Saddlepoint Approximations for Spatial Panel Data Models

Jiang, Chaonan; La Vecchia, Davide; Ronchetti, Elvezio; Scaillet, Olivier

How to cite

JIANG, Chaonan et al. Saddlepoint Approximations for Spatial Panel Data Models. In: Journal of the American Statistical Association, 2021, vol. 118, n° 542, p. 1164–1175. doi: 10.1080/01621459.2021.1981913

This publication URL: https://archive-ouverte.unige.ch/unige:169309

Publication DOI: 10.1080/01621459.2021.1981913

Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Saddlepoint Approximations for Spatial Panel Data Models

Chaonan Jiang, Davide La Vecchia, Elvezio Ronchetti & Olivier Scaillet

To cite this article: Chaonan Jiang, Davide La Vecchia, Elvezio Ronchetti & Olivier Scaillet (2023) Saddlepoint Approximations for Spatial Panel Data Models, Journal of the American Statistical Association, 118:542, 1164-1175, DOI: 10.1080/01621459.2021.1981913

To link to this article: https://doi.org/10.1080/01621459.2021.1981913

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 118, NO. 542, 1164–1175: Theory and Methods
https://doi.org/10.1080/01621459.2021.1981913

Saddlepoint Approximations for Spatial Panel Data Models

Chaonan Jiang[a], Davide La Vecchia[a], Elvezio Ronchetti[a], and Olivier Scaillet[b,c]

[a]Research Center for Statistics and Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland; [b]Geneva Finance Research Institute, Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland; [c]Swiss Finance Institute, Geneva, Switzerland

ABSTRACT
We develop new higher-order asymptotic techniques for the Gaussian maximum likelihood estimator in a spatial panel data model, with fixed effects, time-varying covariates, and spatially correlated errors. Our saddlepoint density and tail area approximation feature *relative error* of order $O(1/(n(T-1)))$ with n being the cross-sectional dimension and T the time-series dimension. The main theoretical tool is the tilted-Edgeworth technique in a nonidentically distributed setting. The density approximation is always nonnegative, does not need resampling, and is accurate in the tails. Monte Carlo experiments on density approximation and testing in the presence of nuisance parameters illustrate the good performance of our approximation over first-order asymptotics and Edgeworth expansion. An empirical application to the investment–saving relationship in OECD (Organisation for Economic Co-operation and Development) countries shows disagreement between testing results based on the first-order asymptotics and saddlepoint techniques. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

ARTICLE HISTORY
Received July 2020
Accepted September 2021

KEYWORDS
Higher-order asymptotics; Investment-saving; Random field; Tail area

1. Introduction

Accounting for spatial dependence is of interest from both an applied and a theoretical point of view. Indeed, panel data with spatial cross-sectional interaction enable empirical researchers to take into account the temporal dimension and, at the same time, control for the spatial dependence. From a theoretical point of view, the special features of panel data with spatial effects present the challenge to develop new methodological tools.

Much of the machinery for conducting statistical inference on panel data models has been established under the simplifying assumption of cross-sectional independence. This assumption may be inadequate in many cases. For instance, correlation across spatial data comes typically from competition, spillovers, or aggregation. The presence of such a correlation might be anticipated in observable variables and/or in the unobserved disturbances in a statistical model and ignoring it can have adverse effects on routinely applied inferential procedures. See, for example, Gaetan and Guyon (2010), Rosenblatt (2012), Cressie (2015), Cressie and Wikle (2015), and recently Wikle, Zammit-Mangion, and Cressie (2019) for book-length discussions in the statistical literature. In the econometric literature, see, for example, Kapoor, Kelejian, and Prucha (2007), Lee and Yu (2010), Robinson and Rossi (2014b), Robinson and Rossi (2015), and, for book-length presentations (Baltagi 2008, ch. 13; Anselin 1988; Kelejian and Piras 2017).

Different nonparametric, semiparametric, and parametric approaches have been proposed to incorporate cross-sectional dependence in panel data models. The choice on the modeling approach depends on the time series (T) and cross-sectional (n) dimensions. A nonparametric approach is only feasible when T is large relative to n. In other situations, typically when T is very small (e.g., $T = 2$) and n is large, semiparametric models have been employed, including time-varying regressors (namely factor models) and spatial autoregressive component, when information on spatial distances is available. Least-square and quasi-maximum-likelihood estimator represent the main popular tools for estimation within this setting. When both T and n are small, the parametric approach is the sensible choice and (Gaussian) likelihood-based procedures are applied to define the maximum likelihood estimator (MLE).

There is a vast literature on the MLE for spatial autoregressive models, an early reference being Ord (1975). The derivation of the first-order asymptotics is available in the econometric literature; we refer to the seminal article by Lee (2004). For the class of spatial autoregressive processes, with fixed effects, time-varying covariates, and spatially correlated errors that we consider in this article, the first-order asymptotic results for the Gaussian MLE are available in Lee and Yu (2010), where the authors derived asymptotic approximations (the exact finite-sample distribution being intractable), when the cross-sectional dimension n is large and T is finite or large.

The main issue related to the first-order asymptotic approximations is that, when n is not very large, such approximations may be unreliable: alternatives are highly recommended. Bao and Ullah (2007) provided analytic formulae for the second-order bias and mean squared error of the MLE for the spatial parameter λ, in a Gaussian model. Bao (2013) and Yang (2015) extended these approximations to include also exogenous explanatory variables, which remain valid also when the

CONTACT Chaonan Jiang jcn0127@gmail.com Research Center for Statistics and Geneva School of Economics and Management, University of Geneva, Blv. Pont d'Arve 40, 1211 Geneva, Switzerland.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

process is not Gaussian. Robinson and Rossi (2014a,b) developed Edgeworth-improved tests for no spatial correlation in spatial autoregressive models for pure cross-sectional data based on least-squares estimation and Lagrange multiplier tests. Moreover, Robinson and Rossi (2015) worked on the concentrated likelihood and derived an Edgeworth expansion for MLE of λ in the setting of the first-order spatial autoregressive panel data model, with fixed effects and without covariates. Hillier and Martellosio (2018) (see their Section 6) and Martellosio and Hillier (2020) (see their Section 3.5) proposed saddlepoint approximations for the profile likelihood estimator of λ.

Resampling methods are also available alternatives to improve on the first-order asymptotics, achieving higher-order asymptotic refinements in terms of *absolute* error. However, they require either a bias correction or an asymptotically pivotal statistics; see Hall (1992) and Horowitz (2001) in the iid setting. To the best of our knowledge, for spatial panel models considered in this article, such results are not available.

The aim of this article is to introduce *saddlepoint approximations* for parametric spatial autoregressive panel data models with fixed effects and time-varying covariates. They overcome the problems mentioned above by means of the tilted-Edgeworth technique. For general references on saddlepoint approximations in the iid setting, see the seminal article of Daniels (1954) and the book-length presentations of Field and Ronchetti (1990), Jensen (1995), Kolassa (2006), and Brazzale, Davison, and Reid (2007). For a result about testing on spatial dependence, see Tiefelsdorf (2002), and for developments in time series models, see La Vecchia and Ronchetti (2019).

We remark that we could cast the methodology of this article into the framework of statistical analysis of random fields on a network graph, where the underlying, known, network graph describes the spatial structure; see, for example, Kolaczyk (2009, chap. 8) for a book-length introduction. In Section 2, we briefly comment on this approach. For the ease-of-reference to the extant econometric literature, in the rest of the article, we prefer to stick to the econometric notation and terminology of spatial panel data models.

The article is organized as follows. In Section 2, we provide a motivating example. Section 3 defines the general model setting and the estimation method, whereas the detailed methodology is presented in Section 4. Section 5 provides numerical comparison with other methods and, in Section 5.2, we tackle the problem of testing in the presence of nuisance parameters. In Section 6, we present an empirical application. The online supplementary material contains detailed derivations, technical appendices (algorithms and computational aspects are in Appendix C) and additional numerical experiments.

2. Motivating Example

We motivate our research by a Monte Carlo (MC) exercise illustrating the low accuracy of the routinely applied first-order asymptotics. We consider the model:

$$\begin{aligned} Y_{nt} &= \lambda_0 W_n Y_{nt} + X_{nt}\beta_0 + c_{n0} + E_{nt}, \\ E_{nt} &= \rho_0 M_n E_{nt} + V_{nt}, \qquad t = 1, ..., T, \end{aligned} \tag{1}$$

where $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{nt})'$, X_{nt} is an $n \times k$ matrix of nonstochastic time-varying regressors, c_{n0} is an $n \times 1$ vector of fixed effects, and $V_{nt} = (v_{1t}, v_{2t}, .., v_{nt})'$ are $n \times 1$ vectors with $v_{it} \sim \mathcal{N}(0, \sigma_0^2)$, iid across i and t. The matrices W_n and M_n are weighting matrices describing the spatial dynamics. Following the literature, we label this model SARAR(1,1) to emphasize the spatial dependence in both the response variable Y_{nt} and the error E_{nt}.

As in the MC example in Lee and Yu (2010) p. 172, we generate samples from (1) using $\theta_0 = (\beta_0, \lambda_0, \rho_0, \sigma_0^2)' = (1.0, 0.2, 0.5, 1)'$, $T = 5$, and $k = 4$ covariates. The quantities X_{nt}, c_{n0} and V_{nt} are generated from independent standard normal distributions and, as it is customary in the econometric literature, we set $W_n = M_n$, where the off-diagonal elements are different from zero, while the diagonal elements are all zero. We consider two sample sizes: $n = 24$ (small sample) and $n = 100$ (moderate/large sample). The choice of $n = 24$ is related to the empirical data analysis that we consider in Section 6, where we apply the model in Equation (1) to conduct inference on the investment-saving relation for the 24 OECD (Organisation for Economic Co-operation and Development) countries. Similar sample sizes arise in many real data analyses, where panel datasets contain a limited number of cross-sectional units, for example, because sampling can be expensive and time consuming, as it is often the case in field studies.

