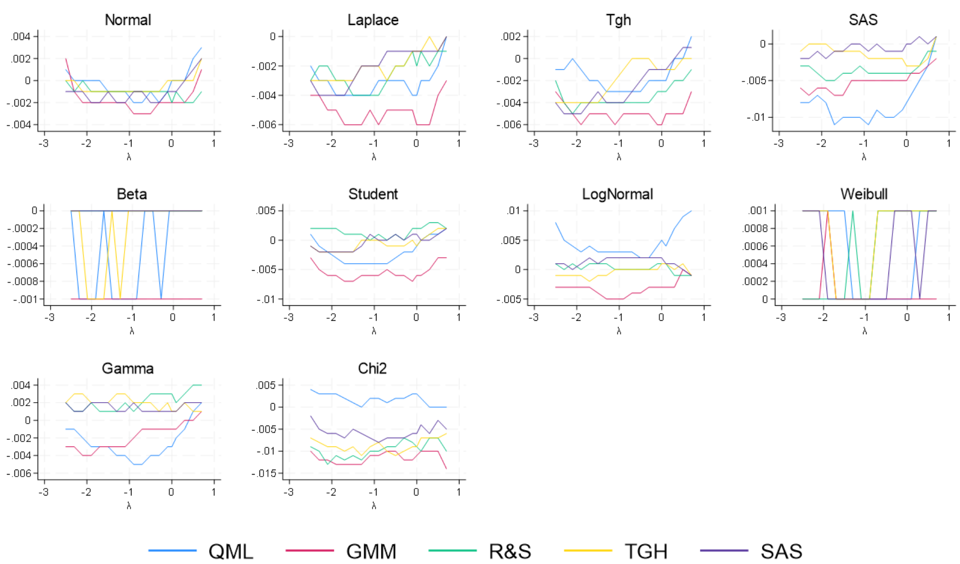

Figure 3: Bias of $\widehat{\beta}_1$, $n = 300$

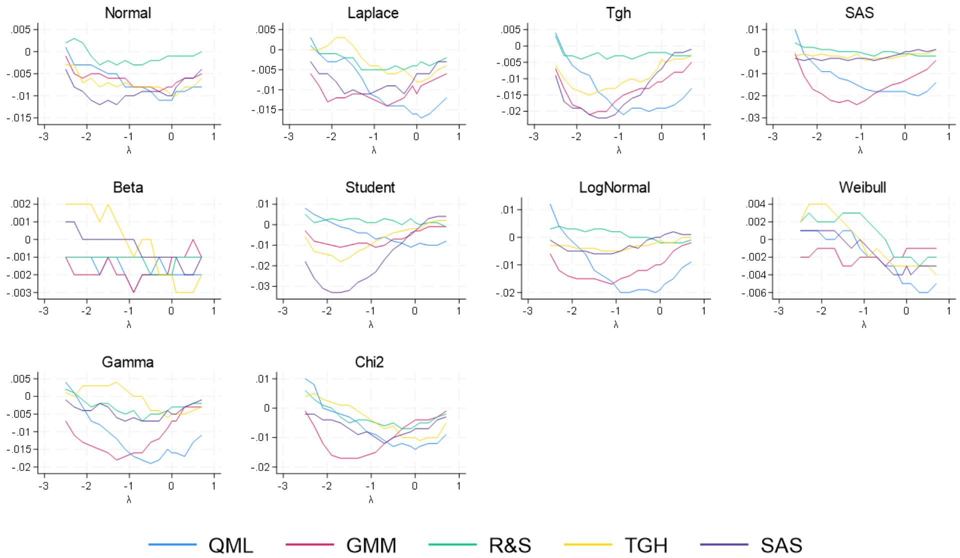

Figure 4: Bias of $\widehat{\lambda}$, $n = 500$

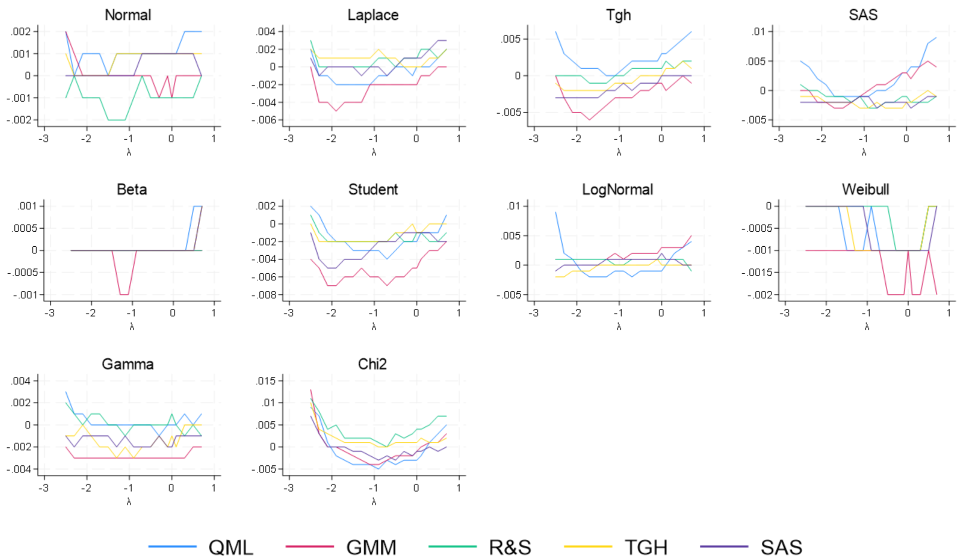

Figure 5: Bias of $\widehat{\beta}_1$, $n = 500$

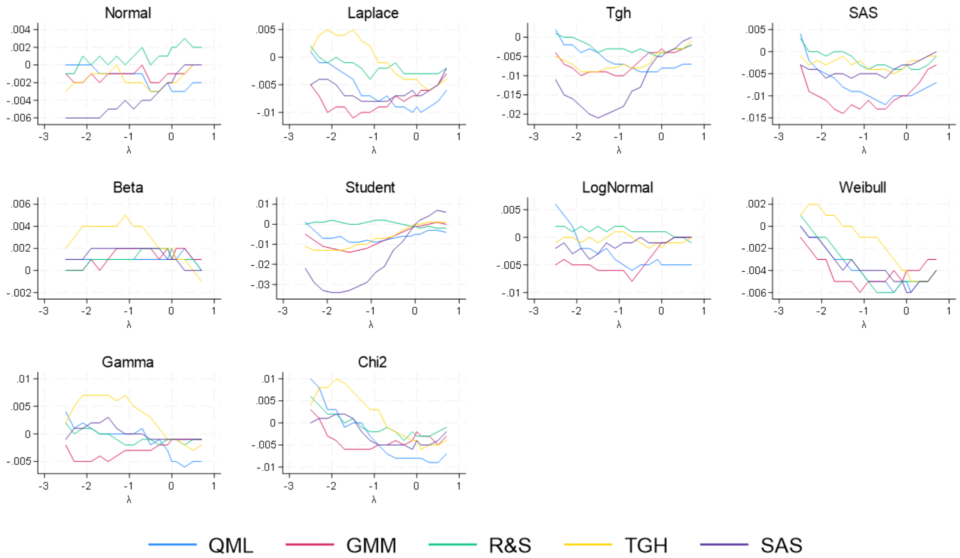

Figure 6: Bias of $\widehat{\lambda}$, $n = 900$

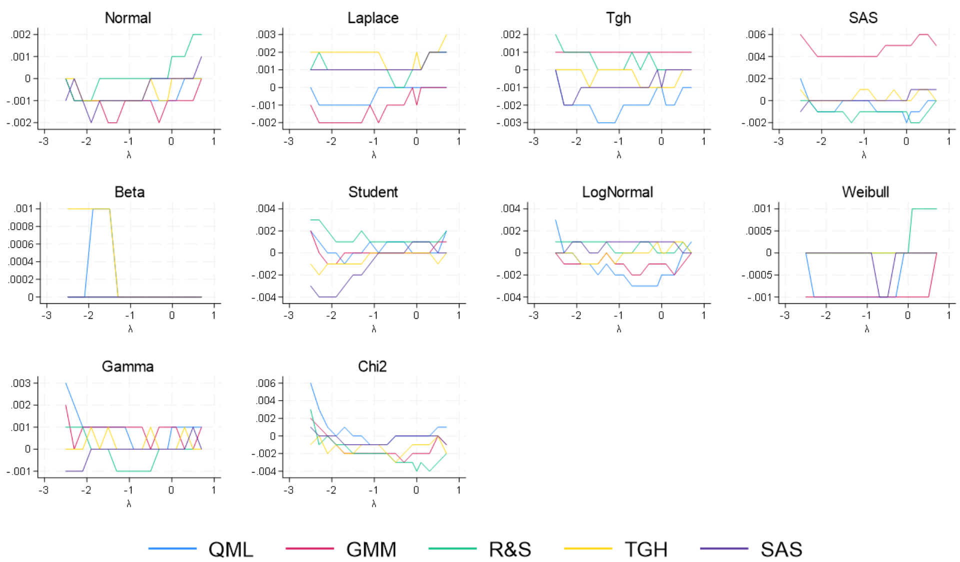

Figure 7: Bias of $\widehat{\beta}_1$, $n = 900$

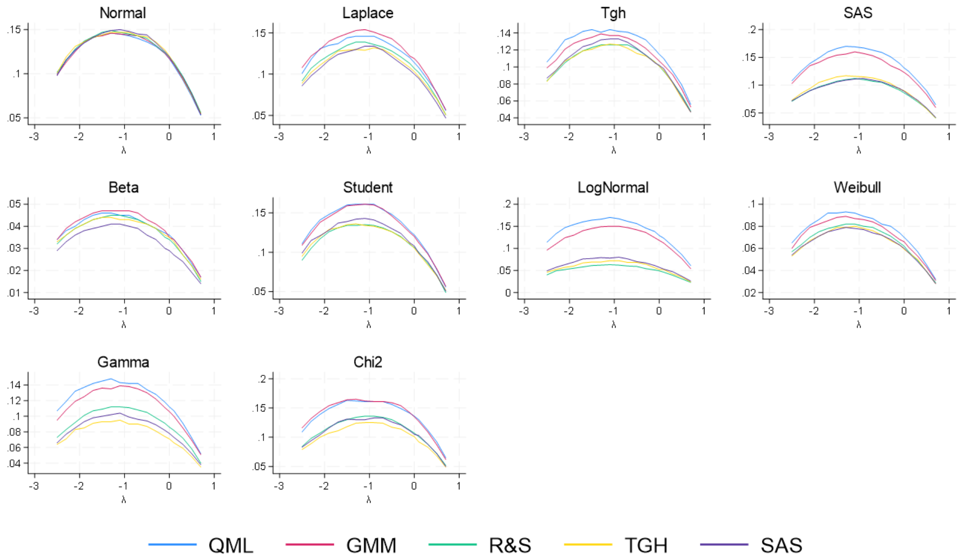

Figure 8: Dispersion of $\widehat{\lambda}$, $n = 300$

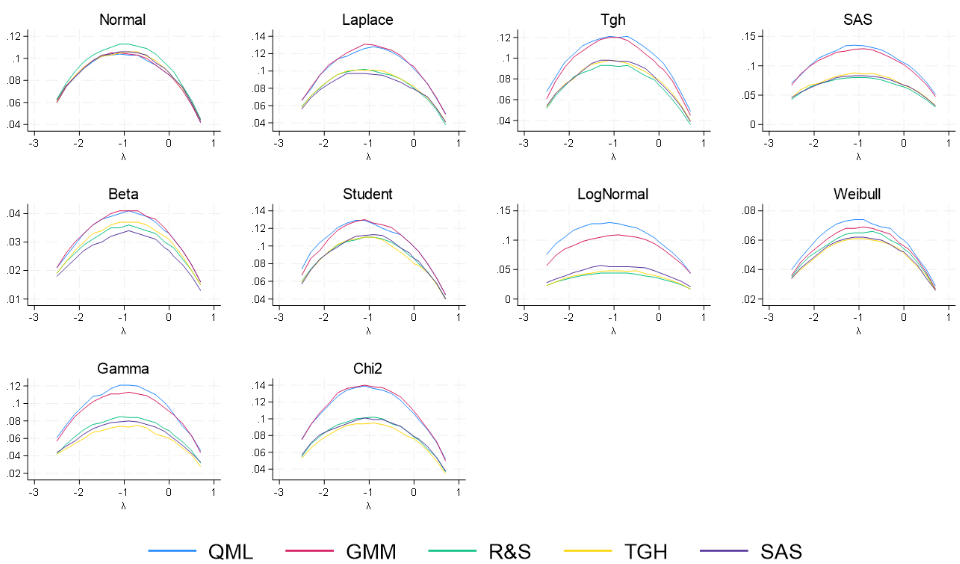

Figure 9: Dispersion of $\widehat{\lambda}$, $n = 500$

