# Addressing endogeneity issues in a spatial autoregressive model using copulas

Yanli Lin[*], Yichun Song[†]

April 5, 2023

**Abstract**

We propose a new semiparametric copula method to tackle possible endogeneity issues in a spatial autoregressive (SAR) model, which might originate from an endogenous spatial weights matrix or endogenous regressors. Using copula endogeneity correction technique, we derive three-stage estimation methods and establish their consistency and asymptotic normality. We then perform Monte Carlo experiments to investigate the finite sample performance of the proposed maximum likelihood (ML) estimator and the instrumental variable (IV) estimator. Moreover, we apply our methods to an empirical study of spatial spillovers in regional productivity with endogenous spatial weights constructed by the proximity of a "meaningful" socioeconomic characteristic - years of education.

**Keywords:** Spatial autoregressive model, Endogenous spatial weights matrix, Endogenous regressors, Copula method

**JEL classification:** C31, C51

## 1 Introduction

The ways in which the outcomes of interconnected spatial units influence each other are usually referred to as spillover effects, which are generally captured by the spatial autoregressive (SAR) model. The SAR model is well-known for its implicit game-theoretical structure and can be regarded as a reaction function. Early research such as Cliff and Ord (1973), Ord (1975), Anselin (1988), Kelejian and Prucha (2001), and Lee (2004, 2007), provides various estimation methods, like maximum likelihood estimation (MLE), instrumental variable (IV) methods, and generalized method of moments (GMM). Some basic assumptions on the traditional SAR model, like an exogenous spatial weight matrix, exogenous regressors, etc., have been relaxed in recent literature to meet the empirical needs. Typical examples include a SAR model with an endogenous spatial weights matrix constructed by "economic distance" (Qu and Lee, 2015), and its further extensions to a panel data setting (Qu et al., 2017) and

[*]Corresponding author: Department of Economics, The Ohio State University, Columbus OH 43210-1120, United States of America. Email: lin.3019@osu.edu.

[†]Department of Economics, The Ohio State University, Columbus OH 43210-1120, United States of America. Email: song.1399@osu.edu.

to the endogenous spatial weights constructed by bilateral variables (Qu et al., 2021), the estimation of peer effects in endogenous social networks (Johnsson and Moon, 2021), a higher-order SAR model with an endogenous regression component (Gupta and Robinson, 2015), and so on.

We consider specifications and estimations of two variants of a SAR model with endogeneity issues. The first variant has an endogenous spatial weights matrix as in Qu and Lee (2015), and the second contains endogenous regressors. The endogenous variables enter into the SAR model in a nonlinear manner in the first variant, but they are linear components in the second one. A standard solution to handle the endogeneity issues in a SAR model is the control function approach proposed in Qu and Lee (2015). They specify two sets of equations: one for the SAR outcome and the other for the endogenous variables, assuming that the endogeneity stems from the correlation between the error terms in the stated equations. As discussed in Section 2, unique structures and correct model specifications (for instance, without omitted variables) for the endogenous variable equations are required for model identification and unbiased estimates.

In contrast, this study directly characterizes the entire dependence structure among the error term, the endogenous variables, and the exogenous variables[1] using an instrument-free parametric copula, rather than specify the additional set of endogenous variable equations. Seminal work of Park and Gupta (2012), Haschka (2022), and Yang et al. (2022) have adopted copula endogeneity correction technique to handle endogenous regressors in a linear regression model. We extend the basic idea of Yang et al. (2022) to spatial settings and propose three-stage estimation methods.

This paper makes two advances compared to the existing literature. Firstly, we extend the copula endogeneity correction technique from linear regression models in Yang et al. (2022) to accommodate more complex spatial econometric models. We propose an alternative instrument-free copula approach to address endogeneity issues in two variants of a SAR model and discuss the differences in their required assumptions and estimation methods. We develop three-stage MLE and three-stage IV estimator for the SAR models. Our method would be appealing, especially when finding valid excluded instruments or constructing endogenous equations for the control function approach becomes challenging in empirical applications. In this situation, identification would be an issue when we have a linear endogenous component (second variant). Moreover, even when the endogenous variables only enter into the SAR model nonlinearly (the first variant), where the identification is no longer problematic[2], our Monte Carlo study shows that the control function approach is less robust against the misspecification than our proposed method when the omitted variables are non-normally distributed.

Secondly, we provide a rigorous justification of the asymptotic properties for our three-stage MLE and IV estimators. Deriving asymptotic results is a central research area in copula correction, as mentioned by Becker et al. (2021) and Haschka (2022), but may lack thorough discussion in previous studies. We show that the sampling errors caused by employing the estimated marginal transformations (for instance, using the proposed estimator in Liu et al. (2012)) instead of the true marginal changes are asymptotically negligible. Under some regularity conditions, the consistency and asymptotic normality for the three-stage estimators can be derived using the asymptotic inference under

1. Those exogenous variables are independent of the error term but may be correlated with the endogenous variables
2. This point has been mentioned in Qu and Lee (2015).

near-epoch dependence (NED) from Jenish and Prucha (2012). However, we detect that asymptotic bias may exist for the three-stage MLE when the endogenous variables also enter into the model linearly as regressors. The IV estimator still works but requires at least one correlated exogenous regressor to be non-normally distributed for identification (Yang et al., 2022).

Then we conduct Monte Carlo experiments to investigate the finite-sample properties of our methods. Gaussian copulas perform well with different sample sizes, various endogenous variables, non-normally distributed disturbances, and non-normal joint distributions of the marginal transformations of endogenous variables and the error term. We verify that the direct extension of Park and Gupta (2012)'s approach to a SAR model causes asymptotic bias whenever there exists correlation between exogenous and endogenous variables. Our simulation also demonstrates that, when endogenous regressors are included, the non-normal distribution of at least one of the exogenous variables (Yang et al., 2022) provides identification. Besides, we compare the performance of our method with Qu and Lee (2015)'s control function approach. When the excluded instruments or the exact functional form are unobserved, simulation results show that our copula estimators are still robust, while the control function may not be. Although Qu and Lee (2015)'s method produces unbiased estimates, the inference might be invalid, i.e., the standard errors increase and the 95% coverage probability deviates from 95%, when the omitted excluded instruments are non-normally distributed.

Furthermore, we provide an empirical application: the regional productivity spillovers in $1,433$ subnational regions from 110 different countries following Gennaioli et al. (2013). We construct the spatial weights matrix by composing Rook contiguity and a possible endogenous economic factor - years of education. The estimate for the spatial coefficient from our copula method is close to that from the control function approach (Qu and Lee, 2015) when the years of education variable is only part of the spatial weights. However, considering a setting where getting excluded instruments is difficult[3], our proposed copula method would be a good alternative to the control function approach.

The rest of the paper is organized as follows. Section 2 sets up the two variants of a SAR model with possible endogeneity issues and explains the copula method in detail. Section 3 introduces the three-stage estimation approaches. We discuss their consistency and asymptotic normality in the subsequent section. Section 5 investigates the finite sample performance by Monte Carlo (MC) simulation. Section 6 applies our proposed methods to empirically study spatial spillovers of regional economic performance. Section 7 concludes the paper. Related statistics of the log pseudo-likelihood function are in Appendix A, and detailed mathematical proofs are collected in Appendix B. Additional proofs, extra MC results, and the application of our copula method to a SAR model with a parametric endogenous heterogeneity component can be found in the online supplementary file.

3. This may probably be because we want to include a rich set of characteristics as regressors.

## 2 Model

We consider a cross-sectional SAR model for the outcome $y$ at the location $i$ as[4]

$$y_i = \lambda \sum_{j \neq i} w_{ij} y_j + x_i' \beta + v_i \tag{1}$$

for $i = 1, .., n$, $w_{ij}$ measures the relative strength of linkage between location $i$ and $j$, $i \neq j$, and $w_{ii} = 0$ for all $i$. $x_i = (x_{1,i}, ..., x_{k,i})' \in \mathbb{R}^k$ is a vector of observed exogenous variables. $v_i$ is the error term. $\lambda$ is a spatial coefficient and $\beta = (\beta_1, ..., \beta_k)'$ is a $k$-dimensional vector of parameters. The SAR model has the matrix form

$$Y_n = \lambda W_n Y_n + X_n \beta + V_n \tag{2}$$

where $Y_n = (y_1, ..., y_n)'$ and $V_n = (v_1, ..., v_n)'$ are $n$-dimensional vectors, $X_n = (x_1, ..., x_n)'$ is an $n \times k$ matrix, $W_n = (w_{ij})$ is an $n \times n$ nonnegative matrix with zero diagonals. When $S_n^{-1}(\lambda)$ exists, where $S_n(\lambda) = I_n - \lambda W_n$, the model has a reduced form,

$$Y_n = S_n^{-1}(\lambda)(X_n \beta + V_n). \tag{3}$$

In general, other than the endogenous $W_n Y_n$, there are two variants of the SAR model that might have endogeneity issues caused by observed endogenous variables $Z_n$, where $Z_n = (z_1, ..., z_n)'$ is an $n \times p$ matrix with $z_i = (z_{1,i}, ..., z_{p,i})'$ being $p$ dimensional column vectors[5]. The first variant is a SAR model (Eq.(1)) with an endogenous spatial weight matrix (Qu and Lee, 2015), i.e., the elements of $W_n$ are constructed by $Z_n : w_{ij} = \psi_{ij}(Z_n, d_{ij})$ for $i, j = 1, ..., n$, $i \neq j$, where $\psi(\cdot)$ is a bounded function. For instance, $w_{ij} = w_{ij}^e w_{ij}^d$, where $w_{ij}^e = 1/|z_i - z_j|$ with $z_i$ being some socioeconomic characteristics (Case et al., 1993), and $w_{ij}^d$ is based on geographic distance. The second variant is a SAR model with endogenous explanatory variables $z_i$,

$$y_i = \lambda \sum_{j \neq i} w_{ij} y_j + x_i' \beta_1 + z_i' \beta_2 + v_i \tag{4}$$

where $\beta_1 = (\beta_{1,1}, ..., \beta_{1,k})'$ and $\beta_2 = (\beta_{2,1}, ..., \beta_{2,p})'$ are $k$ and $p$-dimensional vectors of parameters respectively. This is a direct extension of the linear regression model with one endogenous and one exogenous regressors in Park and Gupta (2012), Haschka (2022) and Yang et al. (2022) to the spatial setting with multiple endogenous and exogenous regressors[6]. In this paper, we discuss both variants[7], and emphasize the differences in estimation methods between the two specifications.

To handle the endogeneity issues in a SAR model, one common method is to employ the control

4. In the scalar form of the model, we omit the subscript $n$, which represents the identity of the $n$th individual, for brevity.

5. In Section S.2 of the supplementary file, we present the specification and the copula estimation method for a third variant, i.e., a SAR model with a parametric endogenous heterogeneity component.

6. See also the higher-order SAR model with an endogenous regression component in Gupta and Robinson (2015).

7. Actually, we focus on the first variant, because the main derivations for the first variant can be applied to the second one, which is less interesting in spatial setting.

function approach[8]. The endogenous variables $Z_n$ can be represented as

$$Z_n = \mu(A_n) + \boldsymbol{\varepsilon}_n \tag{5}$$

where $A_n = (a_1, ..., a_n)'$ is an $n \times k_1$ matrix with $a_i = (a_{1,i}, ..., a_{k_1,i})'$ being $k_1$-dimensional column vectors of observed exogenous variables, $\boldsymbol{\varepsilon}_n = (\varepsilon_1, ..., \varepsilon_n)'$ is an $n \times p$ matrix of disturbances with $\varepsilon_i = (\varepsilon_{1,i}, ..., \varepsilon_{p,i})'$. $\mu(\cdot)$ is a function of $A_n$ and unique structures are imposed on $\mu(A_n)$ for model identification. In Qu and Lee (2015), they consider a linear setting for $\mu(A_n)$, i.e., $Z_n = A_n\Pi + \boldsymbol{\varepsilon}_n$, where $\Pi$ is a $k_1 \times p$ matrix of coefficients. The error terms $v_i$ and $\varepsilon_i$ are assumed to have a joint distribution, which implies conditional homoskedasticity. And then $Z_n - A_n\Pi$ is added as control variables to control the endogeneity of $W_n$. They only consider the nonlinearity of $Z_n$ in $W_n$ case, but the idea can be generalized to the case in Eq.(4). Without observing the structure of $\mu(A_n)$, the linearity assumption could be problematic since the true specification for $\mu(A_n)$ may not be linear[9].

Even if we ignore this issue and implement the control function approach with a linear structure, there are still two notable concerns - identification and model specification. Suppose $A_n$ in the true model setting contains both $X_n$ and some extra variables $X_{2n}$, but due to some reasons, like the unavailability of data[10], $X_{2n}$ may be missing in estimation. When $Z_n$ enter into the SAR model in a nonlinear manner, i.e., the first variant, $A_n$ and $X_n$ are allowed to share common variables. $A_n$ and $X_n$ could even be the same as identification is not an issue (Qu and Lee, 2015). However, a correctly-specified control function is still required; otherwise, biased estimates or invalid asymptotic inference could occur. Simulations in Section 5 indicate the possible invalid inference.

When $Z_n$ linearly enter into a SAR model, i.e., the second variant, apart from the model specification concern, $A_n$ and $X_n$ can not be identical. Some extra exogenous variables, which might not be accessible in practice, are required to satisfy the full-rank assumption for identification.

We handle endogeneity issues in the variants mentioned above of a SAR model and address the aforementioned concerns of the control function approach by directly modeling the correlations between the endogenous variables $z_i$ and the error term $v_i$ using a parametric copula in light of Park and Gupta (2012)[11]. However, as shown in Section S.1 of our supplementary file and Yang et al. (2022), Park and Gupta (2012) requires an implicit assumption of the uncorrelation of exogenous and endogenous regressors to keep the estimator asymptotically correct. Moreover, when $z_i$ are added as linear regressors, Park and Gupta (2012) require the non-normality of $z_{\iota,i}$ ($\iota = 1, ..., p$) for identification, which could be quite restrictive in application.

To tackle these drawbacks, we build our estimator on the modified version in Yang et al. (2022) and

8. Kelejian and Piras (2014) suggest constructing valid instruments by using an estimated weight matrix $\widehat{W}_n$, which is done by directly using the projections of $w_{ij}$ on exogenous variables to construct instruments. However, valid instruments are not easy to find, and high-level assumptions are needed to justify the asymptotic properties of their IV estimators.

9. Or not all the function forms for the $p$ endogenous variables are linear, i.e., $z_{1,n} = \mu_1(A_n) + \varepsilon_{1,n}, ..., z_{p,n} = \mu_p(A_n) + \varepsilon_{p,n}$, where $z_{\iota,n}$ and $\varepsilon_{\iota,n}$ are $n$ dimensional vectors for $\iota = 1, ..., p$, and some (or all) of the $\mu_1(\cdot), ..., \mu_p(\cdot)$ might be nonlinear.

10. All the observed $X_{2n}$ should be exogenous, independent of both $\boldsymbol{\varepsilon}_n$ and $V_n$. Refer to Park and Gupta (2012) for more detailed descriptions of this concern in empirical studies.

11. For nonparametric copula, the construction of an empirical copula (Deheuvels, 1979; Nelsen, 2006) by interpolation or approximation (Li et al., 1997) with selection criteria (e.g., Genest and Rivest, 1993; Barbe et al., 1996) can be considered, which are beyond the scope of this paper.

Haschka (2022). We characterize the entire dependence structure among $x_i$, $z_i$ and $v_i$, while assuming that $x_i$ is independent of $v_i$. Given the known marginal distributions of $x_{1,i}, ..., x_{k,i}$, $z_{1,i}, ..., z_{p,i}$ and $v_i$, denoted as $R_1(x_{1,i}), ..., R_k(x_{k,i})$, $H_1(z_{1,i}), ..., H_p(z_{p,i})$ and $G(v_i)$[12], by Sklar's theorem (Sklar, 1959), a flexible multivariate joint distribution can be constructed,[13]

$$F(x_{1,i}, ..., x_{k,i}, z_{1,i}, ..., z_{p,i}, v_i) = C(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) \tag{6}$$

where $C(\cdot) : [0,1]^{p+k+1} \to [0,1]$ is a $p+k+1$-dimensional copula function, $U_{x_{1,i}} = R_1(x_{1,i}), ..., U_{x_{k,i}} = R_k(x_{k,i}), U_{z_{1,i}} = H_1(z_{1,i}), ..., U_{z_{p,i}} = H_p(z_{p,i}), U_{v_i} = G(v_i)$. From Eq.(6), the joint density function is

$$f(x_{1,i}, ..., x_{k,i}, z_{1,i}, ..., z_{p,i}, v_i) = c(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) g(v_i) \prod_{\iota=1}^{p} h_\iota(z_{\iota,i}) \prod_{\tau=1}^{k} r_\tau(x_{\tau,i}) \tag{7}$$

where $c(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) = \frac{\partial^{p+k+1} C(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i})}{\partial U_{x_{1,i}} \cdots \partial U_{x_{k,i}} \partial U_{z_{1,i}} \cdots \partial U_{z_{p,i}} \partial U_{v_i}}$, $g(\cdot)$, $h_\iota(\cdot)(\iota = 1, ..., p)$ and $r_\tau(\cdot)(\tau = 1, ..., k)$ are marginal density functions of $G(\cdot)$, $H_\iota(\cdot)(\iota = 1, ..., p)$ and $R_\tau(\cdot)(\tau = 1, ..., k)$ respectively. Although there are various copula functions in the statistics literature (Nelsen, 2006; Balakrishnan and Lai, 2009), and in many previous studies (Park and Gupta, 2012; Haschka, 2022, etc), we consider the Gaussian copula for detailed illustration[14]. The $(p+k+1)$-dimensional Gaussian copula with a correlation matrix $P \in [-1,1]^{(k+p+1)\times(k+p+1)}$, where $P = \begin{pmatrix} P_x & P_{xz} & \mathbf{0}_{k\times 1} \\ P_{xz}' & P_z & \rho_{vz} \\ \mathbf{0}_{k\times 1}' & \rho_{vz}' & 1 \end{pmatrix}$ with $\rho_{vz} = (\rho_{vz_1, ..., \rho_{vz_p}})'$ being a $p$ dimensional vector, $P_{xz}$ being a $k \times p$ matrix, $P_z$ and $P_x$ being $p \times p$ and $k \times k$ matrices respectively, can be written as

$$C_P^{Gaussian}(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) = \Phi_P(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*) \tag{8}$$

where $x_{\tau,i}^* = \Phi^{-1}(U_{x_{\tau,i}})(\tau = 1, ..., k)$, $z_{\iota,i}^* = \Phi^{-1}(U_{z_{\iota,i}})(\iota = 1, ..., p)$, $v_i^* = \Phi^{-1}(U_{v_i})$, $\Phi^{-1}(\cdot)$ is a quantile function of standard Gaussian, $\Phi_P(\cdot)$ is the joint CDF of a multivariate normal distribution with mean

12. $G(v_i)$,$H_1(z_{1,i}), ..., H_p(z_{p,i})$ and $R_1(x_{1,i}), ..., R_k(x_{k,i})$ are uniform(0,1) random variables. $H_1(z_{1,i}), ..., H_p(z_{p,i})$ and $R_1(x_{1,i}), ..., R_k(x_{k,i})$ can be identified from the sample data, and $G(v_i)$ is assumed to be normal. The case of unknown $G(v_i)$ might be obtained by some approximations, for instance, the sieve approximation in Xu and Lee (2018). However, due to its complexity, we leave it for future studies.

13. One alternative appealing modelling approach might be the vector version of Sklar's theorem proposed in Fang and Henry (2023), which accounts for both within-vector and between-vector dependence. However, our detailed analysis below is based upon the parametric Gaussian copula, which violates the comonotonic invariance - an invariance property of vector copulas to transformations that leave ranks unchanged - unless we impose a extremely restrictive assumption that there's no within-vector dependence. Other parametric vector copula families that can employ their Vector Sklar Theorem leave for future studies.

14. As mentioned in Park and Gupta (2012), the Gaussian copula is general and robust for most applications (Song, 2000). It incorporates many desirable properties (Danaher and Smith, 2011). For instance, with the empirical CDFs, the Gaussian copula model only depends on the rank-order of raw data and is invariant to strictly monotonic transformations of variables in $x_i$, $z_i$ and $v_i$. We investigate some properties of the proposed Gaussian copula by Monte Carlo simulations in Section 5.

vector of zero and the covariance matrix $\Sigma$ equal to the correlation matrix $P$. Note that

$$
\begin{aligned}
&c_p^{Gaussian}(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) \\
&\quad = \frac{\partial^{p+k+1} C_P^{Gaussian}(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i})}{\partial U_{x_{1,i}} \cdots \partial U_{x_{k,i}} \partial U_{z_{1,i}} \cdots \partial U_{z_{p,i}} \partial U_{v_i}} \\
&\quad = \frac{\phi_P(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)}{\phi(v_i^*) \prod_{\iota=1}^p \phi(z_{\iota,i}^*) \prod_{\tau=1}^k \phi(x_{\tau,i}^*)} \\
&\quad = (2\pi)^{-\frac{p+k+1}{2}} |P|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*) P^{-1} (x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)'\right] \\
&\qquad \cdot (2\pi)^{\frac{1}{2}} \exp\left[\frac{1}{2}(v_i^*)^2\right] \prod_{\iota=1}^p (2\pi)^{\frac{p}{2}} \exp\left[\frac{1}{2}(z_{\iota,i}^*)^2\right] \prod_{\tau=1}^k (2\pi)^{\frac{k}{2}} \exp\left[\frac{1}{2}(x_{\tau,i}^*)^2\right] \\
&\quad = |P|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)(P^{-1} - I_{k+p+1})(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)'\right],
\end{aligned}
$$

then by (7) and (8), the joint density function of $x_i = (x_{1,i}, \ldots, x_{k,i})'$, $z_i = (z_{1,i}, \ldots, z_{p,i})'$ and $v_i$ is

$$
\begin{aligned}
f(x_i', z_i', v_i) = &\frac{1}{|P|^{\frac{1}{2}}} \exp\left[-\frac{(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)(P^{-1} - I_{k+p+1})(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)'}{2}\right] \\
&\cdot g(v_i) \prod_{\iota=1}^p h_\iota(z_{\iota,i}) \prod_{\tau=1}^k r_\tau(x_{\tau,i})
\end{aligned} \tag{9}
$$

The corresponding log-likelihood function for a SAR model with endogenous $W_n$ is

$$
\begin{aligned}
&\ln L_n(\lambda, \beta, \sigma_v^2, P | \{x_i', z_i', v_i\}) \\
&\quad = -\frac{n}{2} \ln|P| - \frac{1}{2} \sum_{i=1}^n (x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)(P^{-1} - I_{k+p+1})(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)' \\
&\qquad + \sum_{i=1}^n \ln \phi_{(0,\sigma_v^2)}\left(v_i(\lambda, \beta)\right)
\end{aligned} \tag{10}
$$

where $\phi_{(0,\sigma_v^2)}(\cdot)$ is the normal density with mean 0 and variance $\sigma_v^2$. Note that the nonparametric densities $h_\iota(z_{\iota,i})$ $(\iota = 1, ..., p)$ and $r_\tau(x_{\tau,i})$ $(\tau = 1, ..., k)$ disappear from the log-likelihood function because they do not contain any unknown parameters. The pseudo-maximum likelihood estimation (PMLE) method based on Eq.(10) is provided in Section 3.