We consider three different spatial weighting matrices: Rook, Queen, and Queen with torus. They are commonly applied in statistics/econometrics; see, for example, the real-data examples in Bivand et al. (2008, chap. 9), the numerical examples in Lee and Yu (2010) and references therein. Besides, those matrices are implemented in the statistical/econometric software. For example, they are readily available in the R packages `spml` and `spdep` that we apply in our MC exercices.

In Figure 1, we display the geometry of Y_{nt} as implied by each considered spatial matrix: the plots highlight that different matrices imply different spatial relations. For instance, we see that the Rook matrix implies fewer links than the Queen matrix. Indeed, the Rook criterion defines neighbors by the existence of a common edge between two spatial units, whilst the Queen criterion is less rigid and defines neighbors as spatial units sharing an edge or a vertex. Besides, we may interpret $\{Y_{nt}\}$ as a n-dimensional random field on the network graph which describes the known underlying spatial structure. Then, W_n represents the weighted adjacency matrix (in the spatial econometrics literature, W_n is called contiguity matrix). In Figure 1, we display the geometry of a random field on a regular lattice (undirected graph). In the real data example of Section 6, we consider a random field over a manifold (a sphere), providing two additional examples for W_n.

For each type of W_n, we generate a sample of n observations. Since c_{n0} creates an incidental parameter issue, we eliminate it by the standard differentiation procedure, and we estimate the model parameter θ by MLE using the transformation approach of Lee and Yu (2010) for each MC run. We set the MC size equal to 5000 and compare the distribution of $\hat{\lambda}$ to the Gaussian asymptotic distribution (see Section 4.1 for details). Via QQ-plot analysis, Figure 2 shows that the Gaussian approximation can be either too thin or too thick in the tails with respect to the

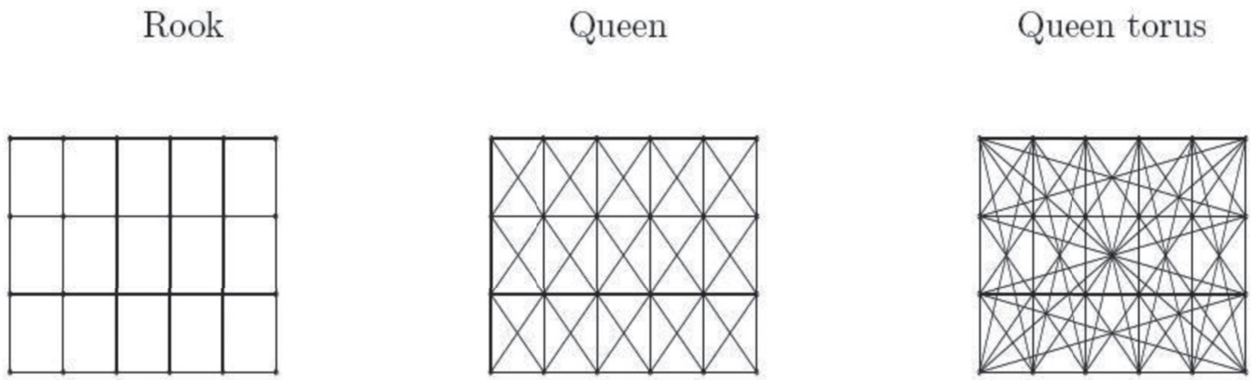

Figure 1. Different types of neighboring structure for Y_{nt}, as implied by different types of W_n matrix, for $n = 24$.

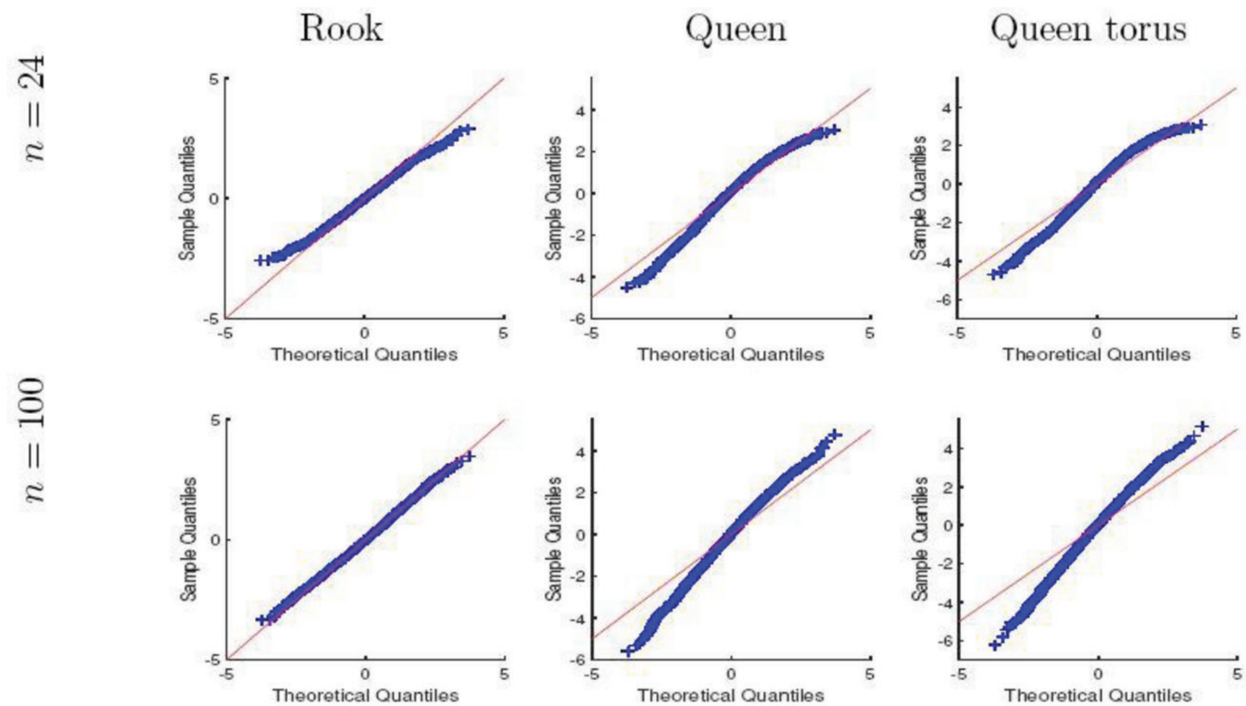

Figure 2. SARAR(1,1) model: QQ-plot vs normal of the MLE $\hat{\lambda}$, for different sample sizes ($n = 24$ and $n = 100$), $\lambda_0 = 0.2$, and different types of W_n matrix.

"exact" distribution. For instance, when $n = 24$ and W_n is Rook, the Gaussian quantiles are larger than the "exact" ones in the left tail, while we observe the opposite phenomenon in the right tail. Similar considerations hold for the other types of W_n. The more complex is the geometry of W_n (e.g., W_n has Queen structure) the more pronounced are the departures from the Gaussian. For $n = 100$, and W_n Rook, the MLE displays a distribution which is in line with the Gaussian one. However, when W_n becomes more complex (e.g., Queen with torus), larger departures in the tails are still evident. In Appendix D.1, we illustrate that similar conclusions are available also for the simpler SAR(1) model:

$$Y_{nt} = \lambda_0 W_n Y_{nt} + c_{n0} + V_{nt}, \quad \text{for} \quad t = 1, 2, \qquad (2)$$

where $\theta_0 = (\lambda_0, \sigma_0^2)'$. More generally, unreported results suggest that, in the considered SARAR setting, the "exact" and the asymptotic distribution, as well as the saddlepoint approximation, agree for the considered types of W_n, when $n \geq 250$.

3. Model Setting and Estimation Method

Let us consider a random field described by the SARAR(1,1) model in Equation (1). We label by $P_{\theta_0} \in \mathcal{P}$, with $\theta_0 \in \Theta \subset \mathbb{R}^d$, the actual underlying distribution, which is characterized by $\theta_0 = (\beta_0, \lambda_0, \rho_0, \sigma_0^2)'$, the true parameter value. The matrix W_n is an $n \times n$ nonstochastic spatial weight matrix that generates the spatial dependence on y_{it} among cross-sectional units. The matrix X_{nt} is an $n \times k$ matrix of nonstochastic time varying regressors, and c_{n0} is an $n \times 1$ vector of fixed effects. Similarly, M_n is an $n \times n$ spatial weight matrix for the disturbances—quite often $W_n = M_n$. Moreover, we define $S_n(\lambda) = I_n - \lambda W_n$, and analogously $R_n(\rho) = I_n - \rho M_n$.

The vector c_{n0} introduces an incidental parameter problem; see Lee and Yu (2010) and Robinson and Rossi (2015). To cope with this issue, we follow the standard approach, and we transform the model in order to derive a consistent estimator for the model parameter $\theta = (\beta', \lambda, \rho, \sigma^2)'$ and $\theta \in \Theta \subset \mathbb{R}^d$. To achieve the goal, we first eliminate the individual effects by the deviation from the time-mean operator $J_T = (I_T - \frac{1}{T} l_T l_T')$, where I_T is the $T \times T$ identity matrix, and $l_T = (1, ..., 1)'$, namely the $T \times 1$ vector of ones. Without creating linear dependence in the resulting disturbances, we adopt the transformation introduced by Lee and Yu (2010). First, let the orthonormal eigenvector matrix of J_T be $[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T]$, where $[\cdot]$ represents a matrix horizontal concatenation and $F_{T,T-1}$ is the $T \times (T-1)$ submatrix corresponding to the unit eigenvalues. Then, for any $n \times T$ matrix $[Z_{n1}, ..., Z_{nT}]$, we define the transformed $n \times (T-1)$ matrix $[Z^*_{n1}, ..., Z^*_{nT}] = [Z_{n1}, ..., Z_{nT}]F_{T,T-1}$. Similarly, we define $X^*_{nt} = [X^*_{nt,1}, X^*_{nt,2}, ..., X^*_{nt,k}]$. Then, we transform the model in Equation (1) and we obtain

$$\begin{aligned} Y^*_{nt} &= \lambda_0 W_n Y^*_{nt} + X^*_{nt}\beta_0 + E^*_{nt}, \\ E^*_{nt} &= \rho_0 M_n E^*_{nt} + V^*_{nt}, \qquad t = 1, 2, ..., T. \end{aligned} \tag{3}$$