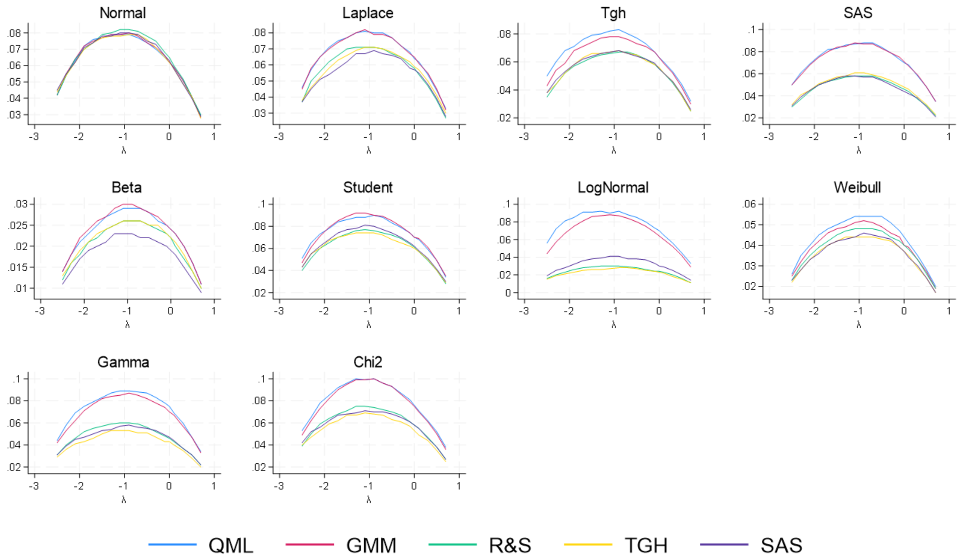

Figure 10: Dispersion of $\widehat{\lambda}$, $n = 900$

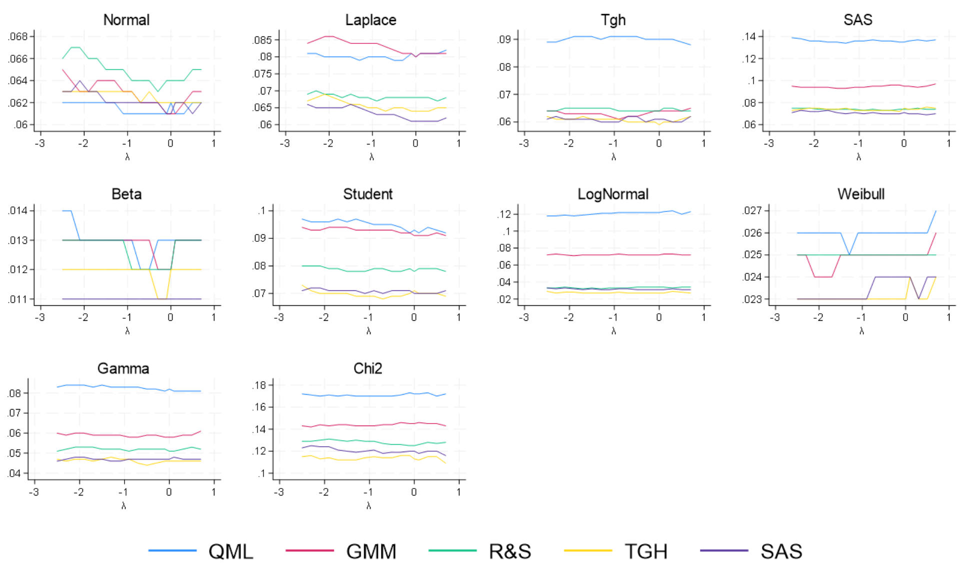

Figure 11: Dispersion of $\widehat{\beta}_1$, $n = 300$

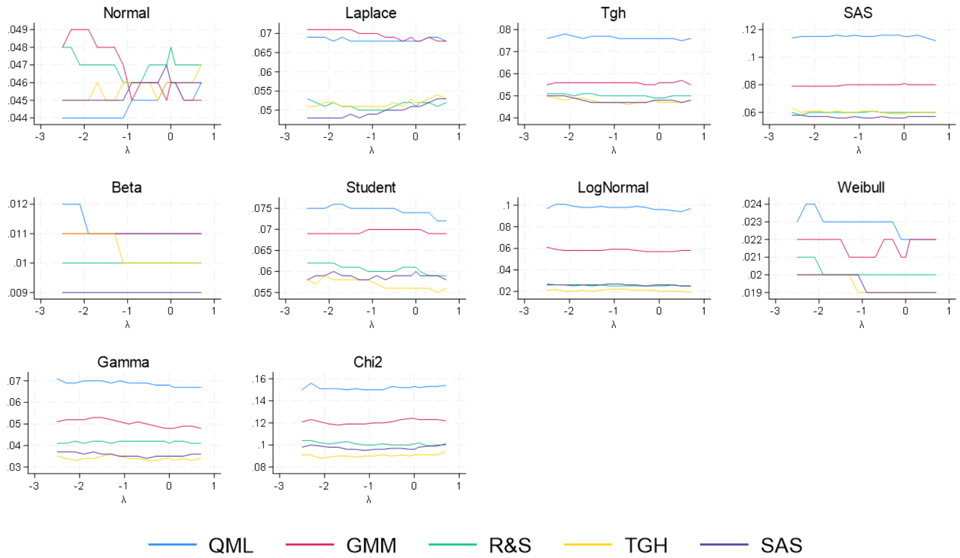

Figure 12: Dispersion of $\widehat{\beta}_1$, $n = 500$

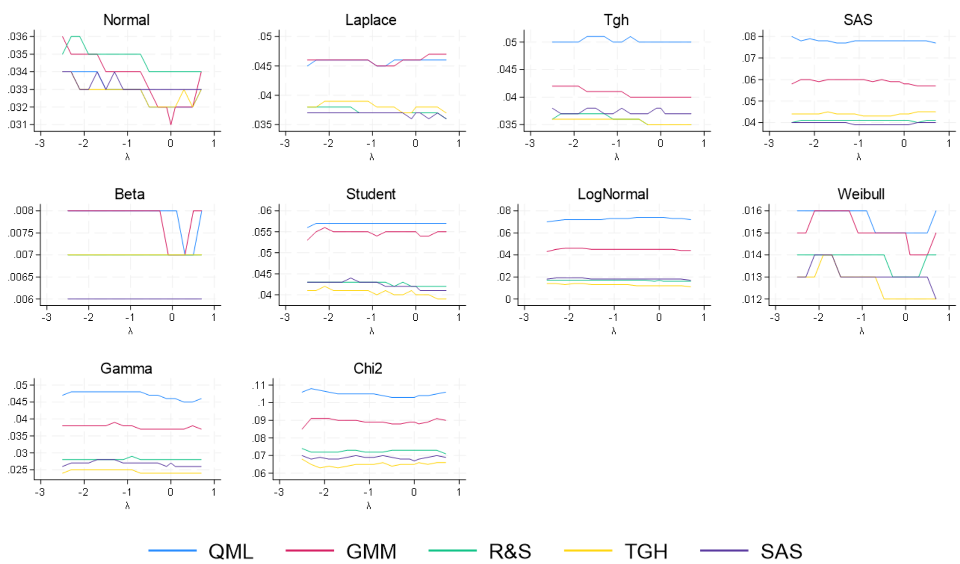

Figure 13: Dispersion of $\widehat{\beta}_1$, $n = 900$

7.3 Pseudo-bias of standard errors

Figures 14 to 16 assess the discrepancy in the estimated standard errors of the coefficient λ when the error term is assumed to be $\text{SAS}_{\epsilon,\delta}(\xi, \sigma)$-distributed. As mentioned above, this pseudo-bias is obtained by comparing the median standard errors in repeated samples with the dispersion of the estimated coefficient (measured by the standardized interquartile range). We generally do not observe significant differences between the two curves, regardless of the true distribution for the error terms. Only when the errors follow a Log-Normal distribution do we note a slight discrepancy for the smallest sample size ($n = 300$). However, this gap narrows for $n = 500$ and completely vanishes on the largest sample size ($n = 900$). Of course, both median standard errors and empirical dispersion decrease as the sample size increases.