The Gaussian copula models $(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)'$ as the standard multivariate normal distribution with the correlation matrix $P$. When there is one exogenous variable $x_{1,i}^*$ and one endogenous variable $z^*_{1,i}$, $\begin{pmatrix} x^*_{1,i} \\ z^*_{1,i} \\ v^*_i \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}_{3\times 1}, \begin{bmatrix} 1 & \rho_{x_1z_1} & 0 \\ \rho_{x_1z_1} & 1 & \rho_{vz_1} \\ 0 & \rho_{vz_1} & 1 \end{bmatrix}\right)$, by Cholesky decomposition,

$$\begin{pmatrix} x^*_{1,i} \\ z^*_{1,i} \\ v^*_i \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{x_1z_1} & \sqrt{1-\rho^2_{x_1z_1}} & 0 \\ 0 & \frac{\rho_{vz_1}}{\sqrt{1-\rho^2_{x_1z_1}}} & \sqrt{1-\frac{\rho^2_{vz_1}}{1-\rho^2_{x_1z_1}}} \end{pmatrix} \cdot \begin{pmatrix} \varpi_{1,i} \\ \varpi_{2,i} \\ \varpi_{3,i} \end{pmatrix}$$

with $(\varpi_{1,i}, \varpi_{2,i}, \varpi_{3,i})' \sim N(\mathbf{0}_{3\times 1}, I_3)$, where $I_3$ is a $3\times 3$ identity matrix. Given the joint distribution and $v_i = \Phi^{-1}_{(0,\sigma^2_v)}(\Phi(v^*_i)) = \sigma_v v^*_i$, where $\Phi_{(0,\sigma^2_v)}(\cdot)$ is the normal distribution with mean 0 and variance $\sigma^2_v$, we have

$$z^*_{1,i} = \rho_{x_1z_1} x^*_{1,i} + \sqrt{1-\rho^2_{x_1z_1}}\varpi_{2,i} = x^*_{1,i}\tilde{\beta}_1 + u_i$$

where $\tilde{\beta}_1 = \rho_{x_1z_1}$, $u_i = \sqrt{1-\rho^2_{x_1z_1}}\varpi_{2,i}$. Then Eq.(1) with endogeneous spatial weights can be rewritten as

$$\begin{aligned} y_i &= \lambda \sum_{j\neq i} w_{ij}y_j + x_{1,i}\beta_1 + \sigma_v v^*_i \\ &= \lambda \sum_{j\neq i} w_{ij}y_j + x_{1,i}\beta_1 + \frac{\sigma_v\rho_{vz_1}}{1-\rho^2_{x_1z_1}}(z^*_{1,i} - \rho_{x_1z_1}x^*_{1,i}) + \sigma_v\sqrt{1-\frac{\rho^2_{vz_1}}{1-\rho^2_{x_1z_1}}}\varpi_{3,i} \\ &= \lambda \sum_{j\neq i} w_{ij}y_j + x_{1,i}\beta_1 + \gamma_1 u_i + \sigma_v\sqrt{1-\frac{\rho^2_{vz_1}}{1-\rho^2_{x_1z_1}}}\varpi_{3,i} \end{aligned}$$

where $\gamma_1 = \frac{\sigma_v\rho_{vz_1}}{1-\rho^2_{x_1z_1}}$. As suggested in Yang et al. (2022), we can use $u_i$ as an additional regressor to correct for the endogeneity bias[15], then $W_n$ can be treated as predetermined or exogenous. The above approach can be easily extended to accommodate high dimensional exogenous variables $x^*_i = (x^*_{1,i}, ..., x^*_{k,i})'$ and endogenous variables $z^*_i = (z^*_{1,i}, ..., z^*_{p,i})'$, i.e., $(x^{*\prime}_i, z^{*\prime}_i, v^*_i)' \sim N(\mathbf{0}_{(p+k+1)\times 1}, P)$. By Cholesky decomposition,

$$P = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \cdot \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix}' = \begin{pmatrix} L_{11}L'_{11} & * & * \\ L_{21}L'_{11} & L_{21}L'_{21} + L_{22}L'_{22} & * \\ L_{31}L'_{11} & L_{31}L'_{21} + L_{32}L'_{22} & L_{31}L'_{31} + L_{32}L'_{32} + L_{33}L'_{33} \end{pmatrix}$$

From this, we have that

$$\begin{pmatrix} x^*_i \\ z^*_i \\ v^*_i \end{pmatrix} = \begin{pmatrix} L_{11} = \mathrm{Chol}(P_x) & \mathbf{0}_{k\times p} & \mathbf{0}_{k\times 1} \\ L_{21} = P'_{xz}(L'_{11})^{-1} & L_{22} = \mathrm{Chol}(P_z - L_{21}L'_{21}) & \mathbf{0}_{p\times 1} \\ L_{31} = \mathbf{0}'_{k\times 1} & L_{32} = \rho'_{vz}(L'_{22})^{-1} & L_{33} = \mathrm{Chol}(1 - L_{32}L'_{32}) \end{pmatrix} \cdot \begin{pmatrix} \varpi_{1,i} \\ \varpi_{2,i} \\ \varpi_{3,i} \end{pmatrix}$$

with $(\varpi_{1,i}, \varpi_{2,i}, \varpi_{3,i})' \sim N(\mathbf{0}_{(k+p+1)\times 1}, I_{k+p+1})$, where $\mathrm{Chol}(\cdot)$ represents the Cholesky decomposition.

15. Park and Gupta (2012) only account for correlation among $z^*_{1,i}$ and $v_i$, and $z^*_{1,i}$ is added rather than $u_i$, but as shown in Yang et al. (2022), $z^*_{1,i}$ can be correlated with $\varpi_{3,i}$.

Given the joint distribution, we have

$$z_i^* = L_{21}L_{11}^{-1}x_i^* + L_{22}\varpi_{2,i} = \Gamma' x_i^* + u_i \tag{11}$$

where $\Gamma = (P_{xz}' P_x^{-1})'$ and $u_i = L_{22}\varpi_{2,i}$. Then Eq.(1) can be rewritten as

$$\begin{aligned} y_i &= \lambda \sum_{j \neq i} w_{ij} y_j + x_i'\beta + \sigma_v v_i^* \\ &= \lambda \sum_{j \neq i} w_{ij} y_j + x_i'\beta + \sigma_v (z_i^* - \Gamma' x_i^*)'(L_{22}^{-1})' L_{32}' + \sigma_v L_{33}\varpi_{3,i} \\ &= \lambda \sum_{j \neq i} w_{ij} y_j + x_i'\beta + u_i'\gamma + \epsilon_i \end{aligned} \tag{12}$$

where $\gamma = \sigma_v (L_{22}^{-1})' L_{32}'$, $\epsilon_i = \eta\varpi_{3,i}$ with $\eta = \sigma_v L_{33}$[16]. Similarly, the specification for Eq.(4) is

$$y_i = \lambda \sum_{j \neq i} w_{ij} y_j + x_i'\beta_1 + z_i^{'}\beta_2 + u_i'\gamma + \epsilon_i. \tag{13}$$

We show in Section 3 that the same technique can be applied in the PMLE approach. By the above construction, we do not need to impose the implicit assumption that $x_i$ and the pseudo-observations $z_i^*$ are uncorrelated as we model the dependence structure between $x_i^*$ and $z_i^*$ directly. Moreover, the non-normality assumption in Park and Gupta (2012) is relaxed for the second variant, where $z_i$ can be normally distributed. We achieve identification because $u_i$ would not be perfectly collinear with $z_i$ and $x_i$ as long as one of the $x_i$s correlated with $z_i$ is not normally distributed[17]. (12) and (13) indicate that the models can also be estimated by an IV estimation approach to account for the endogenous $W_nY_n$. The derivations are provided in the next section.

To summarize, the constructed log pseudo-likelihood function (10) and the two equations (12)-(13) with $u_i$ added to control for the endogeneity impose the two assumptions below:

**Assumption 1.** *For each $n$, $v_i$'s are i.i.d. $N(0, \sigma_v^2)$ random variables*[18].

**Assumption 2.** *(i) $x_i$, $z_i$, and $v_i$ follow a Guassian copula, i.e., $(x_i^{*\prime}, z_i^{*\prime}, v_i^*)' \sim \mathrm{N}(\mathbf{0}_{(p+k+1)\times 1}, P)$. (ii) When endogenous $z_i$ are added as regressors, one of the correlated exogenous regressors $x_i$ should be non-normally distributed.*

16. This implied model differs from the partially linear SAR model (e.g., Su and Jin, 2010), i.e., $y_i = \lambda \sum_{j \neq i} w_{ij} y_j + x_i'\beta + m_0(z_i) + v_i$, where $m_0(\cdot)$ is an unknown function and the model can be estimated by a profile QMLE based on local polynomial procedure. While $u_i$ are residuals by regressing $z_i^*$ on $x_i^*$, where $x_i^*$ and $z_i^*$ are inverse (standard) normal distribution of unknown marginal distributions.

17. Refer to the proof of Theorem 3 in Yang et al. (2022).

18. Although this assumption might seem restrictive, our MC results show that the proposed estimator is robust against non-normally distributed disturbances.

# 3 Estimation

Because $x_i$ and $z_i$ are allowed to have high dimensions, and $v_i$ is a scalar, it is more convenient to use an equivalent version of (10) by shifting the orders of the three variables,

$$\begin{aligned}
&\ln L_n(\lambda, \beta, \sigma_v^2, \tilde{P} | \{v_i, z_i', x_i'\}) \\
&= -\frac{n}{2} \ln|\tilde{P}| - \frac{1}{2} \sum_{i=1}^{n} (v_i^*, z_i^{*\prime}, x_i^{*\prime})(\tilde{P}^{-1} - I_{k+p+1})(v_i^*, z_i^{*\prime}, x_i^{*\prime})' + \sum_{i=1}^{n} \ln \phi_{(0,\sigma_v^2)}(v_i(\lambda, \beta))
\end{aligned} \tag{14}$$

where the correlation matrix is redefined as $\tilde{P} = \begin{pmatrix} 1 & \rho_{vz}' & \mathbf{0}_{k\times 1}' \\ \rho_{vz} & P_z & P_{xz}' \\ \mathbf{0}_{k\times 1} & P_{xz} & P_x \end{pmatrix}$. By the partitioned quadratic formulation[19], we obtain

$$\begin{aligned}
(v_i^*, z_i^{*\prime}, x_i^{*\prime})\tilde{P}^{-1}(v_i^*, z_i^{*\prime}, x_i^{*\prime})' = \frac{1}{\kappa} &\left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*)\right]' \left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*)\right] \\
&+ (z_i^* - P_{xz}'P_x^{-1}x_i^*)'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*) + x_i^{*\prime}P_x^{-1}x_i^*
\end{aligned} \tag{15}$$

where $\Xi = P_z - P_{xz}'P_x^{-1}P_{xz}$, $\kappa = 1 - \rho_{vz}'\Xi^{-1}\rho_{vz}$. From (11), $\Gamma = (P_{xz}'P_x^{-1})'$, then

$$\begin{aligned}
(v_i^*, z_i^{*\prime}, x_i^{*\prime})\tilde{P}^{-1}(v_i^*, z_i^{*\prime}, x_i^{*\prime})' = \frac{1}{\kappa} &\left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - \Gamma'x_i^*)\right]' \left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - \Gamma'x_i^*)\right] \\
&+ (z_i^* - \Gamma'x_i^*)'\Xi^{-1}(z_i^* - \Gamma'x_i^*) + x_i^{*\prime}P_x^{-1}x_i^*,
\end{aligned} \tag{16}$$

Alternatively, the above log-likelihood function (14) can be written as

$$\begin{aligned}
&\ln L_n\left(\lambda, \beta, \sigma_v^2, \kappa, \rho_{vz}, P_x, \Xi \mid \{v_i, z_i', x_i'\}\right) \\
&= -\frac{n}{2}\ln\kappa - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| - \frac{1}{2}\sum_{i=1}^{n}\left[\kappa^{-1}\left(v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - \Gamma'x_i^*)\right)'\left(v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - \Gamma'x_i^*)\right)\right. \\
&\quad \left. + (z_i^* - \Gamma'x_i^*)'\Xi^{-1}(z_i^* - \Gamma'x_i^*) + x_i^{*\prime}(P_x^{-1} - I_k)x_i^* - v_i^{*2} - z_i^{*\prime}z_i^*\right] + \sum_{i=1}^{n}\ln\phi_{(0,\sigma_v^2)}(v_i(\lambda, \beta)).
\end{aligned} \tag{17}$$

One can refer to Appendix A.2 for the derivation of the determinant equality $|\tilde{P}| = |\kappa| \cdot |P_x| \cdot |\Xi|$. Since the unobserved transformed pseudo-observations $x_i^*$, $z_i^*$ and the residuals $u_i = z_i^* - \Gamma'x_i^*$ in Eq.(12)-(13) and (17) are unobservables but can be consistently estimated, we propose a three-stage estimation method for the model settings.

## 3.1 The first stage estimation

In the first stage, we get estimates for the marginal transformations $\bar{r}_\tau(x) = \Phi^{-1}(R_\tau(x))\,(\tau = 1, ..., k)$ and $\hbar_\iota(z) = \Phi^{-1}(H_\iota(z))\,(\iota = 1, ..., p)$. For the purpose of this paper, any estimation method that yields estimators $\hat{x}_{\tau,i}^*$ and $\hat{z}_{\iota,i}^*$ satisfying $\sup_{x_{\tau,i}}|\hat{x}_{\tau,i}^* - x_{\tau,i}^*| = o_p(1)$ and $\sup_{z_{\iota,i}}|\hat{z}_{\iota,i}^* - z_{\iota,i}^*| = o_p(1)$

19. The derivation is provided in Appendix A.

can be chosen[20]. Let $\hat{R}_\tau(x) = \frac{1}{n}\sum_{i=1}^n I(x_{\tau,i} \le x)$ and $\hat{H}_\iota(z) = \frac{1}{n}\sum_{i=1}^n I(z_{\iota,i} \le z)$ be the empirical distribution functions of $x_\tau$ and $z_\iota$. We consider the estimator proposed in Liu et al. (2012):

$$\hat{\bar{r}}_\tau(x) := \Phi^{-1}\left(T_{1/(2n)}[\hat{R}_\tau(x)]\right) \text{ and } \hat{\hbar}_\iota(z) := \Phi^{-1}\left(T_{1/(2n)}[\hat{H}_\iota(z)]\right) \tag{18}$$

where $T_{1/(2n)}[x] := \frac{1}{2n}\cdot I(x < \frac{1}{2n}) + x\cdot I(\frac{1}{2n} \le x \le 1-\frac{1}{2n}) + (1-\frac{1}{2n})\cdot I(x > 1-\frac{1}{2n})$ is a Winsorization (or truncation) operator. The truncation level $\frac{1}{2n}$ is chosen to control the trade-off of bias and variance in high dimensions. Therefore, $\hat{x}^*_{\tau,i} = \hat{\bar{r}}_\tau(x_{\tau,i})(\tau = 1, ..., k)$ and $\hat{z}^*_{\iota,i} = \hat{\hbar}_\iota(z_{\iota,i})(\iota = 1, ..., p)$ for $i = 1, ..., n$.

### 3.2 The second stage estimation

Denote $X^*_n = (x^*_1, ..., x^*_n)'$, $Z^*_n = (z^*_1, ..., z^*_n)'$, and $U_n = (u_1, ..., u_n)'$ with $u_i = (u_{1,i}, ..., u_{p,i})'$, given $\hat{X}^*_n$ and $\hat{Z}^*_n$ from the first stage estimation, by substituting $\hat{X}^*_n$ for $X^*_n$ and $\hat{Z}^*_n$ for $Z^*_n$ in (11), we have

$$\hat{Z}^*_n = \hat{X}^*_n\Gamma + \hat{U}_n \tag{19}$$

where $\hat{U}_n = U_n - (Z^*_n - \hat{Z}^*_n) + (X^*_n - \hat{X}^*_n)\Gamma$. By the ordinary least squares (OLS) method, we can obtain the estimates for $\Gamma$,

$$\hat{\Gamma} = (\hat{X}^{*\prime}_n\hat{X}^*_n)^{-1}\hat{X}^{*\prime}_n\hat{Z}^*_n. \tag{20}$$

Then $\hat{Z}^*_n - \hat{X}^*_n\hat{\Gamma} = \hat{\mathscr{O}}^\perp_n\hat{Z}^*_n = \hat{\mathscr{O}}^\perp_n\hat{U}_n$, where $\hat{\mathscr{O}}^\perp_n = I_n - \hat{\mathscr{O}}_n$ with $\hat{\mathscr{O}}_n = \hat{X}^*_n(\hat{X}^{*\prime}_n\hat{X}^*_n)^{-1}\hat{X}^{*\prime}_n$.

### 3.3 The third stage estimation

#### 3.3.1 Three-stage pseudo-maximum likelihood estimation

Denote $\omega = (\lambda, \beta')'$, as $v^*_i = \Phi^{-1}\left(\Phi_{(0,\sigma^2_v)}(v_i(\omega))\right) = \frac{v_i(\omega)}{\sigma_v}$, given the estimates from the first and second stage estimations, note that

$$\begin{aligned}
&-\frac{1}{2}\sum_{i=1}^n\left[\frac{1}{\kappa}\left(v^*_i - \rho'_{vz}\Xi^{-1}(\hat{z}^*_i - \hat{\Gamma}'\hat{x}^*_i)\right)'\left(v^*_i - \rho'_{vz}\Xi^{-1}(\hat{z}^*_i - \hat{\Gamma}'\hat{x}^*_i)\right) - v^{*2}_i\right] \\
&= -\frac{1}{2\kappa\sigma^2_v}\sum_{i=1}^n\left(v_i(\omega) - \sigma_v\rho'_{vz}\Xi^{-1}(\hat{z}^*_i - \hat{\Gamma}'\hat{x}^*_i)\right)'\left(v_i(\omega) - \sigma_v\rho'_{vz}\Xi^{-1}(\hat{z}^*_i - \hat{\Gamma}'\hat{x}^*_i)\right) + \frac{1}{2\sigma^2_v}\sum_{i=1}^n v_i(\omega)^2 \\
&= -\frac{1}{2\kappa\sigma^2_v}\left[V_n(\omega) - (\hat{\mathscr{O}}^\perp_n\hat{Z}^*_n)\varsigma\sigma_v\right]'\left[V_n(\omega) - (\hat{\mathscr{O}}^\perp_n\hat{Z}^*_n)\varsigma\sigma_v\right] + \frac{1}{2\sigma^2_v}V_n(\omega)'V_n(\omega),
\end{aligned}$$

where $\varsigma = \Xi^{-1}\rho_{vz}$, and

$$\sum_{i=1}^n \ln\phi_{(0,\sigma^2_v)}(v_i(\omega)) = \ln\phi\left(\frac{Y_n - S_n^{-1}(\lambda)X_n\beta}{\sqrt{S_n^{-1}(\lambda)S_n^{-1\prime}(\lambda)\sigma^2_v}}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2_v + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2_v}V_n(\omega)'V_n(\omega),$$

20. Park and Gupta (2012) recommend using kernel density estimators to estimate all the marginal density functions, for example, $\hat{h}_\iota(z_\iota) = n^{-1}\sum_{i=1}^n K_{h_\iota}(z_\iota - z_{\iota,i}), \iota = 1, \ldots, p$, where $K_{h_\iota}(z_\iota) = h_\iota^{-1}K_\iota(z_\iota/h_\iota)$ with $K_\iota(\cdot)$ being a kernel function on $\mathbb{R}$ and $h_\iota = h_{\iota,n}$ being a bandwidth. Then $\hat{U}_{z_{\iota,i}} = \hat{H}_\iota(z_{\iota,i}) = \int_{-\infty}^{z_{\iota,i}}\hat{h}_\iota(z_\iota)\,dz_\iota$ and $\hat{z}^*_{\iota,i} = \Phi^{-1}\left(\hat{U}_{z_{\iota,i}}\right) = \sqrt{2}erf^{-1}(2\hat{U}_{z_{\iota,i}} - 1)$, where $erf(x) = 2\pi^{-\frac{1}{2}}\int_0^x e^{-t^2}dt$ is the error function. In order to achieve the uniform convergence requirement of $z^*_{\iota,i}$ (and $x^*_{\tau,i}$), a set of assumptions on the kernel functions, bandwidth sequences, and density functions $h_\iota(\cdot)$ should be imposed.

where $V_n(\omega) = S_n(\lambda)Y_n - X_n\beta$. Therefore, the log-likelihood function (17) can be rewritten as

$$
\begin{aligned}
\ln L_n(\theta_{ML}) = & -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|S_n(\lambda)| \\
& - \frac{1}{2\sigma_\xi^2}\left[V_n(\omega) - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right]'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right] \\
& - \frac{1}{2}\sum_{i=1}^n(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*)'\Xi^{-1}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*) - \frac{1}{2}\sum_{i=1}^n \hat{x}_i^{*\prime}(P_x^{-1} - I_k)\hat{x}_i^* + \frac{1}{2}\sum_{i=1}^n \hat{z}_i^{*\prime}\hat{z}_i^*
\end{aligned}
\tag{21}
$$

where $\theta_{ML} = (\omega', \sigma_\xi^2, \chi', \alpha', \delta')'$ with $\sigma_\xi^2 = \kappa\sigma_v^2$, $\chi = \sigma_v\varsigma$, $\alpha$ and $\delta$ being $J_1(=\frac{(k-1)(k-2)}{2})$ and $J_2$-dimensional column vectors of distinct elements in $P_x$ and $\Xi$ respectively. The 3SPML estimator $\hat{\theta}_{ML} = \arg\max \ln L_n(\theta_{ML})$.

**Remark** (Potential bias of 3SPML estimator for the second variant) In the second variant, $V_n(\omega) = S_n(\lambda)Y_n - X_n\beta_1 - Z_n\beta_2$, where $\omega = (\lambda, \beta_1', \beta_2')'$. From the first order derivatives[21] at $\theta_{ML,0}$ and the reduced form of (4), denote $G_n = W_nS_n^{-1}$, we have $\frac{\ln L_n(\theta_{ML,0})}{\partial\theta_{ML}} = \frac{\ln L_n^u(\theta_{ML,0})}{\partial\theta_{ML}} + \Lambda_n^{mle}$, where

$$
\frac{\ln L_n^u(\theta_{ML,0})}{\partial\theta_{ML}} = \begin{pmatrix}
\frac{1}{\sigma_{\xi,0}^2}\left[G_n(X_n\beta_{1,0} + V_n)\right]'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right] - \operatorname{tr}(G_n) \\
\frac{1}{\sigma_{\xi,0}^2}X_n'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right] \\
\mathbf{0}_{p\times 1} \\
-\frac{n}{2\sigma_{\xi,0}^2} + \frac{1}{2\sigma_{\xi,0}^4}\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right]'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right] \\
-\frac{n}{2}\frac{\partial\ln|P_{x,0}|}{\partial\alpha} - \frac{1}{2}\frac{\partial}{\partial\alpha}\operatorname{tr}\left[P_{x,0}^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right] \\
-\frac{n}{2}\frac{\partial\ln|\Xi_0|}{\partial\delta} - \frac{1}{2}\frac{\partial}{\partial\delta}\operatorname{tr}\left[\Xi_0^{-1}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\right]
\end{pmatrix},
$$

and

$$
\Lambda_n^{mle} = \left(\frac{1}{\sigma_{\xi,0}^2}(G_nZ_n\beta_{2,0})'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right], \mathbf{0}_{k\times 1}', \frac{1}{\sigma_{\xi,0}^2}Z_n'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right], 0, \mathbf{0}_{J_1\times 1}', \mathbf{0}_{J_2\times 1}'\right)'.
$$

Note that $\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{\ln L_n^u(\theta_{ML,0})}{\partial\theta_{ML}}\right) = \mathbf{0}_{(k+p+J_1+J_2+2)\times 1}$, however, it's possible that $\mathrm{E}(\frac{1}{\sqrt{n}}\Lambda_n^{mle}) \neq \mathbf{0}_{(k+p+J_1+J_2+2)\times 1}$ if $\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}(G_nZ_n\beta_{2,0})'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right]\right) \neq 0$ and $\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}Z_n'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right]\right) \neq \mathbf{0}_{p\times 1}$. To see this, suppose $z_i$ is normally distributed, $z_i = \Phi_{\sigma_z}^{-1}(\Phi(z_i^*)) = \sigma_z z_i^*$, then $\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}(G_nZ_n\beta_{2,0})'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right]\right) = \frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}(\sigma_{v,0} - 1)\sigma_{z,0}\beta_{2,0}'\rho_{vz,0}\operatorname{tr}[\mathrm{E}(G_n)]$ and $\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}Z_n'\left[V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0\right]\right) = \frac{\sqrt{n}}{\sigma_{\xi,0}^2}(\sigma_{v,0} - 1)\sigma_{z,0}\rho_{vz,0}$. The two expectation components can deviate from zero unless $\sigma_{v,0} = 1$.

### 3.3.2 Three-stage IV estimation

Given the first and second stage estimations, for (12), we have

$$
Y_n = \lambda W_nY_n + X_n\beta + (\hat{Z}_n^* - \hat{X}_n^*\hat{\Gamma})\gamma + \hat{\boldsymbol{\epsilon}}_n = \lambda W_nY_n + X_n\beta + \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*\gamma + \hat{\boldsymbol{\epsilon}}_n \tag{22}
$$

21. Similar to those provided in Appendix A.3.2.

Denote $\theta_{IV} = (\lambda, \beta', \gamma')'$ , $\hat{M}_n = \left(W_nY_n, X_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ and $\hat{T}_n = \left(Q_n, X_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ [22], where $Q_n$ is an instrument variable matrix for the endogenous $W_nY_n$. As $W_nY_n = W_n(I_n - \lambda_0 W_n)^{-1}(X_n\beta + U_n\gamma + \boldsymbol{\epsilon}_n)$[23] and $(I_n - \lambda_0 W_n)^{-1} = \sum_{l=0}^{\infty} \lambda_0^l W_n^l$, the column vectors of $Q_n$ can be linear combinations of $X_n, W_nX_n, W_n^2X_n, ...$ and columns in $\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*$. The 3SIV estimator of $\theta_{IV}$ is

$$\hat{\theta}_{IV} = \left[\hat{M}_n'\hat{T}_n\left(\hat{T}_n'\hat{T}_n\right)^{-1}\hat{T}_n'\hat{M}_n\right]^{-1}\hat{M}_n'\hat{T}_n\left(\hat{T}_n'\hat{T}_n\right)^{-1}\hat{T}_n'Y_n. \tag{23}$$

$\eta^2$ is a scalar and can be estimated by the sample average of the estimated residuals,

$$\hat{\eta}^2 = \frac{1}{n}(Y_n - \hat{\lambda}W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma})'(Y_n - \hat{\lambda}W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma}). \tag{24}$$

# 4 Asymptotic analysis

## 4.1 The first stage estimator

The result that $\hat{\bar{r}}_\tau(x)$ and $\hat{\hbar}_\iota(z)$ converge respectively to $\bar{r}_\tau(x)$ and $\hbar_\iota(z)$ uniformly over some expanding intervals is established in Liu et al. (2012) (Theorem 4.6) and is stated in the proposition below.

