Multi-Dimensional Spatial Panel Data Models with Fixed Effects: Formulation, Estimation and Inference*

Alain Pirotte[a] and Zhenlin Yang[b]

[a] *CRED, University Paris-Panthéon-Assas, France*

[b] *School of Economics, Singapore Management University, Singapore*

December 8, 2025

Abstract

We consider formulation, estimation, and inference for multidimensional spatial panel data models with multidimensional fixed effects (FEs) that may appear in the model additively or interactively. Methods of formulating the FEs so that all FE parameters can be identified are given. To estimate the common parameters, a general M-estimation method is proposed in which the unbiased estimating equations are obtained by adjusting the concentrated quasi score function to remove the effects of estimating the *incidental* FE parameters. The resulting M-estimator is shown to be consistent and asymptotically normal. A corrected plug-in method is proposed for inference. A distinguishing feature of the proposed methods is that they allow the estimation and inference for the effects of space- or time-invariant covariates (e.g., *gender* and *policy*). In addition, the proposed methods allow (i) both spatial weights and coefficients to vary with time, (ii) the error variance to vary in all dimensions, and (iii) the panels to be nested or unbalanced. Finite sample performance is assessed by Monte Carlo simulations. A practical guide is provided.

Key Words: Adjusted quasi score; Observable/unobservable fixed effects; Time-varying spatial effects; Multidimensional panel; Unknown heteroskedasticity.

JEL classifications: C10, C13, C21, C23, C15.

*Correspondence: Zhenlin Yang, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: zlyang@smu.edu.sg

1. Introduction

The rapid and massive emergence of multidimensional (mD) panel data, as part of the big data revolution in recent years, has led to an increasing demand for more sophisticated model formulations compared to the well-known 2D panel data to properly address heterogeneity in the data (Balázsi et al., 2024). Traditional panel data sets, by nature, are two-dimensional labeled by space and time. In this context, the methodologies and specifications developed are mainly related to two-dimensional approaches which refer to observations on a cross section of households, or firms, or countries, etc., over several time periods. However, with the availability of more powerful tools in data collection and storage, the space dimension can easily go up to two, three, and even higher so that each observation comes with many labels, e.g., individual employee, firm, industry, and time as in a typical economic study on production; individual person, social network, ethnic group, and time as in a typical study on social interactions. While the heterogeneity in observations on the response is mostly captured by the included regressors, it is almost certain that some intrinsic features are left out (unobserved heterogeneity). Unobserved heterogeneity may be correlated with the time-varying regressors in an arbitrary manner, and thus must be treated as fixed effects. Unobserved heterogeneity may occur in any dimension and may be interactive, and these present serious issues in the formulation, estimation, and inference of mD panel models. "Oddly, applications have rushed ahead of theory in this field", to quote Mátyás (2017, p. vii). See Hornok and Koren (2017) for a review of advances in trade with comprehensive three-dimensional data sets, and Mátyás (2024) for extended coverage on multidimensional panels.

These issues remain and become more challenging for mD spatial panel data (SPD) models due to the coexistence of endogenous spatial spillover effects, contextual effects, and correlated heterogeneity, which gives rise to two major problems: *identification* (Manski, 1993) and *incidental parameters* (Neyman and Scott, 1948). The standard estimation and inference methods for regular mD panels are largely invalid. In addition, spatial units may interact within a 'space', between 'spaces' and change over time, giving rise to more complicated spatial interaction and spillover structures than do the standard 2D-SPD models, which further amplifies the issues of heterogeneity. Furthermore, there might be a strong interest in identifying and estimating certain types of observable dimension-specific effects, but the standard fixed-effects estimation does not allow for the identification and estimation of the time-invariant or space-invariant covariates effects. Finally, spatial data are typically heteroskedastic due to the large variation in unit sizes, and the errors are cross-correlated due to the unobservable interactions between units. A general method for the specification, estimation and inference for models of this type data is lacking, in particular the mD-SDP models with fixed effects.

Besides the studies on production and social interaction mentioned above, many other major applied studies are increasingly using multi-dimensional data to address more complicated economic issues. Flow data (in migration and trade) are labeled by origins, destinations, regions, and time. Housing price data are classified by districts, cities, and time. Health data involve patients, doctors, hospitals, and time. Student achievement data involve students, teachers, classes, schools, and time. Credit risk spread data involve stocks, sectors, provinces and time. All of these studies are likely to face the same issues as discussed above and a general solution is not available. In non-spatial contexts, Balázsi et al. (2024, p.1) underline that the vast majority of empirical studies conducted on multidimensional panels involve fixed effects models of some form; Baltagi et al. (2024, p.382) state that the explicit econometric treatment of the multi-dimensionality of the data is relatively young, and, for structural theoretical reasons, the dominant approach is the use of multiple fixed effects in estimation.

All of these show that a rigorous study on mD-SPD models with fixed effects is important and of great academic and applied significance. However, if we say that the literature on 2D-SPD models has yet reached its maturity,[1] the literature on mD-SPD models is still in its early childhood. To the best of our knowledge, it only contains the following works mainly on random-effects models. Fingleton et al. (2018a,b) and Ye et al. (2019) introduce a GMM estimation of 3D-SPD models with random effects. Baltagi and Pirotte (2014) describe a maximum likelihood iterative procedure to estimate an unbalanced spatial nested error components panel model, and a best linear unbiased predictor. He and Lin (2015) propose an LM test for spatial effects in a spatial lag model with spatial nested error components. An overview of this literature is provided by Le Gallo and Pirotte (2024).

In this paper, we consider the formulation, estimation, and inference for mD-SPD models with FEs, which allow (i) the identification and estimation of the effects of space or time invariant covariates (STIC), (ii) both spatial weights and coefficients to vary with time, (iii) the error variance to vary in all dimensions, and (iv) the panels to be nested or unbalanced.

[1] Le Gallo and Pirotte (2024) underline that it is only since the 2000s that an important literature has developed to deal with spatial panel data, see Anselin et al. (2008), LeSage and Pace (2009), Elhorst (2010, 2014), Lee and Yu (2010a, 2015), Kelejian and Piras (2017) for literature reviews.

We first separate the FEs into the observable part, the STIC effects, and the unobservable part, so that the STIC effects can be identified/estimated while the truly unobserved FEs are controlled for. Then, we propose a general M-estimation method for the estimation and inference of the common parameters and the STIC effects in the mD-SPD models. We obtain unbiased estimating equations (EFs) by adjusting the concentrated quasi-score functions with the unobserved FEs being concentrated out. The *adjusted quasi scores* (AQS) 'remove' the effects of estimating the unobserved FEs, and thereby lead to M-estimators that are consistent and asymptotically normal. The proposed estimation methods are then extended to allow the error variance to vary across spaces and time (the *unknown heteroskedasticity*). For this, the concentrated quasi scores need to be adjusted in a different way to give a new set of EFs that are robust against unknown heteroskedasticity. The resulting heteroskedasticity robust M-estimators are shown to be consistent and asymptotically normal.[2]

For statistical inferences, complication occurs in the estimation of variance-covariance (VC) matrix of the AQS functions as it involves the vector of unobserved-FEs which is high-dimensional. We propose a *corrected plug-in* method to correct the possible inconsistency or asymptotic bias caused by plugging in the estimated unobserved-FEs.

All of the proposed methods are given rigorous justifications by showing their large sample properties and finite sample performance through Monte Carlo simulation. To facilitate empirical applications, a guide is provided with Matlab code based on simulated data. The proposed methods are seen to be simple and reliable. They are quite general as well, which can be applied in tackling various network issues such as group fixed effects and can be further extended to allow additional features in a mD-SPD setting such as multiple networks. These constitute a significant methodological contribution to the spatial panel literature.

Section 2 introduces the mD-SPD model and the FE specifications. Section 3 introduces the M-estimation method. Section 4 introduces the inference methods. Section 5 presents Monte Carlo results, and Section 6 concludes. Appendix A collects essential theoretical results. `Supplemental Materials` contain proofs, more Monte Carlo results, and a practical guide.

Notation. `diag`$(\cdot)$ forms a diagonal matrix and `diagv`$(\cdot)$ a column vector, from the diagonal elements of a given matrix, `bdiag`$(\cdot)$ forms a block diagonal matrix from given matrices or vectors, $\otimes$ denotes Kronecker product, and $A^{\dagger} = A + A'$ for a given matrix A.

[2]Note that being able to identify and estimate the STIC effects is an important feature of the current paper, which had not been considered even in regular FE panel models until recently by Pirotte and Yang (2024).

2. The mD-SPD Models and Fixed Effects Specifications

2.1. Multidimensional spatial panel data model

Let Y be the vector of observations on the response variable and X the matrix containing the values of p exogenous regressors, arranged sequentially according to their occurrences in space dimensions $1, 2, \ldots, m-1$, and lastly in time dimension m.[3] The values of the p regressors are assumed to vary with respect to every index of the mD panels. The mD-SPD model with dimension-specific fixed effects can be written in the following general form:

$$Y = \mathbb{W}(\lambda)Y + X\beta + D\phi + U, \quad U = \mathbb{M}(\rho)U + V, \tag{2.1}$$

where $\mathbb{W}(\lambda) = \texttt{bdiag}(\lambda_1 W_1, \ldots, \lambda_T W_T)$ and $\mathbb{M}(\rho) = \texttt{bdiag}(\rho_1 M_1, \ldots, \rho_T M_T)$, W_t and M_t are the given *spatial weight* (*social connectivity*) matrices for observations in the tth time period with λ_t and ρ_t being the corresponding spatial lag and error parameters, $t = 1, \ldots, T$; ϕ is the vector of fixed effects and D the corresponding 'design' matrix assumed to be of full column rank; and β is the $p \times 1$ vector of regression coefficients, $\lambda = (\lambda_1, \ldots, \lambda_T)'$ and $\rho = (\rho_1, \ldots, \rho_T)'$. The spatial weight or social connectivity matrices W_t and M_t are $S_t \times S_t$ in general, allowing the spatial or social interactions among the spatial entities within a space and across spaces. They are allowed to vary with time in both their entries and matrix dimension.

Quite interestingly, the fixed effects, $D\phi$, can be decomposed into two parts:

$$D\phi = D_o\phi_o + D_e\phi_e, \tag{2.2}$$

being the observable and unobservable dimension-specific effects, respectively. The observable part $D_o\phi_o$ contain all the space or time invariant covariates (STIC) effects. It is typical that ϕ_o is of low dimension and ϕ_e is of high dimension. Thus, for model estimation, $D_o\phi_o$ posts no additional difficulty as it can simply be merged into $X\beta$. It is the formulation of D_o and D_e that can be complicated, for which details are given in Section 2.3.

The model we propose is fairly general. It contains the standard 2D-SPD model (Lee and Yu, 2010a; Yang et al., 2016) as a special case. It allows (i) the panel dimension to go to three or higher, (ii) the spatial effects $(\lambda_t W_t, \rho_t M_t)$ to be time-varying in spatial coefficients, the connectivity, and the number of spatial elements, (iii) the dimension-specific fixed effects

[3] In spatial panels, it is more convenient to list the data first according to space and then time. The elements of Y and X are denoted as $Y_{ij\cdots t}$ and $X_{ij\cdots t}$ with m indices: i goes fastest, then $j, \ldots$, and t the slowest.

($D\phi$) to be additive or interactive up to order $m-1$, (iv) idiosyncratic errors (V) to be homoskedastic or heteroskedastic, and (v) the STIC effects to be identified and estimated.

2.2. FE specifications: balanced panels

The main task now is to formulate the fixed effects so that D is of full column rank and ϕ is identifiable. As we mentioned in the introduction, the dimension-specific effects may appear in the model additively or interactively of various order. Accordingly, the form of D and the definition of ϕ would be different, depending on the dimension m of the panel and whether the panel is balanced, nested, or unbalanced. A balanced panel has the index set for any one dimension free from the indices of the other dimensions.

3D-SPD models. To fix ideas, we start with the simplest and the most popular 3D-SPD models with balanced panels and additive fixed effects that are completely unobserved. Let y_{ijt} be the response value for the ith unit in space 1 (`D1`), jth unit in space 2 (`D2`), and tth period in time (`D3`), and x_{ijt} be the corresponding values of the space-time varying covariates, where $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$, and $t = 1, \ldots, T$. Stack the observations y_{ijt} and x'_{ijt} first over i, then j, and then t to give the vector Y and matrix X. Let $\{\phi_{1,i}\}$, $\{\phi_{2,j}\}$ and $\{\phi_{3,t}\}$ be, respectively, the `D1`-, `D2`- and `D3`-specific effects, which are referred to in this paper the *main* FEs. Then, the additive main FEs have the following formulation:

$$D\phi = 1_N\mu + (1_{N_2T} \otimes I^{\circ}_{N_1})\phi_1 + (1_T \otimes I^{\circ}_{N_2} \otimes 1_{nN_1})\phi_2 + (I^{\circ}_T \otimes 1_{N_1N_2})\phi_3, \tag{2.3}$$

where N ($= N_1N_2T$) denotes the total number of observations, $\otimes$ the Kronecker product, 1_m an $m \times 1$ vector of ones, I°_m is obtained from the $m \times m$ identity matrix I_m by deleting its last column, and $\phi_1 = \{\phi_{1,i}\}_{i=1}^{N_1-1}$, $\phi_2 = \{\phi_{2,j}\}_{j=1}^{N_2-1}$, and $\phi_3 = \{\phi_{3,t}\}_{j=1}^{T-1}$ are, respectively, the `D1`-, `D2`-, and `D3`-specific FE vectors, with $\phi_{1,N_1}, \phi_{2,N_2}$, and $\phi_{3,T}$ being dropped and a grand mean parameter μ added to ensure identification. In this case, $\phi = (\mu, \phi'_1, \phi'_2, \phi'_3)'$, which is $(N_1 + N_2 + T - 2) \times 1$, and $D = [1_N,\ 1_{N_2T} \otimes I^{\circ}_{N_1},\ 1_T \otimes I^{\circ}_{N_2} \otimes 1_{N_1},\ I^{\circ}_T \otimes 1_{N_1N_2}]$.

The 3D-SPD model with formulation (2.3) is denoted as (i, j, t). Clearly, the 3D panel allows not only the first-order additive effects, $(\phi_{1,i}, \phi_{2,j}, \phi_{3,t})$, but also the second-order interactive FEs, $(\phi_{12,ij}, \phi_{13,it}, \phi_{23,jt})$. Each of these six effects can be present or absent, giving a total of 2^6 possible FE formulations for a 3D-SPD model, although not all are practically meaningful and some are mathematically similar by symmetry.[4] Let $D = [1_N, D^{\circ}]$. Table 1

[4]In addition, the main FEs can also appear in the model multiplicatively as in the factor model literature.

lists 12 typical FE formulations. Other formulations can be generated from these.

Table A. Dummy Variable Matrices with Full Column Rank: `Balanced 3D-SPD Models`

	FE	Dummy Variable Matrix D°	$r = \text{rank}(D^\circ)$
(1)	i	$[1_T \otimes 1_{N_2} \otimes I^\circ_{N_1}]$	$N_1 - 1$
(2)	i, j	$[1_T \otimes 1_{N_2} \otimes I^\circ_{N_1},\ 1_T \otimes I^\circ_{N_2} \otimes 1_{N_1}]$	$N_1 + N_2 - 2$
(3)	i, j, t	$[1_T \otimes 1_{N_2} \otimes I^\circ_{N_1},\ 1_T \otimes I^\circ_{N_2} \otimes 1_{N_1},\ I^\circ_T \otimes 1_{N_2} \otimes 1_{N_1}]$	$N_1 + N_2 + T - 3$
(4)	ij	$[1_T \otimes I^\circ_{N_2 N_1}]$	$N_1 N_2 - 1$
(5)	i, ij	$[1_{N_2 T} \otimes I^\circ_{N_1},\ 1_T \otimes I^\circ_{N_2} \otimes I_{N_1}]$	$N_1 N_2 - 1$
(6)	i, j, ij	$[1_{N_2 T} \otimes I^\circ_{N_1},\ 1_T \otimes I^\circ_{N_2} \otimes 1_{N_1},\ 1_T \otimes I^\circ_{N_2} \otimes I^\circ_{N_1}]$	$N_1 N_2 - 1$
(7)	ij, t	$[1_T \otimes I^\circ_{N_2 N_1},\ I^\circ_T \otimes 1_{N_2 N_1}]$	$N_1 N_2 + T - 2$
(8)	i, j, jt	$[1_T \otimes 1_{N_2} \otimes I^\circ_{N_1},\ 1_T \otimes I^\circ_{N_2} \otimes 1_{N_1},\ I^\circ_T \otimes I_{N_2} \otimes 1_{N_1}]$	$N_1 + N_2 + N_2(T-1) - 2$
(9)	ij, it	$[1_T \otimes I^\circ_{N_2 N_1},\ I^\circ_T \otimes 1_{N_2} \otimes I_{N_1}]$	$N_1 N_2 - 1 + N_1(T-1)$
(10)	ij, jt	$[1_T \otimes I^\circ_{N_2 N_1},\ I^\circ_T \otimes I_{N_2} \otimes 1_{N_1}]$	$N_1 N_2 - 1 + N_2(T-1)$
(11)	ij, it, t	$[1_T \otimes I^\circ_{N_2 N_1},\ I_T \otimes 1_{N_2} \otimes I^\circ_{N_1},\ I^\circ_T \otimes 1_{N_2 N_1}]$	$N_1 N_2 + N_1 T - 2$
(12)	ij, it, jt	$[1_T \otimes I^\circ_{N_2} \otimes I_{N_1},\ I_T \otimes 1_{N_2} \otimes I^\circ_{N_1},\ I^\circ_T \otimes I_{N_2} \otimes 1_{N_1}]$	$N_1 N_2 + N_1 T + N_2 T$ $-N_1 - N_2 - T$

Some details for finding the D° matrix may be helpful. The best way may be to start with the full dummy variable matrix. For example, the full dummy variable matrix for the model (ij, it, t) is $[1_T \otimes I_{N_2} \otimes I_{N_1},\ I_T \otimes 1_{N_2} \otimes I_{N_1},\ I_T \otimes 1_{N_2 N_1}]$. With the allowance of the grand mean dummy $1_{N_1 N_2 T}$, one column of $I_T \otimes 1_{N_2 N_1}$ must be dropped to avoid being multicollinear with $1_{N_1 N_2 T}$, giving $I^\circ_T \otimes 1_{N_2 N_1}$; $1_{N_2} \otimes I_{N_1}$ is multicollinear with $1_{N_2 N_1}$, so one level of i needs to be dropped and the second matrix must be $I_T \otimes 1_{N_2} \otimes I_{N_1^\circ}$; and lastly, one column of $1_T \otimes I_{N_2} \otimes I_{N_1}$ corresponding to one ij level needs to be dropped to avoid multicollinearity with the grand dummy, so that the first matrix becomes $1_T \otimes I^\circ_{N_2 N_1}$. It is interesting to note that the combination of (3) and (11) gives the largest FE specification that contains all first- and second-order FEs, and clearly the rank of the resulting D° is $N_1 N_2 + N_1 T + N_2 T - 3$.

4D-SPD models. The FE specifications for the 3D-SPD models presented above extend naturally to higher-dimensional SPD models. For the 4D-SPD models, we use the indices, i, j, k, for the three space dimensions, and t for the time dimension, where $i \in (1, \ldots, N_1)$, $j \in (1, \ldots, N_2)$, $k \in (1, \ldots, N_3)$, and $t \in (1, \ldots, T)$. A 4D-SPD model allows four main effects

We will stay in the classical ANOVA type of setup for the FE specifications in the mD-SPD models.

$(\phi_{1,i}, \phi_{2,j}, \phi_{3,k}, \phi_{4,t})$, six 2nd-order interactive effects $(\phi_{12,ij}, \phi_{13,ik}, \phi_{14,it}, \phi_{23,jk}, \phi_{24,jt}, \phi_{34,kt})$, and four 3rd-order interactive effects $(\phi_{123,ijk}, \phi_{124,ijt}, \phi_{134,ikt}, \phi_{234,jkt})$. Any of these 14 effects can be present or absent from the model, at least in theory, giving a total of 2^{14} possible FE specifications, although only a small subset of them is practically useful. In our FE formulation, we again allow a grand FE parameter ϕ_0 in the model to give a symmetric D° dummy variable matrix, and choose the "drop-the-last" approach in defining D°.

For the additive 4D FE specification, (i, j, k, t), including all four main effects, we have

$$D^\circ = [1_{TN_2N_3} \otimes I^\circ_{N_1},\ 1_{TN_3} \otimes I^\circ_{N_2} \otimes 1_{N_1},\ 1_T \otimes I^\circ_{N_3} \otimes 1_{N_1N_2},\ I^\circ_T \otimes 1_{N_1N_2N_3}],$$

where $\text{rank}(D^\circ) = N_1 + N_2 + N_3 + T - 4$. For a model with the following second-order interactive FEs, (ij, jk, it, jt), we have

$$D^\circ = [1_{TN_3} \otimes I^\circ_{N_2} \otimes I_{N_1},\ 1_T \otimes I^\circ_{N_3} \otimes I_{N_2} \otimes 1_{N_1},\ I_T \otimes 1_{N_2N_3} \otimes I^\circ_{N_1},\ I^\circ_T \otimes 1_{N_3} \otimes I_{N_2} \otimes 1_{N_1}],$$

which has rank of $(N_2 - 1)N_1 + (N_3 - 1)N_2 + (N_1 - 1)T + (T - 1)N_2$. The D° matrices for other FE specifications can be formed and their rank be determined accordingly.

Beyond 4D. In a similar way, the D° matrices for SPD models of dimensions beyond four can be formed and their ranks determined. To conserve space, we do not give details here.

2.3. Identification of the STIC effects

The dim-specific FEs can be decomposed into two parts: observable (STIC) and unobservable. Consider for example a balanced 3D-SPD model (Model 3 in Table 1) with three main FEs in the model additively. Suppose `Gender` (taking value 1 if male and 0 otherwise) is a D1-specific feature that is observed but was confounded in the main FEs $\phi_{1,i}$ in the previous setup. To estimate the gender effect, one `Gender` dummy is added to the model as a new regressor. In order to identify `Gender` effect, one column in D° corresponding to `D1`-main effects needs to be dropped. We drop the column that corresponds to the last value of 1 of `Gender` **if** the column already dropped (in D° corresponding to `D1`-main effects) is associated with `Gender` $= 0$; otherwise drop the column corresponding to the last value of 0 of `Gender`. Another way to proceed is to specify a general `Gender` variable with value 1 if the individual is a male, 0 otherwise. Unlike the previous case, there is no collinearity. This case is often encountered in applied economics when we estimate, for example, a wage equation. Similarly, an observed `D2`-specific effect can be identified and estimated consistently. To estimate the effect of a Policy intervention (taking value 1 at and after the time of the policy implementation), the last column of D° (corresponding to D3-main effects) is dropped.

More complicated STIC effects can be handled along the same line. For example, for an i-specific categorical covariate with three categories, we created two dummy variables and add them to D_o matrix, and correspondingly drop two columns in the D_e matrix.

2.4. FE specifications: nested panels

A nested panel is such that certain dimension(s) are nested within other(s). A dimension is said to be nested within the other dimension if its levels are observed in conjunction with just one level of the other dimension. A dimension is said to be nested within the other two dimensions if its levels are observed in just one joint level of the other two dimensions, etc.