Finally, Figures 17 to 19 present the same information but when the error term is assumed to be $T_{g,h}(\xi, \sigma)$-distributed. We note small discrepancies between the two curves when the true distribution does not exhibit a left-tail, such as the χ^2_5, the $\Gamma(0, 1, 2)$ or the LogNormal(0,1) distributions. These differences diminish as the sample size increases, but they do not completely disappear. These results can be attributed to the fact that the Tukey g-and-h distribution does not allow for tails lighter than those of the gaussian ($h < 0$), which limits its ability to accurately approximate such parametric distributions. In contrast, the SAS distribution captures this feature better because its elongation parameter allows for lighter-than-normal tails.

8 Conclusion

In this paper, we propose a general approach to obtain an efficient parametric estimator for the SAR model within the framework of Local Asymptotic Normality (LAN). The estimator is designed to accommodate any parametric distribution with an explicitly defined quantile function. In addition to presenting general results, we derive the estimator for two highly flexible distributions, namely the Tukey g-and-h and Sinh-Arcsinh (SAS) distributions, both of which can effectively approximate a wide range of distributions. These distributions are especially useful when the error distribution is unknown or exhibits non-standard features, such as asymmetry or heavy tails. Monte Carlo simulations illustrate that the proposed estimator performs well relative to traditional approaches, particularly when the true error distribution significantly deviates from normality.

Finally, we plan to expand this research in at least two directions. The first one relates to the treatment of heteroskedastic and/or clustered errors. The second pertains to the accounting of potential endogeneity of the determinants and/or of the connectivity matrix. Addressing these issues is crucial as they will expand the practical applicability of these flexible estimators across a broader range of empirical contexts.

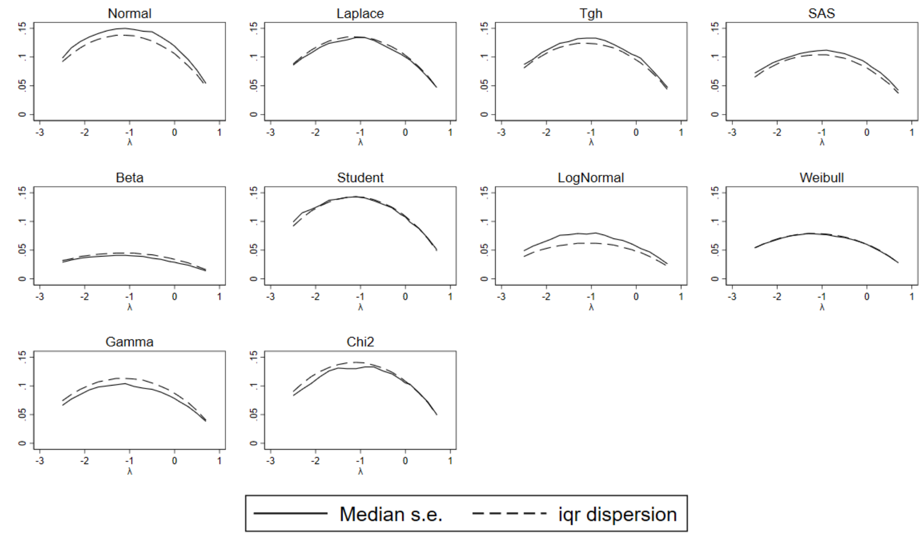

Figure 14: Behavior of standard errors for $\widehat{\lambda}$ assuming $SAS_{\epsilon,\delta}(\xi,\sigma)$-distributed errors, $n = 300$

Acknowledgments

We thank Vincent Boucher, Sophie Dabo, Arnaud Dufays, Abhimanyu Gupta, Xiaodong Liu, Florian Pelgrin and Francesca Rossi, as well as seminar participants at Ohio State University, Pise University, University of Lille, participants of London and Essex workshops, of the 20$^{\text{th}}$ International Spatial Econometrics and Spatial statistics Workshop, those of the Stata workshop organized in Aix-Marseille University and of the Stata Economic symposium. All remaining errors are our responsibility.

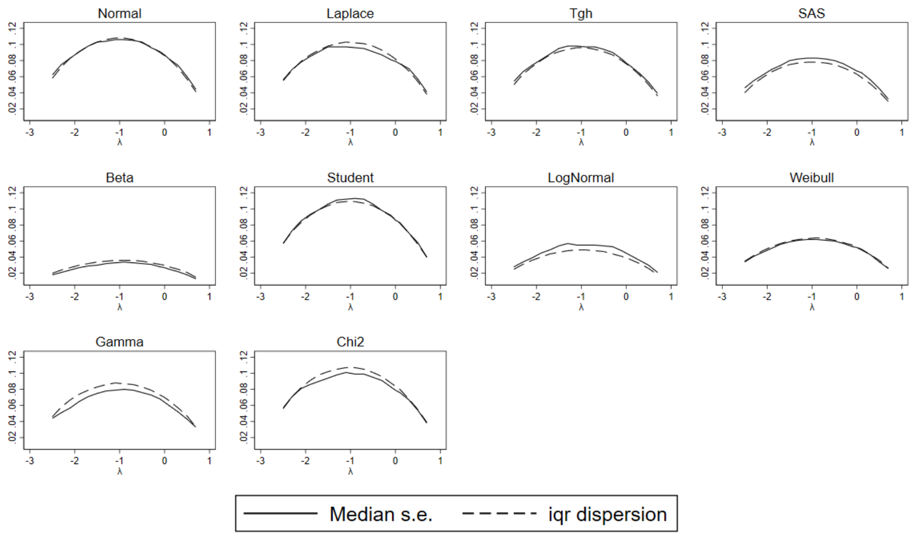

Figure 15: Behavior of standard errors for $\widehat{\lambda}$ assuming $SAS_{\epsilon,\delta}(\xi,\sigma)$-distributed errors, $n = 500$