**Proposition 1.** *Let $\ell_\tau := \bar{r}_\tau^{-1}$ and $\ell_\iota := \hbar_\iota^{-1}$ be the inverse functions of $\bar{r}_\tau$ and $\hbar_\iota$[24]. For any $a_\tau, a_\iota \in (0,1)$, define $\Upsilon_n^\tau := [\ell_\tau(-\sqrt{\frac{7}{4}a_\tau \log n}), \ell_\tau(\sqrt{\frac{7}{4}a_\tau \log n})]$ and $\Upsilon_n^\iota := [\ell_\iota(-\sqrt{\frac{7}{4}a_\iota \log n}), \ell_\iota(\sqrt{\frac{7}{4}a_\iota \log n})]$. Then $\sup_{x\in\Upsilon_n^\tau}|\hat{\bar{r}}_\tau(x) - \bar{r}_\tau(x)| = O_p\left(\sqrt{\frac{\log\log n}{n^{1-a_\tau}}}\right) = o_p(1)$ and $\sup_{z\in\Upsilon_n^\iota}|\hat{\hbar}_\iota(z) - \hbar_\iota(z)| = O_p\left(\sqrt{\frac{\log\log n}{n^{1-a_\iota}}}\right) = o_p(1)$.*

In a follow-up paper (Han et al., 2013), they have a stronger result by extending the regions of $\Upsilon_n^\tau$ and $\Upsilon_n^\iota$ to be optimal, i.e., $\Upsilon_n^\tau := [\ell_\tau(-\sqrt{2a_\tau \log n}), \ell_\tau(\sqrt{2a_\tau \log n})]$ and $\Upsilon_n^\iota := [\ell_\iota(-\sqrt{2a_\iota \log n}), \ell_\iota(\sqrt{2a_\iota \log n})]$. As a result, we have $\sup_{x_{\tau,i}}\left|\hat{x}_{\tau,i}^* - x_{\tau,i}^*\right| = o_p(1)$ and $\sup_{z_{\iota,i}}\left|\hat{z}_{\iota,i}^* - z_{\iota,i}^*\right| = o_p(1)$, where we omit the notations for the expanding intervals for brevity.

## 4.2 The second stage estimator

In this subsection, we show that the sampling errors $\hat{x}_{\tau,i}^* - x_{\tau,i}^*$ and $\hat{z}_{\iota,i}^* - z_{\iota,i}^*$ are asymptotically negligible, i.e., the errors coming from the estimations of $x_{\tau,i}^*$ by $\hat{x}_{\tau,i}^*$ and $z_{\iota,i}^*$ by $\hat{z}_{\iota,i}^*$ are of order $o_p(1)$ in the second (and third) stage estimation.

For any $m \times n$ matrix $A$, denote $\|A\| = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2\right)^{1/2}$ as the Frobenius norm. From the proof of Lemma A.4 in Liu et al. (2012), $\sup_x\|x_i^*\| \leq \wp_0(k)$, $\sup_x\|\hat{x}_i^*\| \leq \hat{\wp}_0(k)$, $\sup_z\|z_i^*\| \leq \wp_0(p)$, and $\sup_z\|\hat{z}_i^*\| \leq \hat{\wp}_0(p)$, where $\wp_0(k)$, $\hat{\wp}_0(k)$, $\wp_0(p)$, $\hat{\wp}_0(p)$ are sequences of constants satisfying $\wp_0^2(k)k/n \to 0$, $\hat{\wp}_0^2(k)k/n \to 0$, $\wp_0^2(p)p/n \to 0$ and $\hat{\wp}_0^2(p)p/n \to 0$ as $n \to \infty$, then we have one supporting lemma (Lemma 1) for the formal result in Proposition 2.

22. $\theta_{IV} = (\lambda, \beta_1', \beta_2', \gamma')'$, $\hat{M}_n = \left(W_nY_n, X_n, Z_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ and $T_n = \left(Q_n, X_n, Z_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ for Eq.(13).
23. $W_nY_n = W_n(I_n - \lambda_0 W_n)^{-1}(X_n\beta_1 + Z_n\beta_2 + U_n\gamma + \boldsymbol{\epsilon}_n)$ for the second variant.
24. Since $\bar{r}_\tau(x) = \Phi^{-1}(R_\tau(x))$ and $\hbar_\iota(z) = \Phi^{-1}(H_\iota(z))$, we have $\ell_\tau(\cdot) = R_\tau^{-1}(\Phi(\cdot))$ and $\ell_\iota(\cdot) = H_\iota^{-1}(\Phi(\cdot))$.

**Lemma 1.** *(i)* $\frac{1}{n}\|X_n^*\|^2 \leq \tilde{C}_0\wp_0^2(k)$, $\frac{1}{n}\|\hat{X}_n^*\|^2 \leq \tilde{C}_0\hat{\wp}_0^2(k)$, $\frac{1}{n}\|Z_n^*\|^2 \leq \tilde{C}_0\wp_0^2(p)$, $\frac{1}{n}\|\hat{Z}_n^*\|^2 \leq \tilde{C}_0\hat{\wp}_0^2(p)$, *where* $\tilde{C}_0$ *is a finite constant.*
*(ii)* $\frac{1}{n}\|\hat{X}_n^* - X_n^*\|^2 = \log\log n \cdot (\sum_{\tau=1}^{k} n^{a_\tau - 1}) = o_p(1)$, $\frac{1}{n}\|\hat{Z}_n^* - Z_n^*\|^2 = \log\log n \cdot (\sum_{\iota=1}^{p} n^{a_\iota - 1}) = o_p(1)$, $a_\tau, a_\iota \in (0,1)$ *for* $\tau = 1, ..., k$ *and* $\iota = 1, ..., p$.

**Proposition 2.** $\frac{1}{n}\left[a'\varphi_n'(\theta)(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)b - a'\varphi_n'(\theta)(\mathscr{O}_n^\perp Z_n^*)b\right] = o_p(1)$, *where* $a$ *and* $b$ *are conformable vectors of constants,* $\varphi_n(\theta)$ *is a bounded vector-value function or a conformable matrix, and* $\mathscr{O}_n^\perp = I_n - X_n^*(X_n^{*\prime}X_n^*)^{-1}X_n^{*\prime}$.

## 4.3 The third stage estimator

### 4.3.1 Assumptions and topological structure

To analyze the asymptotic properties of the 3SPML estimator and the 3SIV estimator, we need the following assumptions and topological specification.

**Assumption 3.** *Individual units in the economy are located or living in a region* $D_n \subset D \subset \mathbb{R}^{d_0}$, *where the cardinality of* $D_n$ *satisfies* $\lim_{n\to\infty}|D_n| = \infty$. *Any two different individual units* $i$ *and* $j$ *are located at distances of at least* $s_0 > 0$ *from each other, w.l.o.g. we assume that* $s_0 = 1$.

**Assumption 4.** *(i) For any* $i, j$ *and* $n$, *the spatial weight* $w_{ij} \geq 0$, $w_{ii} = 0$, *and* $\sup_n\|W_n\|_\infty = c_w < \infty$.
*(ii)* $\Theta$ *denotes a compact parameter space for* $\theta$. *The true parameter* $\theta_0$ *is in the interior of* $\Theta$.
*(iii) The spatial coefficient* $\lambda$ *satisfies* $\sup_{\lambda\in\Theta_\lambda}|\lambda|c_w < 1$, *where* $\Theta_\lambda$ *is the parameter space for* $\lambda$.
*(iv)* $V_n$, $X_n$ *and* $Z_n$ *are stochastic and uniformly bounded,* $\sup_{n.i}\mathrm{E}|v_i|^{4+\mathfrak{d}} < \infty$, $\max_{\tau=1,...,k}\sup_{n.i}\mathrm{E}|x_{\tau,i}|^{4+\mathfrak{d}} < \infty$, *and* $\max_{\iota=1,...,p}\sup_{n.i}\mathrm{E}|z_{\iota,i}|^{4+\mathfrak{d}} < \infty$ *for some* $\mathfrak{d} > 0$; $X_n$ *and* $V_n$ *are independent.* $\lim_{n\to\infty}\frac{1}{n}X_n^{*\prime}X_n^*$ *exists and is nonsingular.*

**Assumption 5.** *The spatial weight* $w_{ij}$ *satisfies* $0 \leq w_{ij} \leq c_1 d_{ij}^{-c_0 d_0}$ *for some* $c_1 \geq 0$ *and* $c_0 > 3 + \frac{1}{d_0}$. *Furthermore, there exist at most* $K(K \geq 1)$ *columns of* $W_n$ *that the column sum exceeds* $c_w$, *where* $K$ *is a fixed nonnegative integer that does not depend on* $n$.

Assumption 3 is the topological specification. The set $D$ is a lattice of (possibly) unevenly placed locations, which can be a geographic or economic space, or the combination of both, for a unit $i$. The distance may refer to physical or economic distance. The minimum distance $s_0$ is used to avoid extreme influence between two units and indicates that our asymptotic analysis is based on inference under spatial near-epoch dependence (NED) using increasing domain rather than infill domain asymptotics. Assumptions 4*(i)*-4*(iii)* are standard assumptions in the spatial econometrics literature, which limits the spatial correlation to a manageable degree.

Assumption 4*(iv)* gives regularity conditions for $V_n$, $Z_n$ and $X_n$. The assumption imposed on $\lim_{n\to\infty}\frac{1}{n}X_n^{*\prime}X_n^*$ is required for the consistency of the second stage OLS estimator. Assumption 5 allows units far apart to correlate with each other, but the spatial weight should decrease sufficiently fast at a certain rate as the distance $d_{ij}$ increases. This assumption accommodates a special case that $w_{ii} = 0$ if $d_{ij}$ exceeds some threshold. The second part of this assumption restricts the number of columns that can have larger magnitudes than the row-sum norm, which allows the existence of some "stars", i.e., larger units that have larger aggregated impacts on other units even when $n$ increases.

#### 4.3.2 Asymptotic properties of the 3SPML estimator

Denote $\ln L_{n0}(\theta_{ML}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|S_n(\lambda)| - \frac{1}{2\sigma_\xi^2}[V_n(\omega) - (\mathscr{O}_n^\perp Z_n^*)\chi]'[V_n(\omega) - (\mathscr{O}_n^\perp Z_n^*)\chi] - \frac{1}{2}\sum_{i=1}^n (z_i^* - \Gamma' x_i^*)'\Xi^{-1}(z_i^* - \Gamma' x_i^*) - \frac{1}{2}\sum_{i=1}^n x_i^{*\prime}(P_x^{-1} - I_k)x_i^* + \frac{1}{2}\sum_{i=1}^n z_i^{*\prime} z_i^*$, where the estimates from the first and second stage estimations are evaluated at their true values. In Lemma 2, we prove that the main difference between $\ln L_{n0}(\theta_{ML})$ and $\ln L_n(\theta_{ML})$ comes from terms related to $\hat{x}_i^*$ versus $x_i^*$, $\hat{z}_i^*$ versus $z_i^*$, and $\hat{\Gamma}$ versus $\Gamma$ (as well as the differences in their sample averages of the first and second order derivatives[25]), which are all $o_p(1)$.

**Lemma 2.** *Under Assumptions 1, 2, 4(ii) - 4(iv),* $\frac{1}{n}[\ln L_n(\theta_{ML}) - \ln L_{n0}(\theta_{ML})] = o_p(1)$, $\frac{1}{n}\left[\frac{\partial \ln L_n(\theta_{ML})}{\partial\theta_l} - \frac{\partial \ln L_{n0}(\theta_{ML})}{\partial\theta_l}\right] = o_p(1)$, *and* $\frac{1}{n}\left[\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\theta_{l_1}\partial\theta_{l_2}} - \frac{\partial^2 \ln L_{n0}(\theta_{ML})}{\partial\theta_{l_1}\partial\theta_{l_2}}\right] = o_p(1)$, *where* $\theta_l, \theta_{l_1}, \theta_{l_2} = \lambda, \beta, \sigma_\xi^2, \chi, \alpha, \delta$ *for the endogenous spatial weights specification.*

**Assumption 6.** *(i)* $\boldsymbol{X}_n'\boldsymbol{X}_n$ *is invertible with probability one, where* $\boldsymbol{X}_n = \left(X_n, \left(\mathscr{O}_n^\perp Z_n^*\right)\right)$.
*(ii)* $S_n(\lambda)'S_n(\lambda)$ *is not proportional to* $S_n'S_n$ *with probability one whenever* $\lambda \neq \lambda_0$, *where* $S_n(\lambda) = I_n - \lambda W_n$ *and* $S_n = S_n(\lambda_0)$.

Assumption 6 is an identification condition[26]. Below we also present a sufficient identification result for finite sample based on Rothenberg (1971).

**Lemma 3.** *Under Assumptions 1, 2 and 4, if* $\boldsymbol{X}_n'\boldsymbol{X}_n$ *is invertible with probability 1,* $W_n + W_n' \neq 0$, *and there exists* $j \neq j'$ *such that* $\sum_{i=1}^n w_{ij}^2 \neq \sum_{i=1}^n w_{ij'}^2$, *then* $\theta_{ML,0} = (\lambda_0, \beta_0', \sigma_\xi^2, \chi_0', \alpha_0', \delta_0')'$ *is identified.*

**Assumption 7.** *(i)* $\{v_i, z_i, x_i\}_{i=1}^n$ *is an* $\alpha$*-mixing random field with* $\alpha$*-mixing coefficient* $\alpha(\mathfrak{p}, \mathfrak{q}, \mathfrak{r}) \leq (\mathfrak{p} + \mathfrak{q})^{\mathfrak{d}}\hat{\alpha}(\mathfrak{r})$ *for some* $\mathfrak{d} \geq 0$, *where* $\hat{\alpha}(\mathfrak{r})$ *satisfies* $\sum_{\mathfrak{r}=1}^\infty \mathfrak{r}^{d_0 - 1}\hat{\alpha}(\mathfrak{r}) < \infty$.
*(ii) For some* $\mathfrak{t} > 0$, *the* $\alpha$*-mixing coefficient of* $\{v_i, z_i, x_i\}_{i=1}^n$ *satisfies* $\sum_{\mathfrak{r}=1}^\infty \mathfrak{r}^{d_0(\mathfrak{d}_* + 1)}\hat{\alpha}^{\frac{\mathfrak{t}}{4+2\mathfrak{t}}}(\mathfrak{r}) < \infty$, *where* $\mathfrak{r} = \mathfrak{t}\mathfrak{d}/(2 + \mathfrak{t})$.
*(iii)* $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta_{ML}'}\right)$ *exists and is nonsingular.*

With the regularity conditions in Assumption 7, we show the asymptotic properties of the 3SPML estimator in Theorem 1 by using the LLN (Theorem 1) and CLT (Theorem 2 and Corollary 1) in Jenish and Prucha (2012).

**Theorem 1.** *Under Assumptions 1-7, the 3SPML estimator* $\hat{\theta}_{ML}$ *is a consistent estimator of* $\theta_{ML,0}$ *and* $\sqrt{n}(\hat{\theta}_{ML} - \theta_{ML,0}) \xrightarrow{d} \mathrm{N}(0, \mathcal{J}_{ML,0}^{-1})$, *where* $\mathcal{J}_{ML,0}^{-1} = -\left(\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta_{ML}'}\right)\right)^{-1}$[27].

#### 4.3.3 Asymptotic properties of the 3SIV estimator

**Assumption 8.** *Let* $M_n = (W_nY_n, X_n, \mathscr{O}_n^\perp Z_n^*)$, $T_n = (Q_{n,w}, X_n, \mathscr{O}_n^\perp Z_n^*)$, *where* $Q_n$ *is an instrument variable matrix for* $W_nY_n$. $\mathrm{plim}_{n\to\infty}\frac{1}{n}[M_n'T_n(T_n'T_n)^{-1}T_n'M_n]$ *exists and is nonsingular.*

25. The first and second order derivatives of $\ln L_n(\theta_{ML})$ can be found in Appendix A.3.2.
26. A sufficient condition is that matrices $I_n$, $(W_n + W_n')$ and $W_n'W_n$ are linearly independent. Refer to the proof in footnote 9 in Qu and Lee (2015).
27. Expressions for $\mathcal{J}_{ML}$ are in Appendix A.

Assumption 8 requires that all regressors, including the additional regressor $\mathscr{O}_n^{\perp} Z_n^*$ are asymptotically linearly independent such that the IV estimation in the third stage will not suffer from the collinearity problem. Then we have the asymptotic properties for the 3SIV estimator in Theorem 2.

**Theorem 2.** *Under Assumptions 1-5, 7(i)-7(ii) and 8, the 3SIV estimator $\hat{\theta}_{IV}$ is a consistent estimator of $\theta_{IV,0}$ and $\sqrt{n}(\hat{\theta}_{IV} - \theta_{IV,0}) \xrightarrow{d} \mathrm{N}(0, \mathcal{U}_{IV})$, where*

$$\mathcal{U}_{IV} = \mathrm{plim}_{n\to\infty}\frac{1}{n}[M_n' T_n(T_n'T_n)^{-1}T_n'M_n]^{-1}M_n'T_n(T_n'T_n)^{-1}T_n'\mathscr{G}_n T_n(T_n'T_n)^{-1}T_n'M_n[M_n'T_n(T_n'T_n)^{-1}T_n'M_n]^{-1},$$

*the variance of the composite error $\hat{\boldsymbol{\epsilon}}_n$ is $\mathscr{G}_n = \eta_0^2 I_n + \gamma_0'\Xi_0\gamma_0\hat{\mathscr{O}}_n$, where $\Xi_0 = P_{z,0} - P_{xz,0}'P_{x,0}^{-1}P_{xz,0}$.*

As $\eta^2$ is not directly estimated in the IV approach, the result below shows its consistent estimator (as constructed in (24)) and the consistently estimated asymptotic variance of $\hat{\theta}_{IV}$.

**Lemma 4.** *Suppose $\hat{\theta}_{IV}$ is a consistent estimator of $\theta_{IV,0}$, then $\hat{\eta}^2 = \frac{1}{n}\tilde{\boldsymbol{\epsilon}}_n'\tilde{\boldsymbol{\epsilon}}_n$ is a consistent estimator of $\eta_0^2$, where $\tilde{\boldsymbol{\epsilon}}_n = Y_n - \hat{\lambda}W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma}$. If $\theta_{IV,0}$ is replaced with $\hat{\theta}_{IV,0}$ in $\hat{M}_n$ and $\mathscr{O}_n^{\perp}Z_n^*$ with $\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*$, the consistently estimated asymptotic variance-covariance matrix for the 3SIV estimator $\hat{\theta}_{IV}$ is*

$$\begin{aligned}\frac{1}{n}\hat{\mathcal{U}}_{IV} =& \left[\hat{M}_n'\hat{T}_n\left(\hat{T}_n'\hat{T}_n\right)^{-1}\hat{T}_n'\hat{M}_n\right]^{-1}\hat{M}_n'\hat{T}_n\left(\hat{T}_n'\hat{T}_n\right)^{-1}\hat{T}_n'\hat{\mathscr{G}}_n \\ &\cdot \hat{T}_n\left(\hat{T}_n'\hat{T}_n\right)^{-1}\hat{T}_n'\hat{M}_n\left[\hat{M}_n'\hat{T}_n\left(\hat{T}_n'\hat{T}_n\right)^{-1}\hat{T}_n'\hat{M}_n\right]^{-1},\end{aligned}$$

*where $\hat{\mathscr{G}}_n = \hat{\eta}_0^2 I_n + \hat{\gamma}_0'\hat{\Xi}_0\hat{\gamma}_0\hat{\mathscr{O}}_n$ with $\hat{\Xi}_0 = \frac{1}{n}\hat{Z}_n^{*\prime}\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*$.*

## 5 Monte Carlo simulation

In this section, we conduct some Monte Carlo experiments to investigate our proposed estimators' finite sample properties and robustness. In the basic experiment design, $x_i$ is a vector of dimension $k$ with $x_{1,i} = 1$. $x_i^*$, $z_i^*$ and $v_i^*$ follow a multivariate normal distribution

$$\begin{pmatrix} x_i^* \\ z_i^* \\ v_i^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P_x & P_{xz} & \mathbf{0}_{k\times 1} \\ P_{xz}' & P_z & \rho_{vz} \\ \mathbf{0}_{k\times 1}' & \rho_{vz}' & 1 \end{bmatrix}\right), \tag{25}$$

$$\begin{aligned} x_{\tau,i} &= R_\tau^{-1}(U_{x_{\tau,i}}) = R_\tau^{-1}(\Phi(x_{\tau,i}^*)),\ z_{\iota,i} = H_\iota^{-1}(U_{z_{\iota,i}}) = H_\iota^{-1}(\Phi(z_{\iota,i}^*)), \\ v_i &= G^{-1}(U_{v_i}) = G^{-1}(\Phi(v_i^*)) = \Phi^{-1}(\Phi(v_i^*)) = v_i^*. \end{aligned}$$

The data generating process is $y_i = \lambda\sum_{j\neq i} w_{ij}y_j + x_i'\beta + v_i$ with a row-normalized weights matrix generated following Qu and Lee (2015), where $W_n = W_n^d \circ W_n^e$, i.e. $w_{ij} = w_{ij}^d w_{ij}^e$. $W_n^d = W_{R,n}^q$ is a predetermined queen-based contiguity weights matrix. $W_{R,n}^q = (w_{R,ij}^q)$ with $w_{R,ij}^q = w_{ij}^q/(\sum_{k=1}^n w_{ik}^q)$. $W_n^e$ is the endogenous part based on socio-economic similarity with $w_{ij}^e = 1/|z_i - z_j|$ if $i \neq j$ and $w_{ii}^e = 0$. Unless stated otherwise, the number of cross-sectional units $n = 361$, and the endogenous variables $z_i$ only enter into the model nonlinearly[28]. The true model settings assume moderate to high correlation between $z_i$ and $v_i$, or $z_i^*$ and $v_i^*$. The exact level of correlation varies with settings, as indicated by $\rho_{vz}$ in the tables. The number of sample repetitions is 1,000 for each experiment in order to obtain the empirical mean, empirical standard deviation (Std), and 95% coverage probability (CP)[29]. In the following tables, empirical standard errors are reported in parentheses, while square brackets contain the averaged asymptotic analytical errors (see Appendix A.1). Table 2 to Table 8 also report $p10$, $p30$, $p50$, $p70$ and $p90$, which represent the percentiles of differences between estimated parameters and the true values, illustrating the degree of symmetry and the range[30].

**Different sample sizes.** In Table 1, the performances of our estimators are evaluated through various sample sizes $n = 49,\ 144,\ 361,\ 529,\ 1024$. In this experiment, we have $x_i = (x_{1,i}, x_{2,i}, x_{3,i})'$. We consider a scalar $z_i$, which correlates with $x_{3,i}$ (and $v_i$). That is, both $x_i^* (= x_{3,i}^*)$ and $z_i^*$ are scalars. The underlying determination structure then follows

$$\begin{pmatrix} x_i^* \\ z_i^* \\ v_i^* \end{pmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix} \right). \tag{26}$$

As for regressors, $x_{1,i} = 1$, $x_{2,i} \sim N(0,1)$, and $x_{3,i} = R_3^{-1}(\Phi(x_{3,i}^*)) = \Phi^{-1}_{(0,2^2)}(\Phi(x_{3,i}^*))$. $z_i = H^{-1}(\Phi(z_i^*)) = \Phi^{-1}(\Phi(z_i^*)) = z_i^*$. Our estimators perform well across a series of sample sizes with correct specifications. Besides, the IV estimator has a lower bias, especially in small to moderate sample sizes, but is less efficient than MLE.