To clarify, consider a 3D panel model where i is nested in j and panel is balanced in t, a typical FE specification would be $(i(j), j, t)$ with $\phi_{12,i(j)}$ denoting the nested FEs for D1, $\phi_{2,j}$ the main FEs for D2, and $\phi_{3,t}$ the main FEs for D3. In this case, a different set of $N_{1(j)}$ entities in D1 are observed at different level j in D2, $j = 1, 2, \dots, N_2$. Therefore, a total of $S = \sum_{j=1}^{N_2} N_{1(j)}$ spatial entities is observed repeatedly at each time period. This contrasts with the balanced 3D panels, where N_1 entities in D1 are repeatedly observed at each level j in D2. The D° matrix for the 3D-SPD model with FE specification, $(i(j), j, t)$, takes the form:

$$D^\circ = [\texttt{bdiag}(I^\circ_{N_{1(j)}}, j = 1, \dots, N_2),\ \texttt{bdiag}^\circ(1_{N_{1(j)}}, j = 1, \dots, N_2),\ I^\circ_T \otimes 1_S],$$

where $\texttt{bdiag}^\circ(\cdot)$ is obtained from $\texttt{bdiag}(\cdot)$ by dropping its last (set) column(s) and $\text{Rank}(D^\circ) = S + T - 2$. The other specifications with nested FEs can be worked out in a similar manner. Our methods allow S to change with t, i.e., $S_t = \sum_{j=1}^{N_{2t}} N_{1(j)t}$, and more, as in next section.

2.5. FE specifications: unbalanced panels

A general unbalanced panel is such that the index set for one dimension is not constant with respect the indices of the other dimensions. This contains as a special but typical case the nested panels discussed in Section 2.3. Unbalanced panels become common in the era of *big data*, and therefore it is important to provide econometric methods to analyze these data.

Consider an unbalanced 3D-SPD model with observations indices $i = 1, 2, \dots, N_{1jt}$, for each of $j = 1, 2, \dots, N_{2t}$, and for each of $t = 1, 2, \dots, T$. Let $\mathcal{N}_{1jt}$ be the set of i-units given j and t and $\mathcal{N}_1$ be their union over j and t with cardinality N_1 ($\geq N_{1jt}$); $\mathcal{N}_{2t}$ be the set of

j-units given t and $\mathcal{N}_2$ be their union over t with cardinality N_2 ($\geq N_{2t}$); and let $\mathcal{S}_t$ be the set of spatial (i-j) entities at time t and $\mathcal{S}$ be their union with cardinality S ($\geq S_t$). An interesting issue is the identification/control of the full vectors of the D_1-FEs ϕ_1 ($N_1 \times 1$), the D_2-FEs ϕ_2 ($N_2 \times 1$), the space (D_1 and D_2) FEs ϕ_{12} ($S \times 1$), and the time FEs ϕ_3 ($T \times 1$), and their possible interactions, given the fact that $N_{1jt} \leq N_1$, $N_{2t} \leq N_2$, and $S_t \leq S$.

To keep the discussion simple, assume for now the FEs are completely unobservable. The simplest scenario may be the model (ij, t), where the D-matrix in Model (2.1) can be

$$D = \{1_N,\ \texttt{bcoln}(D_{12,1}, \dots, D_{12,T}),\ \texttt{bdiag}^{\circ}(1_{S_1}, \dots, 1_{S_{T-1}}, 1_{S_T})\},$$

where $\texttt{bcoln}(\cdots)$ stacks the block matrices column wise, and $D_{12,t}$ is an $S_t \times S$ matrix obtained from the $S \times S$ identity matrix by deleting the rows corresponding to the non-present (ij) spatial entities and the last column, $t = 1, \dots, T$.

The above specification is equivalent to an unbalanced 2D-SDP model. An extended model can be (ij, j, t) when one is interested in the j-specific main effects beside the ij-interactions, and this contains as a special case the nested model with time-varying S_t discussed above:

$$D = \{1_N,\ \texttt{bcoln}(D^*_{12,1}, \dots, D^*_{12,T}),\ \texttt{bdiag}^{\circ}(1_{S_1}, \dots, 1_{S_{T-1}}, 1_{S_T})\},$$

where $D^*_{12,t}$ is obtained from $D_{12,t}$ by separating out the j-main effects, i.e., for each $j = 1, 2, \dots, N_{2t}$, drop the column corresponding to the last i unit and then add a column vector taking value 1 corresponding to the given j and 0 elsewhere.

The above discussions lays out clearly the ideas for the FE specifications. Along the same spirit, the other specifications containing it or jt interactions can be made, and unbalanced higher-order SDP models can be handled.

3. M-Estimation of mD-SPD Models with Fixed Effects

We introduce M-estimation of the mD-SPD models with FEs under four scenarios depending on whether the errors are homoskedastic or heteroskedastic, and whether the spatial parameters are homogeneous or time-varying. **Recall:** S_t denotes the total number of spatial entities appeared in the tth period, $t = 1, \dots, T$, so that the total number of observations $N = \sum_{t=1}^{T} S_t$. Let further N_s be the number of entities in the sth space dimension which may change with the indices of (all) other dimensions, $s = 1, 2, \dots, m-1$. We suppress such dependence for notational simplicity. In case of a balanced mD panel, N_s does not depend on the indices of other dimensions, $S_1 = \ldots = S_T = S = \prod_{s=1}^{m-1} N_s$, and $N = ST$.

3.1. QML-estimation

For Model (2.1), denoting the set of common parameters by $\theta = (\beta', \sigma^2, \delta')'$ with $\delta = (\lambda', \rho')'$, the quasi Gaussian loglikelihood, **as if** $V \sim N(0, \sigma^2 I_N)$, takes the following form:

$$\ell_N(\theta, \phi) = -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbb{A}(\lambda)| + \ln|\mathbb{B}(\rho)| - \tfrac{1}{2\sigma^2}V'(\beta, \delta, \phi)V(\beta, \delta, \phi), \qquad (3.1)$$

where $\mathbb{A}(\lambda) = I_N - \mathbb{W}(\lambda)$, $\mathbb{B}(\rho) = I_N - \mathbb{M}(\rho)$, and $V(\beta, \delta, \phi) = \mathbb{B}(\rho)[\mathbb{A}(\lambda)Y - X\beta - D\phi]$. Maximizing $\ell_N(\theta, \phi)$ gives the quasi maximum likelihood (QML) estimators $\hat{\theta}_{\mathtt{QML}}$ and $\hat{\phi}_{\mathtt{QML}}$ of θ and ϕ, which are consistent (under mild regularity conditions) if all dimensions are expanding when sample size N increases. The problem in this case is that the QML estimators will be asymptotically biased, and a bias-correction must be carried out for valid statistical inferences. Such a bias-correction may not be straightforward, in particular in our mD-SPD modelling framework. The problem with QML method becomes more serious when one or more dimensions do not expand so that the asymptotics depend only on the expansion of other dimensions: the consistency of QML estimators of certain common parameters (such as λ, ρ and σ^2) may not even be achieved, giving rise to the well-known *incidental parameters problem* of Neyman and Scott (1948). For example, when T is fixed and other dimensional FEs are fully interactive, a case of primary interest in microeconometrics, consistency are not achieved for all the QMLEs of the common parameters. In case when the errors in the model are heteroskedastic, the QML estimators will generally be inconsistent. Therefore, alternative methods are highly desirable for the estimation of mD-SPD models.

3.2. M-estimation under homoskedastic errors

Our aim is to provide a general method for consistent estimation and inference for the common parameters θ while controlling the effect of estimating the fixed effects parameters ϕ, or the *incidental parameters* as its dimension grows with the increase of sample size N. See Neyman and Scott (1948). The conventional transformation method (Lee and Yu, 2010a,b) for handling the additive fixed effects in balanced 2D-SPD model does not apply to the general mD-SPD model. Efforts of seeking a general transformation to wipe out the fixed effects $D\phi$ and yet a valid quasi likelihood function can still be formulated may be futile. Even if there is such a transformation in special cases, e.g., a balanced 2D-PD or 2D-SPD model, it will wipe out all the dimension-specific effects so that their observable part cannot be identified.

Our method starts by concentrating out ϕ from $\ell_N(\theta,\phi)$ and then adjusting the concentrated quasi scores to remove the effects of estimating ϕ. Define $\mathbb{D}(\rho) = \mathbb{B}(\rho)D$. Given θ, $\ell_N(\theta,\phi)$ is partially maximized, by an OLS regression of $\mathbb{B}(\rho)[\mathbb{A}(\lambda)Y - X\beta]$ on $\mathbb{D}(\rho)$, at

$$\hat{\phi}(\beta,\delta) = [\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\mathbb{B}(\rho)[\mathbb{A}(\lambda)Y - X\beta], \tag{3.2}$$

leading to the concentrated quasi loglikelihood (CQL) function of θ:

$$\ell_N^c(\theta) = -\frac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbb{A}(\lambda)| + \ln|\mathbb{B}(\rho)| - \frac{1}{2\sigma^2}\tilde{V}'(\beta,\delta)\tilde{V}(\beta,\delta), \tag{3.3}$$

where $\tilde{V}(\beta,\delta) = \mathbb{Q}(\rho)\mathbb{B}(\rho)[\mathbb{A}(\lambda)Y - X\beta]$ and $\mathbb{Q}(\rho) = I_N - \mathbb{D}(\rho)[\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)$.

Case Ia: Time-invariant spatial parameters. To fix ideas, we first consider the case where the spatial parameters are time-invariant (so that λ and ρ are both scalars), and all the FEs are unobservable. Let $\mathbf{W} = \mathtt{bdiag}(W_1,\ldots,W_T)$ and $\mathbf{M} = \mathtt{bdiag}(M_1,\ldots,M_T)$. The concentrated quasi score (CQS), $S_N^c(\theta) = \frac{\partial}{\partial\theta}\ell_N^c(\theta)$, takes the form:

$$S_N^c(\theta) = \begin{cases} \frac{1}{\sigma^2}X'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta), \\ \frac{1}{2\sigma^4}[\tilde{V}'(\beta,\delta)\tilde{V}(\beta,\delta) - N\sigma^2], \\ \frac{1}{\sigma^2}Y'\mathbf{W}'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta) - \mathtt{tr}[\mathbb{F}(\lambda)], \\ \frac{1}{\sigma^2}\tilde{V}'(\beta,\delta)\mathbb{G}(\rho)\tilde{V}(\beta,\delta) - \mathtt{tr}[\mathbb{G}(\rho)], \end{cases} \tag{3.4}$$

where $\mathbb{F}(\lambda) = \mathbf{W}\mathbb{A}^{-1}(\lambda)$ and $\mathbb{G}(\rho) = \mathbf{M}\mathbb{B}^{-1}(\rho)$.[5]

The direct quasi maximum likelihood (QML) estimator $\hat{\theta}_{\mathtt{QML}}$ of θ maximizes $\ell_N^c(\theta)$. Equivalently, $\hat{\theta}_{\mathtt{QML}}$ is obtained by solving the CQS equations: $S_N^c(\theta) = 0$. The QML estimator $\hat{\theta}_{\mathtt{QML}}$ cannot be generally consistent (although its β-components is) due to the well known *incidental parameters problem* of Neyman and Scott (1948). This can be seen by showing that $\frac{1}{N}S_N^c(\theta_0)] \nrightarrow 0$ as $N \to \infty$ with T being fixed, and thus a necessary condition for consistency of estimation is violated. When all dimensions grow with N, one can show that $\hat{\theta}_{\mathtt{QML}}$ is consistent, but the asymptotic distribution of $\frac{1}{\sqrt{N}}S_N^c(\theta_0)]$ is not centered and therefore the asymptotic distribution of $\sqrt{N}(\hat{\theta}_{\mathtt{QML}} - \theta_0)$ is not centered, giving rise the so-called *asymptotic bias*. A bias correction is necessary for valid inference. Common practice is to bias-correct the QMLE $\hat{\theta}_{\mathtt{QML}}$ but it is often complicated. In case of fixed T, the QML method fails to provide a valid method unless a transformation method is available to wipe out the FEs and

[5] In deriving the ρ-component of $S_N^c(\theta)$, write $\tilde{V}'(\beta,\delta)\tilde{V}(\beta,\delta) = [\mathbb{A}(\lambda)Y - X\beta]'\mathbb{C}(\rho)[\mathbb{A}(\lambda)Y - X\beta]$, where $\mathbb{C}(\rho) = \mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathbb{B}(\rho)$. We show that $\frac{\partial}{\partial\rho}\mathbb{C}(\rho) = -\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathbb{G}^\dagger(\rho)\mathbb{Q}(\rho)\mathbb{B}(\rho)$.

the resulting transformed model leads a valid quasi likelihood function, which are clearly not the case for the mD-SPD model that we consider in this paper.

We do bias correction on $S_N^c(\theta_0)$. As shown below, the CQS function can be adjusted as $S_N^c(\theta_0) - \mathrm{E}[S_N^c(\theta_0)]$, to give a set of unbiased and consistent *estimation functions*, leading possibly to a consistent M-estimation of the common parameters θ. This method is valid whether T (or any dimension) is fixed or not.

We have, $\mathrm{E}[S_N^c(\theta_0)] = \big(0,\ -\frac{r+1}{2\sigma_0^2},\ \mathtt{tr}[\mathbb{Q}(\rho_0)\bar{\mathbb{F}}(\delta_0)] - \mathtt{tr}[\mathbb{F}(\lambda_0)],\ \mathtt{tr}[\mathbb{Q}(\rho_0)\mathbb{G}(\rho_0)] - \mathtt{tr}[\mathbb{G}(\rho_0)]\big)'$, where $\bar{\mathbb{F}}(\delta_0) = \mathbb{B}(\rho_0)\mathbb{F}(\lambda_0)\mathbb{B}^{-1}(\rho_0)$. The (λ, ρ) components simplify to $\mathtt{tr}[\mathbb{P}(\rho_0)\bar{\mathbb{F}}(\delta_0)]$ and $\mathtt{tr}[\mathbb{P}(\rho_0)\mathbb{G}(\rho_0)]$, where $\mathbb{P}(\rho) = I_N - \mathbb{Q}(\rho)$. As $\mathrm{rank}[\mathbb{P}(\rho_0)] = r + 1$ and $\bar{\mathbb{F}}(\delta_0)$ and $\mathbb{G}(\rho_0)$ are positive definite, these two components are of order $O(r)$. It follows that $\frac{1}{N}\mathrm{E}[S_N^c(\theta_0)] = O(\frac{r}{N})$, which may not converge to 0. Hence $\mathrm{plim}_{N\to\infty}\frac{1}{N}S_N^c(\theta_0)$ may not be 0, unless $r = O(N)$.

Define the *adjusted quasi scores* (AQS) (EFs or moment conditions) as $S_N^*(\theta_0) = S_N^c(\theta_0) - \mathrm{E}[S_N^c(\theta_0)]$, which takes the form at the general parameter value θ:

$$S_N^*(\theta) = \begin{cases} \frac{1}{\sigma^2}X'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta), \\ \frac{1}{2\sigma^4}[\tilde{V}'(\beta,\delta)\tilde{V}(\beta,\delta) - N^*\sigma^2], \\ \frac{1}{\sigma^2}Y'\mathbf{W}'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}(\rho)\bar{\mathbb{F}}(\delta)], \\ \frac{1}{\sigma^2}\tilde{V}'(\beta,\delta)\mathbb{G}(\rho)\tilde{V}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}(\rho)\mathbb{G}(\rho)], \end{cases} \tag{3.5}$$

where $N^* = N - r - 1$, being the adjusted or effective sample size. Clearly $\mathrm{E}[S_N^*(\theta_0)] = 0$ and it is easy to show that $\mathrm{plim}_{N\to\infty}\frac{1}{N^*}S_N^*(\theta_0) = 0$. Therefore, the AQS functions, $S_N^*(\theta)$, gives a set of *EFs* that are unbiased and consistent at the true parameters θ_0. This opens up a way for consistent estimation of θ by solving the estimating equations, $S_N^*(\theta) = 0$. The solution $\hat{\theta}_{\mathtt{M}}^*$ is referred to as M-estimator or the method of moment estimator.

Theorem 1. *If Assumptions 1-7 (Appendix A) hold, then, as $N^* \to \infty$, $\hat{\theta}_{\mathtt{M}}^* \xrightarrow{p} \theta_0$, and*

$$\sqrt{N^*}(\hat{\theta}_{\mathtt{M}}^* - \theta_0) \xrightarrow{D} N\big[0,\ \Sigma_*^{-1}(\theta_0)\Omega_*(\theta_0)\Sigma_*'^{-1}(\theta_0)\big], \tag{3.6}$$

where $\Sigma_*(\theta_0) = -\lim_{N^*\to\infty}\frac{1}{N^*}\frac{\partial}{\partial\theta'}S_N^*(\theta_0)$ and $\Omega_*(\theta_0) = \lim_{N^*\to\infty}\frac{1}{N^*}\mathrm{E}[S_N^*(\theta_0)S_N^{*\prime}(\theta_0)]$, both assumed to exist and $\Sigma_*(\theta_0)$ further assumed to be positive definite.

Case Ib: Time-varying spatial parameters. The M-estimation using AQS as EFs can be extended to many more general and complicated scenarios. We now allow spatial parameters to change over time so that $\delta = (\lambda', \rho')'$ is a $2T \times 1$ vector. Denote $\delta_t = (\lambda_t, \rho_t)'$.

Assume again all the FEs are unobservable. The constraint QML estimate $\hat{\phi}(\beta, \delta)$ given in (3.2) and the CQL function $\ell_N^c(\theta)$ given in (3.3) remain in the same forms.

To deal with the time-varying feature of the spatial parameters, it is convenient to introduce the following special matrices, $\mathcal{W}_t$, $\mathcal{F}_t(\lambda_t)$ and $\mathcal{G}_t(\rho_t)$, which are $N \times N$ with tt-blocks being W_t, $F_t(\lambda_t)$ and $G_t(\rho_t)$, respectively, and other blocks being zero, where $F_t(\lambda_t)$ and $G_t(\rho_t)$ are the tth diagonal blocks of $\mathbb{F}(\lambda)$ and $\mathbb{G}(\rho)$ in (3.4). With these special matrices, the CQS vector, $S_N^{cc}(\theta) = \frac{\partial}{\partial \theta}\ell_N^c(\theta)$, takes a form similar to (3.4):

$$S_N^{cc}(\theta) = \begin{cases} \frac{1}{\sigma^2} X'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta), \\ \frac{1}{2\sigma^4}[\tilde{V}'(\beta,\delta)\tilde{V}(\beta,\delta) - N\sigma^2], \\ \frac{1}{\sigma^2} Y'\mathcal{W}_t'\, \mathbb{B}'(\rho)\tilde{V}(\beta,\delta) - \mathtt{tr}[\mathcal{F}_t(\lambda_t)], \;\; t = 1,\ldots,T, \\ \frac{1}{\sigma^2}\tilde{V}'(\beta,\delta)\mathcal{G}_t(\rho_t)\tilde{V}(\beta,\delta) - \mathtt{tr}[\mathcal{G}_t(\rho_t)], \;\; t = 1,\ldots,T, \end{cases} \tag{3.7}$$

where the ρ_t-components are obtained by showing $\frac{\partial}{\partial \rho_t}\mathbb{C}(\rho) = -\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathcal{G}_t^{\dagger}(\rho_t)\mathbb{Q}(\rho)\mathbb{B}(\rho)$.

The first two components of $\mathrm{E}[S_N^{cc}(\theta_0)]$ remain the same as those of $\mathrm{E}[S_N^c(\theta_0)]$. Noting $\mathbb{B}\mathcal{W}_t Y = \bar{\mathcal{F}}_t(\mathbb{B}X\beta_0 + \mathbb{D}\phi_0 + V)$ and $\tilde{V}(\beta_0, \delta_0) = \mathbb{Q}V$, where $\bar{\mathcal{F}}_t = \mathbb{B}\mathcal{F}_t\mathbb{B}^{-1}$, we have the (λ_t, ρ_t)-components of $\mathrm{E}[S_N^{cc}(\theta_0)]$ as: $\mathtt{tr}[\bar{\mathcal{F}}_t(\delta_{t0})\mathbb{Q}(\rho_0)] - \mathtt{tr}[\mathcal{F}_t(\lambda_{t0})]$ and $\mathtt{tr}[\mathcal{G}_t(\rho_{t0})\mathbb{Q}(\rho_0)] - \mathtt{tr}[\mathcal{G}_t(\rho_{t0})]$. Thus, the AQS functions for the time-heterogeneous mD-SDP models are:

$$S_N^{**}(\theta) = \begin{cases} \frac{1}{\sigma^2} X'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta), \\ \frac{1}{2\sigma^4}[\tilde{V}'(\beta,\delta)\tilde{V}(\beta,\delta) - N^*\sigma^2], \\ \frac{1}{\sigma^2} Y'\mathcal{W}_t'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta) - \mathtt{tr}[\bar{\mathcal{F}}_t(\delta_t)\mathbb{Q}(\rho)], \;\; t = 1,\ldots,T, \\ \frac{1}{\sigma^2}\tilde{V}'(\beta,\delta)\mathcal{G}_t(\rho_t)\tilde{V}(\beta,\delta) - \mathtt{tr}[\mathcal{G}_t(\rho_t)\mathbb{Q}(\rho)], \;\; t = 1,\ldots,T. \end{cases} \tag{3.8}$$

The M-estimator $\hat{\theta}_{\mathtt{M}}^{**}$ now solves $S_N^{**}(\theta) = 0$, which nests the case of homogeneous spatial parameters considered above. Under mild regularity conditions, we show that $(\hat{\beta}_{\mathtt{M}}^{**}, \hat{\sigma}_{\mathtt{M}}^{**2})$ is $\sqrt{N^*}$-consistent and asymptotically normal (centered). As T may increase with the total sample size N, the dimensions of λ and ρ may grow with the increase of N. In this case, the asymptotic properties of $\hat{\delta}_{t,\mathtt{M}}^{**}$ would be different. In particular, $\hat{\delta}_{t,\mathtt{M}}^{**}$ is $\sqrt{S_t}$-consistent, and $\sqrt{S_t}(\hat{\delta}_{t,\mathtt{M}}^{**} - \delta_{t0})$ has a centered limiting normal distribution, for $t = 1, \ldots, T$.

To study the asymptotic properties of $\hat{\theta}_{\mathtt{M}}$ together, we propose the following general framework to deal with the issue of different convergence rates for the components of $\hat{\theta}_{\mathtt{M}}^{**}$, and the issue of growing dimension in the parameter set $\delta = (\lambda_1, \ldots, \lambda_T, \rho_1, \ldots, \rho_T)'$. Let C be a

$2T \times k$ matrix representing k (a fixed number) linear contraints in δ. Let $\mathbf{C} = \mathtt{bdiag}(I_{p+1}, C)$. Let $R = \mathtt{bdiag}(N^* I_{p+1},\ S_1, \ldots, S_T,\ S_1, \ldots, S_T)$. We have the following theorem.

Theorem 2. *If Assumptions 1-7 (Appendix A) hold, then as* $N^* \to \infty$, $(\hat{\beta}^{**}_{\mathtt{M}},\ \hat{\sigma}^{**2}_{\mathtt{M}}) \xrightarrow{p} (\beta_0,\ \sigma_0^2)$, $\hat{\delta}^{**}_{t,\mathtt{M}} \xrightarrow{p} \delta_{t0}$ *for* $t = 1, 2, \ldots, T$, *and*

$$\mathbf{C}'\sqrt{R}(\hat{\theta}^{**}_{\mathtt{M}} - \theta_0) \xrightarrow{D} N\big[0,\ \lim_{N\to\infty} \mathbf{C}'\Sigma^{-1}_{**}(\theta_0)\Omega_{**}(\theta_0)\Sigma'^{-1}_{**}(\theta_0)\mathbf{C}\big], \tag{3.9}$$

where $\Sigma_{**}(\theta_0) = -R^{-1/2}\frac{\partial}{\partial\theta'}S^{**}_N(\theta_0)R^{-1/2}$ *and* $\Omega_{**}(\theta_0) = R^{-1/2}\mathrm{E}[S^{**}_N(\theta_0)S^{**\prime}_N(\theta_0)]R^{-1/2}$, *both assumed to exist and* $\Sigma_{**}(\theta_0)$ *further assumed to be positive definite.*[6]

Remark 1. The processes of solving $S^*_N(\theta) = 0$ and $S^{**}_N(\theta) = 0$ can be simplified by first analytically solving the first two sets of equations for β and σ^2 given δ, which gives

$$\hat{\beta}^*_N(\delta) = [\mathbb{Z}'(\rho)\mathbb{Z}(\rho)]^{-1}\mathbb{Z}'(\rho)\mathbb{Y}(\delta) \quad \text{and} \quad \hat{\sigma}^{*2}_N(\delta) = \tfrac{1}{N^*}\tilde{V}'(\hat{\beta}^*_N(\delta), \delta)\tilde{V}(\hat{\beta}^*_N(\delta), \delta), \tag{3.10}$$

where $\mathbb{Z}(\rho) = \mathbb{Q}(\rho)\mathbb{B}(\rho)X$, $\mathbb{Y}(\delta) = \mathbb{B}(\rho)\mathbb{A}(\lambda)Y$ and $\tilde{V}(\beta, \delta)$ is defined below (3.3), and then solving the concentrated equations for δ. This is helpful when there are many β parameters.