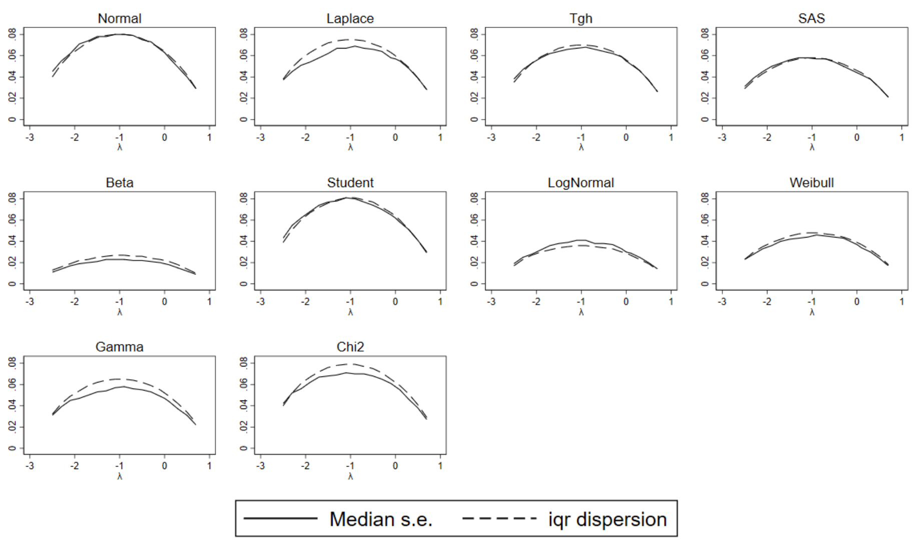

Figure 16: Behavior of standard errors for $\widehat{\lambda}$ assuming $SAS_{\epsilon,\delta}(\xi,\sigma)$-distributed errors, $n = 900$

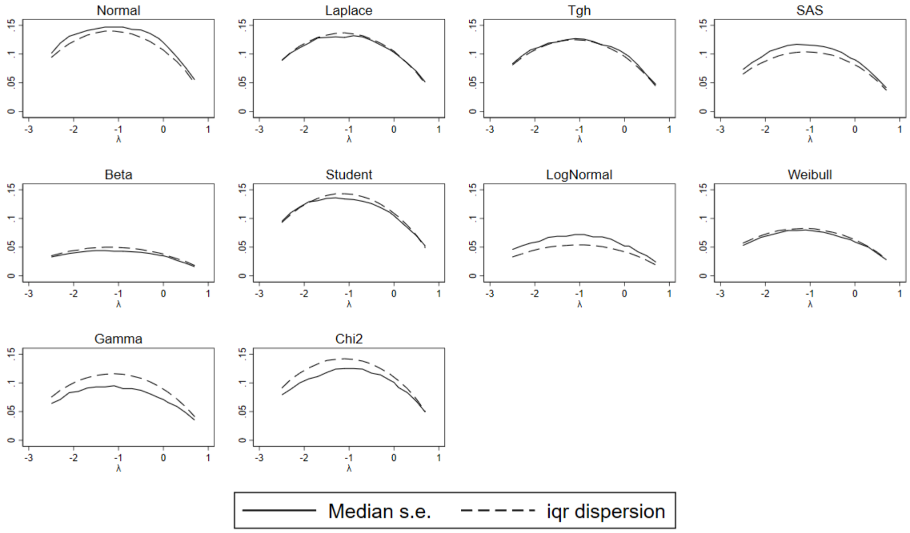

Figure 17: Behavior of standard errors for $\widehat{\lambda}$ assuming $T_{g,h}(\xi, \sigma)$ -distributed errors, $n = 300$

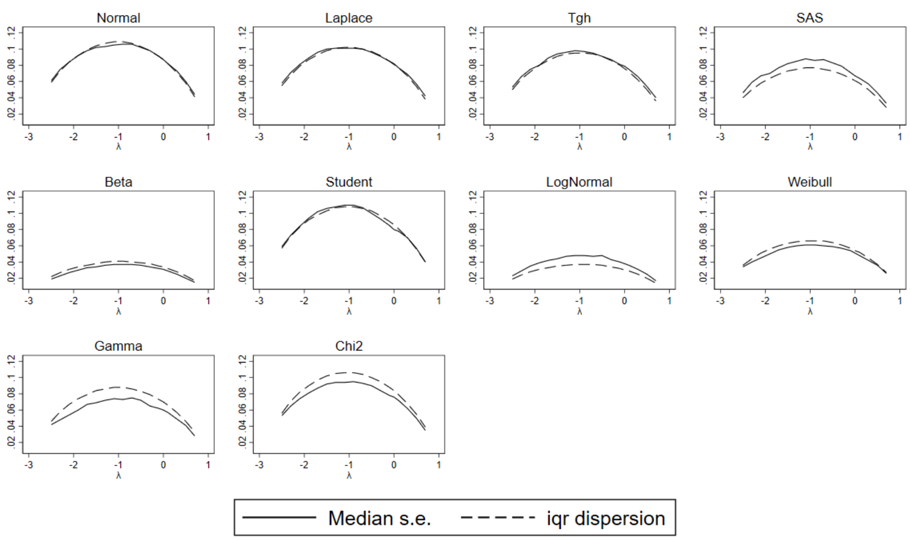

Figure 18: Behavior of standard errors for $\widehat{\lambda}$ assuming $T_{g,h}(\xi, \sigma)$-distributed errors, $n = 500$

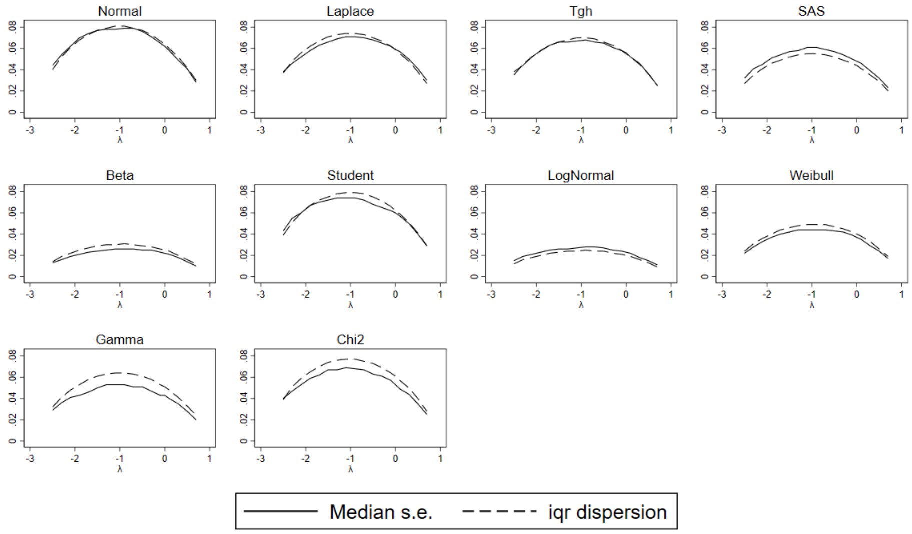

Figure 19: Behavior of standard errors for $\widehat{\lambda}$ assuming $T_{g,h}(\xi, \sigma)$-distributed errors, $n = 900$

References

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**, 41–55.