**Multiple $x_i^*$ and $z_i^*$.** Table 2 contains the case where we have multiple $x_i^*$ and $z_i^*$. In this simulation, we have $x_i = (x_{1,i}, x_{2,i}, x_{3,i})'$. $z_i = (z_{1,i}, z_{2,i})'$, which correlates with $x_{2,i}$ and $x_{3,i}$ (and also $v_i$). $x_i^* = (x_{2,i}^*, x_{3,i}^*)'$, $z_i^* = (z_{1,i}^*, z_{2,i}^*)'$, and $v_i^*$ are generated from

$$\begin{pmatrix} x_{2,i}^* \\ x_{3,i}^* \\ z_{1,i}^* \\ z_{2,i}^* \\ v_i^* \end{pmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0 \\ 0 & 1 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 1 & 0.5 & 0.25 \\ 0.5 & 0.5 & 0.5 & 1 & 0.25 \\ 0 & 0 & 0.25 & 0.25 & 1 \end{bmatrix} \right). \tag{27}$$

We have $x_{1,i} = 1$, $x_{2,i} = R_2^{-1}(\Phi(x_{2,i}^*)) = \Phi^{-1}(\Phi(x_{2,i}^*)) = x_{2,i}^*$, $x_{3,i} = R_3^{-1}(\Phi(x_{3,i}^*)) = \Phi^{-1}_{(0,2^2)}(\Phi(x_{3,i}^*))$. $z_{\iota,i} = H^{-1}(\Phi(z_{\iota,i}^*)) = \Phi^{-1}(\Phi(z_{\iota,i}^*)) = z_{\iota,i}^*$, $\iota \in \{1,2\}$. We also compare our estimator with several other methods. In Table 2, "SAR" refers to the standard SAR model, which incorrectly assumes no endogeneity in $W_n$. "Original" is the estimator proposed in the seminal work of Park and Gupta (2012), which, according to Yang et al. (2022) and as shown in Section S.1 of our online supplementary file, would be inconsistent due to the correlation between $x_i^*$ and $z_i^*$. Thus, we expect both of them to display some bias. As expected, the misspecified SAR and Original display considerable bias, while

28. That is, $z_i$ would be part of the spatial weights but would not be the regressors.

29. 95% CP represents the proportion of the 95% asymptotic-distribution-based confidence intervals that contain the true parameters.

30. For IV estimators, we do not include $\rho_{vz}$ and $\sigma_v$ as they are not part of the parameters.

our MLE or IV estimator successfully corrects the endogeneity. In contrast, Table 3 contains the result for the case where there is no endogeneity in the spatial weights matrix in the true data generating process. Then all the estimators are well-behaved.

**$z_i$ also added as linear regressors.** Table 4 shows the result with endogenous $z_i$ also serving as a linear regressor. We follow the setting as in Table 1, and have $z_i$ as part of linear regressors. Yang et al. (2022) indicate that at least one of $x_i$ needs to be non-Normal to avoid multicollinearity. Here we still have a Normally distributed $x_i^*$ that enters into the determination of $z_i^*$, but the original variable $x_{3,i} = R_3^{-1}(U_{x_{3,i}}) = R_3^{-1}(\Phi(x_{3,i}^*))$ needs to be non-Normal – we assume $R_3$ to be either Exponential (with mean 2) or Gamma(0.5, 2) distribution[31]. We find that all other parameters are precisely estimated, except for a slight bias in $\beta_z$. A larger sample size is needed to guarantee better performance.

Then, consider the situation that in a spatial case, we have regressors also generated following a spatial setting. Or in other words, $X_{3,n} = \rho_x W_n X_{3,n} + R_3^{-1}(U_{X_{3,n}})$ and $Z_n = \rho_z W_n^d Z_n + H^{-1}(\Phi(Z_n)) = \rho_z W_n^d Z_n + Z_n^*$[32], where $\rho_x = 0.3$ and $\rho_z = 0.2$. $R_3$ is the same CDF as the previous discussion. Then there is misspecification in the copulas model. The simulated results in Table 5 show that while this misspecification does not influence the point estimates, the inference becomes imprecise, especially for the MLE estimator[33].

**Misspecifications.** First, we check the robustness of our estimators against non-normally distributed disturbances $v_i$. We run the simulations with two cases. Case 1 has $v_i = v_i^* + U(-2, 2)$, where $U(-2, 2)$ is a Uniform distribution between $-2$ and 2. In case 2, $v_i = \Psi_v^{-1}(\Phi(v_i^*))$, and $\Psi_v^{-1}$ is the inverse CDF of Beta distribution with $a = 2$ and $b = 1$. Both error terms deviate from normal distribution after this adjustment. The generation of random terms follow directly from (26). Table 6 displays the simulated results. We observe robust simulated results even with the non-Normal distribution of the error term presented.

Second, we investigate the performance of our estimators under the misspecification of the joint distribution of $v_i^*$ and $z_i^*$. Three non-Normal joint distributions for $(v_i^*, z_i^*)$ are considered: (1) Wishart

31. Although theoretically, any non-Normal distributed $x_{3,i}$ would satisfy the condition, in practice, for a small to moderate sample size, the distribution of $x_{3,i}$ needs to be very distinct from Normal. For instance, $Gamma(0.5, 2)$ has a shape similar to exponential, completely different from Normal. However, the identification comes after the bell-shaped $Gamma(9, 0.5)$ could still be problematic. With high-level multicollinearity, the coefficients in front of $z_i$ and its correlated $x_{3,i}$ would be influenced. So we recommend a large sample size to justify an instrumental-free estimation of linear parameters.

32. Here, $Z_n$ only depends on $W_n^d$ since $W_n^e$ is generated by $z_i$.

33. This type of misspecification also exists in Qu and Lee (2015), where the underlying spatial structure of $z_i$ is not accounted for in the control function method. Studies that allow spatial structures to sink in are left for further investigation.

distribution[34], (2) t distribution[35], and (3) Gaussian Mixture distribution[36]. We set $z_i = \Phi^{-1}(\hat{H}(z_i^* + 0.5 \times x_{3,i}^*))$, where $\hat{H}$ is the empirical CDF calculated as in Section 3.1, $x_{2,i} \sim N(0,1)$, and $x_{3,i} = R_3^{-1}(\Phi(x_{3,i}^*)) = \Phi_{(0,2^2)}(\Phi(x_{3,i}^*))$. Table 7 illustrates that all the estimates are still robust[37].

Third, we examine another type of misspecification. The baseline data is still generated by equation (26), but now $z_i = H^{-1}(\Phi(z_i^*)) + 0.5 \times \varepsilon_{1,i}^2 + 10 \times \varepsilon_{2,i} + v_i^*$, where $H$ is the CDF for the exponential distribution with a rate parameter $\mu = 2$. $\varepsilon_{1,i}$ follows a Gaussian Mixture distribution[38] and $\varepsilon_{2,i} \sim U(0,1)$. We assume that we do not observe either $\varepsilon_{1,i}$, $\varepsilon_{2,i}$ or the exact functional form.

As discussed in previous sections, if $z_i$ enters only in a nonlinear manner, then even without excluded instruments, regressing $z_i$ on $x_i$ and inserting the error term as a control variable would not cause any identification problem. However, this estimator is less robust with a nonlinear function form or the non-Normally distributed omitted variables. Table 8 illustrates this. Both "SAR-with IV" and "SAR-no IV" refer to the estimator proposed in Qu and Lee (2015), where the potentially endogenous $z_i$ is expressed and estimated as a linear function. The difference between them is whether or not we observe the excluded IV, $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$, in the control function. We can see that despite the misspecification, our Copula estimator manages to grasp a reasonable estimate for structure parameters $\beta_0$, $\beta_1$, $\beta_2$ and $\lambda$. In contrast, Table 8 shows that the standard errors increase and the coverage deviates from 95% using Qu and Lee (2015)'s estimator, which indicates potential inference problems. Nevertheless, this simulation only offers an example of the superior performance of our Copula estimator. Whether or not this can be supported theoretically leaves for further study.

34. The probability density of Wishart for a random matrix $X$ is

$$f(X, \Sigma, \nu) = \frac{|X|^{((\nu-d-1)/2)}\exp(0.5\mathrm{tr}(\Sigma^{-1}X))}{2^{\nu d/2}\pi^{(d(d-1))/4}|\Sigma|^{\nu/2}\Gamma(\nu/2)\cdots\Gamma(\nu-(d-1)/2)}.$$

In the simulation, we set $d = 2$, $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and $\nu = 4$. For each $i \in n$, we generate a random matrix and pick the diagonal elements as $(v_i^*, z_i^*)$.

35. Multivariate t distribution for vector $(v_i, \varepsilon_i)$ has the density

$$f(x, \Sigma, \nu) = \frac{1}{|\Sigma|^{1/2}} \frac{1}{\sqrt{(\nu\pi)^2}} \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)} \left(1 + \frac{x'\Sigma^{-1}x}{\nu}\right)^{-(\nu+d)/2}.$$

In the simulation, we set $d = 2$, $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ and $\nu = 3$.

36. The Gassian Mixture distribution takes the form

$$0.4 \times N\left( \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right) + 0.4 \times N\left( \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right) + 0.2 \times N\left( \begin{bmatrix} -8/3 \\ -8/3 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right).$$

37. $\rho_{vz}$ are omitted here since the underlying process is no longer joint Normal distribution, so there is no true $\rho_{vz,0}$ to compare with.

38. The Gaussian Mixture is $0.4 \times N(1,1) + 0.4 \times N(1/3,1) + 0.2 \times N(-8/3,1)$.

Table 1: Estimates under different sample sizes.

| | | | n = 49 | | n = 144 | | n = 361 | | n = 529 | | n = 1024 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | True | | MLE | IV | MLE | IV | MLE | IV | MLE | IV | MLE | IV |
| $\beta_0$ | 1 | Mean | 0.9913 | 0.9671 | 0.9974 | 0.9925 | 0.9979 | 0.9950 | 0.9976 | 0.9954 | 0.9968 | 0.9960 |
| | | Std | (0.1996) | (0.2144) | (0.0995) | (0.1039) | (0.0597) | (0.0624) | (0.0472) | (0.0494) | (0.0354) | (0.0378) |
| | | | [0.1385] | [0.1690] | [0.0697] | [0.0892] | [0.0475] | [0.0544] | [0.0423] | [0.0446] | [0.0310] | [0.0320] |
| | | Coverage | 0.9510 | 0.9470 | 0.9480 | 0.9530 | 0.9420 | 0.9420 | 0.9530 | 0.9510 | 0.9470 | 0.9490 |
| $\beta_1$ | 4 | Mean | 4.0035 | 3.9986 | 4.0007 | 3.9995 | 4.0022 | 4.0012 | 4.0000 | 3.9994 | 3.9994 | 3.9991 |
| | | Std | (0.1238) | (0.1238) | (0.0728) | (0.0740) | (0.0451) | (0.0457) | (0.0364) | (0.0369) | (0.0255) | (0.0256) |
| | | | [0.0977] | [0.1223] | [0.0689] | [0.0707] | [0.0447] | [0.0445] | [0.0374] | [0.0368] | [0.0259] | [0.0264] |
| | | Coverage | 0.9460 | 0.9460 | 0.9550 | 0.9570 | 0.9500 | 0.9460 | 0.9480 | 0.9430 | 0.9420 | 0.9500 |
| $\beta_2$ | -2 | Mean | -2.0018 | -1.9980 | -2.0028 | -2.0019 | -2.0010 | -2.0003 | -2.0005 | -2.0001 | -1.9997 | -1.9995 |
| | | Std | (0.0785) | (0.0795) | (0.0439) | (0.0441) | (0.0270) | (0.0276) | (0.0223) | (0.0226) | (0.0160) | (0.0162) |
| | | | [0.0523] | [0.0750] | [0.0326] | [0.0433] | [0.0215] | [0.0274] | [0.0181] | [0.0226] | [0.0130] | [0.0162] |
| | | Coverage | 0.9490 | 0.9480 | 0.9500 | 0.9590 | 0.9520 | 0.9530 | 0.9540 | 0.9500 | 0.9450 | 0.9470 |
| $\lambda$ | 0.8 | Mean | 0.7974 | 0.8013 | 0.7990 | 0.7999 | 0.7997 | 0.8003 | 0.7999 | 0.8003 | 0.8001 | 0.8002 |
| | | Std | (0.0182) | (0.0210) | (0.0091) | (0.0107) | (0.0053) | (0.0065) | (0.0044) | (0.0053) | (0.0032) | (0.0040) |
| | | | [0.0170] | [0.0191] | [0.0081] | [0.0104] | [0.0054] | [0.0064] | [0.0046] | [0.0053] | [0.0033] | [0.0038] |
| | | Coverage | 0.9470 | 0.9420 | 0.9440 | 0.9420 | 0.9490 | 0.9450 | 0.9560 | 0.9470 | 0.9440 | 0.9550 |
| $\rho_{vz}$ | 0.5 | Mean | 0.4802 | - | 0.4925 | - | 0.4962 | - | 0.4964 | - | 0.4987 | - |
| | | Std | (0.0977) | - | (0.0552) | - | (0.0357) | - | (0.0298) | - | (0.0211) | - |
| | | | [0.1100] | - | [0.0654] | - | [0.0415] | - | [0.0343] | - | [0.0247] | - |
| | | Coverage | 0.9410 | - | 0.9450 | - | 0.9450 | - | 0.9480 | - | 0.9490 | - |
| $\sigma_v$ | 1 | Mean | 0.9635 | - | 0.9878 | - | 0.9948 | - | 0.9958 | - | 0.9978 | - |
| | | Std | (0.1025) | - | (0.0596) | - | (0.0371) | - | (0.0298) | - | (0.0215) | - |
| | | | [0.0977] | - | [0.0583] | - | [0.0371] | - | [0.0307] | - | [0.0221] | - |
| | | Coverage | 0.9320 | - | 0.9500 | - | 0.9480 | - | 0.9450 | - | 0.9510 | - |

Table 2: Estimates with multiple $x_i^*$ and $z_i^*$. n = 361.

| | | True | Mean | Std | Coverage | p10 | p30 | p50 | p70 | p90 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_0$ | SAR | 1 | 0.4838 | (1.0741) | 0.8990 | -2.1239 | -0.0947 | -0.0304 | 0.0187 | 0.0787 |
| | Original | | 0.9925 | (0.0761) | 0.9460 | -0.1096 | -0.0467 | -0.0086 | 0.0299 | 0.0921 |
| | MLE | | 1.0006 | (0.0666) [0.6000] | 0.9560 | -0.0834 | -0.0338 | 0.0017 | 0.0371 | 0.0859 |
| | IV | | 0.9959 | (0.0708) [0.0651] | 0.9470 | -0.0939 | -0.0403 | -0.0044 | 0.0342 | 0.0856 |
| $\beta_1$ | SAR | 4 | 3.8605 | (0.3375) | 0.9390 | -0.5528 | -0.0754 | -0.0196 | 0.0141 | 0.0557 |
| | Original | | 3.5047 | (0.0619) | 0.0000 | -0.5783 | -0.5290 | -0.4918 | -0.4615 | -0.4181 |
| | MLE | | 3.9971 | (0.0529) [0.0463] | 0.9380 | -0.0714 | -0.0280 | -0.0021 | 0.0239 | 0.0652 |
| | IV | | 3.9960 | (0.0533) [0.0537] | 0.9410 | -0.0730 | -0.0305 | -0.0031 | 0.0223 | 0.0640 |
| $\beta_2$ | SAR | -2 | -1.9399 | (0.1496) | 0.9400 | -0.0296 | -0.0084 | 0.0104 | 0.0332 | 0.2356 |
| | Original | | -2.2491 | (0.0315) | 0.0000 | -0.2899 | -0.2658 | -0.2487 | -0.2337 | -0.2099 |
| | MLE | | -2.0017 | (0.0258) [0.0231] | 0.9500 | -0.0342 | -0.0159 | -0.0011 | 0.0123 | 0.0311 |
| | IV | | -2.0012 | (0.0259) [0.0267] | 0.9550 | -0.0335 | -0.0151 | -0.0010 | 0.0131 | 0.0319 |
| $\lambda$ | SAR | 0.8 | 0.9042 | (0.2121) | 0.9040 | -0.0104 | -0.0022 | 0.0034 | 0.0119 | 0.4060 |
| | Original | | 0.7995 | (0.0078) | 0.9490 | -0.0096 | -0.0045 | -0.0006 | 0.0036 | 0.0092 |
| | MLE | | 0.7995 | (0.0078) [0.0075] | 0.9510 | -0.0097 | -0.0045 | -0.0007 | 0.0035 | 0.0093 |
| | IV | | 0.8004 | (0.0090) [0.0090] | 0.9450 | -0.0109 | -0.0041 | 0.0003 | 0.0047 | 0.0117 |
| $\rho_{vz_1}$ | SAR | 0.25 | - | - | - | - | - | - | - | - |
| | Original | | 0.6102 | (0.0398) | 0.0000 | 0.3084 | 0.3420 | 0.3602 | 0.3819 | 0.4085 |
| | MLE | | 0.2492 | (0.0352) [0.0356] | 0.9530 | -0.0454 | -0.0200 | 0.0006 | 0.0176 | 0.0450 |
| | IV | | - | - | - | - | - | - | - | - |
| $\rho_{vz_2}$ | SAR | 0.25 | - | - | - | - | - | - | - | - |
| | Original | | 0.6095 | (0.0392) | 0.0000 | 0.3099 | 0.3416 | 0.3608 | 0.3802 | 0.4088 |
| | MLE | | 0.2473 | (0.0339) [0.0355] | 0.9530 | -0.0488 | -0.0194 | -0.0023 | 0.0148 | 0.0408 |
| | IV | | - | - | - | - | - | - | - | - |
| $\sigma_v$ | SAR | 1 | 2.9105 | (5.2318) | 0.9440 | -0.0976 | -0.0378 | 0.0134 | 0.0922 | 6.1358 |
| | Original | | 1.2172 | (0.0620) | 0.0560 | 0.1432 | 0.1822 | 0.2155 | 0.2471 | 0.3027 |
| | MLE | | 0.9926 | (0.0370) [0.0370] | 0.9500 | -0.0536 | -0.0277 | -0.0083 | 0.0123 | 0.0408 |
| | IV | | - | - | - | - | - | - | - | - |

Table 3: Estimates when there's no endogeneity in the spatial weight matrix. n = 361.

| | | True | Mean | Std | Coverage | p10 | p30 | p50 | p70 | p90 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_0$ | SAR | 1 | 1.0023 | (0.0643) | 0.9480 | -0.0754 | -0.0324 | 0.0017 | 0.0334 | 0.0836 |
| | Original | | 1.0025 | (0.0649) | 0.9500 | -0.0776 | -0.0325 | 0.0020 | 0.0349 | 0.0840 |
| | MLE | | 1.0025 | (0.0648) [0.0581] | 0.9500 | -0.0773 | -0.0325 | 0.0021 | 0.0349 | 0.0845 |
| | IV | | 0.9982 | (0.0686) [0.0667] | 0.9480 | -0.0863 | -0.0384 | -0.0030 | 0.0310 | 0.0881 |
| $\beta_1$ | SAR | 4 | 4.0027 | (0.0551) | 0.9500 | -0.0718 | -0.0231 | 0.0051 | 0.0324 | 0.0715 |
| | Original | | 4.0027 | (0.0552) | 0.9510 | -0.0721 | -0.0227 | 0.0053 | 0.0326 | 0.0713 |
| | MLE | | 4.0027 | (0.0552) [0.0547] | 0.9510 | -0.0720 | -0.0223 | 0.0052 | 0.0324 | 0.0714 |
| | IV | | 4.0011 | (0.0560) [0.0545] | 0.9450 | -0.0748 | -0.0256 | 0.0035 | 0.0319 | 0.0694 |
| $\beta_2$ | SAR | -2 | -2.0009 | (0.0274) | 0.9540 | -0.0368 | -0.0153 | -0.0010 | 0.0138 | 0.0343 |
| | Original | | -2.0009 | (0.0308) | 0.9460 | -0.0090 | -0.0035 | -0.0002 | 0.0030 | 0.0077 |
| | MLE | | -2.0009 | (0.0275) [0.0263] | 0.9540 | -0.0372 | -0.0153 | -0.0008 | 0.0139 | 0.0341 |
| | IV | | -1.9998 | (0.0283) [0.0282] | 0.9530 | -0.0369 | -0.0150 | 0.0002 | 0.0146 | 0.0366 |
| $\lambda$ | SAR | 0.8 | 0.7996 | (0.0064) | 0.9470 | -0.0087 | -0.0034 | -0.0002 | 0.0029 | 0.0073 |
| | Original | | 0.7996 | (0.0065) | 0.9450 | -0.0090 | -0.0035 | -0.0002 | 0.0030 | 0.0077 |
| | MLE | | 0.7996 | (0.0065) [0.0066] | 0.9460 | -0.0573 | -0.0244 | -0.0006 | 0.0245 | 0.0590 |
| | IV | | 0.8005 | (0.0079) [0.0079] | 0.9470 | -0.0092 | -0.0034 | 0.0001 | 0.0045 | 0.0106 |
| $\rho_{vz}$ | SAR | 0 | - | - | - | - | - | - | - | - |
| | Original | | 0.0004 | (0.0611) | 0.9530 | -0.0774 | -0.0323 | -0.0008 | 0.0326 | 0.0785 |
| | MLE | | 0.0004 | (0.0453) [0.0454] | 0.9580 | -0.0573 | -0.0244 | -0.0006 | 0.0245 | 0.0590 |
| | IV | | - | - | - | - | - | - | - | - |
| $\sigma_v$ | SAR | 1 | 0.9908 | (0.0728) | 0.9470 | -0.1014 | -0.0499 | -0.0091 | 0.0284 | 0.0848 |
| | Original | | 0.9953 | (0.0366) | 0.9490 | -0.0516 | -0.0248 | -0.0040 | 0.0144 | 0.0421 |
| | MLE | | 0.9948 | (0.0366) [0.0371] | 0.9480 | -0.0522 | -0.0252 | -0.0046 | 0.0143 | 0.0417 |
| | IV | | - | - | - | - | - | - | - | - |

Table 4: Estimates when $z_i$ also enters as a regressor. n = 361.

| | | | True | Mean | Std | Coverage | p10 | p30 | p50 | p70 | p90 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_0$ | Gamma | MLE | 1 | 1.0057 | (0.0829) [0.0639] | 0.9550 | -0.1026 | -0.0372 | 0.0044 | 0.0540 | 0.1110 |
| | | IV | | 1.0092 | (0.0849) [0.0769] | 0.9500 | -0.1019 | -0.0340 | 0.0081 | 0.0571 | 0.1173 |
| | Exponential | MLE | | 1.0218 | (0.1358) [0.1001] | 0.9510 | -0.1515 | -0.0464 | 0.0264 | 0.0968 | 0.1921 |
| | | IV | | 1.0309 | (0.1438) [0.1357] | 0.9360 | -0.1543 | -0.0352 | 0.0326 | 0.1051 | 0.2097 |
| $\beta_1$ | Gamma | MLE | 4 | 4.0020 | (0.0453) [0.0447] | 0.9510 | -0.0583 | -0.0193 | 0.0032 | 0.0259 | 0.0590 |
| | | IV | | 4.0004 | (0.0460) [0.0448] | 0.9460 | -0.0603 | -0.0222 | 0.0025 | 0.0257 | 0.0584 |
| | Exponential | MLE | | 4.0019 | (0.0457) [0.0445] | 0.9470 | -0.0582 | -0.0183 | 0.0033 | 0.0270 | 0.0589 |
| | | IV | | 4.0007 | (0.0457) [0.0444] | 0.9440 | -0.0600 | -0.0203 | 0.0030 | 0.0258 | 0.0577 |
| $\beta_2$ | Gamma | MLE | -2 | -2.0103 | (0.0581) [0.0317] | 0.9450 | -0.0840 | -0.0405 | -0.0111 | 0.0164 | 0.0651 |
| | | IV | | -2.0095 | (0.0582) [0.0564] | 0.9440 | -0.0836 | -0.0393 | -0.0116 | 0.0175 | 0.0657 |
| | Exponential | MLE | | -2.0135 | (0.0517) [0.0229] | 0.9430 | -0.0785 | -0.0404 | -0.0150 | 0.0114 | 0.0528 |
| | | IV | | -2.0129 | (0.0517) [0.0504] | 0.9420 | -0.0781 | -0.0404 | -0.0143 | 0.0118 | 0.0537 |
| $\beta_z$ | Gamma | MLE | -1 | -0.9702 | (0.1686) [0.0454] | 0.9500 | -0.1891 | -0.0574 | 0.0346 | 0.1237 | 0.2325 |
| | | IV | | -0.9687 | (0.1694) [0.1665] | 0.9480 | -0.1868 | -0.0601 | 0.0362 | 0.1263 | 0.2336 |
| | Exponential | MLE | | -0.9454 | (0.2095) [0.0460] | 0.9460 | -0.2080 | -0.0508 | 0.0661 | 0.1724 | 0.3172 |
| | | IV | | -0.9440 | (0.2102) [0.2060] | 0.9490 | -0.2119 | -0.0507 | 0.0670 | 0.1746 | 0.3135 |
| $\lambda$ | Gamma | MLE | 0.8 | 0.7998 | (0.0060) [0.0059] | 0.9470 | -0.0078 | -0.0030 | -0.0001 | 0.0028 | 0.0072 |
| | | IV | | 0.8007 | (0.0072) [0.0071] | 0.9410 | -0.0079 | -0.0029 | 0.0005 | 0.0045 | 0.0099 |
| | Exponential | MLE | | 0.7999 | (0.0052) [0.0051] | 0.9510 | -0.0070 | -0.0026 | -0.0001 | 0.0024 | 0.0064 |
| | | IV | | 0.8006 | (0.0062) [0.0061] | 0.9420 | -0.0068 | -0.0026 | 0.0006 | 0.0037 | 0.0082 |
| $\rho_{vz}$ | Gamma | MLE | 0.5 | 0.4715 | (0.0979) [0.0417] | 0.9450 | -0.1529 | -0.0749 | -0.0220 | 0.0293 | 0.0887 |
| | | IV | | - | - | - | - | - | - | - | - |
| | Exponential | MLE | | 0.4525 | (0.1216) [0.0419] | 0.9360 | -0.2089 | -0.1008 | -0.0394 | 0.0252 | 0.0953 |
| | | IV | | - | - | - | - | - | - | - | - |
| $\sigma_v$ | Gamma | MLE | 1 | 0.9863 | (0.0884) [0.0368] | 0.9550 | -0.1185 | -0.0641 | -0.0246 | 0.0246 | 0.1011 |
| | | IV | | - | - | - | - | - | - | - | - |
| | Exponential | MLE | | 0.9788 | (0.1037) [0.0365] | 0.9590 | -0.1404 | -0.0818 | -0.0374 | 0.0213 | 0.1139 |
| | | IV | | - | - | - | - | - | - | - | - |