Since $\left(V^{*'}_{n1}, ..., V^{*'}_{n(T-1)}\right)' = \left(F'_{T,T-1} \otimes I_n\right)\left(V'_{n1}, ..., V'_{n(T-1)}\right)'$, and the v_{it} are iid, we have

$$\mathbb{E}\left[\left(V^{*'}_{n1}, ..., V^{*'}_{n(T-1)}\right)'\left(V^{*'}_{n1}, ..., V^{*'}_{n(T-1)}\right)\right] = \sigma_0^2 I_{n(T-1)},$$

where $\mathbb{E}[\cdot]$ represents the expectation taken w.r.t. P_{θ_0}. The Gaussian assumption on the innovation terms implies that v^*_{it} are independent for all i and t—without this assumption, they would be simply uncorrelated; see Lee and Yu (2010, p. 167). Thus, defining $\zeta = (\beta', \lambda, \rho)'$, the log-likelihood is

$$\begin{aligned} \ln L_{n,T}(\theta) = \ell_{n,T}(\theta) = &-\frac{n(T-1)}{2}\ln(2\pi\sigma^2) \\ &+ (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] \\ &- \frac{1}{2\sigma^2}\sum_{t=1}^{T-1} V^{*'}_{nt}(\zeta)V^*_{nt}(\zeta), \end{aligned}$$

where $V^*_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)Y^*_{nt} - X^*_{nt}\beta]$. As remarked in Lee and Yu (2010), the function $L_{n,T}$ has a conditional likelihood interpretation: it is the likelihood conditional on the time average $\sum_{t=1}^{T} Y_{nt}/T$, which is a sufficient statistic for c_{n0}, under normality.

We rewrite $\ell_{n,T}(\theta)$ in terms of a quadratic form in $\tilde{V}_{nt}(\zeta)$ as

$$\begin{aligned} \ell_{n,T}(\theta) = &-\frac{n(T-1)}{2}\ln(2\pi\sigma^2) \\ &+ (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] \\ &- \frac{1}{2\sigma^2}\sum_{t=1}^{T} \tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta), \end{aligned} \tag{4}$$

where $\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta]$, with

$$\tilde{Y}_{nt} = Y_{nt} - \sum_{t=1}^{T} Y_{nt}/T, \quad \tilde{X}_{nt} = X_{nt} - \sum_{t=1}^{T} X_{nt}/T. \tag{5}$$

The MLE $\hat{\theta}_{n,T}$ for θ is an M-estimator obtained by solving $\hat{\theta}_{n,T} = \arg\max_{\theta \in \Theta} \ell_{n,T}(\theta)$. It implies the system of estimating equations:

$$0 = \frac{\partial \ell_{n,T}(\hat{\theta}_{n,T})}{\partial \theta} = \sum_{t=1}^{T}(T-1)^{-1}\psi_{nt}(\hat{\theta}_{n,T}), \tag{6}$$

where ψ_{nt} is the likelihood score function

$$\psi_{nt}(\theta) = \begin{pmatrix} \frac{T-1}{\sigma^2}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\zeta) \\ \frac{T-1}{\sigma^2}\left(\left(\ddot{G}_n\ddot{X}_{nt}\beta\right)'\tilde{V}_{nt}(\zeta) + \tilde{V}'_{nt}\ddot{G}'_n\tilde{V}_{nt}\right) \\ -\frac{(T-1)^2}{T}\mathrm{tr}(G_n(\lambda)) \\ \frac{T-1}{\sigma^2}(H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) - \frac{(T-1)^2}{T}\mathrm{tr}(H_n(\rho)) \\ \frac{T-1}{2\sigma^4}\left(\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) - \frac{n(T-1)}{T}\sigma^2\right) \end{pmatrix}, \tag{7}$$

where $G_n(\lambda) = W_n S_n^{-1}$, $H_n(\rho) = M_n R_n^{-1}$, $\ddot{G}_n(\lambda) = R_n G_n R_n^{-1}$, and $\ddot{X}_{nt} = R_n\tilde{X}_{nt}$.

4. Methodology

We assume $n \gg T$, so we deal with the so-called *micro panels*. Within this setting for T being fixed, the standard asymptotic arguments rely crucially on the number n of individuals tending to infinity; see Lee and Yu (2010). In contrast, in our development, we consider small sample cross-sectional asymptotics (Field and Ronchetti 1990), and we still leave T fixed (possibly small). However, we will keep T in the notation of normalizing factors to demonstrate the improved rate of convergence that would result if $T \to \infty$ or it is large. The derivation of our higher-order techniques relies on three steps: (i) defining a second-order asymptotic (von Mises) expansion for the MLE, see Section 4.1; (ii) identifying the corresponding U-statistic, see Section 4.2; (iii) deriving the Edgeworth expansion for the U-statistic and deriving the saddlepoint density by means of the tilted-Edgeworth technique, see Sections 4.3 and 4.4. Similar approaches are available for the standard setting of iid random variables in Easton and Ronchetti (1986), Barndorff-Nielsen and Cox (1989), and Gatto and Ronchetti (1996).

4.1. The M-Functional Related to the MLE and Its First-Order Asymptotics

The likelihood score function in Equation (7) is a vector in $\mathbb{R}^d$, and each lth element of this vector, for $l = 1, ..., d$, is a sum of n terms. In what follows, for $i = 1, ..., n$, we denote by $\psi_{i,t,l}(\theta)$ the ith term, at time t, of this sum for the lth component of the score. To specify $\psi_{i,t,l}(\theta)$, we set $R_n(\rho) = \left(r'_1(\rho), r'_2(\rho), \cdots, r'_n(\rho)\right)'$,

$$\tilde{X}_{nt} = \left[\tilde{X}_{nt,1}, \tilde{X}_{nt,2}, \cdots, \tilde{X}_{nt,k}\right],$$

$\tilde{V}_{nt}(\zeta) = \left(\tilde{v}_{1t}(\zeta), \tilde{v}_{2t}(\zeta), \cdots, \tilde{v}_{nt}(\zeta)\right)'$ and $H_n(\rho) = (h'_1(\rho), h'_2(\rho), \cdots, h'_n(\rho))'$, where $r_i(\rho)$ and $h_i(\rho)$ are the i_{th} row of $R_n(\rho)$ and $H_n(\rho)$, g_{ii} and h_{ii} are i_{th} element of the diagonal of $G_n(\lambda)$ and $H_n(\rho)$, respectively. Then, from Equation (7), it follows:

$$\psi_{i,t}(\theta) = \begin{pmatrix} \psi_{i,t,1}(\theta), \\ \psi_{i,t,2}(\theta) \\ \vdots \\ \psi_{i,t,d}(\theta) \end{pmatrix}_{d \times 1}$$

$$
= \begin{pmatrix}
\frac{T-1}{\sigma^2} r_i(\rho)\tilde{X}_{nt,1}\tilde{v}_{it}(\zeta) \\
\frac{T-1}{\sigma^2} r_i(\rho)\tilde{X}_{nt,2}\tilde{v}_{it}(\zeta) \\
\vdots \\
\frac{T-1}{\sigma^2} r_i(\rho)\tilde{X}_{nt,k}\tilde{v}_{it}(\zeta) \\
\frac{T-1}{\sigma^2} r_i(\rho)\left(G_n\tilde{X}_{nt}\beta + G_nR_n^{-1}(\rho)\tilde{V}_{nt}(\zeta)\right)\tilde{v}_{it}(\zeta) \\
-\frac{(T-1)^2}{T} g_{ii} \\
\frac{T-1}{\sigma^2} h_i(\rho)\tilde{V}_{nt}(\zeta)\tilde{v}_{it}(\zeta) - \frac{(T-1)^2}{T} h_{ii} \\
\frac{T-1}{2\sigma^4}\left(\tilde{v}_{it}(\zeta)^2 - \frac{T-1}{T}\sigma^2\right)
\end{pmatrix}_{d\times 1}. \tag{8}
$$

Thus, for every $t = 1, 2, ..., T$, we have

$$
\psi_{nt}(\theta) = \left(\sum_{i=1}^{n} \psi_{i,t,1}(\theta), ..., \sum_{i=1}^{n} \psi_{i,t,d}(\theta)\right)',
$$

and, from Equation (6), it follows that the MLE is the solution to

$$
\frac{1}{n}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}(T-1)^{-1}\psi_{i,t,1}(\hat{\theta}_{n,T}), ..., \sum_{i=1}^{n}(T-1)^{-1}\psi_{i,t,d}(\hat{\theta}_{n,T})\right)' = 0. \tag{9}
$$

The M-functional ϑ related to the MLE is implicitly defined as the unique functional root of

$$
\mathbb{E}\left\{\sum_{t=1}^{T}(T-1)^{-1}\,\psi_{nt}\left[\vartheta(P_{\theta_0})\right]\right\} = 0, \tag{10}
$$

or equivalently via the asymptotic maximization $\theta_0 = \arg\max_{\theta\in\Theta}\mathbb{E}[\ell_{n,T}(\theta_0)]$; see, for example, Lee (2004). In what follows, we write $\theta_0 = \vartheta(P_{\theta_0})$ to emphasize the dependence of the functional on the measure P_{θ_0}. The finite sample version of the M-functional in Equation (10) is the M-estimator defined in Equation (9), or equivalently via the finite sample maximization $\hat{\theta}_{n,T} = \arg\max_{\theta\in\Theta}\ell_{n,T}(\theta)$. In what follows, we write $\hat{\theta}_{n,T} = \vartheta(P_{n,T})$, where $P_{n,T}$ is the measure associated to the n-dimensional sample. We can check the uniqueness of the M-estimator on a case-by-case basis, using Assumption A (see below) and working on the Gaussian log-likelihood. For instance, in the case of the SAR model, we can compute the second derivative of $\ell_{n,T}$ w.r.t. λ and check that $\ell_{n,T}$ is a concave function, admitting a unique maximizer. Alternatively, we can solve the estimating equations implied by first-order conditions related to $\ell_{n,T}$ resorting on a one-step procedure and using for instance the GMM estimator (see Lee and Yu 2010 and reference therein) as a preliminary estimator; for a book-length description of one-step procedure, see, for example, Van der Vaart (1998, chap. 5).