Debarsy, N., Verardi, V. & Vermandele, C. (2024), Semiparametrically efficient estimation of regression models with spillovers, Working Paper DT2024-04, Lille Economics and Management, UMR 9921, Université de Lille.

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

Hallin, M. (1996), éléments de la théorie asymptotique des expériences statistiques, *in* J.-J. Droesbeke & J. Fine, eds, 'Inférence non paramétrique fondée sur les Rangs', Editions del'Université' de Bruxelles et Ellipses, pp. 129–166.

Horn, R. A. & Johnson, C. (1985), *Matrix Analysis*, Cambridge University Press, London.

Jiménez Moscoso, J. & Arunachalam, V. (2011), 'Using tukey's g and h family of distributions to calculate value-at-risk and conditional value-at-risk', *Journal of Risk* **13**, 95–116.

Jones, C. & Pewsey, A. (2009), 'Sinh-arcsinh distributions', *Biometrika* **96**, 761–780.

Jones, M. C. (2015), 'On families of distributions with shape parameters', *International Statistical Review* **83**, 175–192.

Kelejian, H. H. & Prucha, I. R. (1998), 'A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *Journal of Real Estate Finance and Economics* **17**, 99–121.

Kelejian, H. H. & Prucha, I. R. (1999), 'A generalized moments estimator for the autoregressive parameter in a spatial model', *International Economic Review* **40**, 509–533.

Kelejian, H. H. & Prucha, I. R. (2010), 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics* **157**, 53–67.

Le Cam, L. (1960), 'Locally asymptotically normal families of distributions', *University of California Publications in Statistics* **3**, 37–98.

Le Cam, L. (1970), 'On the assumptions used to prove asymptotic normality of maximum likelihood estimates', *Annals of Mathematical Statistics* **41**, 802–828.

Le Cam, L. & Yang, G. L. (2000), *Asymptotics in Statistics; Some Basic Concepts, 2nd Edition*, Springer-Verlag, New York.

Lee, J. & Robinson, P. M. (2020), 'Adaptive inference on pure spatial models', *Journal of Econometrics* **216**, 375–393.

Lee, L.-F. (2002), 'Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models', *Econometric Theory* **18**, 252–277.

Lee, L.-f. (2003), 'Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances', *Econometric Reviews* **22**, 307–335.

Lee, L.-f. (2004), 'Asymptotic distributions of Quasi-Maximum Likelihood estimators for spatial autoregressive models', *Econometrica* **72**, 1899–1925.

Lee, L.-f. (2007), 'GMM and 2SLS estimation of mixed regressive, spatial autoregressive models', *Journal of Econometrics* **137**, 489–514.

Liu, X., Lee, L.-f. & Bollinger, C. R. (2010), 'An efficient GMM estimator of spatial autoregressive models', *Journal of Econometrics* **159**, 302–319.

MacGillivray, H. L. (1992), 'Shape properties of the g-and-h and johnson families', *Communications in Statistics - Theory and Methods* **21**, 123–1250.

Martinez, J. & Iglewicz, B. (1984), 'Some properties of the tukey g-and-h family of distributions', *Communications in Statistics - Theory and Methods* **13**, 353–369.

Neumayer, E. & Plümper, T. (2016), 'W', *Political Science Research and Methods* **4**, 175–193.

Ord, J. (1975), 'Estimation methods for models of spatial interaction', *Journal of the American Statistical Association* **70**, 120–126.

Patacchini, E. & Zenou, Y. (2012), 'Juvenile delinquency and conformism', *The Journal of Law, Economics and Organization* **28**, 1–31.

Rayner, G. & MacGillivray, H. (2002), 'Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions', *Statistics and Computing* **12**, 57–75.

Ricci, L., Verardi, V. & Vermandele, C. (2018), Efficient estimation of a linear regression model with skewed and/or heavy-tailed distributed errors. mimeo.

Robinson, P. M. (2010), 'Efficient estimation of the semiparametric spatial autoregressive model', *Journal of Econometrics* **157**, 6–17.

Tukey, J. W. (1977), *Modern techniques in data analysis*, pp. 761–780.

Van der Vaart, A. W. (1998), *Asymptotic Statistics, first edition*, Cambridge University Press.

White, H. (1984), *Asymptotic theory for econometricians*, Academic Press, Orlando.

Xu, G. & Genton, M. G. (2015), 'Efficient maximum approximated likelihood inference for tukey's g-and-h distribution', *Computational Statistics and Data Analysis* **91**, 78–91.

Xu, Y., Iglewicz, B. & Chervoneva, I. (2014), 'Robust estimation of the parameters of g and h distributions, with applications to outlier detection', *Computational Statistics and Data Analysis* **75**, 66–80.

A Appendix: Fisher information matrix under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$

Under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, the residuals $e_1^{(n)}(\boldsymbol{\beta},\lambda),\ldots,e_n^{(n)}(\boldsymbol{\beta},\lambda)$ are i.i.d. with density function $f_{\boldsymbol{\gamma}}$ and quantile function $Q_{\boldsymbol{\gamma}}$. Consequently, under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, the random variables $u_i^{(n)}(\boldsymbol{\theta}) = Q_{\boldsymbol{\gamma}}^{-1}\left(e_i^{(n)}(\boldsymbol{\beta},\lambda)\right)$, $i=1,\ldots,n$, are i.i.d. $U(0,1)$. Note that all the expectations considered hereafter will be computed under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$.

In this Appendix, to avoid making the notations unnecessarily cumbersome, we simply write $e_i^{(n)}$, $u_i^{(n)}$ and $\mathbf{G}^{(n)}$ for $e_i^{(n)}(\boldsymbol{\beta},\lambda)$, $u_i^{(n)}(\boldsymbol{\theta})$ and $\mathbf{G}^{(n)}(\lambda)$, respectively.