Table 5: Estimates when $z_i$ also enters as a regressor and with a spatial correlated regressor. $n = 361$.

| | | | True | Mean | Std | Coverage | p10 | p30 | p50 | p70 | p90 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_0$ | Gamma | MLE | 1 | 1.0338 | 0.1009 [0.0848] | 0.9350 | -0.0986 | -0.0175 | 0.0322 | 0.0821 | 0.1623 |
| | | IV | | 1.0651 | 0.1055 [0.1011] | 0.9100 | -0.0698 | 0.0127 | 0.0643 | 0.1118 | 0.1967 |
| | Exponential | MLE | | 1.0672 | 0.1560 [0.1326] | 0.9290 | -0.1270 | -0.0095 | 0.0618 | 0.1417 | 0.2672 |
| | | IV | | 1.1283 | 0.1689 [0.1616] | 0.8840 | -0.0769 | 0.0419 | 0.1264 | 0.2042 | 0.3470 |
| $\beta_1$ | Gamma | MLE | 4 | 3.9928 | 0.0552 [0.0542] | 0.9480 | -0.0765 | -0.0370 | -0.0087 | 0.0227 | 0.0631 |
| | | IV | | 3.9849 | 0.0558 [0.0538] | 0.9400 | -0.0860 | -0.0448 | -0.0157 | 0.0157 | 0.0566 |
| | Exponential | MLE | | 3.9951 | 0.0549 [0.0540] | 0.9500 | -0.0747 | -0.0333 | -0.0062 | 0.0253 | 0.0646 |
| | | IV | | 3.9892 | 0.0553 [0.0536] | 0.9450 | -0.0806 | -0.0387 | -0.0120 | 0.0201 | 0.0604 |
| $\beta_2$ | Gamma | MLE | -2 | -1.9935 | 0.0472 [0.0343] | 0.9510 | -0.0534 | -0.0155 | 0.0067 | 0.0311 | 0.0652 |
| | | IV | | -1.9884 | 0.0474 [0.0484] | 0.9500 | -0.0464 | -0.0101 | 0.0113 | 0.0370 | 0.0701 |
| | Exponential | MLE | | -1.9940 | 0.0356 [0.0245] | 0.9440 | -0.0361 | -0.0099 | 0.0062 | 0.0234 | 0.0486 |
| | | IV | | -1.9896 | 0.0358 [0.0364] | 0.9410 | -0.0317 | -0.0054 | 0.0103 | 0.0281 | 0.0545 |
| $\beta_z$ | Gamma | MLE | -1 | -0.9807 | 0.1037 [0.0152] | 0.9340 | -0.0976 | -0.0257 | 0.0120 | 0.0546 | 0.1433 |
| | | IV | | -0.9581 | 0.1060 [0.0957] | 0.9210 | -0.0735 | -0.0045 | 0.0328 | 0.0790 | 0.1665 |
| | Exponential | MLE | | -0.9881 | 0.1092 [0.0144] | 0.9440 | -0.1056 | -0.0353 | 0.0068 | 0.0495 | 0.1445 |
| | | IV | | -0.9707 | 0.1102 [0.0986] | 0.9340 | -0.0864 | -0.0175 | 0.0201 | 0.0655 | 0.1587 |
| $\lambda$ | Gamma | MLE | 0.8 | 0.8049 | 0.0055 [0.0049] | 0.8580 | -0.0019 | 0.0019 | 0.0048 | 0.0077 | 0.0118 |
| | | IV | | 0.8091 | 0.0067 [0.0065] | 0.7440 | 0.0003 | 0.0056 | 0.0089 | 0.0124 | 0.0173 |
| | Exponential | MLE | | 0.8037 | 0.0048 [0.0045] | 0.8850 | -0.0023 | 0.0011 | 0.0037 | 0.0063 | 0.0097 |
| | | IV | | 0.8068 | 0.0058 [0.0057] | 0.7950 | -0.0006 | 0.0039 | 0.0067 | 0.0096 | 0.0139 |
| $\rho_{vz}$ | Gamma | MLE | 0.5 | 0.0363 | 0.4296 [0.0422] | 0.7710 | -1.0572 | -0.7122 | -0.4457 | -0.1975 | 0.1073 |
| | | IV | | - | - | - | - | - | - | - | - |
| | Exponential | MLE | | 0.0523 | 0.4445 [0.0428] | 0.7910 | -1.0795 | -0.7062 | -0.4185 | -0.1617 | 0.1461 |
| | | IV | | - | - | - | - | - | - | - | - |
| $\sigma_v$ | Gamma | MLE | 1 | 1.1314 | 0.2120 [0.0496] | 0.9170 | -0.0232 | 0.0166 | 0.0576 | 0.1379 | 0.3811 |
| | | IV | | - | - | - | - | - | - | - | - |
| | Exponential | MLE | | 1.1482 | 0.2310 [0.0494] | 0.9130 | -0.0198 | 0.0189 | 0.0653 | 0.1660 | 0.4129 |
| | | IV | | - | - | - | - | - | - | - | - |

Table 6: Estimates under non-normally distributed $v_i$s. $n = 361$

| | | | True | Mean | Std | Coverage | p10 | p30 | p50 | p70 | p90 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_0$ | Case 1 | MLE | 1 | 0.9998 | 0.0911 [0.0808] | 0.9480 | -0.1180 | -0.0473 | -0.0005 | 0.0453 | 0.1097 |
| | | IV | | 0.9933 | 0.0963 [0.0934] | 0.9400 | -0.1245 | -0.0573 | -0.0069 | 0.0396 | 0.1064 |
| | Case 2 | MLE | | 0.9993 | 0.0144 [0.0114] | 0.9470 | -0.0190 | -0.0080 | -0.0004 | 0.0069 | 0.0180 |
| | | IV | | 0.9992 | 0.0150 [0.0130] | 0.9470 | -0.0201 | -0.0080 | -0.0010 | 0.0072 | 0.0181 |
| $\beta_1$ | Case 1 | MLE | 4 | 4.0010 | 0.0760 [0.0763] | 0.9510 | -0.0983 | -0.0402 | 0.0006 | 0.0404 | 0.0999 |
| | | IV | | 3.9982 | 0.0770 [0.0769] | 0.9520 | -0.1043 | -0.0415 | -0.0012 | 0.0419 | 0.0971 |
| | Case 2 | MLE | | 4.0005 | 0.0108 [0.0107] | 0.9540 | -0.0146 | -0.0046 | 0.0009 | 0.0069 | 0.0140 |
| | | IV | | 4.0005 | 0.0110 [0.0106] | 0.9540 | -0.0147 | -0.0047 | 0.0009 | 0.0067 | 0.0138 |
| $\beta_2$ | Case 1 | MLE | -2 | -2.0039 | 0.0409 [0.0372] | 0.9460 | -0.0559 | -0.0234 | -0.0051 | 0.0156 | 0.0482 |
| | | IV | | -2.0024 | 0.0417 [0.0423] | 0.9420 | -0.0555 | -0.0231 | -0.0041 | 0.0179 | 0.0518 |
| | Case 2 | MLE | | -2.0002 | 0.0064 [0.0051] | 0.9540 | -0.0084 | -0.0037 | -0.0002 | 0.0030 | 0.0083 |
| | | IV | | -2.0002 | 0.0065 [0.0065] | 0.9540 | -0.0084 | -0.0036 | -0.0003 | 0.0032 | 0.0083 |
| $\lambda$ | Case 1 | MLE | 0.8 | 0.7990 | 0.0089 [0.0090] | 0.9470 | -0.0121 | -0.0053 | -0.0008 | 0.0037 | 0.0104 |
| | | IV | | 0.8002 | 0.0108 [0.0111] | 0.9480 | -0.0136 | -0.0052 | 0.0003 | 0.0060 | 0.0141 |
| | Case 2 | MLE | | 0.8000 | 0.0013 [0.0013] | 0.9510 | -0.0017 | -0.0007 | -0.0001 | 0.0006 | 0.0016 |
| | | IV | | 0.8000 | 0.0016 [0.0015] | 0.9510 | -0.0019 | -0.0008 | -0.0000 | 0.0008 | 0.0020 |
| $\rho_{vz}$ | Case 1 | MLE | 0.5 | 0.3247 | 0.0410 [0.0438] | 0.0090 | -0.2280 | -0.1965 | -0.1745 | -0.1536 | -0.1249 |
| | Case 2 | MLE | | 0.4828 | 0.0353 [0.0417] | 0.9270 | -0.0629 | -0.0351 | -0.0174 | 0.0018 | 0.0280 |
| $\sigma_v$ | Case 1 | MLE | 1.5275 | 1.5191 | 0.0512 [0.0569] | 0.0000 | -0.0730 | -0.0370 | -0.0071 | 0.0190 | 0.0572 |
| | Case 2 | MLE | 0.2357 | 0.2345 | 0.0074 [0.0087] | 0.9500 | -0.0112 | -0.0051 | -0.0011 | 0.0032 | 0.0082 |

Table 7: Estimates from non-Normal joint distributions of $z_i^*$ and $v_i^*$. n = 361.

| | | | True | Mean | Std | Coverage | p10 | p30 | p50 | p70 | p90 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_0$ | Wishart | MLE | 1 | 1.0096 | (0.0726) [0.1543] | 0.9430 | -0.0747 | -0.0227 | 0.0057 | 0.0373 | 0.0947 |
| | | IV | | 0.9906 | (0.1018) [0.1735] | 0.9410 | -0.1245 | -0.0548 | -0.0125 | 0.0322 | 0.1102 |
| | t | MLE | | 1.0095 | (0.1099) [0.0973] | 0.9580 | -0.1249 | -0.0481 | 0.0031 | 0.0595 | 0.1483 |
| | | IV | | 0.9801 | (0.1136) [0.1016] | 0.9410 | -0.1526 | -0.0754 | -0.0249 | 0.0351 | 0.1256 |
| | Mixture | MLE | | 0.9754 | (0.1180) [0.0859] | 0.9510 | -0.1734 | -0.0884 | -0.0291 | 0.0329 | 0.1306 |
| | | IV | | 0.9753 | (0.1213) [0.0985] | 0.9440 | -0.1765 | -0.0870 | -0.0282 | 0.0360 | 0.1303 |
| $\beta_1$ | Wishart | MLE | 4 | 4.0042 | (0.1503) [0.1545] | 0.9510 | -0.1923 | -0.0710 | 0.0104 | 0.0888 | 0.1917 |
| | | IV | | 3.9948 | (0.1538) [0.1491] | 0.9530 | -0.2126 | -0.0805 | 0.0016 | 0.0755 | 0.0755 |
| | t | MLE | | 4.0044 | (0.0834) [0.0830] | 0.9500 | -0.0976 | -0.0365 | 0.0008 | 0.0443 | 0.1141 |
| | | IV | | 3.9947 | (0.0842) [0.0826] | 0.9490 | -0.1098 | -0.0467 | -0.0090 | 0.0371 | 0.1028 |
| | Mixture | MLE | | 3.9954 | (0.0718) [0.0733] | 0.9480 | -0.0975 | -0.0404 | -0.0024 | 0.0341 | 0.0853 |
| | | IV | | 3.9958 | (0.0724) [0.0736] | 0.9490 | -0.0970 | -0.0414 | -0.0022 | 0.0355 | 0.0856 |
| $\beta_2$ | Wishart | MLE | -2 | -2.0024 | (0.0749) [0.0714] | 0.9570 | -0.0985 | -0.0434 | 0.0001 | 0.0364 | 0.0934 |
| | | IV | | -1.9974 | (0.0755) [0.0767] | 0.9510 | -0.0963 | -0.0374 | 0.0070 | 0.0433 | 0.0981 |
| | t | MLE | | -2.0067 | (0.0461) [0.0438] | 0.9450 | -0.0687 | -0.0268 | -0.0035 | 0.0161 | 0.0491 |
| | | IV | | -2.0007 | (0.0459) [0.0465] | 0.9450 | -0.0620 | -0.0214 | 0.0009 | 0.0224 | 0.0577 |
| | Mixture | MLE | | -1.9976 | (0.0474) [0.0363] | 0.9500 | -0.0573 | -0.0224 | 0.0004 | 0.0274 | 0.0654 |
| | | IV | | -1.9977 | (0.0477) [0.0491] | 0.9510 | -0.0602 | -0.0238 | 0.0007 | 0.0279 | 0.0640 |
| $\lambda$ | Wishart | MLE | 0.8 | 0.7963 | (0.0166) [0.0152] | 0.9450 | -0.0250 | -0.0116 | -0.0031 | 0.0058 | 0.0170 |
| | | IV | | 0.8019 | (0.0232) [0.0225] | 0.9490 | -0.0283 | -0.0095 | 0.0037 | 0.0148 | 0.0292 |
| | t | MLE | | 0.7968 | (0.0112) [0.0098] | 0.9490 | -0.0158 | -0.0075 | -0.0024 | 0.0022 | 0.0088 |
| | | IV | | 0.8026 | (0.0125) [0.0119] | 0.9460 | -0.0128 | -0.0027 | 0.0031 | 0.0083 | 0.0166 |
| | Mixture | MLE | | 0.8026 | (0.0089) [0.0093] | 0.9420 | -0.0091 | -0.0019 | 0.0027 | 0.0071 | 0.0143 |
| | | IV | | 0.8024 | (0.0113) [0.0110] | 0.9550 | -0.0124 | -0.0037 | 0.0024 | 0.0083 | 0.0168 |
| $\sigma_v$ | Wishart | MLE | 2.8227 | 2.8122 | (0.1661) [0.1063] | 0.9490 | -0.2183 | -0.1043 | -0.0138 | 0.0638 | 0.2066 |
| | t | MLE | 1.7467 | 1.6964 | (0.3460) [0.0636] | 0.9730 | -0.3113 | -0.1981 | -0.1147 | -0.0039 | 0.2071 |
| | Mixture | MLE | 1.8417 | 1.8220 | (0.0705) [0.0682] | 0.9350 | -0.1135 | -0.0584 | -0.0203 | 0.0190 | 0.0691 |

Table 8: Estimates from linear model misspecification. n = 361.

| | | True | Mean | Std | Coverage | p10 | p30 | p50 | p70 | p90 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_0$ | SAR - with IV | 1 | 1.0016 | (0.0599) | 0.9460 | -0.0699 | -0.0283 | 0.0006 | 0.0324 | 0.0768 |
| | SAR – no IV | | 1.0010 | (0.0883) | 0.9900 | -0.0760 | -0.0305 | 0.0003 | 0.0338 | 0.0788 |
| | MLE | | 0.9981 | (0.0621) [0.0520] | 0.9470 | -0.0776 | -0.0333 | -0.0030 | 0.0300 | 0.0745 |
| | IV | | 0.9939 | (0.0661) [0.0611] | 0.9420 | -0.0878 | -0.0398 | -0.0053 | 0.0251 | 0.0789 |
| $\beta_1$ | SAR - with IV | 4 | 4.0028 | (0.0558) | 0.9460 | -0.0672 | -0.0251 | 0.0023 | 0.0315 | 0.0741 |
| | SAR – no IV | | 4.0026 | (0.0691) | 0.9810 | -0.0681 | -0.0255 | 0.0022 | 0.0321 | 0.0741 |
| | MLE | | 4.0029 | (0.0490) [0.0489] | 0.9440 | -0.0570 | -0.0233 | 0.0026 | 0.0289 | 0.0645 |
| | IV | | 4.0016 | (0.0494) [0.0487] | 0.9430 | -0.0599 | -0.0235 | -0.0006 | 0.0295 | 0.0642 |
| $\beta_2$ | SAR - with IV | -2 | -2.0005 | (0.0272) | 0.9500 | -0.0350 | -0.0142 | -0.0001 | 0.0126 | 0.0346 |
| | SAR – no IV | | -2.0005 | (0.0293) | 0.9670 | -0.0354 | -0.0144 | -0.0005 | 0.0122 | 0.0359 |
| | MLE | | -2.0003 | (0.0268) [0.0235] | 0.9480 | -0.0351 | -0.0145 | -0.0005 | 0.0123 | 0.0349 |
| | IV | | -1.9997 | (0.0271) [0.0271] | 0.9520 | -0.0347 | -0.0143 | -0.0007 | 0.0136 | 0.0377 |
| $\lambda$ | SAR - with IV | 0.8 | 0.7994 | (0.0052) | 0.9450 | -0.0073 | -0.0033 | -0.0004 | 0.0020 | 0.0060 |
| | SAR – no IV | | 0.7997 | (0.0090) | 0.9900 | -0.0080 | -0.0033 | -0.0003 | 0.0025 | 0.0069 |
| | MLE | | 0.7996 | (0.0059) [0.0059] | 0.9450 | -0.0079 | -0.0033 | -0.0003 | 0.0026 | 0.0069 |
| | IV | | 0.8003 | (0.0073) [0.0071] | 0.9530 | -0.0089 | -0.0034 | 0.0005 | 0.0040 | 0.0096 |
| $\sigma_v$ | SAR - with IV | 1 | 1.0021 | (0.0625) | 0.9679 | -0.0542 | -0.0227 | -0.0021 | 0.0177 | 0.0542 |
| | SAR – no IV | | 0.9939 | (0.0438) | 0.9770 | -0.0540 | -0.0242 | -0.0052 | 0.0141 | 0.0422 |
| | MLE | | 0.9946 | (0.0369) [0.0371] | 0.9470 | -0.0532 | -0.0240 | -0.0051 | 0.0141 | 0.0417 |
| | IV | | - | - | - | - | - | - | - | - |

# 6 Empirical application

This section provides an illustrative empirical application on the potential spatial spillovers in regional productivity. Studies on this topic include a regional development model in Gennaioli et al. (2013) and its spatial extension by Sanso-Navarro et al. (2017). Here we deviate from the contiguity-based exogenous spatial weights in Sanso-Navarro et al. (2017) and proceed with a potentially endogenous spatial weights matrix. We then combine the data set collected in and made public by Gennaioli et al. (2013) with the location data from Sanso-Navarro et al. (2017)[39]. For more details regarding the data, readers can refer to these two papers.

In this application, we use all the regional characteristics without a large fraction of missing values in Gennaioli et al. (2013) [40], including (1) average temperature, (2) inverse distance to coast, (3) ln oil production per capita, (4) ln population, (5) years of education, (6) ln number of ethnic groups, (7) the probability of a country's residents speaking the same language, (8) the electricity power grid density, and (9) ln travel times.

The model we estimated is Eq.(2) with an endogenous $W_n$. The outcome $Y_n$ is the regional productivity measured by ln income per capita. The spatial weights matrix is defined by $W_n = W_n^d \circ W_n^e$, with $W_n^d$ being constructed by the Rook contiguity in Hoshino (2022) and $W_n^e = 1/|z_i - z_j|$ if $i \neq j$ and $w_{ii}^e = 0$. The potential endogenous variable $z_i$, in this case, is the years of education. We believe that, beyond the physical distance, spillovers may also depend on socioeconomic status or economic distance, which could be roughly captured by education status. This is consistent with findings in Moretti (2004), Stoyanov and Zubanov (2012) and He et al. (2018). For instance, Moretti (2004) uses the accumulation of social knowledge like patents to measure economic distances between industries to study the spillover effects of productivity. The patent data is not available in our empirical application, so the level of education may serve as a proxy. Stoyanov and Zubanov (2012) show that productivity spillovers occur when workers move out from a high-productive firm. One may expect workers to move to a region with similar socioeconomic conditions, for which education level is a representative factor among the available variables[41].

Furthermore, years of education is potentially endogenous, especially when the country fixed effects are not involved in the regression[42]. To illustrate the performance of our proposed copula endogeneity correction method, we consider two settings. For Setting 1, we have variables (1) to (6) as regressors (as in Hoshino (2022)), with variables (7) - (9) serving as excluded instruments. Thus, the control function approach in Qu and Lee (2015) applies here. Moreover, in Setting 2, we construct a model where all the variables serve as regressors. It is hard to find excluded instruments when such a

39. The original data used in Gennaioli et al. (2013) contains 1,569 subnational regions in 110 different countries, comprising 97% of the world's GDP.

40. Gennaioli et al. (2013) contain a large variety of variables. However, some of them, like the variable trust in others, cannot be used with the primary data set unless one accepts a considerable sample size shrinkage.

41. Years of education, in this case, serves for an illustrative propose. If a richer data set is available, economic distance can also be defined using UPS shipping rates for large packages (Conley and Ligon, 2002), or Pearson's correlation coefficient of interest rates between two countries (Mazurek, 2012). For the studies with a different dependent variable, productivity could be a measure of economic difference (Tsang and Yip, 2007).

42. This, following Gennaioli et al. (2013), is the only prospective source among control variables that may cause the endogeneity issue in this data set.

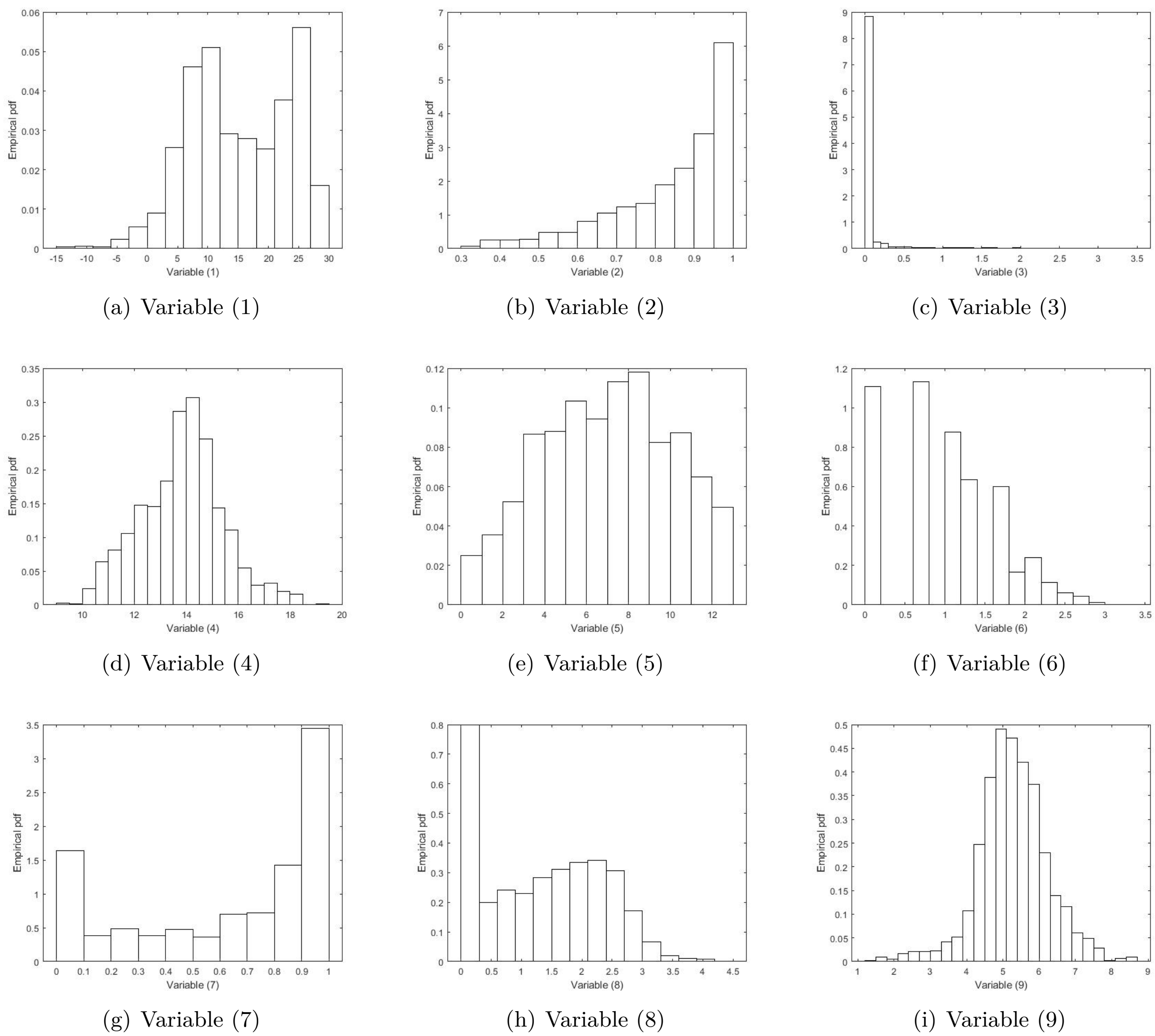
model is adopted. Consequently, Qu and Lee (2015)'s method may break down, but technically the Copula-based method would still work.