In what follows, we set $m := n(T-1)$, with $m \to \infty$, as $n \to \infty$. Then, we introduce

Assumption A.

(i) The elements $\omega_{n,ij}$ of W_n and the elements $m_{n,ij}$ of M_n in Equation (1) are at most of order $\tilde{h}_n^{-1}$, denoted by $O(1/\tilde{h}_n)$, uniformly in all i, j, where the rate sequence $\{\tilde{h}_n\}$ is bounded, and $\tilde{h}_n$ is bounded away from zero for all n. As a normalization, we have $\omega_{n,ii} = m_{n,ii} = 0$, for all i.

(ii) n diverges, while $T \geq 2$ and it is finite.

(iii) Assumptions 2–5 and Assumption 7 in Lee and Yu (2010) are satisfied.

(iv) Denote $C_n = \ddot{G}_n - n^{-1}tr(\ddot{G}_n)I_n$ and $D_n = H_n - n^{-1}tr(H_n)I_n$, where $\ddot{G}_n = R_nG_nR_n^{-1}$ and $H_n = M_nR_n^{-1}$. Then $C_n^s = C_n + C_n'$ and $D_n^s = D_n + D_n'$. The limit of $n^{-2}\left[tr(C_n^sC_n^s)tr(D_n^sD_n^s) - tr^2(C_n^sD_n^s)\right]$ is strictly positive as $n \to \infty$.

Assumptions A(i) characterizes the behavior of W_n and M_n in terms of n, and W_n and M_n are row-normalized. It means $\omega_{n,ij} = d_{ij}/\sum_{j=1}^{n} d_{ij}$, where d_{ij} is the spatial distance of the ith and the jth units in some (characteristic) space. For each i, the weight $\omega_{n,ij}$ defines an average of neighboring values. In what follows, we consider spatial weight matrices (like, e.g., Rook and Queen) such that $\sum_{j=1}^{n} d_{ij} = O(\tilde{h}_n)$ uniformly in i and the row-normalized weight matrix satisfies Assumption A(i); see, for example, Lee (2004). For instance, W_n as Rook creates a square tessellation with $\tilde{h}_n = 4$ for the inner fields on the chessboard, and $\tilde{h}_n = 2$ and $\tilde{h}_n = 3$ for the corner and border fields, respectively. Assumption A(ii) defines the asymptotic scheme of our theoretical development, in which we consider n cross-sectional units and we leave T fixed. Assumption A(iii) refers to Lee and Yu (2010), who develop the first-order asymptotic theory. All W_n, M_n, $S_n^{-1}(\lambda)$, $R_n^{-1}(\rho)$ are uniformly bounded by Assumption A(iv), which guarantees the convergence of the asymptotic variance, see below. Assumption A(iv) states the identification conditions of the model and the conditions for the nonsingularity of the limit of the information matrix. In particular, it implies that the $(d \times d)$-matrix

$$
M_{i,T}(\psi, P_{\theta_0}) = \mathbb{E}\left[-(T-1)^{-1}\sum_{t=1}^{T}\partial\psi_{i,t}(\theta)/\partial\theta\Big|_{\theta=\theta_0}\right] \tag{11}
$$

is non-singular. Under Assumption A(i)–(iv), Theorem 1 part(ii) in Lee and Yu (2010) showed that $\lim_{n\to\infty}\hat{\theta}_{n,T} = \theta_0$. Furthermore, Theorem 2 point (ii) in Lee and Yu (2010) implied, as $n \to \infty$, that $\hat{\theta}_{n,T}$ satisfies $\sqrt{m}(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{0,T}^{-1}\right)$, and $\Sigma_{0,T} = \text{plim}_{n\to\infty}\Sigma_{0,n,T}$. The operator plim stands for the limit in probability and the expression of $\Sigma_{0,n,T}$ is available in the online supplementary material (see Appendix B). The first-order asymptotics is obtained letting $n \to \infty$, while there is no need for $T \to \infty$.

4.2. Second-Order von-Mises Expansion

To derive a higher-order asymptotic expansion of the MLE, we introduce the following assumption.

Assumption B.

(i) $\partial^2\psi_{i,t,l}(\theta)/\partial\theta\partial\theta'$ exists at $\theta = \theta_0$, for every $i = 1, .., n$, $t = 1, ..., T$ and $l = 1, .., d$.

(ii) The $d\times d$-matrix $\mathbb{E}\left[(T-1)^{-1}\sum_{t=1}^{T}\partial^2\psi_{i,t,l}(\theta)/\partial\theta\partial\theta'\Big|_{\theta=\theta_0}\right]$ is positive semi-definite, for every $l = 1, .., d$.

Then, we state the following

Lemma 1. Let the MLE be defined as in Equation (6). Under Assumptions A and B, the following expansion holds:

$$\vartheta(P_{n,T}) - \vartheta(P_{\theta_0}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}_{i,T}(\psi, P_{\theta_0}) + \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \varphi_{i,j,T}(\psi, P_{\theta_0}) + O_P(m^{-3/2}), \tag{12}$$

where

$$\mathrm{IF}_{i,T}(\psi, P_{\theta_0}) = M_{i,T}^{-1}(\psi, P_{\theta_0})(T-1)^{-1}\sum_{t=1}^{T}\psi_{i,t}(\theta_0), \tag{13}$$

and

$$\begin{aligned}&\varphi_{i,j,T}(\psi, P_{\theta_0})\\ &= \mathrm{IF}_{i,T}(\psi, P_{\theta_0}) + \mathrm{IF}_{j,T}(\psi, P_{\theta_0}) + M_{i,T}^{-1}(\psi, P_{\theta_0})\Gamma_{i,j,T}(\psi, P_{\theta_0})\\ &\quad + M_{i,T}^{-1}(\psi, P_{\theta_0})\left\{(T-1)^{-1}\sum_{t=1}^{T}\frac{\partial \psi_{j,t}(\theta)}{\partial\theta}\Big|_{\theta_0}\mathrm{IF}_{i,T}(\psi, P_{\theta_0})\right.\\ &\quad \left. + (T-1)^{-1}\sum_{t=1}^{T}\frac{\partial\psi_{i,t}(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\mathrm{IF}_{j,T}(\psi, P_{\theta_0})\right\},\end{aligned} \tag{14}$$

where

$$\Gamma_{i,j,T}(\psi, P_{\theta_0})' = \begin{pmatrix} \mathrm{IF}'_{j,T}(\psi, P_{\theta_0}) & \mathbb{E}\left[\sum_{t=1}^{T}\frac{\partial^2\psi_{i,t,1}(\theta_0)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}\right] & \mathrm{IF}_{i,T}(\psi, P_{\theta_0})\\ & \vdots & \\ \mathrm{IF}'_{j,T}(\psi, P_{\theta_0}) & \mathbb{E}\left[\sum_{t=1}^{T}\frac{\partial^2\psi_{i,t,d}(\theta_0)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}\right] & \mathrm{IF}_{i,T}(\psi, P_{\theta_0})\end{pmatrix}, \tag{15}$$

and $M_{i,T}(\psi, P_{\theta_0})$ is defined by Equation (11).

In Equation (12), we interpret the quantities $\mathrm{IF}_{i,T}(\psi, P_{\theta_0})$, the first-order von Mises kernel, and $\varphi_{i,j,T}(\psi, P_{\theta_0})$, the second-order von Mises kernel, as functional derivatives of the M-functional related to the MLE; see Fernholz (2001). Specifically, the first term, of order $m^{-1} \propto n^{-1}$, is the influence function (IF) and represents the standard tool applied to derive the first-order (Gaussian) asymptotic theory of the MLE; see, for example, Van der Vaart (1998) and Baltagi (2008). The second term in Equation (12), of order $m^{-2} \propto n^{-2}$, plays a pivotal role in our derivation of higher-order approximations.

4.3. Approximation Via U-Statistic

The result of Lemma 1 together with the chain rule define the second-order asymptotic expansion for a real-valued function of the MLE, such as a component of $\vartheta(P_{n,T})$ or a linear contrast. In Lemma 2, we show that we can write the asymptotic expansion in terms of a U-statistic of order two. To this end, we introduce the following assumption.

Assumption C.
Let q be a function from $\mathbb{R}^d$ to $\mathbb{R}$, which has continuous and nonzero gradient at $\theta = \theta_0$ and continuous second derivative at $\theta = \theta_0$.

Then, we have

Lemma 2. Under Assumptions A–C, the following expansion holds:

$$q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})] = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} h_{i,j,T}(\psi, P_{\theta_0}) + O_P(m^{-3/2}),$$

where

$$\begin{aligned}&h_{i,j,T}(\psi, P_{\theta_0})\\ &= g_{i,T}(\psi, P_{\theta_0}) + g_{j,T}(\psi, P_{\theta_0}) + \gamma_{i,j,T}(\psi, P_{\theta_0})\\ &= \frac{1}{2}\left\{\mathrm{IF}'_{i,T}(\psi, P_{\theta_0}) + \mathrm{IF}'_{j,T}(\psi, P_{\theta_0}) + \varphi'_{i,j,T}(\psi, P_{\theta_0})\right\}\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0}\\ &\quad + \frac{1}{2}\mathrm{IF}'_{i,T}(\psi, P_{\theta_0})\frac{\partial^2 q(\vartheta)}{\partial\vartheta\partial\vartheta'}\Big|_{\theta=\theta_0}\mathrm{IF}_{j,T}(\psi, P_{\theta_0}),\end{aligned} \tag{16}$$

with

$$g_{i,T}(\psi, P_{\theta_0}) = \frac{1}{2}\left(\mathrm{IF}'_{i,T}(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0}\right), \tag{17}$$

$$\begin{aligned}\gamma_{i,j,T}(\psi, P_{\theta_0}) = \frac{1}{2}\Bigg(&\varphi'_{i,j,T}(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0}\\ &+ \mathrm{IF}'_{i,T}(\psi, P_{\theta_0})\frac{\partial^2 q(\vartheta)}{\partial\vartheta\partial\vartheta'}\Big|_{\theta=\theta_0}\mathrm{IF}_{j,T}(\psi, P_{\theta_0})\Bigg).\end{aligned} \tag{18}$$

The function q may select, for example, a single component of the vector θ_0. In many empirical applications, the most interesting parameter is the spatial correlation coefficient λ_0, and the null hypothesis is zero correlation versus the alternative hypothesis of positive spatial correlation—the aim being to check whether there is a contagion effect.