Let us consider the following expressions for the different components of the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ (see Section 4.2):

- Component relative to $\boldsymbol{\beta}$:

$$\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\mathbf{x}_i^{(n)};$$

- Component relative to λ:

$$\Delta_{\lambda}^{(n)}(\boldsymbol{\theta}) = -\frac{1}{\sqrt{n}}\mathrm{tr}\left(G^{(n)}\right) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\mathbf{W}_{i\boldsymbol{.}}^{(n)}\mathbf{y}^{(n)}.$$

Since, in view of (3),

$$\begin{aligned}\mathbf{W}_{i\boldsymbol{.}}^{(n)}\mathbf{y}^{(n)} &= \mathbf{G}_{i\boldsymbol{.}}^{(n)}\left(\mathbf{X}^{(n)}\boldsymbol{\beta} + \mathbf{e}^{(n)}\right)\\ &= \mathbf{G}_{i\boldsymbol{.}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta} + \sum_{j=1}^{n} G_{ij}^{(n)} e_j^{(n)}\\ &= \mathbf{G}_{i\boldsymbol{.}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta} + G_{ii}^{(n)} e_i^{(n)} + \sum_{\substack{j=1\\ j\neq i}}^{n} G_{ij}^{(n)} e_j^{(n)},\end{aligned}$$

we get

$$\Delta_{\lambda}^{(n)}(\boldsymbol{\theta}) = L_1^{(n)}(\boldsymbol{\theta}) + L_2^{(n)}(\boldsymbol{\theta}) + L_3^{(n)}(\boldsymbol{\theta}) + L_4^{(n)}(\boldsymbol{\theta}),$$

with

$$\begin{aligned}L_1^{(n)}(\boldsymbol{\theta}) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta}\right),\\ L_2^{(n)}(\boldsymbol{\theta}) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)Q_{\boldsymbol{\gamma}}(u_i^{(n)})G_{ii}^{(n)},\\ L_3^{(n)}(\boldsymbol{\theta}) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)Q_{\boldsymbol{\gamma}}(u_j^{(n)})G_{ij}^{(n)},\\ L_4^{(n)}(\boldsymbol{\theta}) =& -\frac{1}{\sqrt{n}}\mathrm{tr}\left(G^{(n)}\right);\end{aligned}$$

- Component relative to γ_r ($r \in \{1,\dots,R\}$):

$$\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H_{\boldsymbol{\gamma};r}\left(u_i^{(n)}\right).$$

We have, for all $i = 1,\dots,n$ and for $r, s \in \{1,\dots,R\}$, $r \neq s$:

- $\mathbb{E}\left[Q_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\right] = \mathbb{E}\left[e_i^{(n)}\right] = \int_0^1 Q_{\boldsymbol{\gamma}}(u)du \overset{\text{not}}{=} \mu_{\boldsymbol{\gamma}}$;

- $\mathbb{E}\left[Q^2_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\right] = \mathbb{E}\left[\left(e_i^{(n)}\right)^2\right] = \int_0^1 Q^2_{\boldsymbol{\gamma}}(u)du \overset{\text{not}}{=} \nu_{\boldsymbol{\gamma}};$

- $\mathbb{E}\left[\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\right] = \mathbb{E}\left[\phi_{f_{\boldsymbol{\gamma}}}\left(e_i^{(n)}\right)\right] = \int_{-\infty}^{\infty}\phi_{f_{\boldsymbol{\gamma}}}(e)f_{\boldsymbol{\gamma}}(e)de = -\int_{-\infty}^{\infty} f'_{\boldsymbol{\gamma}}(e)de = -\left[f_{\boldsymbol{\gamma}}(e)\right]_{-\infty}^{\infty} = 0;$

- $\mathbb{E}\left[\widetilde{Q}^2_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\right] = \mathbb{E}\left[\phi^2_{f_{\boldsymbol{\gamma}}}\left(e_i^{(n)}\right)\right] = \int_0^1 \overline{Q}^2_{\boldsymbol{\gamma}}(u)du \overset{\text{not}}{=} \mathcal{I}_{\boldsymbol{\gamma}};$

- $\mathbb{E}\left[\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)Q_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\right] = \mathbb{E}\left[\phi_{f_{\boldsymbol{\gamma}}}\left(e_i^{(n)}\right)e_i^{(n)}\right] = \int_{-\infty}^{\infty}\phi_{f_{\boldsymbol{\gamma}}}(e)ef_{\boldsymbol{\gamma}}(e)de = -\int_{-\infty}^{\infty} f'_{\boldsymbol{\gamma}}(e)ede = -\left[f_{\boldsymbol{\gamma}}(e)e\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f_{\boldsymbol{\gamma}}(e)de = 0 + 1 = 1;$

- $\mathbb{E}\left[\widetilde{Q}^2_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)Q_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\right] = \int_0^1 \widetilde{Q}^2_{\boldsymbol{\gamma}}(u)Q_{\boldsymbol{\gamma}}(u)du \overset{\text{not}}{=} \mathcal{K}_{\boldsymbol{\gamma}};$

- $\mathbb{E}\left[\widetilde{Q}^2_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)Q^2_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)\right] = \int_0^1 \widetilde{Q}^2_{\boldsymbol{\gamma}}(u)Q^2_{\boldsymbol{\gamma}}(u)du \overset{\text{not}}{=} \mathcal{L}_{\boldsymbol{\gamma}};$

- $\mathbb{E}\left[H_{\boldsymbol{\gamma};r}\left(u_i^{(n)}\right)\right] = 0;$

- $\mathbb{E}\left[H^2_{\boldsymbol{\gamma};r}\left(u_i^{(n)}\right)\right] = \int_0^1 H^2_{\boldsymbol{\gamma};r}(u)du \overset{\text{not}}{=} \mathcal{J}_{\boldsymbol{\gamma};r};$

- $\mathbb{E}\left[H_{\boldsymbol{\gamma};r}\left(u_i^{(n)}\right)H_{\boldsymbol{\gamma};s}\left(u_i^{(n)}\right)\right] = \int_0^1 H_{\boldsymbol{\gamma};r}(u)H_{\boldsymbol{\gamma};s}(u)du \overset{\text{not}}{=} \mathcal{J}_{\boldsymbol{\gamma};r,s};$

- $\mathbb{E}\left[\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)H_{\boldsymbol{\gamma};r}\left(u_i^{(n)}\right)\right] = \int_0^1 \widetilde{Q}_{\boldsymbol{\gamma}}(u)H_{\boldsymbol{\gamma};r}(u)du \overset{\text{not}}{=} \mathcal{H}_{\boldsymbol{\gamma};r};$

- $\mathbb{E}\left[\widetilde{Q}_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)Q_{\boldsymbol{\gamma}}\left(u_i^{(n)}\right)H_{\boldsymbol{\gamma};r}\left(u_i^{(n)}\right)\right] = \int_0^1 \widetilde{Q}_{\boldsymbol{\gamma}}(u)Q_{\boldsymbol{\gamma}}(u)H_{\boldsymbol{\gamma};r}(u)du \overset{\text{not}}{=} \mathcal{M}_{\boldsymbol{\gamma};r}.$

Straightforward calculations give us the following results:

a)
$$\mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta})\left(\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta})\right)^{\mathrm{T}}\right] = \mathcal{I}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(n)}(\mathbf{x}_i^{(n)})^{\mathrm{T}}\right\};$$

b)