For the copula estimation, we have all the exogenous characteristics as the correlated regressors $x_i^*$. Figure 1 demonstrates that this requirement of non-normal distribution for identification is satisfied since most variables have empirical distributions that are distinct from Normal.

Table 9 and 10 report the estimated results from different models, all using the IV estimation method. $\gamma$ is the coefficient for the added residual term discussed in Section 3.3.2 for Copula-IV and also represents the coefficient associated with the control variable in Qu and Lee (2015) (the $\delta$ parameter in their Eq.(3.2)). Table 9 shows that the spatial spillover decreases when we include the country fixed-effects. Although the $\hat{\lambda}$ is significant using a standard SAR model (SAR-IV)[43], the estimates for the spatial coefficient are insignificant under both Qu-IV and Copula-IV. Similar to Yang et al. (2022), while our Copula-based estimator corrects the possible endogeneity through a significant $\gamma$ and a set of different estimates from SAR, the results may not be the same as those using the method with excluded IV (Qu-IV). In this case, we reach a small discrepancy between Qu-IV and Copula-IV for $\hat{\lambda}$. However, we do not know the actual data-generating process, so we cannot tell which estimator has the slightest bias.

In Table 10, we compare the estimates from SAR-IV to those from the Copula-IV in Setting 2, while the control-function-based method drops out due to the lack of excluded instruments. Although the differences in the parameters may not be compelling, Copula-IV still points to a significant $\hat{\gamma}$, indicating the probability of an endogenous $W_n$ constructed by the endogenous variable - years of education. Therefore, our Copula method turns out to be a valid alternative to solve the endogeneity issue in a SAR model, especially in the absence of excluded instruments.

43. Here SAR-IV denotes the SAR model estimated by IV estimator to account for the spatial term, but does not consider any endogenous $z_i$.




(a) Variable (1)

(b) Variable (2)

(c) Variable (3)

(d) Variable (4)

(e) Variable (5)

(f) Variable (6)

(g) Variable (7)

(h) Variable (8)

(i) Variable (9)

Figure 1: Distributions of variables.

Table 9: Estimation results for Setting 1.

| | SAR - IV | Qu - IV | Copula-IV | SAR - IV | Qu - IV | Copula-IV |
|---|---|---|---|---|---|---|
| Intercept | 2.6332*** | 2.3709*** | 3.1602*** | - | - | - |
| | (0.2767) | (0.2761) | (0.3387) | - | - | - |
| (1) Temperature | -0.0044* | 0.0139*** | -0.0257*** | -0.0117*** | -0.0111*** | -0.0209*** |
| | (0.0025) | (0.0046) | (0.0064) | (0.0028) | (0.0030) | (0.0047) |
| (2) Inverse distance to coast | 0.9975*** | 0.7298*** | 1.3920*** | 0.3458 | 0.3167*** | 0.5124*** |
| | (0.1345) | (0.1427) | (0.1849) | (0.0979) | (0.1071) | (0.1146) |
| (3) ln Oil production per capita | 0.2947*** | 0.2684*** | 0.3347*** | 0.1879*** | 0.1923*** | 0.2109*** |
| | (0.0394) | (0.0409) | (0.0429) | (0.0223) | (0.0232) | (0.0234) |
| (4) ln Population | 0.0332*** | 0.0138 | 0.0575*** | 0.0139 | -0.0012 | 0.0242** |
| | (0.0090) | (0.0102) | (0.0120) | (0.0094) | (0.0114) | (0.0109) |
| (5) Years of Education | 0.1735*** | 0.2486*** | 0.0981*** | 0.2578*** | 0.3426*** | 0.2406*** |
| | (0.0148) | (0.0246) | (0.0238) | (0.0119) | (0.0303) | (0.0197) |
| (6) ln No. ethnic groups | -0.0500 | 0.0289 | -0.1437*** | -0.0369** | -0.0097 | -0.0737*** |
| | (0.0220) | (0.0290) | (0.0350) | (0.0148) | (0.0188) | (0.0237) |
| $\lambda$ | 0.4216*** | 0.4073*** | 0.3933*** | 0.2012*** | 0.0795 | 0.0715 |
| | (0.0477) | (0.0486) | (0.0505) | (0.0651) | (0.0717) | (0.0716) |
| $\gamma$ | - | -0.0857*** | 0.2663*** | - | -0.0823*** | 0.0991** |
| | - | (0.0186) | (0.0727) | - | (0.0319) | (0.0502) |
| Country FE | N | N | N | Y | Y | Y |
| Sample sizes | 1433 | 1433 | 1433 | 1433 | 1433 | 1433 |

Standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Estimation results for Setting 2.

| | SAR - IV | Copula-IV | SAR - IV | Copula-IV |
|---|---|---|---|---|
| Intercept | 2.0865*** | 2.5661*** | - | - |
| | (0.2478) | (0.2905) | - | - |
| (1) Temperature | -0.0017 | -0.0182*** | -0.0129*** | -0.0188*** |
| | (0.0022) | (0.0048) | (0.0028) | (0.0043) |
| (2) Inverse distance to coast | 0.7808*** | 1.0163*** | 0.3377*** | 0.3977*** |
| | (0.1180) | (0.1409) | (0.0988) | (0.1040) |
| (3) ln Oil production per capita | 0.2547*** | 0.2649*** | 0.1889*** | 0.1930*** |
| | (0.0342) | (0.0349) | (0.0222) | (0.0222) |
| (4) ln Population | 0.0330*** | 0.0502*** | 0.0045 | 0.0135 |
| | (0.0087) | (0.0100) | (0.0100) | (0.0111) |
| (5) Years of Education | 0.1362*** | 0.0558*** | 0.2422*** | 0.2117*** |
| | (0.0123) | (0.0209) | (0.0125) | (0.0211) |
| (6) ln No. ethnic groups | -0.0039 | -0.0380 | -0.0131 | -0.0279 |
| | (0.0221) | (0.0230) | (0.0166) | (0.0184) |
| (7) Prob same language | 0.2036*** | 0.3206*** | 0.1092*** | 0.1444** |
| | (0.0469) | (0.0587) | (0.0587) | (0.0618) |
| (8) ln power density | 0.0386 | 0.1042*** | 0.0007 | 0.0245 |
| | (0.0206) | (0.0269) | (0.0147) | (0.0197) |
| (9) ln travel time | -0.0070 | -0.0221 | -0.0388** | -0.0437*** |
| | (0.0202) | (0.0199) | (0.0153) | (0.0154) |
| $\lambda$ | 0.5098*** | 0.4934*** | 0.2519*** | 0.2484*** |
| | (0.0410) | (0.0427) | (0.0640) | (0.0640) |
| $\gamma$ | - | 0.2717*** | - | 0.1028** |
| | - | (0.0682) | - | (0.0571) |
| Country FE | N | N | Y | Y |
| Sample sizes | 1433 | 1433 | 1433 | 1433 |

Standard errors in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

# 7 Conclusion

This paper studies the specifications and estimations for two variants of a cross-sectional SAR model subject to possible endogeneity issues. In the first variant, i.e., a SAR model with an endogenous spatial weights matrix, the endogenous variables enter into the SAR model nonlinearly, while in the second variant, they are added as linear regressors. Unlike the control function approach, our method directly models the correlations among the endogenous variables, the error term in the SAR outcome equation, and the exogenous variables that are correlated with the endogenous variables using a Gaussian copula. We propose three-stage estimation methods (3SPML and 3SIV) and establish their asymptotic properties through theory under NED. We conduct Monte Carlo simulations to investigate the finite sample properties of our estimators and verify their robustness against different settings and misspecifications. We then apply our estimation methods to study the spatial spillovers in regional productivity and demonstrate the usefulness of our proposed Copula method in empirical applications.

For future research directions, we might use other parametric copulas that can employ the Vector Sklar Theorem (Fang and Henry, 2023), extend the cross-sectional SAR model to spatial panel data setting, or relax the assumptions of Gaussian copulas by considering nonparametric copulas and the normally-distributed errors by approximation techniques such as the sieve approximation. Moreover, we may consider combining the copula endogeneity correction technique with IV methods to address the omitted variable problem in a control function approach[44].


## Acknowledgements

We are grateful to Professors Lung-fei Lee and Rui Wang, our colleague Yang Yang, and the participants from the Econometrics seminar at The Ohio State University and the "Paelinck seminars" V hosted by Spatial Econometric Association for their helpful comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. All errors are our own.


## Appendix A Expressions related to the statistics

### A.1 Partitioned quadratic formulation

$$
\begin{aligned}
&(v_i^*, z_i^{*\prime}, x_i^{*\prime})\,\tilde{P}^{-1}\,(v_i^*, z_i^{*\prime}, x_i^{*\prime})' \\
&= \left(v_i^* - (\rho_{vz}', \mathbf{0}_{k\times 1}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \hat{z}_i^* \\ \hat{x}_i^* \end{pmatrix}\right)' \left(1 - (\rho_{vz}', \mathbf{0}_{k\times 1}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \rho_{vz} \\ \mathbf{0}_{k\times 1} \end{pmatrix}\right)^{-1} \\
&\quad \cdot \left(v_i^* - (\rho_{vz}', \mathbf{0}_{k\times 1}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \hat{z}_i^* \\ \hat{x}_i^* \end{pmatrix}\right) + (\hat{z}_i^{*\prime}, \hat{x}_i^{*\prime}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \hat{z}_i^* \\ \hat{x}_i^* \end{pmatrix} \\
&= \left(v_i^* - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*)\right)' \left(1 - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}\rho_{vz}\right)^{-1} \\
&\quad \cdot \left(v_i^* - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*)\right) + (\hat{z}_i^{*\prime}, \hat{x}_i^{*\prime}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \hat{z}_i^* \\ \hat{x}_i^* \end{pmatrix} \\
&= \left(v_i^* - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*)\right)' \left(1 - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}\rho_{vz}\right)^{-1} \\
&\quad \cdot \left(v_i^* - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*)\right) \\
&\quad + (\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*)'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*) + \hat{x}_i^* P_x^{-1}\hat{x}_i \\
&= \frac{1}{\kappa}\left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*)\right]' \left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*)\right] \\
&\quad + (z_i^* - P_{xz}'P_x^{-1}x_i^*)'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*) + x_i^{*\prime}P_x^{-1}x_i^*
\end{aligned}
$$

44. Diegert et al. (2022) access the impacts of the omitted variable bias by providing a new approach to sensitivity analysis that allows for endogenous controls and calibrates sensitivity parameters by comparing the magnitude of selection on observables with the magnitude of selection on unobservables.

where $\kappa = 1 - \rho_{vz}'\Xi^{-1}\rho_{vz}$, $\Xi = P_z - P_{xz}'P_x^{-1}P_{xz}$. The first and third equality holds by the partitioned quadratic formulation and the inverse of block matrix

$$\begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} = \begin{pmatrix} (P_z - P_{xz}'P_xP_{xz})^{-1} & -(P_z - P_{xz}'P_xP_{xz})^{-1}P_{xz}'P_x^{-1} \\ -P_x^{-1}P_{xz}(P_z - P_{xz}'P_xP_{xz})^{-1} & P_x^{-1} + P_x^{-1}P_{xz}(P_z - P_{xz}'P_xP_{xz})^{-1}P_{xz}P_x^{-1} \end{pmatrix}^{-1}.$$

## A.2 Determinant of block matrix

$$\begin{aligned}
\begin{vmatrix} 1 & \rho_{vz}' & \mathbf{0}_{k\times 1}' \\ \rho_{vz} & P_z & P_{xz}' \\ \mathbf{0}_{k\times 1} & P_{xz} & P_x \end{vmatrix} &= \left| \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix} - \begin{pmatrix} \rho_{vz} \\ \mathbf{0}_{k\times 1} \end{pmatrix} \begin{pmatrix} \rho_{vz}', \mathbf{0}_{k\times 1}' \end{pmatrix} \right| \\
&= \begin{vmatrix} P_z - \rho_{vz}\rho_{vz}' & P_{xz}' \\ P_{xz} & P_z \end{vmatrix} \\
&= |P_z - \rho_{vz}\rho_{vz}'| \cdot \left| P_x - P_{xz}(P_z - \rho_{vz}\rho_{vz}')^{-1}P_{xz}' \right| \\
&= |P_z - \rho_{vz}\rho_{vz}'| \cdot |P_x| \cdot \left| I_k - P_x^{-1}P_{xz}(P_z - \rho_{vz}\rho_{vz}')P_{xz}' \right| \\
&= |P_x| \cdot \left| P_z - P_{xz}'P_x^{-1}P_{xz} - \rho_{vz}\rho_{vz}' \right| \\
&= |P_x| \cdot \left| P_z - P_{xz}'P_x^{-1}P_{xz} \right| \cdot \left| 1 - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}\rho_{vz} \right|,
\end{aligned}$$

where the fourth and last equalities hold by applying the Sylvester's determinant theorem.

## A.3 PMLE derivation

### A.3.1 Expectation of the log pseudo-likelihood function

For a SAR model with endogenous $W_n$, note that $V_n(\omega) - (\mathscr{O}_n^{\perp}Z_n^*)\chi = S_n(\lambda)Y_n - X_n\beta - (\mathscr{O}_n^{\perp}Z_n^*)\chi = (\lambda_0 - \lambda)G_nX_n\beta_0 + X_n(\beta_0 - \beta) + (\mathscr{O}_n^{\perp}Z_n^*)(\chi_0 - \chi) + S_n(\lambda)S_n^{-1}V_n - (\mathscr{O}_n^{\perp}Z_n^*)\chi_0$, where $G_n = W_nS_n^{-1}$ with $S_n = I_n - \lambda_0W_n$, thus

$$\begin{aligned}
&\frac{1}{n}[V_n(\omega) - (\mathscr{O}_n^{\perp}Z_n^*)\chi]'[V_n(\omega) - (\mathscr{O}_n^{\perp}Z_n^*)\chi] \\
&\quad = \frac{1}{n}V_n'S_n'^{-1}S_n'(\lambda)S_n(\lambda)S_n^{-1}V_n - \frac{1}{n}\chi_0'(\mathscr{O}_n^{\perp}Z_n^*)'(\mathscr{O}_n^{\perp}Z_n^*)\chi_0 - \frac{2}{n}[S_n(\lambda)S_n^{-1}V_n - (\mathscr{O}_n^{\perp}Z_n^*)\chi_0]'(\mathscr{O}_n^{\perp}Z_n^*)\chi \\
&\qquad + (\lambda_0 - \lambda, \beta_0' - \beta', \chi_0' - \chi')\mathscr{A}_n(\lambda_0 - \lambda, \beta_0' - \beta', \chi_0' - \chi')'
\end{aligned}$$

where $\mathscr{A}_n = \frac{1}{n}(G_nX_n\beta_0, X_n, \mathscr{O}_n^{\perp}Z_n^*)'(G_nX_n\beta_0, X_n, \mathscr{O}_n^{\perp}Z_n^*)$. By Lemma 2, $\frac{1}{n}\mathrm{E}(\ln L_n(\theta_{ML})) = \frac{1}{n}\mathrm{E}(\ln L_{n0}(\theta_{ML})) + o_p(1)$, the expectation of the log pseudo-likelihood function in Eq.(21) is

$$\begin{aligned}
&\frac{1}{n}\mathrm{E}(\ln L_n(\theta_{ML})) \\
&\quad = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_\xi^2 - \frac{1}{2}\ln|P_x| - \frac{1}{2}\ln|\Xi| + \frac{1}{n}\mathrm{E}(\ln|S_n(\lambda)|) - \frac{1}{2}\mathrm{tr}(P_x^{-1}P_{x,0}) - \frac{1}{2}\mathrm{tr}(\Xi^{-1}\Xi_0) \\
&\qquad - \frac{1}{2n}\sum_{i=1}^{n} x_i^{*\prime}(\Gamma_0 - \Gamma)\Xi^{-1}(\Gamma_0 - \Gamma)'x_i^* - \frac{1}{2\sigma_\xi^2}(\lambda_0 - \lambda, \beta_0' - \beta', \chi_0' - \chi')\mathrm{E}(\mathscr{A}_n)(\lambda_0 - \lambda, \beta_0' - \beta', \chi_0' - \chi')' \\
&\qquad - \frac{1}{2n}\frac{\sigma_{v,0}^2}{\sigma_\xi^2}\mathrm{E}[\mathrm{tr}(S_n'^{-1}S_n'(\lambda)S_n(\lambda)S_n^{-1})] + \frac{\sigma_{v,0}^2(1 - k_0)}{2\sigma_\xi^2} + \frac{1}{n}\frac{\sigma_{v,0}}{\sigma_\xi^2}\rho_{vz,0}'\chi\mathrm{E}[\mathrm{tr}(S_n(\lambda)S_n^{-1})] - \frac{\sigma_{v,0}}{\sigma_\xi^2}\rho_{vz,0}'\chi \\
&\qquad - \frac{1}{2}P_{x,0} + \frac{1}{2}P_{z,0} + o_p(1).
\end{aligned}$$

### A.3.2 First and second order derivatives

For a SAR model with endogenous spatial weights, the first order derivatives are

$$
\begin{aligned}
\frac{\partial \ln L_n(\theta_{ML})}{\partial \lambda} =& \frac{1}{\sigma_\xi^2}(W_nY_n)'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right] - \text{tr}[W_nS_n^{-1}(\lambda)];\\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \beta} =& \frac{1}{\sigma_\xi^2}X_n'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right];\\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \sigma_\xi^2} =& -\frac{n}{2\sigma_\xi^2}+\frac{1}{2\sigma_\xi^4}\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right]'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right];\\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \chi} =& \frac{1}{\sigma_\xi^2}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right];\\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \alpha} =& -\frac{n}{2}\frac{\partial \ln|P_x|}{\partial \alpha}-\frac{1}{2}\frac{\partial}{\partial\alpha}\text{tr}\left[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right];\\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \delta} =& -\frac{n}{2}\frac{\partial \ln|\Xi|}{\partial \delta}-\frac{1}{2}\frac{\partial}{\partial\delta}\text{tr}\left[\Xi^{-1}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\right],
\end{aligned}
$$

where $V_n(\omega) = S_n(\lambda)Y_n - X_n\beta$, $S_n(\lambda) = I_n - \lambda W_n$. $\alpha$ is a $J_1$-dimensional column vector of distinct elements in $P_x$, the $J_1$-dimensional vector $\frac{\partial \ln|P_x|}{\partial\alpha}$ has the $j_1$th element $\text{tr}(P_x^{-1}\frac{\partial P_x}{\partial \alpha_{j_1}})$ and $\frac{\partial}{\partial\alpha}\text{tr}[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*]$ has its $j_1$th element $-\text{tr}(P_x^{-1}\frac{\partial P_x}{\partial\alpha_{j_1}}P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*)$ for $j_1 = 1, ..., J_1$. $\delta$ is a $J_2$-dimensional column vector of distinct elements in $\Xi$, the $J_2$-dimensional vector $\frac{\partial \ln|\Xi|}{\partial\delta}$ has the $j_2$th element $\text{tr}(\Xi^{-1}\frac{\partial \Xi}{\partial\delta_{j_2}})$ and $\frac{\partial}{\partial\delta}\text{tr}[\Xi^{-1}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)]$ has its $j_2$th element $-\text{tr}[\Xi^{-1}\frac{\partial\Xi}{\partial\delta_{j_2}}\Xi^{-1}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)]$ for $j_2 = 1, ..., J_2$.

The second order derivatives are

$$
\begin{aligned}
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\lambda\partial\lambda} = -\frac{1}{\sigma_\xi^2}(W_nY_n)'(W_nY_n) - \text{tr}[W_nS_n^{-1}(\lambda)]^2;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\lambda\partial\beta} = -\frac{1}{\sigma_\xi^2}(W_nY_n)'X_n;\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\lambda\partial\sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}(W_nY_n)'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right];\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\lambda\partial\chi} = -\frac{1}{\sigma_\xi^2}(W_nY_n)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*);\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\lambda\partial\alpha} = 0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\lambda\partial\delta} = 0;\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\beta\partial\beta'} = -\frac{1}{\sigma_\xi^2}X_n'X_n;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\beta\partial\sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}X_n'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right];\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\beta\partial\chi'} = -\frac{1}{\sigma_\xi^2}X_n'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*);\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\beta\partial\alpha'} = 0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\beta\partial\delta'} = 0;\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\sigma_\xi^2\partial\sigma_\xi^2} = \frac{n}{2\sigma_\xi^4}-\frac{1}{\sigma_\xi^6}\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right]'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right];\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\sigma_\xi^2\partial\chi} = -\frac{1}{\sigma_\xi^4}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi\right];\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\sigma_\xi^2\partial\alpha} = 0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\sigma_\xi^2\partial\delta} = 0;\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\chi\partial\chi'} = -\frac{1}{\sigma_\xi^2}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*);\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\chi\partial\alpha'} = 0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\chi\partial\delta'} = 0;\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\alpha\partial\alpha'} = -\frac{n}{2}\frac{\partial^2 \ln|P_x|}{\partial\alpha\partial\alpha'}-\frac{1}{2}\frac{\partial^2}{\partial\alpha\partial\alpha'}\text{tr}[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*];\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\alpha\partial\delta'} = 0;\\
&\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\delta\partial\delta'} = -\frac{n}{2}\frac{\partial^2 \ln|\Xi|}{\partial\delta\partial\delta'}-\frac{1}{2}\frac{\partial^2}{\partial\delta\partial\delta'}\text{tr}[\Xi^{-1}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)],
\end{aligned}
$$

where $\frac{\partial^2 \ln|P_x|}{\partial\alpha\partial\alpha'}$ is a $J_1 \times J_1$ matrix with the $(j_1, k_1)$th element $\frac{\partial^2 \ln|P_x|}{\partial\alpha_{j_1}\partial\alpha_{k_1}} = -\mathrm{tr}(P_x^{-1}\frac{\partial P_x}{\partial\alpha_{k_1}}P_x^{-1}\frac{\partial P_x}{\partial\alpha_{j_1}})$ and the $(j_1, k_1)$th element of $\frac{\partial^2}{\partial\alpha\partial\alpha'}\,\mathrm{tr}\left[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right]$ is

$$\frac{\partial^2}{\partial\alpha_{j_1}\partial\alpha_{k_1}}\,\mathrm{tr}\left[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right]$$
$$= \mathrm{tr}\left(P_x^{-1}\left(\frac{\partial P_x}{\partial\alpha_{k_1}}P_x^{-1}\frac{\partial P_x}{\partial\alpha_{j_1}} + \frac{\partial P_x}{\partial\alpha_{j_1}}P_x^{-1}\frac{\partial P_x}{\partial\alpha_{k_1}}\right)P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right)$$

for $j_1, k_1 = 1, ..., J_1$. $\frac{\partial^2 \ln|\Xi|}{\partial\delta\partial\delta'}$ is a $J_2 \times J_2$ matrix with the $(j_2, k_2)$th element $\frac{\partial^2 \ln|\Xi|}{\partial\delta_{j_2}\partial\delta_{k_2}} = -\mathrm{tr}(\Xi^{-1}\frac{\partial\Xi}{\partial\delta_{k_2}}\Xi^{-1}\frac{\partial\Xi}{\partial\delta_{j_2}})$ and the $(j_2, k_2)$th element of $\frac{\partial^2}{\partial\delta\partial\delta'}\,\mathrm{tr}\left[\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\right]$ is

$$\frac{\partial^2}{\partial\delta_{j_2}\partial\delta_{k_2}}\,\mathrm{tr}\left[\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\right]$$
$$= \mathrm{tr}\left(\Xi^{-1}\left(\frac{\partial\Xi}{\partial\delta_{k_2}}\Xi^{-1}\frac{\partial\Xi}{\partial\delta_{j_2}} + \frac{\partial\Xi}{\partial\delta_{j_2}}\Xi^{-1}\frac{\partial\Xi}{\partial\delta_{k_2}}\right)\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\right)$$

for $j_2, k_2 = 1, ..., J_2$.