4.4. Higher-Order Asymptotics

Making use of Lemmas 1 and 2, we derive the Edgeworth and the saddlepoint approximation to the distribution of a real-valued function q of the MLE.

Let $f_{n,T}(z)$ be the true density of $q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]$ at the point $z \in \mathcal{A}$, where $\mathcal{A}$ is a compact subset of $\mathbb{R}^d$. Our derivation of the saddlepoint density approximation to $f_{n,T}(z)$ is based on the tilted-Edgeworth expansion for U-statistics of order two. With this regard, a remark is in order. From Equation (8), we see that the terms in the random vector $\psi_{nt}(\theta_0)$ depend on the rows of the weight matrix $W_n(\rho)$ and $M_n(\lambda)$. As a consequence, these terms are independent but not identically distributed random variables, and we need to derive the Edgeworth expansion for our U-statistic taking into account this aspect. To this end, we approximate the cumulant-generating function (c.g.f.) of our U-statistic by *summing (in i and j) the (approximate) c.g.f.* of each $h_{i,j,T}$ kernel. This is an extension of the derivation by Bickel, Götze, and Van Zwet (1986) for iid random variables. To elaborate further, we introduce

Assumption D.
Suppose that there exist positive numbers δ, δ_1, C and positive and continuous functions χ_j: $(0,\infty) \to (0,\infty)$, $j = 1, 2$, satisfying $\lim_{z\to\infty}\chi_1(z) = 0$, $\lim_{z\to\infty}\chi_2(z) \geq \delta_1 > 0$, and a real number α such that $\alpha \geq 2 + \delta > 2$,

(i) $\mathbb{E}\left[|\gamma_{i,j,T}(\psi, P_{\theta_0})|^{\alpha}\right] < C$ for any i and j, $1 \le i < j \le n$,
(ii) $\mathbb{E}\left[g_{i,T}(\psi, P_{\theta_0})^4 1_{[z,\infty)}(|g_{i,T}(\psi, P_{\theta_0})|)\right] < \chi_1(z)$ for all $z > 0$ and any i, $1 \le i \le n$,
(iii) $\left|\mathbb{E}\left[e^{\iota\nu g_{i,T}(\psi, P_{\theta_0})}\right]\right| \le 1 - \chi_2(z) < 1$ for all $z > 0$ and any i, $1 \le i \le n$ and $\iota^2 = -1$,
(iv) $||M_{i,T}(\psi, P_{\theta_0}) - M_{j,T}(\psi, P_{\theta_0})|| = O(n^{-1})$ uniformly in λ and ρ.

A few comments are in order. Assumptions D(i)–(iii) are similar to the technical assumptions in (Bickel, Götze, and Van Zwet 1986, pp. 1465 and 1477). However, there are some differences between our assumptions and theirs. Indeed, to take into account the nonidentical distribution of $\psi_{i,t}$ and $\psi_{j,t}$, for $i \neq j$, we consider the first- and second-order von Mises kernels for each i (as in D(i)–(iii)). It is different from Bickel, Götze, and Van Zwet (1986): compare, for example, our D(ii) to their eq. (1.17). D(iv) is not considered in Bickel, Götze, and Van Zwet (1986): it is a peculiar assumption needed for our higher-order asymptotics (the technical aspects are available in Lemma A.1 and its proof in Appendix A). In Appendix D.2, we illustrate that, in the case of the SAR(1) model, the validity of D(iv) is related to more primitive expressions involving the entries of (some powers of) W_n. For other models, one should derive such primitive expressions on a case-by-case basis. For the sake of generality, here we provide an intuition on D(iv). Let us consider two different locations i and j. From Equation (11), we see that D(iv) imposes a structure on the information available at different locations. Indeed, $M_{i,T}(\psi, P_{\theta_0})$ and $M_{j,T}(\psi, P_{\theta_0})$ contribute to the asymptotic variance of the MLE. Since $M_{i,T}(\psi, P_{\theta_0})$ is related to the information available at the ith location, D(iv) essentially assumes that there exists an informative content which is common to location i and j, whilst the (Frobenius norm of the) information content specific to each location is of order $O(n^{-1})$.

Proposition 3. Under Assumptions A–D, the Edgeworth expansion $\Lambda_m(z)$ for the c.d.f. F_m of $\sigma_{n,T}^{-1}\{q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]\}$ is

$$\Lambda_m(z) = \Phi(z) - \phi(z)\left\{n^{-1/2}\frac{\kappa_{n,T}^{(3)}}{3!}(z^2-1) + n^{-1}\frac{\kappa_{n,T}^{(4)}}{4!}(z^3-3z) + n^{-1}\frac{\kappa_{n,T}^{(3)}}{72}(z^5-10z^2+15z)\right\}, \tag{19}$$

where $z \in \mathcal{A}$, $\sigma_{n,T}$ is the standard deviation of $q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]$, $\Phi(z)$ and $\phi(z)$ are the c.d.f. and p.d.f of a standard normal r.v. respectively, $\kappa_{n,T}^{(3)} n^{-1/2}$ and $\kappa_{n,T}^{(4)} n^{-1}$ are the approximate third and fourth cumulants of $\sigma_{n,T}^{-1}\{q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]\}$, as defined in Equations (A.15) and (A.18), respectively. Then $\sup_z |F_m(z) - \Lambda_m(z)| = o(m^{-1})$.

We can get the saddlepoint density approximation by exponentially tilting the Edgeworth expansion.

Proposition 4. Under Assumptions A–D, the saddlepoint density approximation to the density of $q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]$ at the point $z \in \mathcal{A}$ is

$$p_{n,T}(z) = \left[\frac{n}{2\pi\tilde{\mathcal{K}}''_{n,T}(\nu)}\right]^{1/2} \exp\left\{n\left[\tilde{\mathcal{K}}_{n,T}(\nu) - \nu z\right]\right\}, \tag{20}$$

with relative error of order $O(m^{-1})$, $\nu := \nu(z)$ is the saddlepoint defined by

$$\tilde{\mathcal{K}}'_{n,T}(\nu) = z, \tag{21}$$

the function $\tilde{\mathcal{K}}_{n,T}$ is the approximate c.g.f. of $\sqrt{n}(q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})])$, as defined in equation (A.42), while $\tilde{\mathcal{K}}'_{n,T}$ and $\tilde{\mathcal{K}}''_{n,T}$ represent the first and second derivative of $\tilde{\mathcal{K}}_{n,T}$, respectively. Moreover,

$$P\left\{q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})] > z\right\} = \left[1 - \Phi(r) + \phi(r)\left(\frac{1}{c} - \frac{1}{r}\right)\right]\left[1 + O(m^{-1})\right], \tag{22}$$

$$c = \nu\left[\tilde{\mathcal{K}}''_{n,T}(\nu)\right]^{1/2} \text{ and } r = \operatorname{sgn}(\nu)\left\{2n\left[\nu z - \tilde{\mathcal{K}}_{n,T}(\nu)\right]\right\}^{1/2}.$$

The proofs of these propositions are available in Appendix A. Following Durbin (1980), we can further normalize $p_{n,T}$ by dividing the right-hand side of Equation (20) by its integral with respect to z. This normalization typically improves even further the accuracy of the approximation. In Appendix C.1, we provide an algorithm (see Algorithm 1) in which we itemize the main computational steps needed to implement the saddlepoint tail area approximation, for a given transformation q, and for a given reference parameter θ_0, for example, the parameter characterizing the null hypothesis in a simple hypothesis testing, where the tail area probability yields an approximate p-value.

4.5. Links With the Econometric Literature

The expansions in Propositions 3 and Proposition 4 are connected with the results on higher-order expansions available in the spatial econometric literature, as cited in Section 1. However, some key differences are worth a mention.

(i) The Edgeworth expansion in Robinson and Rossi (2015) is for the concentrated MLE of λ and it is based on a higher-order Taylor expansion of the concentrated likelihood score; see also Robinson and Rossi (2014a,b, 2015), Hillier and Martellosio (2018), and Martellosio and Hillier (2020). In contrast, our method is based on a von Mises expansion of the MLE functional of the whole model parameter and we resort on a marginalization procedure to obtain the saddlepoint density approximation of the parameter(s) of interest. Therefore, Lemmas 1 and 2 give generality and flexibility to our approach: not only we may focus on λ, but also, for example, on ρ (which contains information on the spatial dependence of the innovation terms) and/or on β (which convey information on the significance of the time-varying covariates).

(ii) Our saddlepoint approximation is more general than the Edgeworth-based approximations available in the econometric literature, since we work with a larger class of models, which includes the model in Robinson and Rossi (2015) as a special case.

(iii) Although the inference (e.g., testing) derived using the Edgeworth expansion improves on the standard first-order asymptotics, it is well known (see, e.g., Field and Ronchetti 1990) that, in general, this technique provides a good approximation in the center of the distribution, but can be inaccurate in the tails, where the Edgeworth expansion can even become negative. It can lead to inaccurate approximations. Our saddlepoint approximation is a density-like object and is always nonnegative.

(iv) Our saddlepoint approximation yields a tail-area approximation via a Lugannani-Rice type formula. A similar result is not available for the Edgeworth expansion of the concentrated MLE derived in Robinson and Rossi (2015). Recently, Martellosio and Hillier (2020) studied the adjusted profile likelihood estimation method and obtained a result similar to our tail-area approximation. Their formula is derived for the spatial autoregressive model with covariates. However, they do not prove the higher-order properties of their approximation. In Proposition 3, we prove that our saddlepoint density approximation features *relative error* of order $O(1/(n(T-1)))$. This has to be contrasted with the extant Edgeworth expansion, which entails an absolute error of lower order—more precisely, the error order is $o((nT)^{-1/2})$, when the entries of the spatial matrix are $O(1)$; see Eq. (2.15) in Robinson and Rossi (2015). Achieving a small relative error is appealing in tail areas where the probabilities are small.