3.3. M-estimation under heteroskedastic errors

When errors are heteroskedastic, the estimation and inference methods introduced above are no longer valid in general. However, the idea of adjusting the quasi scores in order to obtain unbiased and consistent EFs still proceed. Liu and Yang (2020) introduce heteroskedasticity-robust estimation and inference of a balanced 2D-SPD model with fixed effects.

Assume that $V \sim (0, H)$, where H is a diagonal matrix with all diagonal elements being strictly positive. Now, $h = \mathtt{diagv}(H)$ constitutes another set of incidental parameters which may even be of a higher dimension than ϕ as we allow the variance to change across all observations. As h are the scale parameters, direct estimation/concentration of h as the way we do for the linear parameters ϕ can no longer work. We endeavor to find a set of EFs (moment conditions) for the common parameters $\vartheta = (\beta', \lambda', \rho')'$ that are robust against unknown heteroskedasticity h. From GMM point of view, there may be more than one way of doing so. We do this by adjusting the CQS components for ϑ from the above homoskedastic scenarios, which way may lead to a set of moments that are potentially efficient, as they naturally contain quadratic moments in addition to linear moments. This is important in the

[6] Note that introducing the contrast matrix $\mathbf{C}$ and normalizing factors R is purely for proper asymptotic analyses due to the existence of growing-dimension vector of parameters. In practical applications, the standard errors of $\hat{\theta}^{**}_{\mathtt{M}}$ are simply the square roots of the diagonals $[\frac{\partial}{\partial\theta'}S^{**}_N(\hat{\theta}_{\mathtt{M}})]^{-1}\widehat{\mathrm{Var}}[S^{**}_N(\theta_0)][\frac{\partial}{\partial\theta'}S^{**}_N(\hat{\theta}_{\mathtt{M}})]'^{-1}$.

sense that λ and ρ enter the model in a highly nonlinear manner.

Case IIa: Time-invariant spatial parameters. Consider first the simpler case where the spatial parameters do not change with time. We proceed to modify the ϑ-component of the CQS function given in (3.4) as follows. First, the β-component is linear in V and hence its expectation at true parameters is zero under h (robust against h).

For the λ-component, we see that $\mathrm{E}[Y'\mathbf{W}'\mathbb{B}'(\rho_0)\tilde{V}(\beta_0,\delta_0)] = \mathtt{tr}[\bar{\mathbb{F}}'(\delta_0)\mathbb{Q}(\rho_0)H])$, which can be written as either,

$$\mathrm{E}[Y'\mathbb{A}'(\lambda_0)\mathbb{B}'(\rho_0)\bar{\mathbb{F}}'(\delta_0)\mathbb{Q}(\rho_0)V] \ \text{ or } \ \mathrm{E}[Y'\mathbb{A}'(\lambda_0)\mathbb{B}'(\rho_0)F_d(\delta_0)\mathbb{Q}(\rho_0)V],$$

where $F_d(\delta) = \mathtt{diag}(\bar{\mathbb{F}}'(\delta)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}$. Taking difference between the two quantities inside the expectation, we obtain an EF for λ that is unbiased and consistent under h:

$$Y'\mathbb{A}'(\lambda)\mathbb{B}'(\rho)[\bar{\mathbb{F}}'(\delta) - F_d(\delta)]\tilde{V}(\beta,\delta). \tag{3.11}$$

Similarly, for the ρ-component, $\mathrm{E}[\tilde{V}'(\beta_0,\delta_0)\mathbb{G}(\rho_0)\tilde{V}(\beta_0,\delta_0)] = \mathrm{E}[V'\mathbb{Q}(\rho_0)\mathbb{G}(\rho_0)\mathbb{Q}(\rho_0)V] = \mathtt{tr}[\bar{\mathbb{G}}(\rho_0)\mathbb{Q}(\rho_0)H]$, where $\bar{\mathbb{G}}(\rho) = \mathbb{Q}(\rho)\mathbb{G}(\rho)$, which can be written as either:

$$\mathrm{E}\{[\mathbb{A}(\lambda_0)Y - X\beta_0]'\mathbb{B}'(\rho_0)\bar{\mathbb{G}}(\rho_0)\mathbb{Q}(\rho_0)V\} \ \text{ or } \ \mathrm{E}\{[\mathbb{A}(\lambda_0)Y - X\beta_0]'\mathbb{B}'(\rho_0)G_d(\rho_0)\mathbb{Q}(\rho_0)V\},$$

where $G_d(\rho) = \mathtt{diag}(\bar{\mathbb{G}}(\rho)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}$. Taking difference between the two quantities inside the expectation, we obtain an EF for ρ that is unbiased and consistent under h:

$$[\mathbb{A}(\lambda)Y - X\beta]'\mathbb{B}'(\rho)[\bar{\mathbb{G}}(\rho) - G_d(\rho)]\tilde{V}(\beta,\delta). \tag{3.12}$$

Putting the components together, we obtain the vector of EFs for the parameters β, λ and ρ, for the case of time-invariant spatial parameters:

$$S_N^r(\vartheta) = \begin{cases} X'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta), \\ Y'\mathbb{A}'(\lambda)\mathbb{B}'(\rho)[\bar{\mathbb{F}}'(\delta) - F_d(\delta)]\tilde{V}(\beta,\delta), \\ [\mathbb{A}(\lambda)Y - X\beta]'\mathbb{B}'(\rho)[\bar{\mathbb{G}}(\rho) - G_d(\rho)]\tilde{V}(\beta,\delta). \end{cases} \tag{3.13}$$

Solving the estimating equations, $S_N^r(\vartheta) = 0$, we obtain the robust M-estimator $\hat{\vartheta}_{\mathtt{M}}^r$ of ϑ_0, which is shown to be consistent and asymptotically normal under unknown h.

Theorem 3. *If Assumptions 1*, and 2-7 (Appendix A) hold, then,* $\hat{\vartheta}_{\mathtt{M}}^r \xrightarrow{p} \vartheta_0$, *and*

$$\sqrt{N^*}(\hat{\vartheta}_{\mathtt{M}}^r - \vartheta_0) \xrightarrow{D} N\big[0,\ \Sigma_r^{-1}(\vartheta_0)\Omega_r(\vartheta_0)\Sigma_r'^{-1}(\vartheta_0)\big], \ \ \textit{as } N^* \to \infty, \tag{3.14}$$

where $\Sigma_r(\vartheta) = -\lim_{N\to\infty}\frac{1}{N^*}\frac{\partial}{\partial\vartheta'}S_N^r(\vartheta)$ *and* $\Omega_r(\vartheta_0) = \lim_{N\to\infty}\frac{1}{N^*}\mathrm{E}[S_N^r(\vartheta_0)S_N^{r\prime}(\vartheta_0)]$, *both assumed to exist and* $\Sigma_r(\theta_0)$ *further assumed to be positive definite.*

Case IIb: Time-varying spatial parameters. For the case of time-varying spatial parameters and heteroskedastic errors, we proceed to modify the ϑ-component of the CQS function given in (3.7) to give a set of EFs robust against h. The β-component is robust against h. Write the stochastic parts of the λ and ρ components as

$$Y'\mathbb{A}'(\lambda_0)\mathbb{B}'(\rho_0)\bar{\mathcal{F}}_t'(\delta_{t0})\tilde{V}(\beta_0,\delta_0) \text{ and } \tilde{V}'(\beta_0,\delta_0)\mathcal{G}_t(\rho_{t0})\tilde{V}(\beta_0,\delta_0)$$

where $\bar{\mathcal{F}}_t'(\delta_{t0})$ is an $N \times N$ matrix with entries $\bar{F}_t'(\delta_{t0})$ in its tt-block and 0 elsewhere, and $\mathcal{G}_t(\rho_{t0})$ is an $N\times N$ matrix with entries $G_t(\rho_{t0})$ in its tt-block and 0 elsewhere. Then, following the same arguments leading to (3.13), we have a set of EFs robust against unknown h:

$$S_N^{rr}(\vartheta) = \begin{cases} X'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta), \\ Y'\mathbb{A}'(\lambda)\mathbb{B}'(\rho)[\bar{\mathcal{F}}_t'(\delta_t) - \mathcal{F}_t^d(\delta_t,\rho)]\tilde{V}(\beta,\delta), \quad t=1,\ldots,T, \\ [\mathbb{A}(\lambda)Y - X\beta]'\mathbb{B}'(\rho)[\bar{\mathcal{G}}_t(\rho_t,\rho) - \mathcal{G}_t^d(\rho_t,\rho)]\tilde{V}(\beta,\delta), \quad t=1,\ldots,T, \end{cases} \tag{3.15}$$

where $\mathcal{F}_t^d(\delta_t,\rho) = \mathtt{diag}(\bar{\mathcal{F}}_t'(\delta_t)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}$, $\mathcal{G}_t^d(\rho_t,\rho) = \mathtt{diag}(\bar{\mathcal{G}}(\rho_t,\rho)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}$, and $\bar{\mathcal{G}}(\rho_t,\rho) = \mathbb{Q}(\rho)\mathcal{G}(\rho_t)$. Solving $S_N^{rr}(\vartheta)=0$, we obtain a robust M-estimator $\hat{\vartheta}_{\mathtt{M}}^{rr}$ for ϑ.

Under regularity conditions, we show that $\hat{\beta}_{\mathtt{M}}^{rr}$ is $\sqrt{N^*}$-consistent and that $(\hat{\lambda}_{t,\mathtt{M}}^{rr}, \hat{\rho}_{t,\mathtt{M}}^{rr})$ is $\sqrt{S_t}$-consistent. The asymptotic distributions can be established separately for $\hat{\beta}_{\mathtt{M}}^{rr}$ and $(\hat{\lambda}_{t,\mathtt{M}}^{rr}, \hat{\rho}_{t,\mathtt{M}}^{rr})$, or jointly along the same line as for the case of homoskedastic errors but heterogeneous spatial parameters. Let C be a $2T\times k$ matrix for finite k linear contrasts in δ. Let $\mathbf{C} = \mathtt{bdiag}(I_p, C)$. Let $R = \mathtt{bdiag}(N^* I_p, S_1,\ldots,S_T,S_1,\ldots,S_T)$. We have the following theorem.

Theorem 4. *If Assumptions 1**, *and 2-7 (Appendix A) hold, then as* $N^*\to\infty$, $\hat{\beta}_{\mathtt{M}}^{rr} \xrightarrow{p} \beta_0$, $\hat{\delta}_{t,\mathtt{M}}^{rr} \xrightarrow{p} \delta_{t0}$ *for* $t=1,2,\ldots,T$, *and*

$$\mathbf{C}'\sqrt{R}(\hat{\vartheta}_{\mathtt{M}} - \vartheta_0) \xrightarrow{D} N\big[0, \lim_{N\to\infty} \mathbf{C}'\Sigma_{rr}^{-1}(\vartheta_0)\Omega_{rr}(\vartheta_0)\Sigma_{rr}'^{-1}(\vartheta_0)\mathbf{C}\big], \tag{3.16}$$

where $\Sigma_{rr}(\vartheta_0) = -R^{-1/2}\frac{\partial}{\partial\theta'}S_N^{rr}(\theta_0)R^{-1/2}$ *and* $\Omega_{rr}(\theta_0) = R^{-1/2}\mathrm{E}[S_N^{rr}(\theta_0)S_N^{rr\prime}(\theta_0)]R^{-1/2}$, *both assumed to exist and* $\Sigma_{rr}(\theta_0)$ *further assumed to be positive definite.*

The expressions of the Σ- and Ω-matrices are derived in Sec. 4, and methods for estimating these matrices for statistical inference are introduced therein. The regularity conditions under which the results of Theorems 1-4 hold are given in Appendix A. Detailed proofs of these theorems are given in Appendix B. Similarly, the processes of solving $S_N^r(\vartheta)=0$ and $S_N^{rr}(\vartheta)=0$ can be simplified by solving the concentrated equations for δ with β being concentrated out.

Remark 2. When there are `STIC` effects that include a constant, decompose $D\phi$ in (2.1) into $\texttt{STIC} + D_e\phi_e$ as discussed in (2.2) and merge `STIC` into $X\beta$. Then, $D_e\phi_e$ acts as $D\phi$.

4. Inference for mD-SPD Models with Fixed Effects

We now present inference methods for the mD-SPD models with FEs. To simplify notation, we denote $\mathbb{A} \equiv \mathbb{A}(\lambda_0)$, $\mathbb{B} \equiv \mathbb{B}(\rho_0)$, $\mathbb{D} \equiv \mathbb{D}(\rho_0)$, $\mathbb{Q} \equiv \mathbb{Q}(\rho_0)$, $\tilde{V} \equiv \tilde{V}(\beta_0, \delta_0)$, etc. Furthermore, we denote generically $\mathbb{X}(\rho) = \mathbb{B}(\rho)X$, $\mathbb{Y}(\delta) = \mathbb{B}(\rho)\mathbb{A}(\lambda)Y$ and thus $\mathbb{Y} \equiv \mathbb{Y}(\delta_0) = \mathbb{B}\mathbb{A}Y$.

Inference for θ or ϑ based on the M-estimation frameworks presented in Sec. 3 amounts to estimate consistently the pairs of matrices, $\Sigma_*(\theta_0)$ and $\Omega_*(\theta_0)$, $\Sigma_{**}(\theta_0)$ and $\Omega_{**}(\theta_0)$, $\Sigma_r(\theta_0)$ and $\Omega_r(\theta_0)$, $\Sigma_{rr}(\theta_0)$ and $\Omega_{rr}(\theta_0)$, respectively in Theoresm 1-4. The Σ-matrices can be estimated by their *sample analogues* or *plug-in* method. The estimation of the Ω-matrices is challenging. A *plug-in with correction* (`PIC`) method is introduced.

4.1. Inference methods under homoskedastic errors

Case Ia: Time-invariant spatial parameters. Note that for the λ-component of $S_N^*(\theta_0)$ given in (3.5), $\mathbb{B}\mathbf{W}Y = \bar{\mathbb{F}}\mathbb{Y}$ and $\mathbb{Y} = \mathbb{B}\mathbb{A}Y = \mathbb{B}X\beta_0 + \mathbb{D}\phi_0 + V$. This means that the VC matrix $\mathrm{Var}[S_N^*(\theta_0)]$ contains terms linear or quadratic in ϕ_0 that may not be consistently estimated as discussed earlier. Therefore, $\Omega_*(\theta_0)$ in (3.6) may not be consistently estimated by the plug-in method, and a *plug-in with correction* (`PIC`) method is proposed.

PIC estimation of $\Omega_*(\theta_0)$. To derive $\mathrm{Var}[S_N^*(\theta_0)]$, write the stochastic elements of the first expression of $S_N^*(\theta_0)$ as $\eta_1'V$, $V'\Phi_1V$, $V'\Phi_2V + \eta_2'V$, and $V'\Phi_3V$, using the facts that $\tilde{V} = \mathbb{Q}V$ and $\mathbb{Y} = \mathbb{B}X\beta_0 + \mathbb{D}\phi_0 + V$, where $\eta_1 = \mathbb{Q}\mathbb{B}X$, $\eta_2 = \mathbb{Q}\bar{\mathbb{F}}(\mathbb{B}X\beta_0 + \mathbb{D}\phi_0)$, $\Phi_1 = \mathbb{Q}$, $\Phi_2 = \bar{\mathbb{F}}'\mathbb{Q}$, and $\Phi_3 = \mathbb{Q}\mathbb{G}\mathbb{Q}$. Applying the results: $\mathrm{Var}(\eta'V) = \sigma_0^2\eta'\eta$, $\mathrm{E}(V'\Phi_1V) = \sigma_0^2\mathrm{tr}(\Phi_1)$, $\mathrm{Cov}(\eta'V, V'\Phi_1V) = \sigma_0^3\gamma_0\eta'\phi_1$, and $\mathrm{Cov}(V'\Phi_1V, V'\Phi_2V) = \sigma_0^4(\mathrm{tr}(\Phi_1\Phi_2^\dagger + \kappa_0\phi_1'\phi_2)$, for real vector η and matrices Φ_r with $\phi_r = \texttt{diag}(\Phi_r), r = 1, 2$, we obtain,

$$\mathrm{Var}[S_N^*(\theta_0)] = \frac{1}{\sigma_0^2}\begin{pmatrix} \eta_1'\eta_1, & \frac{1}{2\sigma_0}\gamma_0\eta_1'\varphi_1, & \eta_1'\eta_2 + \sigma_0\gamma_0\eta_1'\varphi_2, & \sigma_0\gamma_0\eta_1'\varphi_3 \\ \sim, & \frac{1}{4\sigma_0^2}\varpi_{11}, & \frac{1}{2}(\varpi_{12} + \frac{1}{\sigma_0}\gamma_0\eta_2'\varphi_2), & \frac{1}{2}\varpi_{13} \\ \sim, & \sim, & \sigma_0^2\varpi_{22} + \eta_2'\eta_2 + 2\sigma_0\gamma_0\eta_2'\varphi_2, & \sigma_0^2\varpi_{23} + \sigma_0\gamma_0\eta_2'\varphi_3 \\ \sim, & \sim, & \sim, & \sigma_0^2\varpi_{33} \end{pmatrix}$$

where $\varphi_r = \texttt{diagv}(\Phi_r)$ and $\varpi_{rv} = \texttt{tr}(\Phi_r\Phi_v^\dagger) + \kappa_0\varphi_r'\varphi_v,\ r, v = 1, 2, 3.$

All the elements in $\Omega_N^*(\theta_0, \phi_0, \gamma_0, \kappa_0) = \frac{1}{N^*}\text{Var}[S_N^*(\theta_0)]$ can be consistently estimated by the plug-in estimators except $\eta_2'\eta_2$, which contains a term $\phi_0'\mathbb{D}'\bar{\mathbb{F}}'\mathbb{Q}\bar{\mathbb{F}}\mathbb{D}\phi_0$ quadratic in ϕ_0 and thus may not be consistently estimated by the plug-in the parameter if $\dim(\phi_0)$ grows as fast as N. The plug-in estimators of the terms linear in ϕ_0 remain consistent.

Denote $\hat{\phi}_{\mathtt{M}} = \hat{\phi}(\hat{\beta}_{\mathtt{M}}, \hat{\delta}_{\mathtt{M}})$, $\hat{\mathbb{P}} = \mathbb{P}(\hat{\rho}_{\mathtt{M}}^*)$, $\hat{\mathbb{Q}} = \mathbb{Q}(\hat{\rho}_{\mathtt{M}}^*)$, and $\hat{\bar{\mathbb{F}}} = \bar{\mathbb{F}}(\hat{\delta}_{\mathtt{M}}^*)$. It is easy to argue that $\hat{\phi}_{\mathtt{M}}'\hat{\mathbb{D}}'\hat{\bar{\mathbb{F}}}'\hat{\mathbb{Q}}\hat{\bar{\mathbb{F}}}\hat{\mathbb{D}}\hat{\phi}_{\mathtt{M}} = \hat{\phi}'(\beta_0, \delta_0)\mathbb{D}'\bar{\mathbb{F}}'\mathbb{Q}\bar{\mathbb{F}}\mathbb{D}\hat{\phi}(\beta_0, \delta_0) + o_p(N^*) = \phi_0'\mathbb{D}'\bar{\mathbb{F}}'\mathbb{Q}\bar{\mathbb{F}}\mathbb{D}\phi_0 + \sigma_0^2\mathtt{tr}(\mathbb{P}\bar{\mathbb{F}}'\mathbb{Q}\bar{\mathbb{F}}) + o_p(N^*)$, where the first equality is due to the consistency of $\hat{\beta}_{\mathtt{M}}$ and $\hat{\delta}_{\mathtt{M}}$. Therefore,

$$\hat{\phi}_{\mathtt{M}}'\hat{\mathbb{D}}'\hat{\bar{\mathbb{F}}}'\hat{\mathbb{Q}}\hat{\bar{\mathbb{F}}}\hat{\mathbb{D}}\hat{\phi}_{\mathtt{M}} - \hat{\sigma}_{\mathtt{M}}^2\mathtt{tr}(\hat{\mathbb{P}}\hat{\bar{\mathbb{F}}}'\hat{\mathbb{Q}}\hat{\bar{\mathbb{F}}}),$$

provides a consistent estimator of $\phi_0'\mathbb{D}'\bar{\mathbb{F}}'\mathbb{Q}\bar{\mathbb{F}}\mathbb{D}\phi_0$. This leads to a $\mathtt{PIC}$ estimator of $\Omega_*(\theta_0)$ as

$$\widehat{\Omega}_{\mathtt{PIC}}^* = \Omega_N^*(\hat{\theta}_{\mathtt{M}}, \hat{\phi}_{\mathtt{M}}, \hat{\gamma}_{\mathtt{M}}, \hat{\kappa}_{\mathtt{M}}) - \mathtt{bdiag}\big(0_{1\times(p+1)},\ \tfrac{\hat{\sigma}_{\mathtt{M}}^2}{N^*}\mathtt{tr}(\hat{\mathbb{P}}\hat{\bar{\mathbb{F}}}'\hat{\mathbb{Q}}\hat{\bar{\mathbb{F}}}),\ 0\big). \tag{4.1}$$

To derive $\hat{\gamma}_{\mathtt{M}}$ and $\hat{\kappa}_{\mathtt{M}}$, let $\imath \in (1, 2, \ldots, N)$ be the combined index, $v_\imath$ and $\tilde{v}_\imath$ be respectively the $\imath$th elements of V and $\tilde{V}$, and $q_{\imath\imath'}$ the $(\imath, \imath')$th element of $\mathbb{Q}$. Then, $\tilde{v}_\imath = \sum_{\imath'=1}^N q_{\imath\imath'}v_{\imath'}$. Finding $\text{E}(\tilde{v}_\imath^3)$ and $\text{E}(\tilde{v}_\imath^4)$, and then summing over $\imath$ and removing $\text{E}(\cdot)$, we obtain,

$$\hat{\gamma}_{\mathtt{M}} = \frac{\sum_{\imath=1}^N \tilde{v}_\imath^3}{\hat{\sigma}_{\mathtt{M}}^3\sum_{\imath=1}^N\sum_{\imath'=1}^N \hat{q}_{\imath\imath'}^3} \quad \text{and} \quad \hat{\kappa}_{\mathtt{M}} = \frac{\sum_{\imath=1}^N \tilde{v}_\imath^4 - 3\hat{\sigma}_{\mathtt{M}}^4\sum_{\imath=1}^N\sum_{\imath'=1}^N\sum_{\imath''=1}^N \hat{q}_{\imath\imath'}^2\hat{q}_{\imath\imath''}^2}{\hat{\sigma}_{\mathtt{M}}^4\sum_{\imath=1}^N\sum_{\imath'=1}^N \hat{q}_{\imath\imath'}^4},$$

a pair of consistent estimators of the skewness and excess kurtosis measures γ_0 and κ_0.