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta})\Delta^{(n)}_{\lambda}(\boldsymbol{\theta})\right] &= \mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta})L_1^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta})L_2^{(n)}(\boldsymbol{\theta})\right] \\ &+ \mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta})L_3^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\beta}}(\boldsymbol{\theta})L_4^{(n)}(\boldsymbol{\theta})\right],\end{aligned}$$

where

$$\mathbb{E}\left[\mathbf{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})L_1^{(n)}(\boldsymbol{\theta})\right] = \mathcal{I}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(n)}\left(\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta}\right)\right\},$$

$$\mathbb{E}\left[\mathbf{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})L_2^{(n)}(\boldsymbol{\theta})\right] = \mathcal{K}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(n)}G_{ii}^{(n)}\right\},$$

$$\mathbb{E}\left[\mathbf{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})L_3^{(n)}(\boldsymbol{\theta})\right] = \mathcal{I}_{\boldsymbol{\gamma}}\mu_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbf{x}_i^{(n)}G_{ij}^{(n)}\right\},$$

$$\mathbb{E}\left[\mathbf{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})L_4^{(n)}(\boldsymbol{\theta})\right] = 0;$$

c) for $r \in \{1, \ldots, R\}$:

$$\mathbb{E}\left[\mathbf{\Delta}_{\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{H}_{\boldsymbol{\gamma};r}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(n)}\right\};$$

d)

$$\begin{aligned}\mathbb{E}\left[\left(\Delta_{\lambda}^{(n)}(\boldsymbol{\theta})\right)^2\right] &= \mathbb{E}\left[\left(L_1^{(n)}(\boldsymbol{\theta})\right)^2\right] + \mathbb{E}\left[\left(L_2^{(n)}(\boldsymbol{\theta})\right)^2\right] + \mathbb{E}\left[\left(L_3^{(n)}(\boldsymbol{\theta})\right)^2\right] + \mathbb{E}\left[\left(L_4^{(n)}(\boldsymbol{\theta})\right)^2\right]\\ &+ 2\mathbb{E}\left[L_1^{(n)}(\boldsymbol{\theta})L_2^{(n)}(\boldsymbol{\theta})\right] + 2\mathbb{E}\left[L_1^{(n)}(\boldsymbol{\theta})L_3^{(n)}(\boldsymbol{\theta})\right] + 2\mathbb{E}\left[L_1^{(n)}(\boldsymbol{\theta})L_4^{(n)}(\boldsymbol{\theta})\right]\\ &+ 2\mathbb{E}\left[L_2^{(n)}(\boldsymbol{\theta})L_3^{(n)}(\boldsymbol{\theta})\right] + 2\mathbb{E}\left[L_2^{(n)}(\boldsymbol{\theta})L_4^{(n)}(\boldsymbol{\theta})\right] + 2\mathbb{E}\left[L_3^{(n)}(\boldsymbol{\theta})L_4^{(n)}(\boldsymbol{\theta})\right],\end{aligned}$$

where

$$\mathbb{E}\left[\left(L_1^{(n)}(\boldsymbol{\theta})\right)^2\right] = \mathcal{I}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta}\right)^2\right\},$$

$$\mathbb{E}\left[\left(L_2^{(n)}(\boldsymbol{\theta})\right)^2\right] = (\mathcal{L}_{\boldsymbol{\gamma}}-1)\left\{\frac{1}{n}\sum_{i=1}^{n}\left(G_{ii}^{(n)}\right)^2\right\} + \frac{\left(\mathrm{tr}\left(\mathbf{G}^{(n)}\right)\right)^2}{n},$$

$$\mathbb{E}\left[\left(L_3^{(n)}(\boldsymbol{\theta})\right)^2\right] = \mathcal{I}_{\boldsymbol{\gamma}}\nu_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(G_{ij}^{(n)}\right)^2\right\} + \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}G_{ij}^{(n)}G_{ji}^{(n)}\right\}$$

$$+\mathcal{I}_{\boldsymbol{\gamma}}\mu_{\boldsymbol{\gamma}}^2\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i,j}}^{n}G_{ij}^{(n)}G_{ik}^{(n)}\right\},$$

$$\mathbb{E}\left[\left(L_4^{(n)}(\boldsymbol{\theta})\right)^2\right] = \frac{\left(\mathrm{tr}\left(\mathbf{G}^{(n)}\right)\right)^2}{n},$$

$$\mathbb{E}\left[L_1^{(n)}(\boldsymbol{\theta})L_2^{(n)}(\boldsymbol{\theta})\right] = \mathcal{K}_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta}\right)G_{ii}^{(n)}\right\},$$

$$\mathbb{E}\left[L_1^{(n)}(\boldsymbol{\theta})L_3^{(n)}(\boldsymbol{\theta})\right] = \mathcal{I}_{\boldsymbol{\gamma}}\mu_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta}\right)G_{ij}^{(n)}\right\},$$

$$\mathbb{E}\left[L_2^{(n)}(\boldsymbol{\theta})L_3^{(n)}(\boldsymbol{\theta})\right] = \mathcal{K}_{\boldsymbol{\gamma}}\mu_{\boldsymbol{\gamma}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}G_{ii}^{(n)}G_{ij}^{(n)}\right\},$$

$$\mathbb{E}\left[L_1^{(n)}(\boldsymbol{\theta})L_4^{(n)}(\boldsymbol{\theta})\right] = \mathbb{E}\left[L_3^{(n)}(\boldsymbol{\theta})L_4^{(n)}(\boldsymbol{\theta})\right] = 0,$$

and

$$\mathbb{E}\left[L_2^{(n)}(\boldsymbol{\theta})L_4^{(n)}(\boldsymbol{\theta})\right] = -\frac{\left(\mathrm{tr}\left(\mathbf{G}^{(n)}\right)\right)^2}{n};$$

e) for $r \in \{1,\ldots,R\}$:

$$\mathbb{E}\left[\Delta_{\lambda}^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] = \mathbb{E}\left[L_1^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[L_2^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right]$$

$$+\,\mathbb{E}\left[L_3^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[L_4^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right],$$

where

$$\mathbb{E}\left[L_1^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{H}_{\boldsymbol{\gamma};r}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta}\right\},$$

$$\mathbb{E}\left[L_2^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{M}_{\boldsymbol{\gamma};r}\frac{\operatorname{tr}\left(\mathbf{G}^{(n)}\right)}{n},$$

$$\mathbb{E}\left[L_3^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{H}_{\boldsymbol{\gamma};r}\mu_{\boldsymbol{\gamma}}\left\{n\overline{G}_{\boldsymbol{\cdot\cdot}}^{(n)} - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}\right)}{n}\right\},$$

$$\mathbb{E}\left[L_4^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right] = 0;$$

f) for $r \in \{1, \ldots, R\}$:

$$\mathbb{E}\left[\left(\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\right)^2\right] = \mathcal{J}_{\boldsymbol{\gamma};r};$$

g) for $r, s \in \{1, \ldots, R\}$, $r \neq s$:

$$\mathbb{E}\left[\Delta_{\gamma_r}^{(n)}(\boldsymbol{\theta})\Delta_{\gamma_s}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{J}_{\boldsymbol{\gamma};r,s}.$$

The expression for the parametric Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ follows directly from the above results.