(v) In the comparison with the bootstrap, our methodology does not need resampling. Moreover, it does neither require bias correction, nor any studentization.

5. Comparisons With Other Approximations and Testing in the Presence of Nuisance Parameters

We compare the performance of our saddlepoint approximations to other routinely applied asymptotic techniques. To start with, we consider the SAR(1) model, where λ is the only unknown parameter. Then, we move to the SARAR(1,1) model, where we illustrate how to take care of nuisance parameters. We use the same setting as in Section 2; we refer to the online supplementary material (Appendix D) for details and for additional results.

5.1. Comparisons With Other Asymptotic Techniques

Saddlepoint vs first-order asymptotics. For the SAR(1) model, we analyse the behavior of the MLE of λ_0, whose PP-plots are available in Figure 3. For each type of W_n, for $n = 24$ and $n = 100$, the plots show that the saddlepoint approximation is closer to the "exact" probability than the first-order asymptotics approximation. For W_n Rook, the saddlepoint approximation improves on the routinely applied first-order asymptotics. In Figure 3, the accuracy gains are evident also for W_n Queen and Queen with torus, where the first-order asymptotic theory displays large errors essentially over the whole support (specially in the tails). On the contrary, the saddlepoint approximation is close to the 45-degree line.

Saddlepoint vs Edgeworth expansion (testing simple hypotheses). The Edgeworth expansion derived in Proposition 3 represents the natural alternative to the saddlepoint approximation since it is fully analytic. To gain insights into the different behavior of the saddlepoint and Edgeworth approximations, we investigate the size of a hypothesis test based on the approximations. We set $n = 24$ and we assume that σ^2 is known and equal to one. We consider the simple null hypothesis H_0: $\lambda_0 = 0$ for a one-sided test of zero against positive values of spatial correlation. We use 25,000 replications of $\hat{\lambda}_{n,T}$ to get the empirical estimate $\hat{F}_0$ of the c.d.f. F_0 of the estimator under the null hypothesis. We use the generic notation G for the c.d.f. of one of the Edgeworth, or saddlepoint approximations, under the null hypothesis. For the sake of completeness, we also display the results for the Gaussian (first-order) approximation. The empirical rejection probabilities $\hat{\alpha} = 1 - \hat{F}_0(G^{-1}(1-\alpha))$ are shown in Figure 4 for nominal size α ranging from 1% to 10%, and correspond to an estimated size. We have overrejection when we are above the 45-degree line. We observe strong size distortions for the asymptotic and Edgeworth approximations as expected from the previous results. The saddlepoint approximation exhibits only mild size distortions. For example, we get an estimated size $\hat{\alpha}$ of 11.72%, 7.36%, 5.70%, for the Normal, Edgeworth, and saddlepoint approximations, for a nominal size of 5%.

Saddlepoint vs. parametric bootstrap. The parametric bootstrap represents a (computer-based) competitor. To compare our saddlepoint approximation to the one obtained by

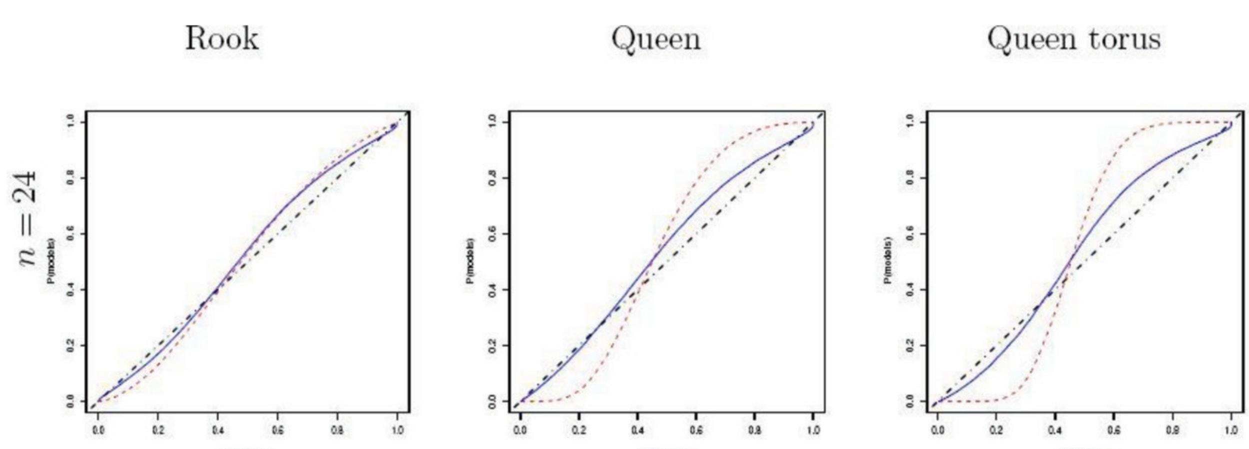

Figure 3. SAR(1) model: PP-plots for saddlepoint (continuous line) vs asymptotic normal (dotted line) probability approximation, for the MLE $\hat{\lambda}$, for $n = 24$ and $n = 100$, $\lambda_0 = 0.2$, and different W_n.

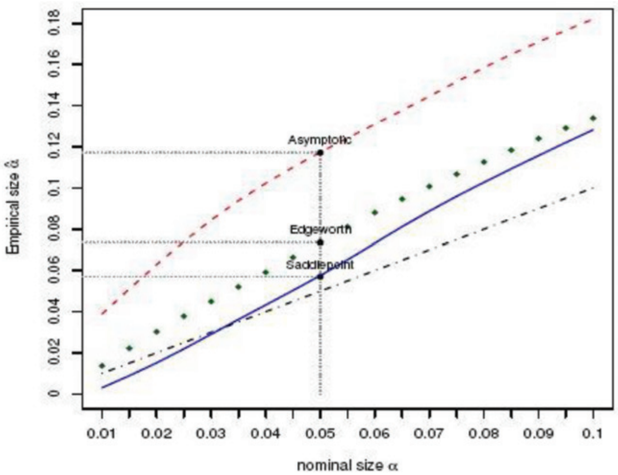

Figure 4. SAR(1) model: Estimated $\hat{\alpha}$ versus nominal size α between 1% and 10% under saddlepoint (continuous line), Edgeworth (dotted line with diamonds) and first-order asymptotic approximation (dotted line). W_n is Rook, $n = 24$ and $\lambda_0 = 0.0$.

bootstrap, we consider different numbers of bootstrap repetitions, labeled as B: we use $B = 499$ and $B = 999$. For space constraints, in Figure 5, we display the results for $B = 499$ (similar plots are available for $B = 999$) showing the functional boxplots (as obtained iterating the procedure 100 times) of the bootstrap approximated density, for sample size $n = 24$ and for W_n is Queen.

To visualize the variability entailed by the bootstrap, we display the first and third quartile curves and the median functional curve; for details about functional boxplots, we refer to Sun and Genton (2011) and to R routine `fbplot`. We notice that, while the bootstrap median functional curve (representing a typical bootstrap density approximation) is close to the actual density (as represented by the histogram), the range between the quartile curves illustrates that the bootstrap approximation has a variability. Clearly, the variability depends on B: the larger is B, the smaller is the variability. However, larger values of B entail bigger computational costs: when $B = 499$, the bootstrap is almost as fast as the saddlepoint density approximation (computation time about 7 minutes, on a 2.3 GHz Intel Core i5 processor), but for $B = 999$, it is three times slower.

5.2. Testing in the Presence of Nuisance Parameters

5.2.1. Saddlepoint Test for Composite Hypotheses

An interesting case (suggested by the Associate Editor and an anonymous referee) that has a strong practical relevance is related to testing a composite null hypothesis. It is a problem which is different from the one considered so far in the article, because it raises the issue of dealing with nuisance parameters.

To tackle this problem, several possibilities are available. For instance, we may fix the nuisance parameters at the MLE estimates. Combined with the saddlepoint density in Equation (20), this yields a ready solution to the nuisance parameter problem. In our numerical experience (see Appendix D.6 for an experiment about the SAR(1)), the use of the maximum likelihood estimates may preserve reasonable accuracy in some cases. Alternatively, one may consider to use profile estimators, as suggested, for example, in Hillier and Martellosio (2018) (see their numerical exercises on p.416) and Martellosio and Hillier (2020). The main theoretical drawback related to the use of both MLE or profile estimates for the nuisance parameter(s) is that the results currently available in the literature do not guarantee that the second-order properties still hold. To cope with this issue, we propose to build on Robinson, Ronchetti, and Young (2003), who derived a saddlepoint test statistic which takes into account explicitly the nuisance parameters, while preserving relative error in normal region. This test statistic is derived going through steps which are similar to ours. Robinson, Ronchetti, and Young (2003) define the statistic in the iid setting, while Lô and Ronchetti (2009) and Czellar and Ronchetti (2010) extend it to non-iid data.