Estimation of Σ_*. The Σ_* matrix can be consistently estimated by its sample analogue $\widehat{\Sigma}_{\mathtt{SA}}^* = \frac{1}{N^*}H_N^*(\hat{\theta}_{\mathtt{M}})$ or by plug-in method $\widehat{\Sigma}_{\mathtt{PI}}^* = \frac{1}{N^*}\text{E}[H_N^*(\theta_0)]|_{\theta_0=\hat{\theta}_{\mathtt{M}}}$, where the negative Hessian, $H_N^*(\theta) = -\frac{\partial}{\partial\theta'}S_N^*(\theta)$, has the following distinct elements:

$$H_{N,\beta\theta}^*(\theta) = \big[\tfrac{1}{\sigma^2}X'\mathbb{C}(\rho)X,\ \tfrac{1}{\sigma^4}X'\mathbb{B}'(\rho)\tilde{V}(\beta,\delta),\ \tfrac{1}{\sigma^2}X'\mathbb{C}(\rho)\mathbf{W}Y,\ \tfrac{1}{\sigma^2}X'\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathbb{G}^\dagger(\rho)\tilde{V}(\beta,\delta)\big],$$

$$H_{N,\sigma^2\theta}^*(\theta) = \big[\ \sim,\ \tfrac{1}{\sigma^6}\tilde{V}'(\beta,\delta)\tilde{V}(\beta,\delta) - \tfrac{N^*}{2\sigma^4},\ \tfrac{1}{\sigma^4}\tilde{V}'(\beta,\delta)\mathbb{B}(\rho)\mathbf{W}Y,\ \tfrac{1}{\sigma^4}\tilde{V}'(\beta,\delta)\mathbb{G}(\rho)\tilde{V}(\beta,\delta)\big],$$

$$H_{N,\lambda\lambda}^*(\theta) = \tfrac{1}{\sigma^2}Y'\mathbf{W}'\mathbb{C}(\rho)\mathbf{W}Y + \mathtt{tr}[\mathbb{Q}(\rho)\bar{\mathbb{F}}^2(\delta)],$$

$$H_{N,\lambda\rho}^*(\theta) = \tfrac{1}{\sigma^2}Y'\mathbf{W}'\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathbb{G}^\dagger(\rho)\tilde{V}(\beta,\delta) - \mathtt{tr}[\mathbb{P}(\rho)\mathbb{G}^\dagger(\rho)\mathbb{Q}(\rho)\bar{\mathbb{F}}(\delta)],$$

$$H_{N,\rho\lambda}^*(\theta) = \tfrac{1}{\sigma^2}Y'\mathbf{W}'\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathbb{G}^\dagger(\rho)\tilde{V}(\beta,\delta),$$

$$H_{N,\rho\rho}^*(\theta) = \tfrac{1}{\sigma^2}\tilde{V}'(\beta,\delta)\big[\mathbb{G}'(\rho)\mathbb{G}(\rho) - \mathbb{G}^\dagger(\rho)\mathbb{P}(\rho)\mathbb{G}^\dagger(\rho)\big]\tilde{V}(\beta,\delta) + \mathtt{tr}[\mathbb{Q}(\rho)\mathbb{G}(\rho)(\mathbb{P}(\rho)\mathbb{G}^\dagger(\rho) + \mathbb{G}(\rho))],$$

where recall $\mathbb{C}(\rho) = \mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathbb{B}(\rho)$ defined in Footnote 5, $\mathbb{P}(\rho) = I_N - \mathbb{Q}(\rho)$, and $\mathbb{Q}(\rho)$ is defined in (3.3).[7]

[7] In deriving $H_N^*(\theta)$ we use: $\frac{\partial}{\partial\rho}\mathbb{Q}(\rho) = \mathbb{Q}(\rho)\mathbb{G}(\rho)\mathbb{P}(\rho) + \mathbb{P}(\rho)\mathbb{G}'(\rho)\mathbb{Q}(\rho)$, $\frac{\partial}{\partial\rho}\mathbb{C}(\rho)$ given in Footnote 5, and $a'\frac{\partial}{\partial\rho}\big[\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathbb{G}(\rho)\mathbb{Q}(\rho)\mathbb{B}(\rho)\big]a = -a'\mathbb{B}'(\rho)\mathbb{Q}(\rho)\big[\mathbb{G}'(\rho)\mathbb{G}(\rho) - \mathbb{G}^\dagger(\rho)\mathbb{P}(\rho)\mathbb{G}^\dagger(\rho)\big]\mathbb{Q}(\rho)\mathbb{B}(\rho)a$ for a vector a.

Corollary 1. *Under the assumptions of Theorem 1, we have, as* $N^* \to \infty$, $\widehat{\Sigma}^*_{\mathtt{SA}} - \Sigma_*(\theta_0) \xrightarrow{p} 0$; $\widehat{\Sigma}^*_{\mathtt{SA}} - \widehat{\Sigma}^*_{\mathtt{PI}} \xrightarrow{p} 0$; *and* $\widehat{\Omega}^*_{\mathtt{PIC}} - \Omega_*(\theta_0) \xrightarrow{p} 0$.

Our Monte Carlo simulation shows that $\widehat{\Sigma}^*_{\mathtt{SA}}$ and $\widehat{\Sigma}^*_{\mathtt{PI}}$ perform equally well. The derivation of $\mathrm{E}[H^*_{N,\beta\theta}(\theta_0)]$ is trivial and is therefore omitted.

Case Ib: Time-varying spatial parameters. When the spatial parameters vary with time, λ and ρ become $T \times 1$ vectors. The special matrices, $\mathcal{W}_t$, $\bar{\mathcal{F}}_t(\delta_t)$ and $\mathcal{G}_t(\rho_t)$, defined respectively by W_t, $\bar{F}_t(\delta_t)$ and $G_t(\rho_t)$, still play an important role in simplifying the derivations.

PIC estimation of $\Omega_{**}(\theta_0)$. The derivation of $\mathrm{Var}[S^{**}_N(\theta_0)]$ is greatly facilitated by new notations $\bar{\mathcal{F}}_t$ and $\mathcal{G}_t$ in that the stochastic parts of $S^{**}_N(\theta_0)$ in (3.8) can all be written in terms of V: $X'\mathbb{B}'\tilde{V} = \eta_1'V, \tilde{V}'\tilde{V} = V'\Phi_1 V$, $Y'\mathcal{W}_t'\mathbb{B}'\tilde{V} = \eta_{2t}'V + V'\Phi_{2t}V$, and $\tilde{V}'\mathcal{G}_t\tilde{V} = V'\Phi_{3t}V$, where $\eta_1 = \mathbb{Q}\mathbb{B}X$, $\eta_{2t} = \mathbb{Q}\bar{\mathcal{F}}_t(\mathbb{B}X\beta_0 + \mathbb{D}\phi_0)$, $\Phi_1 = \mathbb{Q}$, $\Phi_{2t} = \bar{\mathcal{F}}_t'\mathbb{Q}$, and $\Phi_{3t} = \mathbb{Q}\mathcal{G}_t\mathbb{Q}$. Applying the same set of results leading to $\mathrm{Var}[S^*_N(\theta_0)]$, we obtain, for $t, s = 1, \ldots, T$,

$$\mathrm{Var}[S^{**}_N(\theta_0)] =$$

$$\frac{1}{\sigma_0^2}\begin{pmatrix} \eta_1'\eta_1, & \frac{1}{2\sigma_0}\gamma_0\eta_1'\varphi_1, & \{\eta_1'\eta_{2t} + \sigma_0\gamma_0\eta_1'\varphi_{2t}\}, & \{\sigma_0\gamma_0\eta_1'\varphi_{3t}\} \\ \sim, & \{\frac{1}{4\sigma_0^2}\varpi_{11}\}, & \{\frac{1}{2}(\varpi_{12,t} + \frac{1}{\sigma_0}\gamma_0\varphi_1'\eta_{2t})\}, & \{\frac{1}{2}\varpi_{13,t}\} \\ \sim, & \sim, & \{\sigma_0^2\varpi_{22,ts} + \eta_{2t}'\eta_{2s} + \sigma_0\gamma_0(\eta_{2t}'\varphi_{2s} + \varphi_{2t}'\eta_{2s})\}, & \{\sigma_0^2\varpi_{23,ts} + \sigma_0\gamma_0\eta_{2t}'\varphi_{3s}\} \\ \sim, & \sim, & \sim, & \{\sigma_0^2\varpi_{33,ts}\} \end{pmatrix}$$

where $\varphi_1 = \mathtt{diagv}(\Phi_1)$, $\varphi_{rt} = \mathtt{diagv}(\Phi_{rt}), r = 2, 3$, $\varpi_{11} = \mathtt{tr}(\Phi_1\Phi_1^\dagger) + \kappa_0\varphi_1'\varphi_1$, $\varpi_{1r,t} = \mathtt{tr}(\Phi_1\Phi_{rt}^\dagger) + \kappa_0\varphi_1'\varphi_{rt}, r = 2, 3$, $\varpi_{rv,ts} = \mathtt{tr}(\Phi_{rt}\Phi_{vs}^\dagger) + \kappa_0\varphi_{rt}'\varphi_{vs}$, $r, v = 2, 3$, $t, s = 1, \ldots, T$, and $\{\,\cdot\,\}$ denotes a vector or matrix formed by the given elements with $t, s = 1, \ldots, T$.

The plug-in estimate of $\eta_{2t}'\eta_{2s}$ may not be consistent when one or more space dimensions are fixed, as it contains a term quadratic in ϕ_0: $\phi_0'\mathbb{D}'\bar{\mathcal{F}}_t'\mathbb{Q}\bar{\mathcal{F}}_s\mathbb{D}\phi_0$. An argument similar to that leading to (4.1) shows that the bias of the plug-in estimate $\hat{\phi}'\hat{\mathbb{D}}'\hat{\bar{\mathcal{F}}}_t'\hat{\mathbb{Q}}\hat{\bar{\mathcal{F}}}_s\hat{\mathbb{D}}\hat{\phi}$ has a leading term $\sigma_v^2\mathtt{tr}(\mathbb{P}\bar{\mathcal{F}}_t'\mathbb{Q}\bar{\mathcal{F}}_t)$. Therefore, a PIC estimate of $\Omega_{**}(\theta_0)$ is

$$\widehat{\Omega}^{**}_{\mathtt{PIC}} = \Omega_{**}(\hat{\theta}_{\mathtt{M}}, \hat{\gamma}, \hat{\kappa}) - \mathtt{bdiag}\Big(0_{(p+1)\times(p+1)},\ \big\{\tfrac{\hat{\sigma}^2_{v,\mathtt{M}}}{\sqrt{S_t S_s}}\mathtt{tr}(\hat{\mathbb{P}}\hat{\bar{\mathcal{F}}}_t'\hat{\mathbb{Q}}\hat{\bar{\mathcal{F}}}_s)\big\},\ 0\Big), \tag{4.2}$$

where we recall S_t denotes the total number of entities at time t, and $\hat{\gamma}$ and $\hat{\kappa}$ are obtained in a similar way as those for (4.1), but using the new residual estimates $\hat{V}$.

Estimation of Σ_{**}. Again, Σ_{**} can be consistently estimated by either sample analogue or plug-in, $\widehat{\Sigma}^{**}_{\mathtt{SA}} = R^{-1/2}H^{**}_N(\hat{\theta}_{\mathtt{M}})R^{-1/2}$ or $\widehat{\Sigma}^{**}_{\mathtt{PI}} = R^{-1/2}\mathrm{E}[H^{**}_N(\theta_0)]_{\theta_0=\hat{\theta}_{\mathtt{M}}}R^{-1/2}$, where

$H^{**}_N(\theta) = -\frac{\partial}{\partial\theta'}S^{**}_N(\theta)$ and R is defined above Theorem 2. To derive $H^{**}_N(\theta)$, write the stochastic parts of $S^{**}_N(\theta)$ in (3.8) as $X'\mathbb{C}(\rho)U(\beta,\lambda)$, $U'(\beta,\lambda)\mathbb{C}(\rho)U(\beta,\lambda)$, $Y'\mathcal{W}'_t\mathbb{C}(\rho)U(\beta,\lambda)$, and $U'(\beta,\lambda)\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathcal{G}_t(\rho_t)\mathbb{Q}(\rho)\mathbb{B}(\rho)U(\beta,\lambda)$, where $U(\beta,\lambda) = \mathbb{A}(\lambda)Y - X\beta$ (the last component relates to $\frac{\partial}{\partial\rho_t}\mathbb{C}(\rho)$ given below (3.7)). $H^{**}_N(\theta)$ has the distinct elements:

$$H^{**}_{N,\beta\theta}(\theta) = \tfrac{1}{\sigma^2}\big[X'\mathbb{C}(\rho)X,\ \tfrac{1}{\sigma^2}\mathbb{X}'(\rho)\tilde{V}(\beta,\delta),\ \{X'\mathbb{C}(\rho)\mathcal{W}_tY\},\ \{\mathbb{X}'(\rho)\mathbb{Q}(\rho)\mathcal{G}^\dagger_t(\rho_t)\tilde{V}(\beta,\delta)\}\big],$$

$$H^{**}_{N,\sigma^2\theta}(\theta) = \tfrac{1}{\sigma^4}\big[\sim,\ \tfrac{1}{\sigma^2}\tilde{V}'(\beta,\delta)\tilde{V}(\beta,\delta) - \tfrac{N^*}{2},\ \{\tilde{V}'(\beta,\delta)\mathbb{B}(\rho)\mathcal{W}_tY\},\ \{\tilde{V}'(\beta,\delta)\mathcal{G}_t(\rho_t)\tilde{V}(\beta,\delta)\}\big],$$

$$H^{**}_{N,\lambda_t\lambda_s}(\theta) = \big\{\tfrac{1}{\sigma^2}Y'\mathcal{W}'_t\mathbb{C}(\rho)\mathcal{W}_sY + \mu_{\lambda\lambda,ts}\big\},$$

$$H^{**}_{N,\lambda_t\rho_s}(\theta) = \big\{\tfrac{1}{\sigma^2}Y'\mathcal{W}'_t\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathcal{G}^\dagger_s(\rho_t)\tilde{V}(\beta,\delta) + \mu_{\lambda\rho,ts}\big\},$$

$$H^{**}_{N,\rho_t\lambda_s}(\theta) = \big\{\tfrac{1}{\sigma^2}\tilde{V}'(\beta,\delta)\mathcal{G}^\dagger_t(\rho_t)\mathbb{Q}(\rho)\mathbb{B}(\rho)\mathcal{W}_sY\big\},$$

$$H^{**}_{N,\rho_t\rho_s}(\theta) = \big\{\tfrac{1}{\sigma^2}\tilde{V}'(\beta,\delta)\mathcal{K}_{ts}(\rho)\tilde{V}'(\beta,\delta) + \mu_{\rho\rho,ts}\big\}.$$

where $\mu_{\lambda\lambda,ts} = \mathtt{tr}[\bar{\mathcal{F}}^2_t(\delta_t)\mathbb{Q}(\rho)]$ if $s=t$, else 0; $\mu_{\lambda\rho,ts} = \mathtt{tr}[\mathbb{Q}(\rho)\bar{\mathcal{F}}_t(\delta_t)\mathbb{P}(\rho)\mathcal{G}^\dagger_t(\rho_t)]$ if $s=t$, else $\mathtt{tr}[\bar{\mathcal{F}}_t(\delta_t)\frac{\partial}{\partial\rho_s}\mathbb{Q}(\rho)]$; $\mu_{\rho\rho,ts} = \mathtt{tr}[\mathcal{G}^2_t(\rho_t)\mathbb{Q}(\rho) + \mathcal{G}_t(\rho_t)\frac{\partial}{\partial\rho_t}\mathbb{Q}(\rho)]$ if $s=t$, else $\mathtt{tr}[\mathcal{G}_t(\rho_t)\frac{\partial}{\partial\rho_s}\mathbb{Q}(\rho)]$, and $\mathcal{K}_{ts}(\rho) = \mathcal{G}'_t(\rho_t)\mathcal{G}_t(\rho_t) - \mathcal{G}^\dagger_t(\rho_t)\mathbb{P}(\rho)\mathcal{G}^\dagger_t(\rho_t)$ if $s=t$, else $\mathcal{G}^\dagger_t(\rho_t)[\mathcal{G}_s(\rho_s) - \mathbb{P}(\rho)\mathcal{G}^\dagger_s(\rho_s)]$.[8]

Corollary 2. *Under the assumptions of Theorem 2, we have, as $N^*\to\infty$, $\widehat{\Sigma}^{**}_{\mathtt{SA}} - \Sigma_{**}(\theta_0) \xrightarrow{p} 0$; $\widehat{\Sigma}^{**}_{\mathtt{PI}} - \Sigma_{**}(\theta_0) \xrightarrow{p} 0$; and $\widehat{\Omega}^{**}_{\mathtt{PIC}} - \Omega_{**}(\theta_0) \xrightarrow{p} 0$.*

Our Monte Carlo simulation shows that the estimator $\widehat{\Sigma}^{**}_{\mathtt{PI}}$ performs better (numerically more stable) than $\widehat{\Sigma}^{**}_{\mathtt{SA}}$. The derivation of $\mathrm{E}[H^{**}_{N,\beta\theta}(\theta_0)]$ is trivial and is therefore omitted.

4.2. Inferences under unknown heteroskedasticity

Allowing the errors to be heteroskedastic in all dimensions does not complicate much the estimation of Σ_r and Σ_{rr}, but does complicate very much on the estimation of Ω_r and Ω_{rr}. Fortunately, the key matrices in the quadratic forms in the AQS functions $S^r(\vartheta_0)$ and $S^{rr}(\vartheta_0)$ have zero diagonals, and hence $\mathrm{Var}[S^r(\vartheta_0)]$ and $\mathrm{Var}[S^{rr}(\vartheta_0)]$ involve only ϕ and h, not the (possibly varying) third and fourth moments of the errors. This makes it possible to analytically estimate Ω_r and Ω_{rr} and then correct the effects of plugging in $\hat{\phi}_{\mathtt{M}}$ and $\hat{h}$.

Case IIa: Time-invariant spatial parameters.

[8]Derivation of $H^{**}_N(\theta)$ benefits greatly from the results: $\frac{\partial}{\partial\rho_t}\mathbb{Q}(\rho) = \mathbb{Q}(\rho)\mathcal{G}_t(\rho_t)\mathbb{P}(\rho) + \mathbb{P}(\rho)\mathcal{G}'_t(\rho_t)\mathbb{Q}(\rho)$, $\frac{\partial}{\partial\rho_t}\mathbb{C}(\rho)$ given below (3.7), $a'\frac{\partial}{\partial\rho_s}[\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathcal{G}_t(\rho_t)\mathbb{Q}(\rho)\mathbb{B}(\rho)]a = a'\mathbb{B}'(\rho)\mathbb{Q}(\rho)\big[\mathcal{G}'_t(\rho_t)\mathcal{G}_t(\rho_t) - \mathcal{G}^\dagger_t(\rho_t)\mathbb{P}(\rho)\mathcal{G}^\dagger_t(\rho_t)\big]\mathbb{Q}(\rho)\mathbb{B}(\rho)a$ if $s=t$, and $a'\mathbb{B}'(\rho)\mathbb{Q}(\rho)\mathcal{G}^\dagger_t(\rho_t)[\mathcal{G}_s(\rho_s) - \mathbb{P}(\rho)\mathcal{G}^\dagger_s(\rho_s)]\mathbb{Q}(\rho)\mathbb{B}(\rho)a$ if $s\neq t$, for a vector a.

PIC Estimation of Ω_r. To derive $\text{Var}[S_N^r(\vartheta_0)]$, write the RAQS function (3.13) at ϑ_0:

$$S_N^r(\vartheta_0) = \big[X'\mathbb{B}'\mathbb{Q}V;\ \eta_0'\mathbf{F}V + V'\mathbf{F}V;\ \phi_0'\mathbb{D}'\mathbf{G}V + V'\mathbf{G}V\big], \tag{4.3}$$

where $\eta_0 = \mathbb{X}\beta_0 + \mathbb{D}\phi_0$, $\mathbf{F} = [\bar{\mathbb{F}}'(\delta_0) - F_d(\delta_0)]\mathbb{Q}(\rho_0)$, and $\mathbf{G} = [\bar{\mathbb{G}}(\rho_0) - G_d(\rho_0)]\mathbb{Q}(\rho_0)$.

Noting that $\mathbf{F}$ and $\mathbf{G}$ both have zero diagonals, we have, $\text{E}(V'\omega V) = 0$, $\text{Cov}(a'V, V'\omega V) = 0$, and $\text{Cov}(V'\omega V, V'\varpi V) = \texttt{tr}(H\omega H\varpi^\dagger) = h'(\omega \odot \varpi^\dagger)h$, for $\omega, \varpi = \mathbf{F}, \mathbf{G}$, and $\varpi^\dagger = \varpi + \varpi'$. A simple expression for $\text{Var}[S_N^r(\vartheta_0)]$ free of 3rd and 4th moments of the errors takes the form:

$$\text{Var}[S_N^r(\vartheta_0)] =$$
$$\begin{pmatrix} X'\mathbb{B}'\mathbb{Q}H\mathbb{Q}\mathbb{B}X & X'\mathbb{B}'\mathbb{Q}H\mathbf{F}'\eta_0 & X'\mathbb{B}'\mathbb{Q}H\mathbf{G}'\mathbb{D}\phi_0 \\ \sim & \text{tr}(H\mathbf{F}H\mathbf{F}^\dagger) + \eta_0'\mathbf{F}H\mathbf{F}'\eta_0 & \text{tr}(H\mathbf{F}H\mathbf{G}^\dagger) + \eta_0'\mathbf{F}H\mathbf{G}'\mathbb{D}\phi_0 \\ \sim & \sim & \text{tr}(H\mathbf{G}H\mathbf{G}^\dagger) + \phi_0'\mathbb{D}'\mathbf{G}H\mathbf{G}'\mathbb{D}\phi_0 \end{pmatrix}. \tag{4.4}$$

This opens up a possible way for a plug-in estimation of $\text{Var}[S_N^r(\vartheta_0)]$ followed by a correction if necessary. The idea is that we first do a plug-in estimation, i.e., plugging $\hat{\vartheta}_{\mathtt{M}}^r$ for ϑ_0, $\hat{\phi}_{\mathtt{M}}^r$ for ϕ_0, and $\hat{H}$ (to be defined below) for H, and then we correct the possible inconsistency caused by such replacements. We argue that only the terms quadratic in ϕ_0, i.e., $\phi_0'\mathbb{D}'\omega H\varpi'\mathbb{D}\phi_0$ (note $\eta_0 = \mathbb{X}\beta_0 + \mathbb{D}\phi_0$), and the terms of the form $\texttt{tr}(H\omega H\varpi^\dagger)$ may cause an issue. Let $\hat{\phi}$, $\hat{\mathbb{P}}$, $\hat{\mathbb{Q}}$, $\hat{\omega}$, and $\hat{\varpi}$ be the estimates of the quantities based on $\hat{\vartheta}_{\mathtt{M}}^r$.