### A.3.3 The variance-covariance matrix

The variance-covariance matrix $\mathscr{I}_{ML,0}^{-1} = -\left(\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta_{ML}'}\right)\right)^{-1}$, where $\mathscr{I}_{ML,0}$ is the Fisher information matrix. For a SAR model with endogenous $W_n$,

$$\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta_{ML}'}\right)$$
$$= -\frac{1}{\sigma_{\xi,0}^2}\begin{bmatrix} A_{\lambda\lambda} & \mathrm{E}(G_nX_n\beta_0)'X_n & \mathrm{tr}[\mathrm{E}(G_n)] & \sigma_{v,0}\rho_{vz,0}'\,\mathrm{tr}\left[\mathrm{E}\left(G_n\right)\right] & 0 & 0 \\ * & X_n'X_n & 0 & 0 & 0 & 0 \\ * & * & \frac{n}{2\sigma_{\xi,0}^2} & 0 & 0 & 0 \\ * & * & * & n\Xi_0 & 0 & 0 \\ * & * & * & * & A_{\alpha\alpha} & 0 \\ * & * & * & * & * & A_{\delta\delta} \end{bmatrix} \tag{A.1}$$

where $A_{\lambda\lambda} = \mathrm{E}[(G_nX_n\beta_0)'(G_nX_n\beta_0)] + \mathrm{tr}[\mathrm{E}(\sigma_{\xi,0}^2G_n^2 + \sigma_{v,0}^2G_n'G_n)]$; $(A_{\alpha\alpha})_{k_1j_1} = \frac{n\sigma_{\xi,0}^2}{2}\mathrm{tr}(P_{x,0}^{-1}\frac{\partial P_{x,0}}{\partial\alpha_{k_1}}P_{x,0}^{-1}\frac{\partial P_{x,0}}{\partial\alpha_{j_1}})$ for $j_1, k_1 = 1, ..., J_1$; $(A_{\delta\delta})_{k_2j_2} = \frac{n\sigma_{\xi,0}^2}{2}\mathrm{tr}(\Xi_0^{-1}\frac{\partial\Xi_0}{\partial\delta_{k_2}}\Xi_0^{-1}\frac{\partial\Xi_0}{\partial\delta_{j_2}})$ for $j_2, k_2 = 1, ..., J_2$.

## Appendix B Mathematical proofs

**Proof of Lemma 1.** First, we prove part (i). $\frac{1}{n}\|X_n^*\|^2 = \frac{1}{n}\sum_{i=1}^n\|x_i^*\|^2 \le \sup_i\|x_i^*\|^2 = \wp_0^2(k)$, the other three results can be shown in a similar way. Second, we consider part (ii). $\frac{1}{n}\|\hat{X}_n^* - X_n^*\|^2 = \frac{1}{n}\sum_{i=1}^n\sum_{\tau=1}^k\|\hat{x}_{\tau,i}^* - x_{\tau,i}^*\|^2 \le \frac{1}{n}\sum_{i=1}^n\sum_{\tau=1}^k\frac{\log\log n}{n^{1-a_\tau}} = \log\log n\cdot(\sum_{\tau=1}^k n^{a_\tau-1})$. Similarly, we have $\frac{1}{n}\|\hat{Z}_n^* - Z_n^*\|^2 \le \log\log n\cdot(\sum_{\iota=1}^k n^{a_\iota-1})$.

**Proof of Proposition 2.**

$$
\begin{aligned}
&\frac{1}{n}\left[a'\varphi_n'(\theta)(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)b - a'\varphi_n'(\theta)(\mathscr{O}_n^{\perp}Z_n^*)b\right]\\
&= \frac{1}{n}a'\varphi_n'(\theta)\hat{\mathscr{O}}_n^{\perp}(\hat{Z}_n^* - Z_n^*)b - \frac{1}{n}a'\varphi_n'(\theta)(\hat{\mathscr{O}}_n - \mathscr{O}_n)Z_n^*b\\
&= \frac{a'\varphi_n'(\theta)(\hat{Z}_n^* - Z_n^*)b}{n} - \frac{a'\varphi_n'(\theta)\hat{X}_n^*}{n}\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}(\hat{Z}_n^* - Z_n^*)b}{n} - \frac{a'\varphi_n'(\theta)(\hat{X}_n^* - X_n^*)}{n}\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}Z_n^*b}{n}\\
&\quad - \frac{a'\varphi_n'(\theta)X_n^*}{n}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\frac{(\hat{X}_n^* - X_n^*)'Z_n^*b}{n} - \frac{a'\varphi_n'(\theta)X_n^*}{n}\left\{\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1} - \left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\right\}\frac{\hat{X}_n^{*\prime}Z_n^*b}{n}\\
&= \mathscr{I}_1 - \mathscr{I}_2 - \mathscr{I}_3 - \mathscr{I}_4 - \mathscr{I}_5.
\end{aligned}
$$

For any $m \times n$ matrix $A$, denote $\|A\|_o = [\mu_{\max}(A'A)]^{1/2}$ as the operator norm, where $\mu_{\max}$ is the largest eigenvalue of $A$. Note that $\|A\|_o \le \|A\|$ by the spectral radius theorem. The desired result $\frac{1}{n}[a'\varphi_n'(\theta)(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)b - a'\varphi_n'(\theta)(\mathscr{O}_n^{\perp}Z_n^*)b] = o_p(1)$ follows by triangular inequality if we show

$$
|\mathscr{I}_1|, |\mathscr{I}_2|, |\mathscr{I}_3|, |\mathscr{I}_4|, |\mathscr{I}_5| = o_p(1).
$$

For term $\mathscr{I}_1$,

$$
\begin{aligned}
|\mathscr{I}_1| &\le \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}}\cdot\left[b'(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)b\right]^{\frac{1}{2}}}{n} \le \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}}\cdot(b'b)^{\frac{1}{2}}\cdot\left[\mu_{max}(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)\right]^{\frac{1}{2}}}{n}\\
&= \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}}\cdot(b'b)^{\frac{1}{2}}\cdot\left\|\hat{Z}_n^* - Z_n^*\right\|_o}{n} \le \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}}\cdot(b'b)^{\frac{1}{2}}}{\sqrt{n}}\cdot\left\|\frac{\hat{Z}_n^* - Z_n^*}{\sqrt{n}}\right\| = O_p(1)o_p(1) = o_p(1),
\end{aligned}
$$

where the first inequality is from the Cauchy-Schwarz inequality, the second inequality holds as $(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)$ is non-negative definite, the last inequality holds because $\|A\|_o \le \|A\|$, the penultimate equality holds as $\frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]\cdot(b'b)}{n} = O_p(1)$ and by Lemma 1(ii).

For term $\mathscr{I}_2$,

$$
\begin{aligned}
|\mathscr{I}_2| &\le \frac{\left[a'\varphi_n'(\theta)\hat{X}_n^*\hat{X}_n^{*\prime}\varphi_n(\theta)a\right]^{\frac{1}{2}}}{n}\left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o\frac{\left[b'(\hat{Z}_n^* - Z_n^*)'\hat{X}_n^*\hat{X}_n^{*\prime}(\hat{Z}_n^* - Z_n^*)b\right]^{\frac{1}{2}}}{n}\\
&\le \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}}\,\|\hat{X}_n^*\|_o}{n}\left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o\frac{(b'b)^{\frac{1}{2}}\|\hat{X}_n^*\|_o\|\hat{Z}_n^* - Z_n^*\|_o}{n}\\
&\le \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}}\cdot(b'b)^{\frac{1}{2}}}{\sqrt{n}}\cdot\left\|\frac{\hat{X}_n^*}{\sqrt{n}}\right\|^2\left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o\cdot\left\|\frac{\hat{Z}_n^* - Z_n^*}{\sqrt{n}}\right\| = O_p(1)\tilde{C}_0\hat{\wp}_0^2(k)O_p(1)o_p(1) = o_p(1),
\end{aligned}
$$

where the first and second inequalities hold by Cauchy-Schwarz inequality and that $\hat{X}_n^*\hat{X}_n^{*\prime}$ and $(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)$ are non-negative definite, the third inequality holds because $\|A\|_o \le \|A\|$, the penultimate equality holds by Lemma 1. The desired results of term $|\mathscr{I}_3|$ and term $|\mathscr{I}_4|$ follow by similar argument used for term $|\mathscr{I}_2|$.

For term $\mathscr{I}_5$, note that

$$
\begin{aligned}
\mathscr{I}_5 &= \frac{a'\varphi_n'(\theta)X_n^*}{n}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\left\{\left(\frac{X_n^{*\prime}X_n^*}{n}\right) - \left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)\right\}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}Z_n^*b}{n}\\
&= -\frac{a'\varphi_n'(\theta)X_n^*}{n}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\left(\frac{X_n^{*\prime}(\hat{X}_n^* - X_n^*)}{n}\right)\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}Z_n^*b}{n}\\
&\quad - \frac{a'\varphi_n'(\theta)X_n^*}{n}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\left(\frac{(\hat{X}_n^* - X_n^*)'\hat{X}_n^*}{n}\right)\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}Z_n^*b}{n}
\end{aligned}
$$

Then,

$$
\begin{aligned}
|\mathscr{I}_5| &\leq O_p(1)[\tilde{C}_0\wp_0^2(k)]O_p(1)o_p(1)O_p(1)[\tilde{C}_0\hat{\wp}_0^2(k)]^{\frac{1}{2}}[\tilde{C}_0\wp_0^2(p)]^{\frac{1}{2}} \\
&\quad + O_p(1)[\tilde{C}_0\wp_0^2(k)]^{\frac{1}{2}}O_p(1)o_p(1)O_p(1)[\tilde{C}_0\hat{\wp}_0^2(k)][\tilde{C}_0\wp_0^2(p)]^{\frac{1}{2}} \\
&= o_p(1).
\end{aligned}
$$

This finishes the proof.

**Proof of Lemma 2.** $\frac{1}{n}\left[\ln L_n\left(\theta_{ML}\right) - \ln L_{n0}\left(\theta_{ML}\right)\right]$ has three main components. The first component $\mathscr{A}_1$ is

$$
\begin{aligned}
\mathscr{A}_1 &= -\frac{1}{n}\Big([V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi]'[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi] - [V_n(\omega) - (\mathscr{O}_n^{\perp}Z_n^*)\chi]'[V_n(\omega) - (\mathscr{O}_n^{\perp}Z_n^*)\chi]\Big) \\
&= \frac{2}{n}V_n'(\omega)(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)\chi - \frac{1}{n}\Big([(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi]'[(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi] - [(\mathscr{O}_n^{\perp}Z_n^*)\chi]'[(\mathscr{O}_n^{\perp}Z_n^*)\chi]\Big) \\
&= \frac{2}{n}V_n'(\omega)(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)\chi - \frac{1}{n}\chi'(\hat{Z}_n^{*\prime}\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - Z_n^{*\prime}\mathscr{O}_n^{\perp}Z_n^*)\chi \\
&= \frac{2}{n}V_n'(\omega)(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)\chi - \frac{1}{n}\chi'(\hat{Z}_n^* - Z_n^*)'\hat{Z}_n^*\chi - \frac{1}{n}\chi'Z_n^{*\prime}(\hat{Z}_n^* - Z_n^*)\chi \\
&\quad + \frac{1}{n}\chi'(\hat{Z}_n^* - Z_n^*)'\hat{\mathscr{O}}_n\hat{Z}_n^*\chi + \frac{1}{n}\chi'Z_n^{*\prime}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)\chi \\
&= \mathscr{A}_{11} - \mathscr{A}_{12} - \mathscr{A}_{13} + \mathscr{A}_{14} + \mathscr{A}_{15}
\end{aligned}
$$

where $V_n(\omega) = (\lambda_0 - \lambda)G_nX_n\beta_0 + X_n(\beta_0 - \beta) + S_n(\lambda)S_n^{-1}V_n$ for endogenous spatial weights specification. $|\mathscr{A}_{11}|, |\mathscr{A}_{15}| = o_p(1)$ by Proposition 2. For term $\mathscr{A}_{12}$,

$$
|\mathscr{A}_{12}| \leq \frac{[\chi'(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)\chi]^{\frac{1}{2}} \cdot [\chi'\hat{Z}_n^{*\prime}\hat{Z}_n^*\chi]^{\frac{1}{2}}}{n} \leq (\chi'\chi)\left\|\frac{\hat{Z}_n^* - Z_n^*}{\sqrt{n}}\right\|\left\|\frac{\hat{Z}_n^*}{\sqrt{n}}\right\| = O_p(1)o_p(1)[\tilde{C}_0\hat{\wp}_0^2(p)]^{\frac{1}{2}} = o_p(1),
$$

The first inequality is from the Cauchy-Schwarz inequality, the second inequality holds because $(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)$ and $\hat{Z}_n^{*\prime}\hat{Z}_n^*$ are non-negative definite and the spectral radius theorem, the second to last equality is by Lemma 1 and the compact parameter space assumption (Assumption 4(*ii*)). The proof of $|\mathscr{A}_{12}| = o_p(1)$ follow by similar fashion. For term $\mathscr{A}_{14}$, note that

$$
\mathscr{A}_{14} = \frac{\chi'(\hat{Z}_n^* - Z_n^*)'\hat{X}_n^*}{n}\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}\hat{Z}_n^*\chi}{n},
$$

Then,

$$
|\mathscr{A}_{14}| \leq O_p(1)o_p(1)\tilde{C}_0\hat{\wp}_0^2(k)O_p(1)[\tilde{C}_0\hat{\wp}_0^2(p)]^{\frac{1}{2}} = o_p(1).
$$

As a result, we have $|\mathscr{A}_1| = o_p(1)$ by the triangular inequality.

The second component $\mathscr{A}_2$ is

$$
\begin{aligned}
\mathscr{A}_2 &= -\frac{1}{n}\left(\sum_{i=1}^n \hat{x}_i^{*\prime}(P_x^{-1} - I_k)\hat{x}_i^* - \sum_{i=1}^n x_i^{*\prime}(P_x^{-1} - I_k)x_i^*\right) - \frac{1}{n}\left(\sum_{i=1}^n \hat{z}_i^{*\prime}(\Xi^{-1} - I_p)\hat{z}_i^* - \sum_{i=1}^n z_i^{*\prime}(\Xi^{-1} - I_p)z_i^*\right) \\
&= -\mathscr{A}_{21} - \mathscr{A}_{22}
\end{aligned}
$$

For term $\mathscr{A}_{21}$, note that

$$
\begin{aligned}
\mathscr{A}_{21} &= \frac{1}{n}\sum_{i=1}^n(\hat{x}_i^* - x_i^*)'P_x^{-1}\hat{x}_i^* + \frac{1}{n}\sum_{i=1}^n x_i^{*\prime}P_x^{-1}(\hat{x}_i^* - x_i^*) - \frac{1}{n}\sum_{i=1}^n(\hat{x}_i^* - x_i^*)'\hat{x}_i^* - \frac{1}{n}\sum_{i=1}^n x_i^{*\prime}(\hat{x}_i^* - x_i^*) \\
&= \mathscr{A}_{211} + \mathscr{A}_{212} - \mathscr{A}_{213} - \mathscr{A}_{214}
\end{aligned}
$$

and that

$$
\begin{aligned}
|\mathscr{A}_{211}| &\leq \frac{1}{n}\sum_{i=1}^{n}\left[(\hat{x}_i^*-x_i^*)'(\hat{x}_i^*-x_i^*)\right]^{\frac{1}{2}}\left[\hat{x}_i^{*\prime}P_x^{-1}\hat{x}_i^*\right]^{\frac{1}{2}} \leq \frac{1}{n}\sum_{i=1}^{n}\left[(\hat{x}_i^*-x_i^*)'(\hat{x}_i^*-x_i^*)\right]^{\frac{1}{2}}\left[\hat{x}_i^{*\prime}\hat{x}_i^*\right]^{\frac{1}{2}}\|P_x^{-1}\|_o \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{\tau=1}^{k}\sup_{x_{\tau,i}}|\hat{x}_{\tau,i}^*-x_{\tau,i}^*|\sup_{x_{\tau,i}}|x_{\tau,i}^*|\cdot\|P_x^{-1}\|_o \leq \frac{1}{n}\sum_{i=1}^{n}\sum_{\tau=1}^{k}\sqrt{\frac{\log\log n}{n^{1-a_\tau}}}\left(\sqrt{2a_\tau\log n}+4\sqrt{\frac{\log\log n}{n^{1-a_\tau}}}\right)\|P_x^{-1}\|_o \\
&= o_p(1)O_p(1)O_p(1) = o_p(1),
\end{aligned}
$$

where the first inequality comes from the Cauchy-Schwarz inequality, the second inequality holds because $P_x^{-1}$ is non-negative definite, the fourth inequality is by Proposition 1 and the proof of Lemma A.4 in Liu et al. (2012), the second to last equality is due to the compact parameter space assumption. Similarly, we can show $|\mathscr{A}_{212}|, |\mathscr{A}_{213}|, |\mathscr{A}_{214}| = o_p(1)$. Then $|\mathscr{A}_{21}| = o_p(1)$ by the triangular inequality. The desired result of term $|\mathscr{A}_{22}| = o_p(1)$ follow by similar argument used for term $|\mathscr{A}_{21}|$. As a result, $|\mathscr{A}_2| = o_p(1)$.

The third component $|\mathscr{A}_3|$ is

$$
\begin{aligned}
\mathscr{A}_3 &= \frac{2}{n}\left(\sum_{i=1}^{n}\hat{z}_i^{*\prime}\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) - \sum_{i=1}^{n}z_i^{*\prime}\Xi^{-1}(\Gamma'x_i^*)\right) - \frac{1}{n}\left((\hat{\Gamma}'\hat{x}_i^*)'\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) - (\Gamma'x_i^*)'\Xi^{-1}(\Gamma'x_i^*)\right) \\
&= \frac{2}{n}\sum_{i=1}^{n}(\hat{z}_i^*-z_i^*)'\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) + \frac{2}{n}\sum_{i=1}^{n}z_i^{*\prime}\Xi^{-1}[(\hat{\Gamma}-\Gamma)'\hat{x}_i^*] + \frac{2}{n}\sum_{i=1}^{n}z_i^{*\prime}\Xi^{-1}[\Gamma'(\hat{x}_i^*-x_i^*)] \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}[(\hat{\Gamma}-\Gamma)'\hat{x}_i^*]'\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) - \frac{1}{n}\sum_{i=1}^{n}[\Gamma'(\hat{x}_i^*-x_i^*)]'\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) - \frac{1}{n}\sum_{i=1}^{n}(\Gamma'x_i^*)'\Xi^{-1}[(\hat{\Gamma}-\Gamma)'\hat{x}_i^*] \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}(\Gamma'x_i^*)'\Xi^{-1}[\Gamma'(\hat{x}_i^*-x_i^*)]
\end{aligned}.
$$

The three key terms are the ones associated with $\hat{\Gamma}-\Gamma$, the results that other terms are all $o_p(1)$ follow by similar argument for term $|\mathscr{A}_{211}|$. We show the result for the second term, by the second stage estimation,

$$
\begin{aligned}
\|\hat{\Gamma}-\Gamma\|_o &\leq \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o\left\|\frac{\hat{X}_n^{*\prime}U_n}{n}\right\| + \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o\left\|\frac{\hat{X}_n^*}{\sqrt{n}}\right\|\left\|\frac{\hat{Z}_n^*-Z_n^*}{\sqrt{n}}\right\| \\
&\quad + \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o\left\|\frac{\hat{X}_n^*}{\sqrt{n}}\right\|\left\|\frac{(\hat{X}_n^*-X_n^*)\Gamma}{\sqrt{n}}\right\| \\
&\leq O_p(1)o_p(1) + O_p(1)\left[\tilde{C}_0\wp_0^2(k)\right]^{\frac{1}{2}}o_p(1) + O_p(1)\left[\tilde{C}_0\wp_0^2(k)\right]^{\frac{1}{2}}o_p(1) = o_p(1),
\end{aligned}
$$

therefore,

$$
\begin{aligned}
\left|\frac{2}{n}\sum_{i=1}^{n}z_i^{*\prime}\Xi^{-1}[(\hat{\Gamma}-\Gamma)'\hat{x}_i^*]\right| &\leq \frac{2}{n}\sum_{i=1}^{n}(z_i^{*\prime}\Xi'^{-1}\Xi^{-1}z_i^*)^{\frac{1}{2}}\left(\hat{x}_i^{*\prime}(\hat{\Gamma}-\Gamma)(\hat{\Gamma}-\Gamma)'\hat{x}_i^*\right)^{\frac{1}{2}} \\
&\leq \frac{2}{n}\sum_{i=1}^{n}(z_i^{*\prime}z_i^*)^{\frac{1}{2}}(\hat{x}_i^{*\prime}\hat{x}_i^*)^{\frac{1}{2}}\|\Xi^{-1}\|_o\|\hat{\Gamma}-\Gamma\|_o \\
&= O_p(1)O_p(1)O_p(1)o_p(1) = o_p(1),
\end{aligned}
$$

then $|\mathscr{A}_3| = o_p(1)$ and we have that

$$
\frac{1}{n}\left[\ln L_n(\theta_{ML}) - \ln L_{n0}(\theta_{ML})\right] = o_p(1).
$$

The required results for the differences in the first and second order derivatives, i.e., $\frac{1}{n}\left[\frac{\partial \ln L_n(\theta_{ML})}{\partial \theta_l} - \frac{\partial \ln L_{n0}(\theta_{ML})}{\partial \theta_l}\right] = o_p(1)$ and $\frac{1}{n}\left[\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \theta_{l_1}\partial \theta_{l_2}} - \frac{\partial^2 \ln L_{n0}(\theta_{ML})}{\partial \theta_{l_1}\partial \theta_{l_2}}\right] = o_p(1)$, follow by similar fashion, thus we omit the proof.

**Proof of Lemma 3.** For MLE with a finite sample, identification is equivalent to $P(\ln L_n(\theta_{ML,0}) \neq \ln L_n(\theta_{ML})) > 0$ for any $\theta_{ML,0} \neq \theta_{ML}$ (Rothenberg, 1971). Let $\theta_{ML,0} = \left(\omega_0', \sigma_{\xi,0}^2, \chi_0', \alpha_0', \delta_0'\right)'$ be the true parameter vector, and $\theta_{ML} = \left(\omega', \sigma_\xi^2, \chi', \alpha', \delta'\right)'$ be arbitrary parameter values in $\Theta$ defined in Assumption 4(*ii*). Since $\ln x \leq x-1$ for any $x \geq 0$, which implies $\ln x \leq 2\sqrt{x}-2$ for any $x \geq 0$, we have

$$
\begin{aligned}
&\frac{1}{n}\mathrm{E}\ln[L_n(\theta_{ML})/L_n(\theta_{ML,0})] \\
&\quad\leq \frac{2}{n}\mathrm{E}(\sqrt{L_n(\theta_{ML})/L_n\left(\theta_{ML,0}\right)} - 1) = \frac{2}{n}\int \left(\sqrt{L_n(\theta_{ML})/L_n\left(\theta_{LM,0}\right)} - 1\right) L_n\left(\theta_{ML,0}\right) d\mathscr{U}_n \\
&\quad= \frac{2}{n}\int \sqrt{L_n(\theta_{ML})L_n\left(\theta_{ML,0}\right)}d\mathscr{U}_n - \frac{1}{n} = -\frac{1}{n}\int \left[\sqrt{L_n(\theta_{ML})} - \sqrt{L_n\left(\theta_{ML,0}\right)}\right]^2 d\mathscr{U}_n \leq 0
\end{aligned}
$$

where $\mathscr{U}_n = (Y_n', \mathrm{vec}(Z_n)', \mathrm{vec}(X_n)')'$. This implies in particular the information inequality that $\frac{1}{n}\mathrm{E}\ln L_n(\theta_{ML}) \leq \frac{1}{n}\mathrm{E}\ln L_n\left(\theta_{ML,0}\right)$ for all $\theta$. Thus $\theta_{ML,0}$ is a maximizer. Also, this inequality implies that if $\frac{1}{n}\mathrm{E}\ln L_n(\theta_{ML}) = \frac{1}{n}\mathrm{E}\ln L_n(\theta_{ML,0})$, we have $\frac{1}{n}\ln L_n(\theta_{ML}) = \frac{1}{n}\ln L_n(\theta_{ML,0})$ almost surely. By Lemma 2, this is equivalent to $\frac{1}{n}\ln L_{n0}(\theta_{ML}) + o_p(1) = \frac{1}{n}\ln L_{n0}(\theta_{ML,0}) + o_p(1)$ almost surely, i.e.,

$$
\begin{aligned}
&-\frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|I_n - \lambda W_n| - \frac{1}{2}\sum_{i=1}^n \left(z_i^* - \Gamma' x_i^*\right)' \Xi^{-1}\left(z_i^* - \Gamma' x_i^*\right) - \frac{1}{2}\sum_{i=1}^n x_i^{*\prime} P_x^{-1} x_i^* \\
&- \frac{1}{2\sigma_\xi^2}[(I_n - \lambda W_n)Y_n - X_n\beta - \left(\mathscr{O}_n^\perp Z_n^*\right)\chi]'[(I_n - \lambda W_n)Y_n - X_n\beta - \left(\mathscr{O}_n^\perp Z_n^*\right)\chi] \\
= &-\frac{n}{2}\ln\sigma_{\xi,0}^2 - \frac{n}{2}\ln|P_{x,0}| - \frac{n}{2}\ln|\Xi_0| + \ln|I_n - \lambda_0 W_n| - \frac{1}{2}\sum_{i=1}^n \left(z_i^* - \Gamma_0' x_i^*\right)' \Xi_0^{-1}\left(z_i^* - \Gamma_0' x_i^*\right) - \frac{1}{2}\sum_{i=1}^n x_i^{*\prime} P_{x,0}^{-1} x_i^* \\
&- \frac{1}{2\sigma_{\xi,0}^2}[(I_n - \lambda_0 W_n)Y_n - X_n\beta_0 - \left(\mathscr{O}_n^\perp Z_n^*\right)\chi_0]'[(I_n - \lambda_0 W_n)Y_n - X_n\beta_0 - \left(\mathscr{O}_n^\perp Z_n^*\right)\chi_0]
\end{aligned} \tag{B.1}
$$

holds for $Y_n$, $Z_n$, and $X_n$ almost surely. Differentiate (B.1) with respect to $y_j$, we have

$$
\begin{aligned}
&\sigma_\xi^{-2}\left\{y_j - \lambda w_{j\cdot}Y_n - x_j\beta - (\mathscr{O}_n^\perp Z_n^*)_j\chi - \lambda\sum_{i=1}^n\left[y_i - \lambda w_{i\cdot}Y_n - x_i\beta - \left(\mathscr{O}_n^\perp Z_n^*\right)_i\chi\right]w_{ij}\right\} \\
&= \sigma_{\xi,0}^{-2}\left\{y_j - \lambda_0 w_{j\cdot}Y_n - x_j\beta_0 - (\mathscr{O}_n^\perp Z_n^*)_j\chi_0 - \lambda_0\sum_{i=1}^n\left[y_i - \lambda_0 w_{i\cdot}Y_n - x_i\beta_0 - \left(\mathscr{O}_n^\perp Z_n^*\right)_i\chi_0\right]w_{ij}\right\}.
\end{aligned} \tag{B.2}
$$