Let us consider a SARAR model whose parameter is $\theta = (\theta_{10}, \theta_2)'$, where θ_{10} is specified by the null composite hypothesis: typically, the null concerns λ only, while θ_2 contains all the nuisance parameters. More specifically, the parameter is $\theta = (\lambda, \beta, \rho, \sigma^2)'$ and the general function $q(\theta)$ used in the previous sections is simply $q(\theta) = \lambda$. Thus, we have the composite hypothesis

$$\mathcal{H}_0 : \lambda = \lambda_0 = 0 \quad \text{vs} \quad \mathcal{H}_1 : \lambda > 0, \tag{23}$$

where $\theta = (\lambda, \theta_2)'$, with $\theta_{10} = \lambda_0$ and $\theta_2 = (\beta, \rho, \sigma^2)'$. Then, we define the test statistic

$$SAD_n(\hat{\lambda}) = 2n \inf_{\theta_2} \sup_{\nu} -\mathcal{K}_{\psi}(\nu, \hat{\lambda}, \theta_2). \tag{24}$$

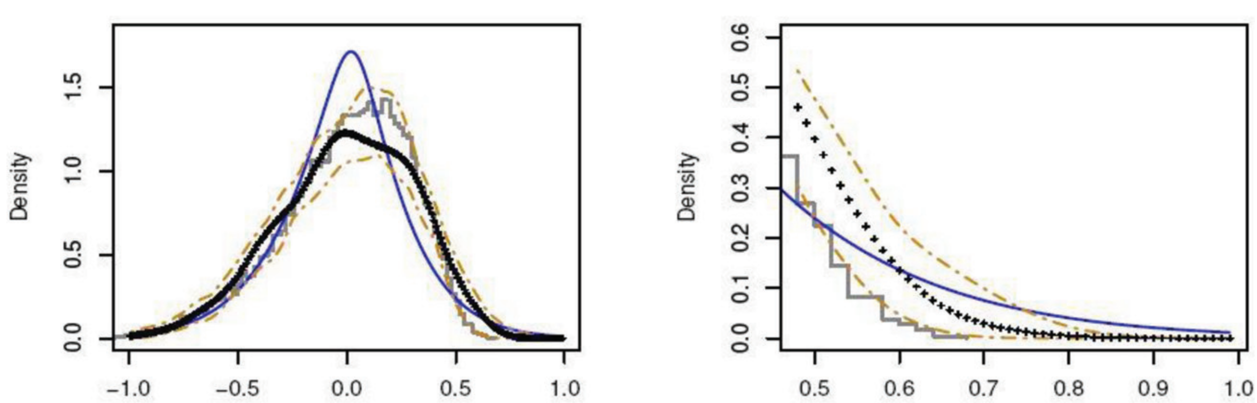

Figure 5. SAR(1) model. Left panel: Density plots for saddlepoint (continuous line) vs the functional boxplot of the parametric bootstrap probability approximation to the exact density (as expressed by the histogram and obtained using MC with size 25000), for the MLE $\hat{\lambda}$ and W_n is Queen. Sample size is $n = 24$, while $\lambda_0 = 0.2$. Right panel: zoom on the right tail. In each plot, we display the functional central curve (dotted line with crosses), the 1_{st} and 3_{rd} functional quartile (two-dash lines).

The function $\mathcal{K}_\psi(\nu, (\lambda, \theta_2))$ is the c.g.f. of the estimating function

$$\mathcal{K}_\psi(\nu, \hat{\lambda}, \theta_2) = n^{-1} \sum_{i=1}^{n} \ln E_{P_{(\lambda_0,\theta_2)}} \exp(\nu^T \psi_i^{(T)}(\hat{\lambda}, \theta_2)), \quad (25)$$

where $\psi_i^{(T)}(\lambda, \theta_2) := \sum_{t=1}^{T}(T-1)^{-1}\psi_{i,t}(\lambda, \theta_2)$ and $\psi_{i,t}$ is as in (4.1). The c.g.f. $\mathcal{K}_\psi$ has a role analogous to the one of the c.g.f. of the U-statistic, that we derived in Section 4. We highlight that the expected value in Equation (25) is taken w.r.t. the probability $P_{(\lambda_0,\theta_2)}$, where λ_0 is specified by the null, while the nuisance parameters are not fixed: the infimum over θ_2 takes care of the nuisance parameters. In our inference procedure, we have that $\hat{\theta}_{n,T} = (\hat{\lambda}, \hat{\theta}_2)'$ is the solution to $\sum_{i=1}^{n} \psi_i^{(T)}(\lambda, \theta_2) = 0$. Under the null hypothesis, the test statistic $SAD_n(\hat{\lambda})$ is asymptotically χ_1^2 distributed with a relative error of order $O(m^{-1})$ in the normal region.

Appendix C.2 outlines an algorithm (Algorithm 2), which itemizes the main computational steps.

5.2.2. Numerical Results

Let us work with a SARAR(1,1) model, having no covariates and known variance $\sigma^2 = 1$ and $n = 24$. It implies that $\theta = (\lambda, \rho)'$ and we consider the problem in Equation (23), with ρ being the nuisance parameter. We set three different values $\rho = 0.25, 0.5, 0.75$ to analyze numerically the impact that the spatial dependence in the innovation term has on SAD_n. We study the behavior of the Wald test, as obtained using the first-order asymptotic theory and making use of the expression of the asymptotic variance as available in Appendix B. We compare the Wald test to SAD_n–to implement (24) we make use of the R routine `nlm`. We consider two types of spatial matrix W_n, the Rook and the Queen, and we set $W_n \equiv M_n$. Both test statistics are asymptotically χ_1^2 distributed under the null hypothesis. To compare them in small samples, we first obtain the 95th and 97.5th quantile of each test statistic; then we compute the corresponding probability as obtained using the χ_1^2. We display the results in Table 1. We see that the Wald test has severe size distortion. For instance, for $\rho = 0.25$, we observe a relative error of about 30%, for the quantile of 95%, when W_n is Rook, while the saddlepoint test entails a relative error of about 1.8%. Looking at the performance of SAD_n, we see that it is uniformly more accurate than the Wald test: considering all cases, we observe a maximal relative error of about 2%, for the quantile of 95%, when $\rho = 0.75$ and W_n is Queen; in the same setting, the Wald test entails a relative error of about 24%. Moreover, the size is fairly constant for the different values of ρ: it illustrates that the test statistic takes care correctly of the nuisance parameter.

Table 1. Wald and SAD_n test for the problem (23)–spatial dependence in the presence of a nuisance parameter, in a SARAR(1,1) model–with no covariates and known variance.

	$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$	
	95.00%	97.50%	95.00%	97.50%	95.00%	97.50%
Rook						
Wald	66.08%	89.89%	98.33%	99.41%	99.99%	99.99%
SAD_n	96.71%	97.18%	96.66%	97.18%	95.55%	96.04%
Queen						
Wald	72.71%	80.20%	90.48%	98.00%	98.36%	99.22%
SAD_n	94.79%	96.94%	95.90%	98.20%	96.94%	97.49%

NOTE: The quantiles are obtained using 100 repetitions of each test statistic.

6. Empirical Application

Feldstein and Horioka (1980) documented empirically that domestic saving rate in a country has a positive correlation with the domestic investment rate. It contrasts with the understanding that, if capital is perfectly mobile between countries, most of any incremental saving is invested to get the highest return regardless of any locations, and that such correlation should actually vanish. Debarsy and Ertur (2010) suggested to use spatial modeling since several articles challenge these findings but under the strong assumption that investment rates are independent across countries. Such an assumption might influence the conclusions of applied spatial economics.

In this empirical exercise, we investigate the presence of spatial autocorrelation in the investment-saving relationship. We consider investment and saving rates for 24 OECD countries between 1960 and 2000 (41 years). Because of macroeconomic reasons (deregulating financial markets), we divide the whole period into shorter sub-periods: 1960–1970, 1971–1985 and 1986–2000, as advocated by Debarsy and Ertur (2010). Since the cross-sectional size is only $n = 24$, the asymptotics may suffer from size distortion as documented in Section 5. Therefore, we resort on a saddlepoint test to investigate whether or not there are inferential issues (coming from finite sample distortions and nuisance parameters) in the use of the first-order asymptotic theory. In line with the econometric literature, we specify the following SARAR(1,1) model for the three sub-periods:

$$\begin{aligned} \text{Inv}_{nt} &= \lambda_0 W_n \,\text{Inv}_{nt} + \beta_0 \text{Sav}_{nt} + c_{n0} + E_{nt}, \\ E_{nt} &= \rho_0 M_n E_{nt} + V_{nt}, \qquad t = 1, 2, \ldots, T \end{aligned} \quad (26)$$

where Inv_{nt} is the $n \times 1$ vector of investment rates for all countries and Sav_{nt} is the $n \times 1$ vector of saving rates. Each element v_{it} in V_{nt} is iid across i and t, having Gaussian distribution with zero mean and variance σ_0^2. c_{n0} is the vector of fixed effects.

We assume $W_n = M_n$ and adopt two different weight matrices as in Debarsy and Ertur (2010). The first one is based on the inverse distance. Each element ω_{ij} in W_n is d_{ij}^{-1}, where d_{ij} is the arc distance between capitals of countries i and j. The second is the binary seven nearest neighbors (7NN) weight matrix. More precisely, ω_{ij}=1, if $d_{ij} \leq d_i$ and $i \neq j$. Otherwise, $\omega_{ij} = 0$, where d_i is the 7th-order smallest arc-distance between countries i and j such that each country i has exactly 7 neighbors. Both weight matrices are row-normalized.

We estimate the parameters using the MLE described in Section 3. Table 2 gathers the point estimates (and their standard errors) that agree with the magnitudes found by Debarsy and Ertur (2010). To investigate the validity of the model (26), we test for spatial dependence, working on $\lambda = 0$ and/or $\rho = 0$. Specifically, our aim is to detect if and in which period(s) the inference yielded by the first-order asymptotic theory differs from the inference obtained using our saddlepoint test. With this goal, in Table 3 we provide the p-values for testing (at the 5% level) three different composite hypotheses: in the first row, we consider the problem of testing for $\lambda = 0$; in the second row, we test for $\rho = 0$; in the third row, we test for $\lambda = \rho = 0$. To

Table 2. SARAR(1,1) model: Maximum likelihood estimates of parameters β, λ, ρ.

	Weight matrix: inverse distance			Weight matrix: 7 nearest neighbors		
	1960–1970	1971–1985	1986–2000	1960–1970	1971–1985	1986–2000
β	0.935(0.05)	0.638(0.04)	0.356(0.07)	0.932(0.05)	0.633(0.04)	0.368(0.07)
λ	0.004(0.10)	0.381(0.11)	0.430(0.30)	−0.016(0.09)	0.340(0.10)	0.437(0.18)
ρ	−0.305(0.22)	0.334(0.16)	0.222(0.40)	−0.219(0.19)	0.258(0.15)	0.025(0.28)

NOTE: Standard errors are between brackets.