First, for a given H, similar arguments to those leading to the $\texttt{PIC}$ estimate (4.1) show that $\hat{\phi}_{\mathtt{M}}'\hat{\mathbb{D}}'\hat{\omega}H\hat{\varpi}'\hat{\mathbb{D}}\hat{\phi}_{\mathtt{M}} = \hat{\phi}'(\beta_0, \delta_0)\mathbb{D}'\omega H\varpi'\mathbb{D}\hat{\phi}(\beta_0, \delta_0) + o_p(N) = \phi_0'\mathbb{D}'\omega H\varpi'\mathbb{D}\phi_0 + \texttt{tr}(\mathbb{P}\omega H\varpi'\mathbb{P}H) + o_p(N)$, where the first equality is due to the consistency of $\hat{\beta}_{\mathtt{M}}^r$ and $\hat{\delta}_{\mathtt{M}}^r$. Therefore,

$$\hat{\phi}_{\mathtt{M}}'\hat{\mathbb{D}}'\hat{\omega}H\hat{\varpi}'\hat{\mathbb{D}}\hat{\phi}_{\mathtt{M}} - \texttt{tr}(H\hat{\mathbb{P}}\hat{\omega}H\hat{\varpi}'\hat{\mathbb{P}}), \tag{4.5}$$

provides a consistent estimator of $\phi_0'\mathbb{D}'\omega H\varpi'\mathbb{D}\phi_0$, for a given H, where $\hat{\omega}, \hat{\varpi} = \hat{\mathbf{F}}, \hat{\mathbf{G}}$.

Next, we examine the costs of replacing $\texttt{tr}(H\omega H\varpi)$ by its plug-in estimate $\texttt{tr}(\hat{H}\hat{\omega}\hat{H}\hat{\varpi})$, for $\omega = \mathbf{F}, \mathbf{G}, \mathbb{P}\mathbf{F}, \mathbb{P}\mathbf{G}$, and $\varpi = \mathbf{F}^\dagger, \mathbf{G}^\dagger, \mathbb{P}\mathbf{F}, \mathbb{P}\mathbf{G}$. We have $\text{E}(\tilde{V}^2) = \text{E}[(\mathbb{Q}V) \odot (\mathbb{Q}V)) = (\mathbb{Q} \odot \mathbb{Q})h$. This naturally leads to an estimate of h:

$$\hat{h} = [\mathbb{Q}(\hat{\rho}_{\mathtt{M}}) \odot \mathbb{Q}(\hat{\rho}_{\mathtt{M}})]^{-1}\ [\hat{V}(\hat{\vartheta}_{\mathtt{M}}) \odot \hat{V}(\hat{\vartheta}_{\mathtt{M}})],$$

and thus $\hat{H} = \texttt{diag}(\hat{h})$, where $\odot$ denotes the Hadamard product. From (4.4) and (4.5), we see that ω and ϖ are formed by $\mathbf{F}$ and $\mathbf{G}$ or $\mathbb{P}\mathbf{F}$ and $\mathbb{P}\mathbf{G}$, containing only the δ_0 parameters. Hence, the transitions from $\texttt{tr}(H\omega H\varpi)$ to $\texttt{tr}(H\hat{\omega}H\hat{\varpi})$ is costless asymptotically.

For the transitions from $\mathtt{tr}(H\omega H\varpi)$ to $\mathtt{tr}(\hat{H}\omega\hat{H}\varpi)$, let $\Pi(\rho) = [\mathbb{Q}(\rho)\odot\mathbb{Q}(\rho)]^{-1}$, and $\pi_i'(\rho)$ be its ith row. By a Taylor expansion of $\hat{h}_i = \pi_i'(\hat{\rho}_{\mathtt{M}})\hat{V}^2(\hat{\vartheta}_{\mathtt{M}})$ (the ith element of $\hat{h}$) at ϑ_0, combined with a first-order expansion of $\hat{\vartheta}_{\mathtt{M}} - \vartheta_0$ obtained through a Taylor expansion of $S_N^r(\hat{\vartheta}_{\mathtt{M}})$ at ϑ_0, we have $\hat{h}_i = h_i + \frac{1}{N^*}\Gamma_i'(\vartheta_0)\Sigma_r^{-1}(\vartheta_0)S_N^r(\vartheta_0) + o_p(1)$, which leads to

$$\hat{h}_i\hat{h}_j = h_ih_j + \tfrac{1}{N^*}\Gamma_i'(\vartheta_0)\Sigma_r^{-1}(\vartheta_0)\Omega_r(\vartheta_0)\Sigma_r'^{-1}(\vartheta_0)\Gamma_j(\vartheta_0) + o_p(1),\ i,j = 1,2,\ldots N, \tag{4.6}$$

where $\Sigma_r(\vartheta_0)$ and $\Omega_r(\vartheta_0)$ are defined in Theorem 3, and $\Gamma_i'(\vartheta_0) = \mathrm{E}[\frac{\partial}{\partial\vartheta_0'}(\pi_i'(\rho_0)\tilde{V}^2(\vartheta_0))] = \mathrm{E}\big\{-2\pi_i'[(\mathbb{Q}\mathbb{X})\odot\tilde{V}],\ -2\pi_i'[(\mathbb{Q}\mathbb{B}WY)\odot\tilde{V}],\ \dot{\pi}_i'\tilde{V}^2 + 2\pi_i'[\tilde{V}\odot((\dot{\mathbb{Q}}-\mathbb{Q}\mathbb{G})(\mathbb{Y}-\mathbb{X}\beta_0))]\big\}$ with $\dot{\pi}_i'$ being the ith row of $\frac{\partial}{\partial\rho_0}\Pi(\rho_0) = -2\Pi(\mathbb{Q}\odot\dot{\mathbb{Q}})\Pi$ and $\dot{\mathbb{Q}} = \frac{\partial}{\partial\rho}\mathbb{Q}(\rho_0)$ given in Footnote 7.

Denote the second term of (4.6) by $\upsilon_{ij}(\vartheta)$ and let $\Upsilon_r \equiv \Upsilon_r(\vartheta_0)$ be the $N\times N$ matrix formed by $\upsilon_{ij}(\vartheta_0)$. Then, the expansion (4.6) leads quickly to

$$\mathtt{tr}(\hat{H}\omega\hat{H}\varpi) = \mathtt{tr}(H\omega H\varpi)] + \mathtt{tr}[(\omega\odot\Upsilon_r)\varpi] + o_p(N^*). \tag{4.7}$$

Therefore, a $\mathtt{PIC}$ estimator of $\Omega_r(\vartheta_0)$ is

$$\widehat{\Omega}^r_{\mathtt{PIC}} = \Omega_N^r(\hat{\vartheta}^r_{\mathtt{M}},\hat{\phi}^r_{\mathtt{M}},\hat{h}) - \mathrm{Bias}^r_{\phi}(\hat{\vartheta}^r_{\mathtt{M}},\hat{\phi}^r_{\mathtt{M}},\hat{h}) - \mathrm{Bias}^r_{H}(\hat{\vartheta}^r_{\mathtt{M}},\hat{\phi}^r_{\mathtt{M}},\hat{h}), \tag{4.8}$$

where the first correction term has entries $\frac{1}{N^*}\mathtt{tr}(\hat{\mathbb{P}}\hat{\omega}\hat{H}\hat{\varpi}'\hat{\mathbb{P}}\hat{H})$ in its δ-δ block, for $\hat{\omega},\hat{\varpi} = \hat{\mathbf{F}},\hat{\mathbf{G}}$, and zero elsewhere; the second correction term has entries $\frac{1}{N^*}\mathtt{tr}[(\hat{\omega}\odot\hat{\Upsilon}_r)\hat{\varpi}^\dagger - (\hat{\mathbb{P}}\hat{\omega}\odot\hat{\Upsilon}_r)\hat{\varpi}'\hat{\mathbb{P}}]$ in its δ-δ block, and $\hat{\omega},\hat{\varpi} = \hat{\mathbf{F}},\hat{\mathbf{G}}$, and zero elsewhere.

Estimation of Σ_r. Similar to the precious cases, Σ_r can be consistently estimated by its sample analogue or plug-in. To derive the negative Hessian matrix, rewrite the RAQS function (3.13) as $S_N^r(\vartheta) = \big[X'\mathbb{C}(\rho)(\mathbb{A}(\lambda)Y - X\beta);\mathbb{Y}'(\delta)\mathbf{F}(\delta)\mathcal{E}(\delta,\beta);\mathcal{E}'(\delta,\beta)\mathbf{G}(\rho)\mathcal{E}(\delta,\beta)\big]$, where $\mathbf{F}(\delta)$ and $\mathbf{G}(\rho)$ are $\mathbf{F}$ and $\mathbf{G}$ at general λ and ρ, and $\mathcal{E}(\delta,\beta) = \mathbb{Y}(\delta) - \mathbb{X}(\rho)\beta$. Then, $\mathcal{H}_N^r(\vartheta) = -\frac{\partial}{\partial\vartheta'}S_N^r(\vartheta)$ has the form:

$$\mathcal{H}_N^r(\vartheta) = \begin{pmatrix} X'\mathbb{C}(\rho)X & X'\mathbb{C}(\rho)\mathbf{W}Y & \mathbb{X}'(\rho)\mathbb{Q}(\rho)\mathbb{G}^\dagger(\rho)\tilde{V}(\vartheta) \\ \mathbb{Y}'(\delta)\mathbf{F}(\delta)\mathbb{X}(\rho) & \mathbb{Y}'(\delta)\Phi_1(\delta)\mathbb{Y}(\delta) + \mathbb{Y}'(\delta)\eta_1 & \mathbb{Y}'(\delta)\Phi_2(\delta)\mathcal{E}(\beta,\delta) \\ \mathcal{E}'(\delta,\beta)\mathbf{G}^\dagger(\rho)\mathbb{X}(\rho) & \mathcal{E}'(\delta,\beta)\mathbf{G}^\dagger(\rho)\mathbb{B}(\rho)\mathbf{W}Y & \mathcal{E}'(\delta,\beta)\Phi_3(\rho)\mathcal{E}(\delta,\beta) \end{pmatrix} \tag{4.9}$$

where $\Phi_1(\delta) = \mathbf{F}^\dagger(\delta)\bar{\mathbb{F}}(\delta) - \mathbf{F}_\lambda(\delta)$, $\Phi_2(\delta) = \mathbb{G}'(\rho)\mathbf{F}(\delta) + \mathbf{F}(\delta)\mathbb{G}(\rho) - \mathbf{F}_\rho(\delta)$, $\Phi_3(\rho) = \mathbf{G}^\dagger(\rho)\mathbb{G}(\rho) - \mathbf{G}_\rho(\rho)$, $\eta_1 = [\mathbf{F}_\lambda(\delta) - \bar{\mathbb{F}}'(\delta)\mathbf{F}(\delta)]\mathbb{X}(\rho)\beta$, $\mathbf{F}_\omega(\delta) = \frac{\partial}{\partial\omega}\mathbf{F}(\delta)$, $\omega = \lambda,\rho$, and $\mathbf{G}_\rho(\rho) = \frac{\partial}{\partial\rho}\mathbf{G}(\rho)$. These partial derivatives are derived in Appendix A. They can also be approximated by numerical derivatives, for example, $\frac{\partial}{\partial\lambda}\mathbf{F}(\delta) \approx \frac{\mathbf{F}(\lambda+\varepsilon,\rho)-\mathbf{F}(\lambda-\varepsilon,\rho)}{2\varepsilon}$, for a small positive number ε.

Therefore, a sample analogue estimate of Σ_r is $\widehat{\Sigma}^r_{\texttt{SA}} = \frac{1}{N^*}\mathcal{H}^r_N(\hat{\vartheta}_{\texttt{M}})$. Under heteroskedasticity, the plug-in estimate is more complicated as the analytical expression of $\mathrm{E}[\mathcal{H}^r_N(\vartheta_0)]$ involves the unknown h. Our Monte Carlo experiments show that $\widehat{\Sigma}^r_{\texttt{SA}}$ works very well.

Corollary 3. *Under the assumptions of Theorem 3, we have, as $N^* \to \infty$, $\widehat{\Sigma}^r_{\texttt{SA}} - \Sigma_r(\vartheta_0) \xrightarrow{p} 0$; $\widehat{\Sigma}^r_{\texttt{SA}} - \widehat{\Sigma}^r_{\texttt{PI}} \xrightarrow{p} 0$; and $\widehat{\Omega}^r_{\texttt{PIC}} - \Omega_r(\vartheta_0) \xrightarrow{p} 0$.*

Case IIb: Time-varying spatial parameters. Similar ideas can be implemented when spatial parameters are allowed to change over time along with spatial weight matrices.

PIC Estimation of Ω_{rr}. Write the RAQS function given in (3.15) at the true parameters as, similarly to (4.3),

$$S^{rr}_N(\vartheta_0) = \big(X'\mathbb{B}'\mathbb{Q}V;\ \{\eta_0'\mathbf{F}_tV + V'\mathbf{F}_tV\};\ \{\phi_0'\mathbb{D}'\mathbf{G}_tV + V'\mathbf{G}_tV\}\big), \tag{4.10}$$

where η_0 is defined in (4.3), $\mathbf{F}_t = (\bar{\mathcal{F}}_t' - \mathcal{F}^d_t)\mathbb{Q}$ and $\mathbf{G}_t = (\bar{\mathcal{G}}_t - \mathcal{G}^d_t)\mathbb{Q}$, with all the quantities being defined in (3.15) and evaluated at the true parameter values. It is easy to see that $\mathbf{F}_t$ and $\mathbf{G}_t$ have zero diagonals, and therefore the same way leading to (4.4) gives the following:

$$\mathrm{Var}[S^{rr}_N(\vartheta_0)] =$$
$$\begin{pmatrix} X'\mathbb{B}'\mathbb{Q}H\mathbb{Q}\mathbb{B}X & \{X'\mathbb{B}'\mathbb{Q}H\mathbf{F}_t'\eta_0\} & \{X'\mathbb{B}'\mathbb{Q}H\mathbf{G}_t'\mathbb{D}\phi_0\} \\ \sim & \{\mathrm{tr}(H\mathbf{F}_tH\mathbf{F}^\dagger_s) + \eta_0'\mathbf{F}_tH\mathbf{F}_s'\eta_0\} & \{\mathrm{tr}(H\mathbf{F}_tH\mathbf{G}^\dagger_s) + \eta_0'\mathbf{F}_tH\mathbf{G}_s'\mathbb{D}\phi_0\} \\ \sim & \sim & \{\mathrm{tr}(H\mathbf{G}_tH\mathbf{G}^\dagger_s) + \phi_0'\mathbb{D}'\mathbf{G}_tH\mathbf{G}_s'\mathbb{D}\phi_0\} \end{pmatrix}. \tag{4.11}$$

To consistently estimate $\mathrm{Var}[S^{rr}_N(\vartheta_0)]$, note that the robust VC matrix (4.11) has a structure identical to that given in (4.4) for the case of time-invariant spatial parameters, except that it is indexed by t and s. Therefore, the correction methods for (4.4) can be directly extended to give a `PIC` estimator of $\mathrm{Var}[S^{rr}_N(\vartheta_0)]$. First, for a given H, we show that

$$\hat{\phi}'\hat{\mathbb{D}}'\hat{\omega}_tH\hat{\varpi}_s'\hat{\mathbb{D}}\hat{\phi} - \mathtt{tr}(\hat{\mathbb{P}}'\hat{\omega}_t'H\hat{\varpi}_s\hat{\mathbb{P}}H) \tag{4.12}$$

provides a consistent estimate of $\phi_0'\mathbb{D}'\omega_tH\varpi_s'\mathbb{D}\phi_0$, for $\omega_t, \varpi_s = \mathbf{F}_t, \mathbf{G_s}$, and $t, s = 1, \ldots, T$.

Next, a valid $\hat{H}$ can be obtained in a similar manner as for Case IIa. Replacing H by $\hat{H}$ in the second term above incurs a non-negligible cost that must be corrected. Furthermore, replacing H by $\hat{H}$ in the terms of the form $\mathrm{tr}(H\hat{\omega}_tH\hat{\varpi}^\dagger_s)$, $\hat{\omega}_t = \mathbf{F}_t, \mathbf{G}_t$, and $\hat{\varpi}^\dagger_s = \mathbf{F}^\dagger_s, \mathbf{G}^\dagger_s$, incurs non-negligible costs that must be corrected as well. For each (t, s), these corrections take the same forms as those for Case IIa. Therefore, the PIC estimator of $\Omega_{rr}(\vartheta_0)$ is defined as

$$\widehat{\Omega}^{rr}_{\mathtt{PIC}} = \Omega^{rr}_N(\hat{\vartheta}^{rr}_{\mathtt{M}}, \hat{\phi}^{rr}_{\mathtt{M}}, \hat{h}) - \text{Bias}^{rr}_{\phi}(\hat{\vartheta}^{rr}_{\mathtt{M}}, \hat{\phi}^{rr}_{\mathtt{M}}, \hat{h}) - \text{Bias}^{rr}_{H}(\hat{\vartheta}^{rr}_{\mathtt{M}}, \hat{\phi}^{rr}_{\mathtt{M}}, \hat{h}), \tag{4.13}$$

where $\text{Bias}^{rr}_{\phi}(\hat{\vartheta}^{rr}_{\mathtt{M}}, \hat{\phi}^{rr}_{\mathtt{M}}, \hat{h})$ has entries $\{\frac{1}{N^*}\mathtt{tr}(\hat{\mathbb{P}}\hat{\omega}_t\hat{H}\hat{\varpi}'_s\hat{\mathbb{P}}\hat{H})\}$ in its δ-δ block ($2T \times 2T$) and zero elsewhere, and $\text{Bias}^{rr}_{H}(\hat{\vartheta}^{rr}_{\mathtt{M}}, \hat{\phi}^{rr}_{\mathtt{M}}, \hat{h})$ has entries $\{\frac{1}{N^*}\mathtt{tr}\big[(\hat{\omega}_t \odot \hat{\Upsilon}_{rr})\hat{\varpi}^{\dagger}_s - (\hat{\mathbb{P}}\hat{\omega}_t \odot \hat{\Upsilon}_{rr})\hat{\varpi}'_s\hat{\mathbb{P}}\big]\}$ in its δ-δ block ($2T \times 2T$) and zero elsewhere, for $\omega_t = \mathbf{F}_t, \mathbf{G}_t$, and $\varpi^{\dagger}_s = \mathbf{F}^{\dagger}_s, \mathbf{G}^{\dagger}_s$.

The Υ_{rr} matrix has the same form as Υ_r defined in (4.6) for the time-invariant case,

$$\Upsilon_{rr} = \big\{\tfrac{1}{N^*}\Gamma'_i(\vartheta_0)\Sigma^{-1}_{rr}(\vartheta_0)\Omega_{rr}(\vartheta_0)\Sigma'^{-1}_{rr}(\vartheta_0)\Gamma_j(\vartheta_0)\big\},$$

where $\Sigma_{rr}(\vartheta_0)$ and $\Omega_{rr}(\vartheta_0)$ are defined in Theorem 4, and $\Gamma'_i(\vartheta_0) = \mathrm{E}[\frac{\partial}{\partial\vartheta'_0}(\pi'_i(\rho_0)\tilde{V}^2(\vartheta_0))] = \mathrm{E}\big\{-2\pi'_i[(\mathbb{QX})\odot\tilde{V}], \big\{-2\pi'_i[(\mathbb{QB}\mathcal{W}_tY)\odot\tilde{V}]\big\}, \big\{\dot{\pi}'_{i,t}\tilde{V}^2 + 2\pi'_i[\tilde{V}\odot((\dot{\mathbb{Q}}_t - \mathbb{Q}\mathcal{G}_t)(\mathbb{Y} - \mathbb{X}\beta_0))]\big\}\big\}$ with $\dot{\pi}'_{i,t}$ being the ith row of $\frac{\partial}{\partial\rho_{t0}}\Pi(\rho_0) = -2\Pi[\mathbb{Q}\odot\dot{\mathbb{Q}}_t]\Pi$ and $\dot{\mathbb{Q}}_t = \frac{\partial}{\partial\rho_{t0}}\mathbb{Q}(\rho_0)$ given in Footnote 8.

Estimation of Σ_{rr}. To derive the negative Hessian matrix, rewrite the RAQS function (3.15) as $S^{rr}_N(\vartheta) = \big[X'\mathbb{C}(\rho)(\mathbb{A}(\lambda)Y - X\beta);\ \{\mathbb{Y}'(\delta)\mathbf{F}_t(\delta_t,\rho)\mathcal{E}(\vartheta)\};\ \{\mathcal{E}'(\vartheta)\mathbf{G}_t(\rho_t,\rho)\mathcal{E}(\vartheta)\}\big]$, where $\mathcal{E}(\vartheta) = \mathbb{Y}(\delta) - \mathbb{X}(\rho)\beta$ defined above (4.9), $\mathbf{F}_t(\delta_t,\rho) = (\bar{\mathcal{F}}'_t(\delta_t) - \mathcal{F}^d_t(\delta_t,\rho))\mathbb{Q}(\rho)$ and $\mathbf{G}_t(\rho_t,\rho) = (\bar{\mathcal{G}}_t(\rho_t,\rho) - \mathcal{G}^d_t(\rho_t,\rho))\mathbb{Q}(\rho)$, which are $\mathbf{F}_t$ and $\mathbf{G}_t$ at general parameter values. Then, $\mathcal{H}^{rr}_N(\vartheta) = -\frac{\partial}{\partial\vartheta}S^{rr}_N(\vartheta)$ has the following expression:

$$\mathcal{H}^{rr}_N(\vartheta) = \begin{pmatrix} X'\mathbb{C}(\rho)X & \{X'\mathbb{C}(\rho)\mathcal{W}_tY\} & \{\mathbb{X}'(\rho)\mathbb{Q}(\rho)\mathcal{G}^{\dagger}_t(\rho_t)\tilde{V}(\vartheta)\} \\ \{\mathbb{Y}'(\delta)\mathbf{F}_t(\delta_t)\mathbb{X}(\rho)\} & \{\mathbb{Y}'(\delta)\Phi_{1,ts}\mathbb{Y}(\delta) + \mathbb{Y}'(\delta)\eta_{1,ts}\} & \{\mathbb{Y}'(\delta)\Phi_{2,ts}\mathcal{E}(\vartheta)\} \\ \{\mathcal{E}'(\vartheta)\mathbf{G}^{\dagger}_t(\rho_t,\rho)\mathbb{X}(\rho)\} & \{\mathcal{E}'(\vartheta)\mathbf{G}^{\dagger}_t(\rho_t,\rho)\mathbb{B}(\rho)\mathcal{W}_sY\} & \{\mathcal{E}'(\vartheta)\Phi_{3,ts}\mathcal{E}(\vartheta)\} \end{pmatrix} \tag{4.14}$$

where $\Phi_{1,ts} = \mathbf{F}^{\dagger}_t(\delta_t,\rho)\bar{\mathcal{F}}_s(\delta_s) - \frac{\partial}{\partial\lambda_s}\mathbf{F}_t(\delta_t,\rho)$, $\eta_{1,ts} = [\frac{\partial}{\partial\lambda_s}\mathbf{F}_t(\delta_t,\rho) - \bar{\mathcal{F}}'_s(\delta_t)\mathbf{F}_t(\delta_t,\rho)]\mathbb{X}(\rho)\beta$, $\Phi_{2,ts} = \mathcal{G}'_s(\rho_s)\mathbf{F}_t(\delta_t,\rho) + \mathbf{F}_t(\delta_t,\rho)\mathcal{G}_s(\rho_s) - \frac{\partial}{\partial\rho_s}\mathbf{F}_t(\delta_t,\rho)$, $\Phi_{3,ts} = \mathcal{G}'_s(\rho_s)\mathbf{G}_t(\rho_t,\rho) + \mathbf{G}_t(\rho_t,\rho)\mathcal{G}_s(\rho_s) - \frac{\partial}{\partial\rho_s}\mathbf{G}_t(\rho_t,\rho)$. Again, the partial derivatives have complicate expressions and can simply be approximated by numerical derivatives. Their detailed expressions are given in Appendix A for users who are more comfortable with analytical derivatives.

Therefore, a sample analogue estimate of Σ_{rr} is, $\widehat{\Sigma}^{rr}_{\mathtt{SA}} = R^{-1/2}\mathcal{H}^{rr}_N(\hat{\vartheta}_{\mathtt{M}})R^{-1/2}$, where R is defined above Theorem 4. Under heteroskedasticity, it is more complicated to use the plug-in estimate as the analytical expression of $\mathrm{E}[\mathcal{H}^{rr}_N(\vartheta_0)]$ involves the unknown h.