Differentiate (B.2) with respect to $y_j$ once more,

$$
\sigma_\xi^{-2}(1 - \lambda w_{jj} + \lambda^2\sum_{i=1}^n w_{ij}^2) = \sigma_{\xi,0}^{-2}(1 - \lambda_0 w_{jj} + \lambda_0^2\sum_{i=1}^n w_{ij}^2),
$$

since $w_{jj} = 0$ and there exists $j \neq j'$ such that $\sum_{i=1}^n w_{ij}^2 \neq \sum_{i=1}^n w_{ij'}^2$, we have that $1/\sigma_\xi^2 = 1/\sigma_{\xi,0}^2$ and $\lambda^2/\sigma_\xi^2 = \lambda_0^2/\sigma_{\xi,0}^2$. Hence, $\sigma_\xi = \sigma_{\xi,0}$ and $|\lambda| = |\lambda_0|$. Differentiate (B.2) with respect to $y_k (k \neq j)$,

$$
\sigma_\xi^{-2}(\lambda^2\sum_{i=1}^n w_{ik}w_{ij} - \lambda(w_{kj} + w_{jk})) = \sigma_{\xi,0}^{-2}(\lambda_0^2\sum_{i=1}^n w_{ik}w_{ij} - \lambda_0(w_{kj} + w_{jk})),
$$

Thus, $\lambda(w_{kj}+w_{jk}) = \lambda_0(w_{kj}+w_{jk})$. Because $W_n + W_n' \neq 0$ and $w_{ii} = 0$, we have $\lambda = \lambda_0$. Eq.(B.2) implies that

$$(x_j\beta - \lambda\sum_{i=1}^{n} w_{ij}x_i\beta) + [(\mathscr{O}_n^{\perp}Z_n^*)_j\chi - \lambda\sum_{i=1} w_{ij}\left(\mathscr{O}_n^{\perp}Z_n^*\right)_i\chi]$$
$$= (x_j\beta_0 - \lambda_0\sum_{i=1}^{n} w_{ij}x_i\beta_0) + [(\mathscr{O}_n^{\perp}Z_n^*)_j\chi_0 - \lambda_0\sum_{i=1} w_{ij}\left(\mathscr{O}_n^{\perp}Z_n^*\right)_i\chi_0],$$

which is equivalent to $(I_n-\lambda_0W_n')X_n\beta+(I_n-\lambda_0W_n')(\mathscr{O}_n^{\perp}Z_n^*)\chi = (I_n-\lambda_0W_n')X_n\beta_0+(I_n-\lambda_0W_n')(\mathscr{O}_n^{\perp}Z_n^*)\chi_0$. As $(I_n-\lambda_0W_n')$ is invertible, $X_n$ and $\mathscr{O}_n^{\perp}Z_n^*$ are not linearly independent, it must be that $X_n\beta = X_n\beta_0$ and $(\mathscr{O}_n^{\perp}Z_n^*)\chi = (\mathscr{O}_n^{\perp}Z_n^*)\chi_0$. Therefore, $\beta = \beta_0$ and $\chi = \chi_0$. With these results, Eq.(B.1) implies that

$$-\frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| - \frac{1}{2}\sum_{i=1}^{n}(z_i^* - \Gamma'x_i^*)'\Xi^{-1}(z_i^* - \Gamma'x_i^*) - \frac{1}{2}\sum_{i=1}^{n} x_i^{*\prime}P_x^{-1}x_i^*$$
$$= -\frac{n}{2}\ln|P_{x,0}| - \frac{n}{2}\ln|\Xi_0| - \frac{1}{2}\sum_{i=1}^{n}(z_i^* - \Gamma_0'x_i^*)'\Xi_0^{-1}(z_i^* - \Gamma_0'x_i^*) - \frac{1}{2}\sum_{i=1}^{n} x_i^{*\prime}P_{x,0}^{-1}x_i^*,$$

which only involves $z_j^*$ and $x_j^*$ for $j = 1, ..., n$ that are strictly increasing functions of $z_j$ and $x_j$. First, by differentiating the above equation with respect to $z_j^*$, we have

$$\left(z_j^* - \Gamma'x_j^*\right)'\Xi^{-1} = \left(z_j^* - \Gamma_0'x_j^*\right)'\Xi_0^{-1}.$$

Differentiate the above equation once more with respect to $z_j^*$, we can get $\Xi^{-1} = \Xi_0^{-1}$. By the uniqueness of a matrix inverse, $\Xi = \Xi_0$. Then we have $(\Gamma - \Gamma_0)'x_j^*\Xi_0^{-1} = 0$, which indicates that $\Gamma = \Gamma_0$ as $x_j^*$ is stochastic.[45] Second, given $\Xi = \Xi_0$ and $\Gamma = \Gamma_0$, differentiating the above equation with respect to $x_i^*$ twice yields $P_x^{-1} = P_{x,0}^{-1}$ , by the uniqueness of a matrix inverse, $P_x = P_{x,0}$. Thus, we have $\theta_{ML} = \theta_{ML,0}$, $\theta_{ML,0}$ is the unique maximizer of $\ln L_n(\theta_{ML,0})$.

**Proof of Theorem 1.** For a SAR model with endogenous $W_n$, all terms in the log pseudo-likelihood function in Appendix A.3.1 can be written in the general formula

$$(\theta_{ML,0} - \theta_{ML})'\frac{1}{n}a\boldsymbol{\varphi}_n^{*\prime}\mathcal{M}_n\boldsymbol{\varphi}_n^* b(\theta_{ML,0} - \theta_{ML}),$$

where $a$, $b$ are some constant vectors, $\boldsymbol{\varphi}_n^* = (\varphi_1^*, ..., \varphi_n^*)'$ with $\varphi_i^* = f_i(v_i, X_n, Z_n, \theta_{ML,0})$ being a vector value function, $\mathcal{M}_n = A_n'B_n$ are either $W_n^{m_1}$ or $G_n^{m_2}$ with $m_1$ and $m_2$ being finite non-negative integers. Therefore, the proofs of consistency and asymptotic normality follow by similar arguments for the proof of Theorem 3 in Qu and Lee (2015) as direct applications of Proposition 1, Corollary 1 and Proposition 2 in their paper.

**Proof of Theorem 2.** We can express

$$\hat{\theta}_{IV} - \theta_{IV,0} = [\hat{M}_n'\hat{T}_n(\hat{T}_n'\hat{T}_n)^{-1}\hat{T}_n'\hat{M}_n]^{-1}\hat{M}_n'\hat{T}_n(\hat{T}_n'\hat{T}_n)^{-1}\hat{T}_n'\boldsymbol{\epsilon}_n$$
$$= [\frac{\hat{M}_n'\hat{T}_n}{n}(\frac{\hat{T}_n'\hat{T}_n}{n})^{-1}\frac{\hat{T}_n'\hat{M}_n}{n}]^{-1}\frac{\hat{M}_n'\hat{T}_n}{n}(\frac{\hat{T}_n'\hat{T}_n}{n})^{-1}\Big(\frac{1}{n}\hat{T}_n'\boldsymbol{\epsilon}_n + \frac{1}{n}\hat{T}_n'(\hat{\mathscr{O}}_nU_n\gamma_0)$$
$$+\frac{1}{n}\hat{T}_n'[\hat{\mathscr{O}}_n^{\perp}(Z_n^* - \hat{Z}_n^*)\gamma_0] - \frac{1}{n}\hat{T}_n'[\hat{\mathscr{O}}_n^{\perp}(X_n^* - \hat{X}_n^*)\Gamma_0\gamma_0]\Big).$$

45. $\Gamma$ can also be identified from the second stage estimation, $\|\Gamma - \Gamma_0\|_o = o_p(1)$ as shown in the proof of Lemma 2.

Note that $\hat{M}_n = M_n + \left(0, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*\right)$ and $\hat{T}_n = T_n + \left(0, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*\right)$, the asymptotic analysis relies on terms of

$$\begin{aligned}&\frac{1}{n}T_n'\boldsymbol{\epsilon}_n,\ \frac{1}{n}T_n'(\hat{\mathscr{O}}_nU_n\gamma_0),\ \frac{1}{n}T_n'[\hat{\mathscr{O}}_n^{\perp}(Z_n^*-\hat{Z}_n^*)\gamma_0],\ \frac{1}{n}T_n'[\hat{\mathscr{O}}_n^{\perp}(X_n^*-\hat{X}_n^*)\Gamma_0\gamma_0],\ \frac{1}{n}T_n'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*),\\ &\frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)'\boldsymbol{\epsilon}_n,\ \frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)'(\hat{\mathscr{O}}_nU_n\gamma_0),\ \frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*),\\ &\frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)'[\hat{\mathscr{O}}_n^{\perp}(Z_n^*-\hat{Z}_n^*)\gamma_0],\ \frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)'[\hat{\mathscr{O}}_n^{\perp}(X_n^*-\hat{X}_n^*)\Gamma_0\gamma_0].\end{aligned} \tag{B.3}$$

For term $\frac{1}{n}T_n'\boldsymbol{\epsilon}_n$, denote $t_i = (q_i', x_i', \mathbb{O}_i')$ with $\mathbb{O}_i = \sum_{j=1}^n \hat{\mathscr{O}}_{ij}^{\perp}\hat{z}_j^{*\prime}$, the key is to consider $q_i = \sum_{j=1}^n w_{ij}x_j$ because other terms in $t_i$ satisfy required moment properties by assumption. For all $i$, $\sum_{j\neq i} c_1 d_{ij}^{-c_0 d_0} < \infty$ by Claim C.1.1 in Qu and Lee (2015), we have

$$\begin{aligned}\mathrm{E}(|q_{\iota i}q_{\iota i}'|) &= \mathrm{E}(|\sum_{j_1=1}^n w_{ij_1}x_{j_1}\sum_{j_2=1}^n w_{ij_2}x_{j_2}|) \leq \mathrm{E}(\sum_{j_1=1}^n\sum_{j_2=1}^n |w_{ij_1}||w_{ij_2}||x_{j_1}x_{j_2}'|)\\ &\leq \sum_{j_1\neq i} c_1 d_{ij_1}^{-c_0d_0}\sum_{j_2\neq i} c_1 d_{ij_2}^{-c_0d_0}\mathrm{sup}_{n.j_1,n.j_2}\mathrm{E}|x_{j_1}x_{j_2}'| < \infty,\end{aligned}$$

then $\mathrm{Var}\left(\frac{1}{n}T_n'\boldsymbol{\epsilon}_n\right) = \frac{\eta_0^2}{n^2}\mathrm{E}(T_n'T_n) \leq \frac{\eta_0^2}{n}\mathrm{sup}_{n.i}\mathrm{E}(t_it_i') = O(\frac{1}{n})$. Therefore, $\frac{1}{n}T_n'\boldsymbol{\epsilon}_n = O_p(\frac{1}{\sqrt{n}})$. For term $\frac{1}{n}T_n'(\hat{\mathscr{O}}_nU_n\gamma_0)$, as

$$\|\frac{1}{n}T_n'(\hat{\mathscr{O}}_nU_n\gamma_0)\|_o \leq \left\|\frac{T_n}{\sqrt{n}}\right\|\left\|\frac{\hat{X}_n^*}{\sqrt{n}}\right\|\left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o\left\|\frac{\hat{X}_n^{*\prime}U_n\gamma_0}{n}\right\| \leq O_p(1)\left[\tilde{C}_0\wp_0^2(k)\right]^{\frac{1}{2}}O_p(1)o_p(1) = o_p(1),$$

then $\frac{1}{n}T_n'(\hat{\mathscr{O}}_nU_n\gamma_0) = o_p(1)$. By the result of Lemma 1 and the proof in Proposition 2, it can be easily shown that all the remaining terms related to $(\hat{X}_n^* - X_n^*)$, $(\hat{Z}_n^* - Z_n^*)$, and $(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)$ in (B.3) are $o_p(1)$. As a result, the first and second stage estimation errors will not affect the asymptotic analysis of the 3rd stage, and we can treat $\hat{X}_n^*$, $\hat{Z}_n^*$ and $\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*$ as the true $X_n^*$, $Z_n^*$ and $\mathscr{O}_n^{\perp}Z_n^*$. Under Assumptions 1-5, 7(*i*)-7(*ii*) and 8, the asymptotic properties can be obtained by applying Proposition 1 in Qu and Lee (2015). Besides, we can apply a related CLT (Proposition 2 in Qu and Lee (2015) to get the results that $\frac{1}{\sqrt{n}}T_n'\boldsymbol{\epsilon}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n t_i'\epsilon_i \xrightarrow{d} \mathrm{N}(0, \eta_0^2T_n'T_n)$ and that $\frac{1}{\sqrt{n}}T_n'(\hat{\mathscr{O}}_nU_n\gamma_0) \xrightarrow{d} \mathrm{N}\left(0, T_n'(\gamma_0'\Xi_0, \gamma_0\hat{\mathscr{O}}_n)T_n\right)$, where $\Xi_0 = P_{z,0} - P_{xz,0}'P_{x,0}^{-1}P_{xz,0}$, then we have the desired asymptotic distribution.

**Proof of Lemma 4.** By Eq.(22), denote $G_n = S_n^{-1}W_n$ with $S_n = I_n - \lambda_0 W_n$, we have

$$\begin{aligned}\tilde{\boldsymbol{\epsilon}}_n &= S_n(\hat{\lambda})Y_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma} = S_n(\hat{\lambda})S_n^{-1}(X_n\beta_0 + \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\gamma_0 + \hat{\boldsymbol{\epsilon}}_n)\\ &= (\lambda_0 - \hat{\lambda})G_n(X_n\beta_0 + \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\gamma_0) + X_n(\beta_0 - \hat{\beta}) + \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*(\gamma_0 - \hat{\gamma}) + (\lambda_0 - \hat{\lambda})G_n\hat{\boldsymbol{\epsilon}}_n + \hat{\boldsymbol{\epsilon}}_n,\end{aligned}$$

where the composite error $\hat{\boldsymbol{\epsilon}}_n = \boldsymbol{\epsilon}_n + \hat{\mathscr{O}}_nU_n\gamma_0 + \hat{\mathscr{O}}_n^{\perp}(Z_n^* - \hat{Z}_n^*)\gamma_0 - \hat{\mathscr{O}}_n^{\perp}(X_n^* - \hat{X}_n^*)\Gamma_0\gamma_0$. We can express $\frac{1}{n}\tilde{\boldsymbol{\epsilon}}_n'\tilde{\boldsymbol{\epsilon}}_n$ in the form below as in Proposition 1 in Qu and Lee (2015) so that

$$\begin{aligned}\frac{1}{n}\tilde{\boldsymbol{\epsilon}}_n'\tilde{\boldsymbol{\epsilon}}_n &= \left(\theta_{IV,0} - \hat{\theta}_{IV}\right)'\frac{1}{n}a_1\varphi_n^{*\prime}\mathcal{M}_n\boldsymbol{\varphi}_n^*b_1\left(\theta_{IV,0} - \hat{\theta}_{IV}\right) + \frac{1}{n}a_2\boldsymbol{\varphi}_n^{*\prime}\mathcal{M}_n\boldsymbol{\varphi}_n^*b_2\left(\theta_{IV,0} - \hat{\theta}_{IV}\right)\\ &\quad + \frac{1}{n}\hat{\boldsymbol{\epsilon}}_n'\hat{\boldsymbol{\epsilon}}_n \xrightarrow{p} \eta_0^2.\end{aligned}$$

Therefore, $\frac{1}{n}\tilde{\boldsymbol{\epsilon}}_n'\tilde{\boldsymbol{\epsilon}}_n \xrightarrow{p} \eta_0^2$.

To show $\hat{u}_{IV} \xrightarrow{p} u_{IV}$, the most complicated term to be analyzed is

$$\frac{1}{n}\left[a_3'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'G_n(\hat{\lambda})'G_n(\hat{\lambda})(\hat{\mathscr{O}}_n^{\perp}Z_n^*)b_3 - \mathrm{E}\left(a_3'(\mathscr{O}_n^{\perp}Z_n^*)'G_n'G_n(\mathscr{O}_n^{\perp}Z_n^*)b_3\right)\right] = o_p(1). \tag{B.4}$$

Similar to the proof in Proposition 2, it can be verified that $\frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*) = o_p(1)$, then we have

$$\frac{1}{n}[a_3'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'G_n(\hat{\lambda})'G_n(\hat{\lambda})(\hat{\mathscr{O}}_n^{\perp}Z_n^*)b_3 - \mathrm{E}(a_3'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'G_n(\hat{\lambda})'G_n(\hat{\lambda})(\hat{\mathscr{O}}_n^{\perp}Z_n^*)b_3)] = o_p(1)$$

from the ULLN (Corollary 1) in Qu and Lee (2015). The term in (B.4) holds from the equicontinuity of $\frac{1}{n}\mathrm{E}[a_3'U_n'(\Gamma)G_n(\lambda)'G_n(\lambda)U_n(\Gamma)b_3]$.

## References


Anselin, L. 1988. *Spatial Econometrics: Methods and Models.* The Netherlands: Kluwer Academic.

Balakrishnan, N., and C. Lai. 2009. *Continuous Bivariate Distributions.* New York: Springer.

Barbe, P., C. Genest, K. Ghoudi, and B. Rémillard. 1996. "On Kendall's Process." *Journal of Multivariate Analysis* 58:197–229.

Becker, J., D. Proksch, and C. Ringle. 2021. "Revisiting Gaussian Copulas to Handle Endogenous Regressors." *Journal of the Academy of Marketing Science* 50:1–21.

Case, A., H. Rosen, and J. R. Hines Jr. 1993. *Budget Spillovers and Fiscal Policy Interdepence: Evidence from the States. In: Giles, D., Ullah, A. (Eds.), Handbook of Applied Economics Statistics.* New York: Marcel Dekker.

Cliff, A., and J. Ord. 1973. *Spatial Autocorrelation.* London: Pion Ltd.

Conley, T., and E. Ligon. 2002. "Economic Distance and Cross-Country Spillovers." *Journal of Economic Growth* 7:157–187.

Danaher, P., and M. Smith. 2011. "Modeling Multivariate Distributions using Copulas: Applications in Marketing." *Marketing Science* 30:4–21.

Deheuvels, P. 1979. "La fonction de dépendance empirique et ses propriétés - Un test non paramétrique d'indépendance." *Académie Royale de Belgique - Bulletin de la Classe des Sciences - 5e Série* 65:274–292.

Diegert, P., M. Masten, and A. Poirier. 2022. "Assessing Omitted Variable Bias when the Controls are Endogenous." ArXiv:2206.02303v2, https://arxiv.org/abs/2206.02303, July.

Fang, Y., and M. Henry. 2023. "Vector copulas." *Journal of Econometrics* 234:128–150.

Genest, C., and L.-P. Rivest. 1993. "Statistical Inference Procedures for Bivariate Archimedean Copulas." *Journal of the American Statistical Association* 88:1034–1043.

Gennaioli, N., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer. 2013. "Human Capital and Regional Development." *The Quarterly Journal of Economics* 128:105–164.

Gupta, A., and P. Robinson. 2015. "Inference on Higher-order Spatial Autoregressive Models with Increasingly Many Parameters." *Journal of Econometrics* 186:19–31.

Han, F., T. Zhao, and H. Liu. 2013. "CODA: High Dimensional Copula Discriminant Analysis." *Journal of Machine Learning Research* 14:629–671.

Haschka, R. 2022. "Handling Endogenous Regressors Using Copulas: a Generalization to Linear Panel Models with Fixed Effects and Correlated Regressors." *Journal of Marketing Research* 59:860–881.

He, M., Y. Chen, and R. Schramm. 2018. "Technological Spillovers in Space and Firm Productivity: Evidence from China's Electric Apparatus Industry." *Urban Studies* 55:2522–2541.

Hoshino, T. 2022. "Sieve IV Estimation of Cross-sectional Interaction Models with Nonparametric Endogenous Effect." *Journal of Econometrics* 229:263–275.

Jenish, N., and I. Prucha. 2012. "On Spatial Processes and Asymptotic Inference under Near-epoch Dependence." *Journal of Econometrics* 170:178–190.

Johnsson, I., and H. Moon. 2021. "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach." *The Review of Economics and Statistics* 103:328–345.

Kelejian, H., and G. Piras. 2014. "Estimation of Spatial Models with Endogenous Weighting Matrices." *Regional Science and Urban Economics* 46:140–149.

Kelejian, H., and I. Prucha. 2001. "On the Asymptotic Distribution of the Moran I Test Statistic with Applications." *Journal of Econometrics* 104:219–257.

Lee, L. 2004. "Asymptitic Distributions of Quasi-maximum Likelihood Estimators for Spatial Econometric Models." *Econometrica* 72:1899–1925.

———. 2007. "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models." *Journal of Econometrics* 137:489–514.

Li, X., P. Mikusinski, H. Sherwood, and M. Taylor. 1997. *On Approximation of Copulas, Distributions with Given Marginals and Moment Problems.* 107–116. Dordrecht: Kluwer Academic Publishers.

Liu, H., F. Han, M. Yuan, J. Lafferty, and L. Wasserman. 2012. "High-dimensional Semiparametric Gaussian Copula Graphical Models." *The Annals of Statistics* 30:2293–2326.

Mazurek, J. 2012. "The Evaluation of An Economic Distance Among Countries: A Novel Approach." Working Paper, https://pep.vse.cz/pdfs/pep/2012/03/02.pdf.

Moretti, E. 2004. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* 94:656–690.

Nelsen, R. 2006. *An Introduction to Copulas.* New York: Springer.

Ord, K. 1975. "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association* 70:120–126.

Park, S., and S. Gupta. 2012. "Handling Endogenous Regressors by Joint Estimation Using Copulas." *Marketing Science* 31:567–586.

Qu, X., L. F. Lee, and C. Yang. 2021. "Estimation of a SAR Model with Endogenous Spatial Weights Constructed by Bilateral Variables." *Journal of Econometrics* 221:180–197.

Qu, X., and L.-F. Lee. 2015. "Estimating a Spatial Autoregressive model with an Endogenous Spatial Weight Matrix." *Journal of Econometrics* 184:209–232.

Qu, X., L.-F. Lee, and J. Yu. 2017. "QML Estimation of Spatial Dynamic Panel Data Models with Endogenous Time Varying Spatial Weights Matrices." *Journal of Econometrics* 197:173–201.

Rothenberg, T. 1971. "Identification in Parametric Models." *Econometrica* 39:577–591.

Sanso-Navarro, M., M. Vera-Cabello, and D. P. Ximénez-de-Embún. 2017. "Human Capital Spillovers and Regional Development." *Journal of Applied Econometrics* 32:923–930.

Sklar, A. 1959. "Fonctions de. répartition $n$ dimensions et leurs marges." *Publications de l′Institut de Statistique de L′Universit de Paris* 8:229–231.

Song, P. 2000. "Multivariate Dispersion Models Generated from Gaussian Copula." *Scandinavian Journal of Statistics* 27:305–320.

Stoyanov, A., and N. Zubanov. 2012. "Productivity Spillovers across Firms through Worker Mobility." *American Economic Journal: Applied Economics* 4:168–198.

Su, L., and S. Jin. 2010. "Profile Quasi-maximum Likelihood Estimation of Partially Linear Spatial Autoregressive Models." *Journal of Econometrics* 157:18–33.

Tsang, E., and P. Yip. 2007. "Economic Distance and The Survival of Foreign Direct Investments." *Academy of Management Journal* 50:1156–1168.

Xu, X., and L.-f. Lee. 2018. "Sieve Maximum Likelihood Estimation of the Spatial Autoregressive Tobit Model." *Journal of Econometrics* 203:96–112.

Yang, F., Y. Qian, and H. Xie. 2022. "Addressing Endogeneity Using a Two-stage Copula Generated Regressor Approach." NBER Working Paper 29708, https://www.nber.org/papers/w29708, July.