Table 3. SARAR(1,1) model: p-values of Saddlepoint (SAD_n) and Wald (ASY) tests for several composite hypotheses.

		Weight matrix: inverse distance			Weight matrix: 7 nearest neighbors		
		1960–1970	1971–1985	1986–2000	1960–1970	1971–1985	1986–2000
$\lambda = 0$	SAD_n	1.0000	0.0096	**0.0000**	0.9998	**0.2248**	0.0000
	ASY	1.0000	0.0116	<u>0.5679</u>	0.9987	<u>0.0130</u>	0.1123
$\rho = 0$	SAD_n	0.1134	0.0024	0.1217	0.3232	**0.0403**	0.9993
	ASY	0.5890	0.2261	0.9578	0.7101	<u>0.3898</u>	0.9998
$\lambda = \rho = 0$	SAD_n	0.1414	0.0000	0.0000	0.2603	0.0000	0.0000
	ASY	0.4615	0.0000	0.0000	0.5042	0.0000	0.0000

perform the tests, we consider the routinely applied Wald test (as obtained using the first-order asymptotic approximation, ASY) and the saddlepoint test (SAD_n). In each testing procedure, we treat the parameters not specified by the null hypothesis as nuisance parameters. In the SAD_n test, we take care of the nuisance as indicated in Equation (24), while in the ASY test we simply plug-in the MLE estimates for the nuisance parameters—as it is customary in the econometric software based on the first-order asymptotic theory.

In the periods 60–70, both ASY and SAD_n yield the same inference, for both the considered types of weight matrix, with conventional significance levels. The other sub-periods display some discrepancies between the inference obtained via ASY and via SAD_n. We do not want to discuss all discrepancies but only briefly comment on some key differences—we highlights the corresponding values in Table 3. In the sub-periods 71–85 under 7NN W_n, the saddlepoint test finds no evidence against no spatial dependence in the investing rates across countries, and vice-versa for the asymptotic approximation. Moreover, the ASY test does not find evidence against $\rho = 0$, while the SAD_n test rejects this composite hypothesis. Thus, the SAD_n test indicates a spillover through the contemporary shocks between countries. This spillover goes through the innovations, that is, through the unexpected part of the model dynamics, a finding not documentable when one relies on the first-order asymptotic theory. This results suggests that a test statistic designed to perform well in small samples and in the presence of nuisance parameters is able to document spatial dependence in the disturbances E_{nt}. Some differences are detectable also in the sub-periods 86–00, under the inverse distance matrix.

Supplementary Material

The online supplementary material includes proofs, lengthy analytical derivations and additional numerical results for the SAR(1) model. All the codes and data are available in our Github repository *https://github.com/ChaonanJiang/Sadd_Panel*.

Acknowledgments

We thank the editor, the associate editor and the two referees for constructive criticism and numerous suggestions which have led to substantial improvements over the previous versions. We thank Luc Anselin and participants at a seminar at the University of Chicago for contributing stimulating suggestions, and Nicolas Debarsy for sharing the empirical dataset with us. We are also grateful to participants at the 2017 conference in honour of Luc Bauwens, EEA-ESEM 2018 conference, and seminars at Lille, Orleans, Lyon, Turin (Collegio Carlo Alberto), and McGill University for comments that greatly improved the article.

Funding

We acknowledge the financial support of the Swiss National Science Foundation, grant 100018_169559.

ORCID

Olivier Scaillet http://orcid.org/0000-0002-3948-6891

References

Anselin, L. (1988), *Spatial Econometrics: Methods and Models* (Vol. 4), New York: Springer. [1164]
Baltagi, B. (2008), *Econometric Analysis of Panel Data* (Vol. 1), New York: Wiley. [1164,1169]
Bao, Y. (2013), "Finite-Sample Bias of the QMLE in Spatial Autoregressive Models," *Econometric Theory*, 29, 68–89. [1164]
Bao, Y., and Ullah, A. (2007), "Finite Sample Properties of Maximum Likelihood Estimator in Spatial Models," *Journal of Econometrics*, 137, 396–413. [1164]
Barndorff-Nielsen, O., and Cox, D. (1989), *Asymptotic Techniques for Use in Statistics*, London: Chapman and Hall. [1167]
Bickel, P., Götze, F., and Van Zwet, W. (1986), "The Edgeworth Expansion for U-statistics of Degree Two," *The Annals of Statistics*, 14, 1463–1484. [1169,1170]
Bivand, R. S., Pebesma, E. J., and Gómez-Rubio, V. (2008), *Applied Spatial Data Analysis With R*, New York: Springer. [1165]
Brazzale, A., Davison, A. C., and Reid, N. (2007), *Applied Asymptotics: Case Studies in Small-Sample Statistics* (Vol. 23), Cambridge: Cambridge University Press. [1165]
Cressie, N. (2015), *Statistics for Spatial Data*, New York: Wiley. [1164]
Cressie, N., and Wikle, C. (2015), *Statistics for Spatio-Temporal Data*, New York: Wiley. [1164]
Czellar, V., and Ronchetti, E. (2010), "Accurate and Robust Tests for Indirect Inference," *Biometrika*, 97, 621–630. [1172]
Daniels, H. E. (1954), "Saddlepoint Approximations in Statistics," *Annals of Mathematical Statistics*, 25, 631–650. [1165]

Debarsy, N., and Ertur, C. (2010), "Testing for Spatial Autocorrelation in a Fixed Effects Panel Data Model," *Regional Science and Urban Economics*, 40, 453–470. [1173]

Durbin, J. (1980), "Approximations for Densities of Sufficient Estimators," *Biometrika*, 67, 311–333. [1170]

Easton, G. S., and Ronchetti, E. (1986), "General Saddlepoint Approximations with Applications to L-Statistics," *Journal of the American Statistical Association*, 81, 420–430. [1167]

Feldstein, M., and Horioka, C. (1980), "Domestic Saving and International Capital Flows," *The Economic Journal*, 90, 314–329. [1173]

Fernholz, L. T. (2001), "On Multivariate Higher-order Von Mises Expansions," *Metrika*, 53, 123–140. [1169]

Field, C. A., and Ronchetti, E. (1990), *Small Sample Asymptotics* (Vol. 13), IMS, Lecture Notes-Monograph Series. [1165,1167,1171]

Gaetan, C., and Guyon, X. (2010), *Spatial Statistics and Modeling* (Vol. 90), New York: Springer. [1164]

Gatto, R., and Ronchetti, E. (1996), "General Saddlepoint Approximations of Marginal Densities and Tail Probabilities," *Journal of the American Statistical Association*, 91, 666–673. [1167]

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer. [1165]

Hillier, G., and Martellosio, F. (2018), "Exact and Higher-order Properties of the MLE in Spatial Autoregressive Models, With Applications to Inference," *Journal of Econometrics*, 205, 402–422. [1165,1170,1172]

Horowitz, J. (2001), "The Bootstrap," *Handbook of Econometrics*, 5, 3159–3228. [1165]

Jensen, J. L. (1995), *Saddlepoint Approximations*, Oxford: Oxford University Press. [1165]

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007), "Panel Data Models With Spatially Correlated Error Components," *Journal of Econometrics*, 140, 97–130. [1164]

Kelejian, H., and Piras, G. (2017), *Spatial Econometrics*, Cambridge: Academic Press. [1164]

Kolaczyk, E. D. r. (2009), *Statistical Analysis of Network Data* (Vol. 65), New York: Springer. [1165]

Kolassa, J. (2006), *Series Approximation Methods in Statistics* (Vol. 88), New York: Springer. [1165]

La Vecchia, D. and Ronchetti, E. (2019), "Saddlepoint Approximations for Short and Long Memory Time Series: A Frequency Domain Approach," *Journal of Econometrics*, 213, 578–592. [1165]

Lee, L. (2004), "Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72, 1899–1925. [1164,1168]

Lee, L., and Yu, J. (2010), "Estimation of Spatial Autoregressive Panel Data Models With Fixed Effects," *Journal of Econometrics*, 154, 165–185. [1164,1165,1166,1167,1168]

Lô, S. N., and Ronchetti, E. (2009), "Robust and Accurate Inference for Generalized Linear Models," *Journal of Multivariate Analysis*, 100, 2126–2136. [1172]

Martellosio, F., and Hillier, G. (2020), "Adjusted QMLE for the Spatial Autoregressive Parameter," *Journal of Econometrics*, 219, 488–506. [1165,1170,1171,1172]

Ord, K. (1975), "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120–126. [1164]

Robinson, J., Ronchetti, E., and Young, G. (2003), "Saddlepoint Approximations and Tests Based on Multivariate M-estimates," *The Annals of Statistics*, 31, 1154–1169. [1172]

Robinson, P. M., and Rossi, F. (2014a), Improved Lagrange multiplier tests in spatial autoregressions, *The Econometrics Journal*, 17, 139–164. [1165,1170]

——— (2014b), "Refined Tests for Spatial Correlation," *Econometric Theory*, 31, 1–32. [1164,1165,1170]

——— (2015), "Refinements in Maximum Likelihood Inference on Spatial Autocorrelation in Panel Data," *Journal of Econometrics*, 189, 447–456. [1164,1165,1166,1170,1171]

Rosenblatt, M. (2012), *Gaussian and Non-Gaussian Linear Time Series and Random Fields*, New York: Springer. [1164]

Sun, Y. and Genton, M. G. (2011), "Functional Boxplots," *Journal of Computational and Graphical Statistics*, 20, 316–334. [1172]

Tiefelsdorf, M. (2002), "The Saddlepoint Approximation of Moran's I's and Local Moran's I's Reference Distributions and Their Numerical Evaluation," *Geographical Analysis*, 34, 187–206. [1165]

Van der Vaart, A. W. (1998), *Asymptotic Statistics* (Vol. 3), Cambridge: Cambridge University Press. [1168,1169]

Wikle, C., Zammit-Mangion, A., and Cressie, N. (2019), *Spatio-Temporal Statistics With R*, London: Chapman and Hall. [1164]

Yang, Z. (2015), "A General Method for Third-order Bias and Variance Corrections on a Nonlinear Estimator," *Journal of Econometrics*, 186, 178–200. [1164]