Corollary 4. *Under the assumptions of Theorem 4, we have, as $N^* \to \infty$, $\widehat{\Sigma}^{rr}_{\mathtt{SA}} - \Sigma_{rr}(\vartheta_0) \xrightarrow{p} 0$; and $\widehat{\Omega}^{rr}_{\mathtt{PIC}} - \Omega_{rr}(\vartheta_0) \xrightarrow{p} 0$.*

Our Monte Carlo results show that $\widehat{\Sigma}^{rr}_{\mathtt{SA}}$ works well in finite samples. The proofs of Corollaries 1-4 are given in `Supplemental Materials` (Appendix B).

5. Monte Carlo Study

Extensive Monte Carlo experiments were conducted to investigate the finite sample performance of the proposed M-estimation and inference methods, in the presence of multidimensional and interactive fixed effects. To illustrate the important ideas and methods presented in Sec. 2.2 about the fixed effects specifications, we consider the following three representative FE specifications in our Monte Carlo experiments:

`FE-1`: The (i, j, t) specification given in Table A, without STIC effect,

`FE-2`: The (i, j, t) specification given in Table A, with STIC,

`FE-3`: The (ij, jt) specification given in Table A, without STIC effect,

where STIC contains a D1-dummy and a time dummy.

Case 1 - Time-invariant spatial parameters

Consider first a 3D-SPD model, where the spatial parameters are constant over time:

$$Y_t = \lambda W_t Y_t + \beta_0 1_{S_t} + X_t \beta_1 + \mathtt{FE\text{-}r} + U_t \tag{5.1}$$

$$U_t = \rho M_t U_t + V_t, \quad t = 1, \ldots, T, \ r = 1, 2, 3, \tag{5.2}$$

in relation to the general model (2.1) but written in vector form for each period.

The elements v_{ijt} of V_t are such that $v_{ijt} = \sigma_v e_{ijt}$ with $\{e_{ijt}\}$ being iid (i) standard normal, (ii) standardized normal mixture (90% $N(0, 1)$ and 10% $N(0, 4)$), and (iii) standardized chi-squared distribution. The spatial weight matrices W_t and M_t are generated according to (i) *circular world* (Kelejian and Prucha, 1999) with r neighbors ahead and r neighbors behind, (ii) *Rook*, and (iii) *Queen* (Yang, 2015). We choose $\beta_0 = 5$, $\beta_1 = 1$, $\sigma_v = 1$, and $\lambda, \rho = 0.2, 0.5, 0.9$. The heteroskedasticity is generated as $h_{ijt} = \sigma_v^2 \exp(\alpha z_{ij})$ normalized to have an average of σ_v^2, where $z_{ij} = \bar{x}_{ij.} = \frac{1}{T}\sum_t x_{ijt}$ and $\alpha = 0.5$.

The values x_{ijt} of the explanatory variable are generated by:

$$x_{ijt} = 1 + \omega_i + \theta_{ij} + \delta_{ijt}, \tag{5.3}$$

where ω_i, θ_{ij} and δ_{ijt} are independent, and are iid normal random variables with means zero and variances .75, .75, .75, respectively. Thus, x_{ijt} has a mean 1 and variance 2.25. In certain cases, a second regressor is added, which is simply generated as iid $N(0,2)$ over i, j, and t.

The fixed effects in `FE-1` and `FE-2` are generated by first setting:

$$\phi_{1,i} = \eta_{1,i} + \phi(\bar{x}_{i..} - \bar{x}_{...}), \;\; \phi_{2,j} = \eta_{2,j} + \phi(\bar{x}_{.j.} - \bar{x}_{...}), \;\; \text{and } \phi_{3,t} = \eta_{3,t} + \phi(\bar{x}_{..t} - \bar{x}_{...}),$$

where $\phi = 0.5$, $\eta_{1,i}$, $\eta_{2,j}$ and $\eta_{3,t}$ are iid $U(0,1)$ (uniformly distributed on the interval $[-1,1]$), $\bar{x}_{i..} = (N_2T)^{-1}\sum_j\sum_t x_{ijt}$, $\bar{x}_{..t} = (N_1N_2)^{-1}\sum_i\sum_j x_{ijt}$, $\bar{x}_{.j.} = (N_1T)^{-1}\sum_i\sum_t x_{ijt}$ and $\bar{x}_{...} = (N_1N_2T)^{-1}\sum_i\sum_j\sum_t x_{ijt}$, and then dropping the last element in each dimension.

The interactive fixed effects in `FE-3` are generated by first generating

$$\phi_{12,ij} = 0.5 * (\bar{x}_{ij.} - \bar{x}_{...}) + \eta_{ij}, \quad \text{and} \;\; \phi_{23,jt} = 0.5 * (\bar{x}_{.jt} - \bar{x}_{...}) + \eta_{jt},$$

where η_{ij} and η_{jt} are both iid $U(0,1)$, and then defining ϕ° and $\mathbf{D}^\circ$ accordingly.

For the STIC effects in `FE-2`, we add a D1-dummy variable with elements being iid *Bernoulli*(0.5), and a time dummy with elements being iid *Bernoulli*(0.3), and drop one column in D1 FE dummies and one column in time FE dummies.

In Tables 1-3, we present partial Monte Carlo results for **Case 1** under, respectively, the three FE specifications `FE-1`, `FE-2`, and `FE-3`. Monte Carlo means (`me`) and standard deviations (`sd`) are reported for the QML estimator (`QMLE`), the M-estimator (`M-Est`), and the robust M-estimator (`RM-Est`). The averages of estimated standard errors (`se`) and the empirical coverage probabilities of the 95% confidence intervals, `CP95` and `CP95r`, are also reported for `M-Est` and `RM-Est`. The results show that the proposed M-estimators and related inference methods perform excellently in finite samples. The `se`'s are very close to the corresponding `sd`'s, and the `CP95` and `CP95r` are very close to their nominal level 0.95. The larger the sample size, the better the `M-Est` and `RM-Est` and their `se`'s perform under homoskedasticity. Under heteroskedasticity, it is true that `RM-Est` is more robust than `M-Est` against heteroskedasticity (the `CP95r` is generally closer to its nominal level than `CP95`). In contrast, `QMLE` can perform poorly: severely underestimate ρ (Tables 1 and 2) and severely underestimate σ_v^2 (Table 3). A complete set of Monte Carlo results (along with `Matlab codes`) is given in `Supplemental Materials`, from which we see that these general observations remain.

Case 2 - Time-varying spatial parameters

We now extend the dgp in **Case I** by letting the spatial parameters λ and ρ to be λ_t and ρ_t, i.e., to vary over time. The spatial matrices W_t and M_t are generated according to the `Rook` or `Queen` contiguity scheme. The spatial parameters λ_t and ρ_t are generated according to the (rounded) exponential decay processes $\exp(-0.3t)$ and $\exp(-0.2t)$, respectively. The same set of configurations for the Monte Carlo experiments (sample sizes, FE specifications, and β and σ_v^2 values, error distributions, heteroskedasticity, and regressors' values) are considered.

The number of cross-sections $S(= N_1N_2)$ varies with $N_1, N_2 = 5, 10, 20, 50$, and the number of time periods $T = 5, 10$. The number of Monte Carlo runs for each experiment is set to 1000. Again, the `me` and `sd` are reported for `QMLE`, `M-Est`, and `RM-Est`. The `se`'s, and the `CP95` and `CP95r` are also reported for `M-Est` and `RM-Est`.

Partial Monte Carlo results are presented in Tables 4-6 for **Case 2** under, respectively, the three FE specifications, `FE-1`, `FE-2` and `FE-3`. A complete set of Monte Carlo results (along with `Matlab codes`) is given in `Supplemental Materials`. In **Case 2**, there are $2T$ nonlinear parameters involved in the optimization processes, and therefore it is much more demanding in computing power and numerical stability, especially under `FE-2`. The results show that both `M-Est` and `RM-Est` perform well in finite samples when errors are homoskedastic, regardless of whether they are normal or nonnormal. The $\sqrt{S}$-consistency of the two M-estimators of time-varying spatial parameters is clearly demonstrated. When errors are heteroskedastic, `RM-Est` performs well consistently, but `M-Est` does not. `QMLE` can perform badly, especially when there are many unobserved FE parameters that need to be controlled (see Table 6).

6. Conclusion

We proposed a strategy to specify the fixed effects in multi-dimensional spatial panel data models so that they can all be identified and estimated, leading to consistent estimation of the structural parameters in the model. A notable feature of this strategy is that it allows for the decomposition of the fixed effects into an observable part and an unobservable part. The observable part contains time-invariant covariates such as gender and race, and space-invariant covariates such as policy interventions. We introduce an M-estimation method and a robust (against unknown heteroskedasticity) M-estimation method for model estimation, which allows both the spatial weight matrices and the spatial coefficients to change over time. We introduced a plug-in with correction method for model inference. We provide a detailed guide coupled with Matlab codes to guide applied researchers to apply our proposed methods.

Appendix A: Basic Lemmas, Assumptions and Derivatives

Appendix A presents some basic lemmas that help appreciating the theoretical results, a sufficient set of assumptions with discussions particularly on the identification uniqueness assumption, and the detailed expressions of the additional partial derivatives required in the expressions of the Hessian matrices that are not reported in Sec. 4.2.

A1. Basic Lemmas.

Lemma A.1. *Let ε_n be an $n \times 1$ vector of independent innovations with its ith element having mean 0, variance $\sigma_{n,i}^2$, skewness $\gamma_{n,i}$, and excess kurtosis $\kappa_{n,i}$. Let D_{rn} be $n \times n$ matrices and c_{rn} $n \times 1$ vectors, both non-stochastic. Define $q_{rn} = \varepsilon_n' D_{rn} \varepsilon_n + c_{rn}' \varepsilon_n$, $r = 1, \ldots, m$. Then,*

(*i*) $\mathrm{E}(q_{rn}) = \mathtt{tr}(H_n D_{rn})$, *where* $H_n = \mathtt{diag}(\sigma_{n,1}^2, \ldots, \sigma_{n,n}^2)$, *and*

(*ii*) $$\mathrm{Cov}(q_{rn}, q_{sn}) = \mathtt{tr}(H_n D_{rn} H_n D_{sn}^{\dagger}) + c_{rn}' H_n c_{sn} + \textstyle\sum_{i=1}^n \left[d_{rn,ii} d_{sn,ii} \sigma_{n,i}^4 \kappa_{n,i} + (d_{rn,ii} c_{sn,i} + d_{sn,ii} c_{rn,i}) \sigma_{n,i}^3 \gamma_{n,i}\right],$$

where $d_{rn,ii}$ is the ith diagonal element of D_{rn} and $c_{rn,i}$ is the ith element of c_{rn}.

When $r = s$, (ii) gives $\mathrm{Var}(q_{rn})$. When $d_{rn,ii} = 0$, then the results simplify to

(*i*) $\mathrm{E}(q_{rn}) = 0$ *and* (*ii*) $\mathrm{Cov}(q_{rn}, q_{sn}) = \mathtt{tr}(H_n D_{rn} H_n D_{sn}^{\dagger}) + c_{rn}' H_n c_{sn}$.

Lemma A.2. *(Kelejian and Prucha, 1999): Let $\{A_n\}$ and $\{B_n\}$ be two sequences of $n \times n$ matrices that are bounded in both ∞-norm and 1-norm. Let C_n be a sequence of conformable matrices whose elements are uniformly $O(h_n^{-1})$. Then,*

(i) the sequence $\{A_n B_n\}$ are uniformly bounded in ∞-norm and 1-norm,

(ii) the elements of A_n are uniformly bounded and $tr(A_n) = O(n)$, and

(iii) the elements of $A_n C_n$ and $C_n A_n$ are uniformly $O(h_n^{-1})$.

Lemma A.3. *(Lemma A6, Meng and Yang, 2021): Let $\{A_n\}$ be a sequence of $n \times n$ matrices such that either $\|A_n\|_\infty$ or $\|A_n\|_1$ is bounded. Suppose that the elements of A_n are uniformly $O(h_n^{-1})$. Let innovation vector ε_n be defined as in Lemma A.1. Let c_n be an $n \times 1$ vector with elements of uniform order $O(h_n^{-1/2})$. Then*

(*i*) $\mathrm{E}(\varepsilon_n' A_n \varepsilon_n) = O(\frac{n}{h_n})$, (*ii*) $\mathrm{Var}(\varepsilon_n' A_N \varepsilon_n) = O(\frac{n}{h_n})$,

(*iii*) $\varepsilon_n' A_n \varepsilon_n = O_p(\frac{n}{h_n})$, (*iv*) $\varepsilon_n' A_n \varepsilon_n - \mathrm{E}(\varepsilon_n' A_n \varepsilon_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$,

(*v*) $c_n' A_n \varepsilon_n = O_p((\frac{n}{h_n})^{\frac{1}{2}})$, **if** $\|A_n\|_1$ *is bounded.*

A2. Regularity Conditions.

To state/discuss the regularity conditions under which the results of the theorems and corollaries hold, it is helpful to introduce the following **notations**. For an $n \times m$ matrix A of elements a_{ij}, 1-norm $\stackrel{\text{def}}{=} \max_{1 \le j \le m} \sum_{i=1}^{n} |a_{ij}|$, ∞-norm $\stackrel{\text{def}}{=} \max_{1 \le i \le n} \sum_{j=1}^{m} |a_{ij}|$, and (Frobenius) F-norm $\stackrel{\text{def}}{=} (\sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2)^{1/2}$. The the minimum and maximum eigenvalues of a real symmetric matrix are denoted, respectively, by $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$.

Assumption 1. The elements $\{v_\jmath\}$ of V are iid with mean zero, variance σ_0^2, and $\mathrm{E}(v_\jmath^{4+\epsilon})$ exists for some $\epsilon > 0$, where $\jmath$ is the combined index for all dimensions.

Assumption 1*. The elements $\{v_\jmath\}$ of V are inid with means zero, variances $\sigma_{\jmath 0}^2$, and $\mathrm{E}(v_\jmath^{4+\epsilon})$ exists for some $\epsilon > 0$, where $\jmath$ is the combined index for all dimensions.

Assumption 2. The space Δ is compact, and the true parameter δ_0 lies in its interior.

Assumption 3. The elements of X are non-stochastic and bounded uniformly in $\jmath$, and $\lim_{N^* \to \infty} \frac{1}{N^*} \mathbb{X}'(\rho)\mathbb{Q}(\rho)\mathbb{X}(\rho)$ exists and is positive definite, uniformly in $\rho \in \Delta_\rho$.

Assumption 4. $\{W_t\}$ and $\{M_t\}$ are known time-varying matrices such that, $\forall t$, (i) their elements are at most of uniform order $h_{S_t}^{-1}$ such that $\frac{h_{S_t}}{S_t} \to 0$ as $S_t \to \infty$; (ii) their diagonal elements are zero; and (iii) they are bounded in both 1-norm and ∞-norm.

Assumption 5. (i) $\mathbb{A}^{-1}(\lambda_0)$ and $\mathbb{B}^{-1}(\rho_0)$ are bounded in both 1-norm and ∞-norm; (ii) $\mathbb{A}^{-1}(\lambda)$ and $\mathbb{B}^{-1}(\rho)$ are bounded in either 1-norm or ∞-norm, uniformly in $\delta \in \Delta$; and (iii) $\gamma_{\min}[\mathbb{A}'(\lambda)\mathbb{A}(\lambda)]$ and $\gamma_{\min}[\mathbb{B}'(\rho)\mathbb{B}(\rho)]$ are bounded from below away from 0 uniformly in $\delta \in \Delta$, and $\gamma_{\max}[\mathbb{A}'(\lambda)\mathbb{A}(\lambda)]$ and $\gamma_{\max}[\mathbb{B}'(\rho)\mathbb{B}(\rho)]$ are bounded from above uniformly in $\delta \in \Delta$.

Assumption 6. (i) As $N \to \infty$, $\frac{S_t}{N} \to c_t$, where $c_t \in (0, 1], \forall t$; (ii) as $T \to \infty$, $\frac{T_\imath}{T} \to d_\imath$, where $d_\imath \in (0, 1], \forall\, \imath$ (a combined index for all space dimensions), and $T_\imath$ is the number of times the $\imath$th space entity shows up in the entire T periods; and (iii) $\min_\imath(T_\imath) \ge 2$ and $\min_t(S_t) \ge 2$.

Assumption 7. $\inf_{\delta: d(\delta, \delta_0) \ge \epsilon} \left\| \bar{S}_N^c(\delta) \right\| > 0$ for every $\epsilon > 0$, where $d(\delta, \delta_0)$ is a measure of distance between δ and δ_0, and $\bar{S}_N^c(\delta)$ is the 'concentrated' $\mathrm{E}[S_N^*(\theta)]$, or $\mathrm{E}[S_N^{**}(\theta)]$, or $\mathrm{E}[S_N^r(\vartheta)]$, or $\mathrm{E}[S_N^{rr}(\vartheta)]$, with (β, σ^2) or β being concentrated out from the expected AQS functions.

Assumptions 1 and 1* are standard assumptions on model errors for QML-type or GMM estimation of models containing nonlinear parameters under, respectively, homoskedasticity and heteroskedasticity. Assumptions 2 is the standard assumption on the space of nonlinear parameters (in our case the spatial parameters), so that a crucial uniform convergence result can be established. Assumptions 3 is similar to the assumptions on regressors in GLS estimation. Assumption 4 extends the classical assumptions on spatial weight matrices (Lee and Yu, 2010a; Yang et al., 2016) to allow for time-varying and unbalancedness features. Assumptions 5 (i) and (ii) are standard in spatial econometrics literature. Assumption 5 (iii) is a mild technical assumption in light of Assumption 5 (ii).

Assumption 6 controls the degree of unbalancedness so that the spatial structure is completely preserved after concentrating the unobservable FEs. Assumption 7 is a generic assumption on identification uniqueness of structural parameters in a model under the M-estimation or the zero-estimation (Van der Vaart, 1998). It is a high-level assumption in a unified M-estimation setting. More primitive assumptions can be derived under a specific setting. Here, we illustrate using the model with time-invariant spatial parameters and homoskedastic errors.

Note that the expectation operator $\mathrm{E}(\cdot)$ corresponds to the true parameters or the true model. The population counterpart of (3.5) is $\mathrm{E}[S_N^{*}\theta)]$. Solving the first two components of the equations $\mathrm{E}[S_N^{*}\theta)] = 0$ for β and σ^2 for a given δ, we have,

$$\bar{\beta}_N^{*}(\delta) = [\mathbb{Z}'(\rho)\mathbb{Z}(\rho)]^{-1}\mathbb{Z}'(\rho)\mathrm{E}[\mathbb{Y}(\delta)], \text{ and} \tag{A.1}$$

$$\bar{\sigma}_N^{*2}(\delta) = \tfrac{\sigma_0^2}{N^*}\mathtt{tr}[\Pi_1(\delta)(\mathbb{A}'\mathbb{B}'\mathbb{B}\mathbb{A})^{-1}] + \tfrac{1}{N^*}\zeta_0'\Pi_2(\rho)\zeta_0, \tag{A.2}$$

where $\Pi_1(\delta)=\mathbb{R}'(\delta)\mathbb{R}(\delta)$, $\Pi_2(\rho)=\mathbb{R}'(\delta)\mathbf{Q}(\rho)\mathbb{R}(\delta)$, $\mathbb{R}(\delta)=\mathbb{Q}(\rho)\mathbb{B}(\rho)\mathbb{A}(\lambda)$, $\zeta_0=\mathbb{A}^{-1}(X\beta_0+D\phi_0)$, $\mathbf{Q}(\rho) = I_N - \mathbb{Z}(\rho)[\mathbb{Z}'(\rho)\mathbb{Z}(\rho)]^{-1}\mathbb{Z}'(\rho)$, and $\mathbb{Z}(\rho) = \mathbb{Q}(\rho)\mathbb{B}(\rho)X$ defined in (3.10). Substituting $\bar{\beta}_N^{*}(\delta)$ and $\bar{\sigma}_N^{*2}(\delta)$ into the δ-components of $\mathrm{E}[S_N^{*}\theta)]$ gives

$$\bar{S}_N^{*c}(\delta) = \begin{cases} \left\{\sigma_0^2\mathtt{tr}[\mathbb{F}'(\lambda)\Pi_1(\delta)(\mathbb{A}'\mathbb{B}'\mathbb{B}\mathbb{A})^{-1}] + \zeta_0'\mathbb{F}'(\lambda)\Pi_2(\rho)\zeta_0\right\}/\bar{\sigma}_N^{*2}(\delta) - \mathtt{tr}[\mathbb{Q}(\rho)\bar{\mathbb{F}}(\delta)], \\ \left\{\sigma_0^2\mathtt{tr}[\Pi_3(\delta)(\mathbb{A}'\mathbb{B}'\mathbb{B}\mathbb{A})^{-1}] + \zeta_0'\Pi_4(\delta)\zeta_0\right\}/\bar{\sigma}_N^{*2}(\delta) - \mathtt{tr}[\mathbb{Q}(\rho)\mathbb{G}(\rho)], \end{cases} \tag{A.3}$$

where $\Pi_3(\delta) = \mathbb{R}'(\delta)\mathbb{G}(\rho)\mathbb{R}(\delta)$ and $\Pi_4(\delta) = \mathbb{R}'(\delta)\mathbf{Q}(\rho)\mathbb{G}(\rho)\mathbf{Q}(\rho)\mathbb{R}(\delta)$.[9] We have $\bar{\beta}_N^{*}(\delta_0) = \beta_0$, $\bar{\sigma}_N^{*2}(\delta_0) = \sigma_0^2$, and $\bar{S}_N^{*c}(\delta_0) = 0$. Therefore, δ is identified if $\bar{S}_N^{c}(\delta) \neq 0, \forall\delta \neq \delta_0$. Furthermore, we show that $\bar{\sigma}_N^{*2}(\delta)$ is bounded from below away from zero, uniformly in $\delta \in \Delta$. Therefore, the identification uniqueness condition for the nonlinear parameters δ is that $\forall\delta \neq \delta_0$, either

$$\sigma_0^2\mathtt{tr}[\mathbb{F}'(\lambda)\Pi_1(\delta)(\mathbb{A}'\mathbb{B}'\mathbb{B}\mathbb{A})^{-1}] + \zeta_0'\mathbb{F}'(\lambda)\Pi_2(\rho)\zeta_0 - \bar{\sigma}_N^{*2}(\delta)\mathtt{tr}[\mathbb{Q}(\rho)\bar{\mathbb{F}}(\delta)] \neq 0, \text{ or} \tag{A.4}$$

$$\sigma_0^2\mathtt{tr}[\Pi_3(\delta)(\mathbb{A}'\mathbb{B}'\mathbb{B}\mathbb{A})^{-1}] + \zeta_0'\Pi_4(\delta)\zeta_0 - \bar{\sigma}_N^{*2}(\delta)\mathtt{tr}[\mathbb{Q}(\rho)\mathbb{G}(\rho)] \neq 0. \tag{A.5}$$

Given the spatial weight matrices, we can verify these conditions analytically and numerically.

[9]In the above derivations, we have used the decomposition: $\tilde{V}(\bar{\beta}(\delta), \delta) = \mathbf{Y}(\delta) - \mathrm{E}(\mathbf{Y}(\delta)) + \mathbf{Q}(\rho)\mathrm{E}(\mathbf{Y}(\delta))$, where $\mathbf{Y}(\delta) = \mathbb{Q}(\rho)\mathbb{Y}(\delta)$, and the easily verified facts: $\mathbb{Q}(\rho)\mathbb{Z}(\rho) = \mathbb{Z}(\rho)$ and $\mathbf{Q}(\rho)\mathbb{Q}(\rho) = \mathbb{Q}(\rho)\mathbf{Q}(\rho)$.

A3. Additional Partial Derivatives for Section 4.2.

For $\mathcal{H}_N^r(\vartheta)$ in (4.9), note $\mathbf{F}_\omega(\delta) = \frac{\partial}{\partial \omega}\mathbf{F}(\delta) = \frac{\partial}{\partial \omega}\big[\big[\bar{\mathbb{F}}'(\delta) - F_d(\delta)\big]\mathbb{Q}(\rho)\big], \omega = \lambda, \rho$, where $\bar{\mathbb{F}}(\delta) = \mathbb{B}(\rho)\mathbb{F}(\lambda)\mathbb{B}^{-1}(\rho)$, $\mathbb{F}(\lambda) = \mathbf{W}\mathbb{A}^{-1}(\lambda)$ and $F_d(\delta) = \texttt{diag}(\bar{\mathbb{F}}'(\delta)\mathbb{Q}(\rho))\texttt{diag}(\mathbb{Q}(\rho))^{-1}$. We have

$$\begin{aligned}
\mathbf{F}_\lambda(\delta) &= \big[\tfrac{\partial}{\partial \lambda}\bar{\mathbb{F}}'(\delta) - \tfrac{\partial}{\partial \lambda}F_d(\delta)\big]\mathbb{Q}(\rho) \\
&= \big\{\big[\mathbb{B}(\rho)\mathbb{F}^2(\lambda)\mathbb{B}^{-1}(\rho)\big]' - \texttt{diag}\big(\big[\mathbb{B}(\rho)\mathbb{F}^2(\lambda)\mathbb{B}^{-1}(\rho)\big]'\mathbb{Q}(\rho)\big)\texttt{diag}(\mathbb{Q}(\rho))^{-1}\big\}\mathbb{Q}(\rho), \\
\mathbf{F}_\rho(\delta) &= \big[\tfrac{\partial}{\partial \rho}\bar{\mathbb{F}}'(\delta) - \tfrac{\partial}{\partial \rho}F_d(\delta)\big]\mathbb{Q}(\rho) + \big[\bar{\mathbb{F}}'(\delta) - F_d(\delta)\big]\big[\tfrac{\partial}{\partial \rho}\mathbb{Q}(\rho)\big] \\
&= E_1'\mathbb{Q}(\rho) - \texttt{diag}(E_1'\mathbb{Q}(\rho))E_2 - \texttt{diag}\big[\bar{\mathbb{F}}'(\delta)\big[\tfrac{\partial}{\partial \rho}\mathbb{Q}(\rho)\big]\big]E_2 \\
&\quad + \texttt{diag}\big[\bar{\mathbb{F}}'(\delta)\mathbb{Q}(\rho)\big]\texttt{diag}(\mathbb{Q}(\rho))^{-1}\big[\tfrac{\partial}{\partial \rho}\mathbb{Q}(\rho)\big]E_2 + \big[\bar{\mathbb{F}}'(\delta) - F_d(\delta)\big]\big[\tfrac{\partial}{\partial \rho}\mathbb{Q}(\rho)\big],
\end{aligned}$$

where $E_1 = [\bar{\mathbb{F}}(\delta)\mathbf{W} - \mathbf{W}\mathbb{F}(\lambda)]\mathbb{B}^{-1}(\rho)$, $E_2 = \texttt{diag}(\mathbb{Q}(\rho))^{-1}\mathbb{Q}(\rho)$, and $\frac{\partial}{\partial \rho}\mathbb{Q}(\rho)$ in Footnote 7.

Now, recall $\mathbf{G}(\rho) = \mathbf{G}(\rho) = \big[\bar{\mathbb{G}}(\rho) - G_d(\rho)\big]\mathbb{Q}(\rho)$, where $\bar{\mathbb{G}}(\rho) = \mathbb{Q}(\rho)\mathbb{G}(\rho)$, $\mathbb{G}(\rho) = \mathbf{M}\mathbb{B}^{-1}(\rho)$ and $G_d(\rho) = \texttt{diag}(\bar{\mathbb{G}}(\rho)\mathbb{Q}(\rho))\texttt{diag}(\mathbb{Q}(\rho))^{-1}$. We have

$$\mathbf{G}_\rho(\rho) = \tfrac{\partial}{\partial \rho}\mathbf{G}(\rho) = \big[\tfrac{\partial}{\partial \rho}\bar{\mathbb{G}}(\rho) - \tfrac{\partial}{\partial \rho}G_d(\rho)\big]\mathbb{Q}(\rho) + \big[\bar{\mathbb{G}}(\rho) - G_d(\rho)\big]\big[\tfrac{\partial}{\partial \rho}\mathbb{Q}(\rho)\big],$$

where $\frac{\partial}{\partial \rho}\bar{\mathbb{G}}(\rho) = \big(\frac{\partial}{\partial \rho}\mathbb{Q}(\rho)\big)\mathbb{G}(\rho) + \mathbb{Q}(\rho)\mathbb{G}^2(\rho)$, and

$$\begin{aligned}
\tfrac{\partial}{\partial \rho}G_d(\rho) = &\big[\texttt{diag}\big(\big(\tfrac{\partial}{\partial \rho}\bar{\mathbb{G}}(\rho)\big)\mathbb{Q}(\rho)\big) + \texttt{diag}\big(\bar{\mathbb{G}}(\rho)\big(\tfrac{\partial}{\partial \rho}\mathbb{Q}(\rho)\big)\big)\big]\texttt{diag}(\mathbb{Q}(\rho))^{-1} \\
&- \texttt{diag}(\bar{\mathbb{G}}(\rho)\mathbb{Q}(\rho))\texttt{diag}(\mathbb{Q}(\rho))^{-1}\texttt{diag}\big(\tfrac{\partial}{\partial \rho}\mathbb{Q}(\rho)\big)\texttt{diag}(\mathbb{Q}(\rho))^{-1}.
\end{aligned}$$

For $\mathcal{H}_N^r(\vartheta)$ in (4.14), the partial derivatives of $\mathbf{F}_t(\delta_t, \rho)$ and $\mathbf{G}_t(\rho_t, \rho)$ are, firstly,

$$\begin{aligned}
\tfrac{\partial}{\partial \lambda_s}\mathbf{F}_t(\delta_t, \rho) &= \tfrac{\partial}{\partial \lambda_s}\big[\big[\bar{\mathcal{F}}_t'(\delta_t) - \mathcal{F}_t^d(\delta_t, \rho)\big]\ \mathbb{Q}(\rho)\big] \\
&= \big[\tfrac{\partial}{\partial \lambda_s}\bar{\mathcal{F}}_t'(\delta_t) - \tfrac{\partial}{\partial \lambda_s}\mathcal{F}_t^d(\delta_t, \rho)\big]\mathbb{Q}(\rho) + \big[\bar{\mathcal{F}}_t'(\delta_t) - \mathcal{F}_t^d(\delta_t, \rho)\big]\big[\tfrac{\partial}{\partial \lambda_s}\mathbb{Q}(\rho)\big] \\
&= \big[\tfrac{\partial}{\partial \lambda_s}\bar{\mathcal{F}}_t'(\delta_t) - \tfrac{\partial}{\partial \lambda_s}\mathcal{F}_t^d(\delta_t, \rho)\big]\mathbb{Q}(\rho),
\end{aligned}$$

where $\bar{\mathcal{F}}_t'(\delta_t)$ and $\mathcal{F}_t^d(\delta_t, \rho)$ are obtained by inserting $\bar{F}_t'(\delta_t)$ and $F_t^d(\delta_t, \rho)$ into the tt-block of the $N \times N$ zero matrix. The derivative components are given by $\frac{\partial}{\partial \lambda_s}\mathbb{Q}(\rho) = 0, \forall s$,

$$\begin{aligned}
\tfrac{\partial}{\partial \lambda_s}\bar{\mathcal{F}}_t'(\delta_t) &= \big[\mathbb{B}(\rho)\mathcal{F}_t^2(\lambda_t)\mathbb{B}^{-1}(\rho)\big]', \text{ if } s = t, \text{ and } 0, \text{ otherwise}, \\
\tfrac{\partial}{\partial \lambda_s}\mathcal{F}_t^d(\delta_t, \rho) &= \texttt{diag}\big(\big(\tfrac{\partial}{\partial \lambda_t}\bar{\mathcal{F}}_t'(\delta_t)\big)\mathbb{Q}(\rho)\big)\texttt{diag}(\mathbb{Q}(\rho))^{-1}, \text{ if } s = t, \text{ and } 0, \text{ otherwise}.
\end{aligned}$$

Secondly,

$$\begin{aligned}
\tfrac{\partial}{\partial \rho_s}\mathbf{F}_t(\delta_t, \rho) &= \tfrac{\partial}{\partial \rho_s}\big[\big[\bar{\mathcal{F}}_t'(\delta_t) - \mathcal{F}_t^d(\delta_t, \rho)\big]\mathbb{Q}(\rho)\big] \\
&= \big[\tfrac{\partial}{\partial \rho_s}\bar{\mathcal{F}}_t'(\delta_t) - \tfrac{\partial}{\partial \rho_s}\mathcal{F}_t^d(\delta_t, \rho)\big]\mathbb{Q}(\rho) + \big[\bar{\mathcal{F}}_t'(\delta_t) - \mathcal{F}_t^d(\delta_t, \rho)\big]\big[\tfrac{\partial}{\partial \rho_s}\mathbb{Q}(\rho)\big],
\end{aligned}$$

where the derivative components are given by

$$\begin{aligned}
\tfrac{\partial}{\partial\rho_s}\mathbb{Q}(\rho) &= \mathbb{Q}(\rho)\mathcal{G}_s(\rho_s)\mathbb{P}(\rho)+\mathbb{P}(\rho)\mathcal{G}_s'(\rho_s)\mathbb{Q}(\rho), \forall s,\\
\tfrac{\partial}{\partial\rho_s}\bar{\mathcal{F}}_t'(\delta_t) &= \big[-\mathcal{W}_s\mathcal{F}_t(\lambda_t)\mathbb{B}^{-1}(\rho)+\mathbb{B}(\rho)\mathcal{F}_t(\lambda_t)\mathbb{B}^{-1}(\rho)\mathcal{W}_s\mathbb{B}^{-1}(\rho)\big]', \forall s,\\
\tfrac{\partial}{\partial\rho_s}\mathcal{F}_t^d(\delta_t,\rho) &= -\big[\mathtt{diag}(\mathcal{W}_s\mathcal{F}_t'(\lambda_t)\mathbb{B}^{-1}(\rho)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\big]\\
&\quad+\big[\mathtt{diag}(\mathbb{B}(\rho)\mathcal{F}_t'(\lambda_t)\mathbb{B}^{-1}(\rho)\mathcal{W}_s\mathbb{B}^{-1}(\rho)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\big]\\
&\quad+\big[\mathtt{diag}(\mathbb{B}(\rho)\mathcal{F}_t'(\lambda_t)\mathbb{B}^{-1}(\rho)\tfrac{\partial}{\partial\rho_s}\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\big]\\
&\quad-\big[\mathtt{diag}(\mathbb{B}(\rho)\mathcal{F}_t'(\lambda_t)\mathbb{B}^{-1}(\rho)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\mathtt{diag}\big(\tfrac{\partial}{\partial\rho_s}\mathbb{Q}(\rho)\big)\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\big]\forall s,
\end{aligned}$$

with $\mathcal{G}_s(\rho_s)$ obtained by inserting $G_s(\rho_s)$ into the ss-block of the $N\times N$ zero matrix, and $\mathcal{W}_s$ obtained by inserting W_t into the ss-block of the $N\times N$ zero matrix. Lastly,

$$\begin{aligned}
\tfrac{\partial}{\partial\rho_s}\mathbf{G}_t(\rho_t,\rho) &= \tfrac{\partial}{\partial\rho_s}\big[\big[\bar{\mathcal{G}}_t(\rho_t,\rho)-\mathcal{G}_t^d(\rho_t,\rho)\big]\mathbb{Q}(\rho)\big]\\
&= \big[\tfrac{\partial}{\partial\rho_s}\bar{\mathcal{G}}_t(\rho_t,\rho)-\tfrac{\partial}{\partial\rho_s}\mathcal{G}_t^d(\rho_t,\rho)\big]\mathbb{Q}(\rho)+\big[\bar{\mathcal{G}}_t(\rho_t,\rho)-\mathcal{G}_t^d(\rho_t,\rho)\big]\big[\tfrac{\partial}{\partial\rho_s}\mathbb{Q}(\rho)\big],
\end{aligned}$$

where $\frac{\partial}{\partial\rho_s}\bar{\mathcal{G}}_t(\rho_t,\rho)=\big(\frac{\partial}{\partial\rho_s}\mathbb{Q}(\rho)\big)\mathcal{G}_t(\rho_t)+\mathbb{Q}(\rho)\mathcal{G}_t^2(\rho_t)$ if $s=t$, and $\big(\frac{\partial}{\partial\rho_s}\mathbb{Q}(\rho)\big)\mathcal{G}_t(\rho_t)$ if $s\neq t$, and

$$\begin{aligned}
\tfrac{\partial}{\partial\rho_s}\mathcal{G}_t^d(\rho_t,\rho) &= \tfrac{\partial}{\partial\rho_s}\big(\mathtt{diag}(\bar{\mathcal{G}}_t(\rho_t,\rho)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\big)\\
&= \mathtt{diag}\big(\big(\tfrac{\partial}{\partial\rho_s}\bar{\mathcal{G}}_t(\rho_t,\rho)\big)\mathbb{Q}(\rho)\big)\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\\
&\quad+\mathtt{diag}(\bar{\mathcal{G}}_t(\rho_t,\rho)\tfrac{\partial}{\partial\rho_s}\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\\
&\quad-\mathtt{diag}(\bar{\mathcal{G}}_t(\rho_t,\rho)\mathbb{Q}(\rho))\mathtt{diag}(\mathbb{Q}(\rho))^{-1}\mathtt{diag}\big(\tfrac{\partial}{\partial\rho_s}\mathbb{Q}(\rho)\big)\mathtt{diag}(\mathbb{Q}(\rho))^{-1}.
\end{aligned}$$

Supplemental Materials

The `Supplemental Materials` contain **Appendix B:** Proofs of Theoretical Results, **Appendix C:** A Complete Set of Monte Carlo Results, and **Appendix D:** A Practical Guide.

Acknowledgments

We thank Giuseppe Arbia, Xiaodong Liu, Xi Qu, and the participants of 23rd International Workshop on Spatial Econometrics and Statistics (SEW2025), Saint-Étienne, France, June 4-6, 2025, 7th International Conference on Econometrics and Statistics (EcoSta2024), Beijing, China, July 17-19, 2024, XVIII World Conference of Spatial Econometric Association (SEA2024), Groningen, the Netherlands, June 2024, and XVII World Conference of Spatial Econometric Association (SEA2023), San Diego, November 2023, for their helpful comments. The main part of this work was carried out during Zhenlin Yang's visits in the past few years to CRED, University Paris-Panthéon-Assas, France.

References

Andrews, D. W. (1992). Generic uniform convergence. *Econometric Theory 8*, 241–257.

Anselin, L., J. L. Gallo, and H. Jayet (2008). Spatial Panel Econometrics. In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Advanced Studies in Theoretical and Applied Econometrics, pp. 625–660. Berlin, Heidelberg: Springer.

Balázsi, L., L. Mátyás, and T. Wansbeek (2024). Fixed effects models. In L. Mátyás (Ed.), *The Econometrics of Multi-Dimensional Panels: Theory and Applications*, pp. 1–37. Springer.

Baltagi, B. H., P. Egger, and K. Erhardt (2024). The econometrics of gravity models in international trade. In L. Mátyás (Ed.), *The Econometrics of Multi-Dimensional Panels: Theory and Applications*, pp. 381–412. Springer.

Baltagi, B. H. and A. Pirotte (2014). Prediction in a spatial nested error components panel data model. *International Journal of Forecasting 30*, 407–414.

Elhorst, J. P. (2010). Spatial panel data models. In F. M. M. and G. A. (Eds.), *Handbook of Applied Spatial Analysis*, pp. 377–405. Springer Verlag.

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer Verlag.

Fingleton, B., J. Le Gallo, and A. Pirotte (2018a). A multidimensional spatial lag panel data model with spatial moving average nested random effects errors. *Empirical Economics 55*, 113–146.

Fingleton, B., J. Le Gallo, and A. Pirotte (2018b). Panel data models with spatially dependent nested random effects. *Journal of Regional Science 58*, 63–80.

He, M. and K.-P. Lin (2015). Testing spatial effects and random effects in a nested panel data model. *Economics Letters 135*, 85–91.

Hornok, C. and M. Koren (2017). Winners and losers of globalization: sixteen challenges for measurement and theory. In *Economics without Borders - Economic Research for European Policy Challenges*, pp. 238–273. Cambridge University Press.

Kelejian, H. and G. Piras (2017). *Spatial Econometrics*. Academic Press.

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

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics 104*, 219–257.

Le Gallo, J. and A. Pirotte (2024). Multi-dimensional models for spatial panels. In *The Econometrics of Multi-dimensional Panels: Theory and Applications*, pp. 353–379. Springer.

Lee, L.-F. and J. Yu (2010a). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*, 165–185.

Lee, L.-F. and J. Yu (2010b). Some recent developments in spatial panel data models. *Regional Science and Urban Economics 40*, 255–271.

Lee, L.-F. and J. Yu (2015). Spatial panel data models. In *The Oxford Handbook of Panel Data*, pp. 363–401. Oxford Academic.

LeSage, J. and R. K. Pace (2009). *Introduction to Spatial Econometrics.* Chapman and Hall/CRC.

Liu, S. F. and Z. Yang (2020). Robust estimation and inference of spatial panel data models with fixed effects. *Japanese Journal of Statistics and Data Science 3*, 257–311.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*, 531–542.

Mátyás, L. (2017). *The Econometrics of Multi-Dimensional Panels.* Springer.

Mátyás, L. (2024). *The Econometrics of Multi-Dimensional Panels, 2nd Ed.* Springer.

Meng, X. and Z. Yang (2021). Unbalanced spatial panel data models with fixed effects. *Working Paper, Singapore Management University*.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Pirotte, A. and Z. Yang (2024). Identification and estimation of space or time invariant covariates effects in multi-dimensional panels with fixed effects. *Working Paper, Singapore Management University*.

Van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Yang, Z. (2015). A general method for third-order bias and variance corrections on a nonlinear estimator. *Journal of Econometrics 186*, 178–200.

Yang, Z., J. Yu, and S. F. Liu (2016). Bias correction and refined inferences for fixed effects spatial panel data models. *Regional Science and Urban Economics 61*, 52–72.

Ye, Q., H. Liang, K.-P. Lin, and Z. Long (2019). Hierarchically spatial autoregressive and moving average error model. *Economic Modelling 76*, 14–30.

Table 1: Monte Carlo Results: `FE-1`, Time-Invariant Spatial Parameters
$W_t = M_t =$ `Circular World` with $r = 10$

Parameter		QMLE Mean	QMLE sd	M-Est Mean	M-Est sd	M-Est se	M-Est CP95	RM-Est Mean	RM-Est sd	RM-Est se	RM-Est CP95r
		Normal Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$									
β_0	1.0	1.0073	0.4606	1.0053	0.4677	0.4715	94.2	1.0057	0.4681	0.4816	94.6
β_1	1.0	0.9993	0.0292	0.9992	0.0292	0.0302	95.7	0.9992	0.0292	0.0300	95.5
λ	0.2	0.1976	0.0924	0.1980	0.0949	0.0941	94.1	0.1979	0.0950	0.0974	93.6
ρ	0.3	0.0639	0.1941	0.2523	0.1693	0.1659	94.3	0.2525	0.1697	0.1693	94.4
σ_v^2	1.0	0.9433	0.0604	0.9891	0.0627	0.0638	94.2	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 5$									
β_0	1.0	1.0031	0.2043	1.0046	0.2047	0.2084	95.8	1.0047	0.2048	0.2091	95.2
β_1	1.0	1.0003	0.0127	1.0003	0.0127	0.0130	95.6	1.0003	0.0127	0.0130	95.5
λ	0.2	0.2000	0.0324	0.1995	0.0326	0.0320	94.9	0.1995	0.0326	0.0327	94.9
ρ	0.3	0.2115	0.0774	0.2911	0.0714	0.0681	94.6	0.2911	0.0714	0.0694	94.6
σ_v^2	1.0	0.9775	0.0324	0.9974	0.0331	0.0319	94.7	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 10$									
β_0	1.0	1.0039	0.1681	1.0060	0.1688	0.1732	95.0	1.0060	0.1689	0.1754	95.0
β_1	1.0	1.0001	0.0090	1.0000	0.0090	0.0094	96.4	1.0000	0.0090	0.0094	96.5
λ	0.2	0.1994	0.0302	0.1988	0.0305	0.0302	94.5	0.1988	0.0305	0.0311	95.6
ρ	0.3	0.2488	0.0516	0.2973	0.0494	0.0498	95.6	0.2973	0.0495	0.0511	96.2
σ_v^2	1.0	0.9878	0.0219	0.9985	0.0222	0.0225	95.7	—	—	—	—
		Normal Errors, Heteroskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$									
β_0	1.0	1.0128	0.5310	1.0091	0.5464	0.4709	89.9	1.0092	0.5465	0.5566	94.4
β_1	1.0	1.0005	0.0305	1.0005	0.0306	0.0302	93.9	1.0005	0.0306	0.0306	94.3
λ	0.2	0.1960	0.1077	0.1969	0.1117	0.0937	89.4	0.1969	0.1115	0.1136	93.4
ρ	0.3	0.0581	0.2092	0.2467	0.1827	0.1662	92.9	0.2442	0.1826	0.1785	94.5
σ_v^2	1.0	0.9434	0.0810	0.9892	0.0845	0.0941	95.0	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 5$									
β_0	1.0	0.9997	0.1803	1.0011	0.1803	0.1917	96.4	1.0005	0.1804	0.1857	96.2
β_1	1.0	0.9998	0.0155	0.9998	0.0155	0.0134	90.8	0.9998	0.0155	0.0154	94.3
λ	0.2	0.1993	0.0385	0.1985	0.0390	0.0385	94.8	0.1988	0.0390	0.0394	95.0
ρ	0.3	0.2106	0.0833	0.2905	0.0770	0.0707	93.6	0.2893	0.0773	0.0768	95.0
σ_v^2	1.0	0.9767	0.0432	0.9965	0.0441	0.0489	96.2	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 10$									
β_0	1.0	1.0044	0.2062	1.0054	0.2064	0.1774	90.2	1.0054	0.2064	0.2109	94.5
β_1	1.0	1.0001	0.0101	1.0001	0.0101	0.0094	92.9	1.0001	0.0101	0.0099	94.6
λ	0.2	0.2010	0.0212	0.2008	0.0212	0.0197	93.0	0.2008	0.0212	0.0215	95.5
ρ	0.3	0.2475	0.0550	0.2957	0.0525	0.0454	90.4	0.2943	0.0524	0.0524	94.6
σ_v^2	1.0	0.9886	0.0328	0.9993	0.0332	0.0363	96.3	—	—	—	—
		Chi-Squared Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$									
β_0	1.0	1.0100	0.4875	1.0099	0.4905	0.4724	93.3	1.0105	0.4905	0.4816	93.8
β_1	1.0	1.0004	0.0293	1.0004	0.0293	0.0301	95.7	1.0004	0.0293	0.0300	95.7
λ	0.2	0.1953	0.0981	0.1954	0.0993	0.0943	93.5	0.1952	0.0994	0.0974	93.1
ρ	0.3	0.0660	0.1956	0.2536	0.1696	0.1659	94.5	0.2538	0.1701	0.1683	94.1
σ_v^2	1.0	0.9419	0.1048	0.9876	0.1098	0.1055	91.0	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 5$									
β_0	1.0	0.9998	0.2139	1.0021	0.2140	0.2083	94.3	1.0021	0.2138	0.2088	94.5
β_1	1.0	0.9996	0.0131	0.9996	0.0131	0.0130	95.3	0.9996	0.0131	0.0130	95.5
λ	0.2	0.1998	0.0331	0.1992	0.0333	0.0320	93.3	0.1992	0.0333	0.0327	93.4
ρ	0.3	0.2088	0.0762	0.2887	0.0704	0.0685	94.6	0.2887	0.0703	0.0697	95.5
σ_v^2	1.0	0.9787	0.0530	0.9987	0.0541	0.0542	94.6	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 10$									
β_0	1.0	0.9959	0.1739	0.9977	0.1746	0.1730	94.4	0.9978	0.1745	0.1749	94.3
β_1	1.0	0.9995	0.0098	0.9995	0.0098	0.0094	94.6	0.9995	0.0098	0.0094	94.7
λ	0.2	0.2010	0.0303	0.2005	0.0306	0.0301	94.0	0.2005	0.0306	0.0311	94.7
ρ	0.3	0.2468	0.0521	0.2955	0.0499	0.0499	94.9	0.2955	0.0499	0.0512	95.7
σ_v^2	1.0	0.9885	0.0372	0.9992	0.0376	0.0386	94.5	—	—	—	—

Table 2: Monte Carlo Results: FE-2, Time-Invariant Spatial Parameters
$W_t = M_t =$ Circular World with $r = 10$

Parameter		QMLE		M-Est				RM-Est			
		Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
		Normal Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 10$									
β_0	1.0	0.9836	0.3002	0.9802	0.3050	0.3085	95.3	0.9801	0.3051	0.3132	95.0
β_1	1.0	1.0005	0.0188	1.0005	0.0188	0.0199	96.1	1.0005	0.0188	0.0198	96.0
β_2	0.5	0.5015	0.1414	0.5015	0.1414	0.1413	95.6	0.5015	0.1414	0.1407	95.2
β_3	0.5	0.4982	0.1998	0.4979	0.1994	0.1966	94.2	0.4979	0.1993	0.1962	94.2
λ	0.2	0.2009	0.0687	0.2019	0.0707	0.0707	93.2	0.2019	0.0707	0.0728	93.6
ρ	0.3	0.1313	0.1239	0.2575	0.1151	0.1150	94.7	0.2576	0.1152	0.1171	94.2
σ_v^2	1.0	0.9710	0.0455	0.9959	0.0466	0.0450	93.1	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 10$									
β_0	1.0	1.0122	0.1455	1.0131	0.1458	0.1536	96.2	1.0131	0.1458	0.1540	96.2
β_1	1.0	0.9997	0.0092	0.9996	0.0092	0.0093	95.7	0.9996	0.0092	0.0093	95.7
β_2	0.5	0.4999	0.0946	0.4999	0.0946	0.0985	95.9	0.4999	0.0946	0.0984	95.8
β_3	0.5	0.4968	0.0994	0.4969	0.0994	0.1015	95.5	0.4969	0.0994	0.1014	95.9
λ	0.2	0.1991	0.0237	0.1988	0.0238	0.0237	94.3	0.1988	0.0238	0.0242	95.0
ρ	0.3	0.2491	0.0487	0.2949	0.0466	0.0468	94.5	0.2949	0.0466	0.0476	95.1
σ_v^2	1.0	0.9878	0.0217	0.9980	0.0220	0.0225	95.9	—	—	—	—
		Normal Errors, Heteroskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 10$									
β_0	1.0	0.9892	0.2853	0.9862	0.2864	0.3094	95.9	0.9869	0.2873	0.2921	94.7
β_1	1.0	0.9994	0.0199	0.9994	0.0199	0.0199	95.7	0.9994	0.0199	0.0205	96.6
β_2	0.5	0.5031	0.1461	0.5031	0.1461	0.1413	94.4	0.5031	0.1461	0.1449	94.8
β_3	0.5	0.4972	0.2027	0.4970	0.2025	0.1963	94.7	0.4971	0.2026	0.1960	94.9
λ	0.2	0.2016	0.0688	0.2025	0.0697	0.0709	95.0	0.2023	0.0698	0.0702	94.6
ρ	0.3	0.1300	0.1243	0.2567	0.1148	0.1157	95.2	0.2585	0.1149	0.1175	95.3
σ_v^2	1.0	0.9713	0.0537	0.9963	0.0551	0.0582	95.4	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 10$									
β_0	1.0	1.0108	0.1655	1.0117	0.1657	0.1537	92.2	1.0117	0.1658	0.1710	95.0
β_1	1.0	0.9997	0.0102	0.9997	0.0102	0.0093	93.1	0.9997	0.0102	0.0101	94.7
β_2	0.5	0.4996	0.0984	0.4996	0.0984	0.0985	94.2	0.4996	0.0984	0.1020	95.3
β_3	0.5	0.4981	0.1033	0.4982	0.1033	0.1015	94.0	0.4982	0.1033	0.1014	94.2
λ	0.2	0.1992	0.0237	0.1990	0.0238	0.0237	95.5	0.1990	0.0238	0.0238	95.6
ρ	0.3	0.2495	0.0503	0.2953	0.0482	0.0468	94.4	0.2943	0.0481	0.0507	95.8
σ_v^2	1.0	0.9876	0.0302	0.9978	0.0305	0.0343	97.0	—	—	—	—
		Chi-Squared Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 10$									
β_0	1.0	0.9886	0.3158	0.9868	0.3196	0.3092	93.9	0.9870	0.3199	0.3122	94.0
β_1	1.0	1.0005	0.0201	1.0004	0.0201	0.0198	94.6	1.0004	0.0201	0.0198	94.8
β_2	0.5	0.4984	0.1389	0.4984	0.1389	0.1409	95.2	0.4984	0.1389	0.1396	95.1
β_3	0.5	0.5088	0.2061	0.5087	0.2063	0.1978	93.9	0.5087	0.2063	0.1962	93.8
λ	0.2	0.1976	0.0707	0.1981	0.0725	0.0709	94.6	0.1981	0.0726	0.0728	94.1
ρ	0.3	0.1368	0.1260	0.2628	0.1166	0.1147	94.8	0.2632	0.1168	0.1167	94.3
σ_v^2	1.0	0.9662	0.0762	0.9909	0.0781	0.0751	93.0	—	—	—	—
		$N_1 = 20,\ N_2 = 20,\ T = 10$									
β_0	1.0	1.0060	0.1554	1.0069	0.1558	0.1536	95.2	1.0069	0.1558	0.1540	95.1
β_1	1.0	1.0002	0.0093	1.0002	0.0093	0.0093	94.0	1.0002	0.0093	0.0093	94.1
β_2	0.5	0.4977	0.0998	0.4977	0.0998	0.0985	95.2	0.4977	0.0998	0.0982	94.8
β_3	0.5	0.4926	0.1033	0.4927	0.1034	0.1015	94.3	0.4927	0.1034	0.1016	94.1
λ	0.2	0.1999	0.0245	0.1996	0.0246	0.0237	94.5	0.1996	0.0246	0.0242	94.8
ρ	0.3	0.2491	0.0488	0.2950	0.0468	0.0469	95.4	0.2950	0.0468	0.0476	96.0
σ_v^2	1.0	0.9877	0.0377	0.9980	0.0381	0.0385	94.9	—	—	—	—

Table 3: Monte Carlo Results: FE-3, Time-Invariant Spatial Parameters
$W_t = M_t =$ Queen Contiguity

Parameter		QMLE		M-Est				RM-Est			
		Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
Normal Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$											
β_0	5.0	5.0233	0.8390	5.0530	0.8583	0.7863	92.9	5.0535	0.8598	0.8269	93.1
β_1	1.0	0.9985	0.0466	0.9982	0.0466	0.0463	95.7	0.9982	0.0466	0.0459	95.0
λ	0.2	0.1980	0.1022	0.1933	0.1061	0.0900	89.6	0.1933	0.1066	0.1045	92.3
λ	0.3	0.2637	0.1451	0.2904	0.1364	0.1149	89.2	0.2905	0.1376	0.1303	91.7
σ_v^2	1.0	0.7100	0.0523	0.9842	0.0726	0.0735	93.2	—	—	—	—
$N_1 = 20,\ N_2 = 20,\ T = 5$											
β_0	5.0	4.9967	0.5900	5.0018	0.5926	0.5618	93.4	5.0023	0.5929	0.5270	89.5
β_1	1.0	0.9987	0.0214	0.9987	0.0214	0.0213	94.9	0.9987	0.0214	0.0213	95.2
λ	0.2	0.1977	0.0530	0.1968	0.0540	0.0477	91.4	0.1967	0.0541	0.0530	94.2
ρ	0.3	0.3019	0.0684	0.3004	0.0673	0.0596	90.9	0.3006	0.0674	0.0655	94.2
σ_v^2	1.0	0.7565	0.0286	0.9956	0.0376	0.0363	94.2	—	—	—	—
$N_1 = 20,\ N_2 = 20,\ T = 10$											
β_0	5.0	5.0079	0.4587	5.0141	0.4610	0.4544	93.9	5.0143	0.4611	0.4559	93.9
β_1	1.0	1.0006	0.0145	1.0005	0.0145	0.0144	96.8	1.0005	0.0145	0.0144	96.6
λ	0.2	0.1991	0.0346	0.1982	0.0351	0.0327	93.0	0.1982	0.0352	0.0365	95.8
ρ	0.3	0.2985	0.0440	0.2988	0.0432	0.0404	93.6	0.2989	0.0432	0.0448	95.8
σ_v^2	1.0	0.8534	0.0211	0.9982	0.0247	0.0243	95.2	—	—	—	—
Normal Errors, Heteroskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$											
β_0	5.0	5.0410	0.8331	5.0717	0.8532	0.7896	92.8	5.0707	0.8519	0.8251	92.8
β_1	1.0	1.0004	0.0454	1.0002	0.0454	0.0464	95.4	1.0002	0.0454	0.0459	95.1
λ	0.2	0.1969	0.1014	0.1921	0.1054	0.0905	89.7	0.1922	0.1054	0.1054	92.1
ρ	0.3	0.2608	0.1482	0.2878	0.1391	0.1154	89.6	0.2870	0.1391	0.1312	91.8
σ_v^2	1.0	0.7132	0.0542	0.9888	0.0751	0.0758	93.7	—	—	—	—
$N_1 = 20,\ N_2 = 20,\ T = 5$											
β_0	5.0	5.0038	0.5429	5.0084	0.5455	0.5618	94.6	5.0095	0.5456	0.4994	90.7
β_1	1.0	0.9999	0.0215	0.9999	0.0215	0.0213	94.4	0.9999	0.0215	0.0214	94.2
λ	0.2	0.2014	0.0510	0.2006	0.0519	0.0475	92.2	0.2004	0.0520	0.0525	94.1
ρ	0.3	0.2972	0.0669	0.2958	0.0658	0.0596	91.6	0.2963	0.0659	0.0653	94.3
σ_v^2	1.0	0.7582	0.0278	0.9978	0.0365	0.0372	95.5	—	—	—	—
$N_1 = 20,\ N_2 = 20,\ T = 10$											
β_0	5.0	4.9808	0.4710	4.9888	0.4745	0.4853	95.4	4.9893	0.4746	0.4956	95.6
β_1	1.0	1.0003	0.0144	1.0002	0.0144	0.0141	94.9	1.0002	0.0144	0.0141	95.0
λ	0.2	0.1997	0.0333	0.1988	0.0340	0.0319	93.3	0.1988	0.0340	0.0354	95.5
ρ	0.3	0.3040	0.0434	0.3010	0.0430	0.0397	92.7	0.3010	0.0430	0.0437	95.3
σ_v^2	1.0	0.8528	0.0211	0.9978	0.0246	0.0249	95.1	—	—	—	—
Chi-Squared Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$											
β_0	5.0	4.9774	0.8324	5.0047	0.8513	0.7812	92.9	5.0084	0.8527	0.8092	92.1
β_1	1.0	0.9988	0.0463	0.9986	0.0464	0.0463	95.0	0.9986	0.0464	0.0457	94.6
λ	0.2	0.2011	0.1003	0.1967	0.1044	0.0891	89.6	0.1961	0.1049	0.1028	91.3
ρ	0.3	0.2551	0.1491	0.2823	0.1399	0.1147	88.9	0.2832	0.1411	0.1288	91.2
σ_v^2	1.0	0.7115	0.0825	0.9865	0.1143	0.1111	91.4	—	—	—	—
$N_1 = 20,\ N_2 = 20,\ T = 5$											
β_0	5.0	5.0209	0.5597	5.0258	0.5621	0.5621	94.6	5.0264	0.5618	0.5231	90.9
β_1	1.0	0.9990	0.0222	0.9989	0.0222	0.0213	94.3	0.9989	0.0222	0.0213	94.2
λ	0.2	0.1983	0.0524	0.1974	0.0533	0.0477	92.1	0.1973	0.0534	0.0529	94.6
ρ	0.3	0.2999	0.0657	0.2984	0.0647	0.0596	93.3	0.2986	0.0646	0.0654	95.5
σ_v^2	1.0	0.7568	0.0434	0.9960	0.0571	0.0567	94.3	—	—	—	—
$N_1 = 20,\ N_2 = 20,\ T = 10$											
β_0	5.0	4.9945	0.4915	5.0022	0.4952	0.4863	95.1	5.0022	0.4952	0.4859	94.1
β_1	1.0	0.9994	0.0146	0.9994	0.0146	0.0141	94.1	0.9994	0.0146	0.0141	93.8
λ	0.2	0.1994	0.0339	0.1986	0.0346	0.0320	93.2	0.1986	0.0346	0.0354	95.7
ρ	0.3	0.3023	0.0433	0.2993	0.0429	0.0400	92.8	0.2993	0.0429	0.0439	95.5
σ_v^2	1.0	0.8530	0.0328	0.9980	0.0384	0.0397	95.2	—	—	—	—

Table 4: Monte Carlo Results: `FE-1`, Time-Variant Spatial Parameters
$W_t = M_t =$ `Rook Contiguity` with $k = 5$

		`QMLE`		`M-Est`				`RM-Est`			
Parameter		Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
		Normal Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$									
β_0	1.0	0.9757	0.3633	0.9811	0.3656	0.3547	92.7	0.9828	0.3649	0.3717	93.8
β_1	1.0	1.0004	0.0275	0.9997	0.0276	0.0272	94.4	0.9996	0.0275	0.0283	94.9
λ_1	0.7	0.7024	0.0840	0.6982	0.0912	0.1004	90.3	0.6975	0.0968	0.1119	92.6
λ_2	0.5	0.5049	0.0861	0.5010	0.0888	0.0854	91.9	0.5002	0.0883	0.0962	94.5
λ_3	0.4	0.3975	0.0898	0.3956	0.0924	0.0994	92.9	0.3956	0.0941	0.1096	95.9
λ_4	0.3	0.3027	0.0802	0.3015	0.0815	0.0859	95.6	0.3017	0.0821	0.0889	95.3
λ_5	0.2	0.2078	0.0980	0.2063	0.0991	0.0967	90.9	0.2056	0.0986	0.1005	92.6
ρ_1	0.8	0.7558	0.1046	0.7665	0.1108	0.1090	92.5	0.7684	0.1125	0.1208	91.9
ρ_2	0.7	0.6640	0.1167	0.6761	0.1185	0.1062	92.0	0.6807	0.1185	0.1195	92.8
ρ_3	0.5	0.4669	0.1437	0.4792	0.1445	0.1427	93.3	0.4851	0.1470	0.1539	94.1
ρ_4	0.4	0.3702	0.1461	0.3833	0.1452	0.1439	92.9	0.3875	0.1479	0.1510	92.9
ρ_5	0.4	0.3566	0.1534	0.3585	0.1496	0.1499	93.6	0.3631	0.1505	0.1570	92.0
σ_v^2	1.0	0.9364	0.0607	0.9753	0.0633	0.0651	92.9	—	—	—	—
		Normal Mixture Errors, Heteroskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$									
β_0	1.0	0.9855	0.3465	0.9910	0.3489	0.3566	94.2	0.9933	0.3499	0.3695	94.2
β_1	1.0	0.9982	0.0411	0.9975	0.0411	0.0271	94.7	0.9977	0.0411	0.0282	95.3
λ_1	0.7	0.7040	0.0930	0.6998	0.1001	0.0985	86.8	0.6993	0.1008	0.1117	92.1
λ_2	0.5	0.4997	0.0883	0.4958	0.0910	0.0851	90.0	0.4959	0.0916	0.0974	93.7
λ_3	0.4	0.4023	0.1001	0.4007	0.1029	0.0993	89.6	0.4019	0.1030	0.1137	94.1
λ_4	0.3	0.2999	0.0952	0.2987	0.0967	0.0852	90.9	0.2990	0.0942	0.0937	94.3
λ_5	0.2	0.2044	0.0976	0.2029	0.0987	0.0978	93.0	0.2023	0.0970	0.0981	93.5
ρ_1	0.8	0.7504	0.1159	0.7605	0.1227	0.1114	89.5	0.7642	0.1178	0.1243	91.2
ρ_1	0.7	0.6623	0.1182	0.6745	0.1199	0.1104	92.6	0.6798	0.1182	0.1203	93.7
ρ_1	0.5	0.4614	0.1496	0.4734	0.1507	0.1467	92.8	0.4823	0.1505	0.1532	92.7
ρ_1	0.4	0.3675	0.1598	0.3806	0.1591	0.1460	91.8	0.3838	0.1566	0.1538	92.5
ρ_1	0.4	0.3568	0.1589	0.3588	0.1550	0.1534	92.7	0.3667	0.1514	0.1525	92.7
σ_v^2	1.0	0.9312	0.1018	0.9699	0.1061	0.1001	88.7	—	—	—	—
		Normal Errors, Homoskedasticity; $N_1 = 20,\ N_2 = 20,\ T = 5$									
β_0	1.0	0.9969	0.2023	0.9983	0.2026	0.2044	94.5	0.9994	0.2030	0.2041	94.3
β_1	1.0	0.9997	0.0136	0.9994	0.0136	0.0137	96.1	0.9995	0.0137	0.0138	95.7
λ_1	0.7	0.7008	0.0473	0.6988	0.0489	0.0490	93.3	0.6997	0.0524	0.0575	95.7
λ_2	0.5	0.5038	0.0450	0.5022	0.0456	0.0460	94.5	0.5026	0.0469	0.0503	95.5
λ_3	0.4	0.3997	0.0427	0.3992	0.0431	0.0439	94.1	0.3995	0.0432	0.0465	95.8
λ_4	0.3	0.3005	0.0438	0.3001	0.0440	0.0429	93.5	0.2997	0.0443	0.0461	95.1
λ_5	0.2	0.2016	0.0444	0.2011	0.0446	0.0447	93.6	0.2007	0.0449	0.0445	94.5
ρ_1	0.8	0.7915	0.0485	0.7927	0.0501	0.0494	93.1	0.7920	0.0536	0.0571	94.0
ρ_2	0.7	0.6924	0.0507	0.6928	0.0514	0.0524	94.7	0.6924	0.0554	0.0576	95.5
ρ_3	0.5	0.4962	0.0668	0.4957	0.0669	0.0666	94.2	0.4963	0.0687	0.0702	95.0
ρ_4	0.4	0.3905	0.0721	0.3903	0.0718	0.0712	94.6	0.3921	0.0727	0.0755	95.0
ρ_5	0.4	0.3948	0.0708	0.3933	0.0701	0.0708	94.7	0.3954	0.0727	0.0717	94.1
σ_v^2	1.0	0.9735	0.0323	0.9946	0.0330	0.0330	95.0	—	—	—	—
		Normal Mixture Errors, Heteroskedasticity; $N_1 = 20,\ N_2 = 20,\ T = 5$									
β_0	1.0	0.9947	0.1926	0.9961	0.1928	0.2035	96.2	0.9964	0.1944	0.1943	95.1
β_1	1.0	0.9992	0.0146	0.9990	0.0147	0.0137	93.4	0.9991	0.0147	0.0144	95.1
λ_1	0.7	0.7001	0.0465	0.6980	0.0481	0.0489	92.2	0.6994	0.0519	0.0571	94.8
λ_2	0.5	0.5039	0.0438	0.5023	0.0444	0.0463	94.8	0.5034	0.0467	0.0513	95.9
λ_3	0.4	0.4007	0.0444	0.4002	0.0448	0.0439	93.6	0.4005	0.0446	0.0468	95.4
λ_4	0.3	0.3009	0.0431	0.3005	0.0434	0.0431	93.9	0.3006	0.0436	0.0459	96.3
λ_5	0.2	0.2024	0.0428	0.2019	0.0430	0.0447	95.7	0.2019	0.0430	0.0439	95.8
ρ_1	0.8	0.7911	0.0517	0.7924	0.0533	0.0513	91.1	0.7922	0.0544	0.0561	93.4
ρ_2	0.7	0.6898	0.0538	0.6901	0.0546	0.0558	95.1	0.6909	0.0557	0.0579	96.2
ρ_3	0.5	0.4956	0.0707	0.4951	0.0708	0.0691	93.3	0.4971	0.0701	0.0697	94.2
ρ_4	0.4	0.3902	0.0732	0.3899	0.0729	0.0732	94.4	0.3917	0.0721	0.0743	94.9
ρ_5	0.4	0.3899	0.1000	0.4000	0.1000	0.0733	95.4	0.3908	0.0694	0.0715	94.7
σ_v^2	1.0	0.9706	0.0522	0.9917	0.0533	0.0556	93.4	—	—	—	—

Table 5: Monte Carlo Results: `FE-2`, Time-Variant Spatial Parameters
$W_t = M_t =$ `Rook Contiguity` with $k = 5$

		`QMLE`		`M-Est`				`RM-Est`			
Parameter		Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
		Normal Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 6$									
β_0	1.0	0.9849	0.5414	0.9925	0.5454	0.5750	93.0	0.9962	0.5510	0.6531	96.1
β_1	1.0	0.9974	0.0254	0.9969	0.0254	0.0254	94.7	0.9968	0.0255	0.0264	95.3
β_2	0.5	0.5104	0.1690	0.5114	0.1689	0.1727	95.0	0.5115	0.1692	0.1752	95.2
β_3	0.5	0.5272	0.6381	0.5215	0.6431	0.6647	93.4	0.5177	0.6486	0.6771	93.8
λ_1	0.7	0.6986	0.0834	0.6946	0.0893	0.0915	90.4	0.6959	0.0944	0.1323	95.0
λ_2	0.5	0.5003	0.0902	0.4986	0.0913	0.0992	93.4	0.4977	0.0924	0.1195	97.5
λ_3	0.4	0.3946	0.0977	0.3928	0.1003	0.1022	91.4	0.3942	0.1024	0.1355	96.4
λ_4	0.3	0.2986	0.0770	0.2977	0.0781	0.0766	92.3	0.2976	0.0781	0.1013	97.5
λ_5	0.2	0.2041	0.0971	0.2019	0.0989	0.1004	93.1	0.2021	0.0993	0.1315	97.2
λ_6	0.2	0.1949	0.0988	0.1946	0.0995	0.0986	91.1	0.1945	0.0994	0.1075	93.5
ρ_1	0.8	0.7572	0.1017	0.7706	0.1053	0.1012	89.2	0.7734	0.1081	0.1327	93.3
ρ_2	0.7	0.6686	0.1164	0.6637	0.1163	0.1132	92.4	0.6664	0.1174	0.1341	95.6
ρ_3	0.5	0.4686	0.1402	0.4851	0.1406	0.1437	93.4	0.4882	0.1442	0.1801	95.1
ρ_4	0.4	0.3638	0.1464	0.3816	0.1452	0.1387	91.9	0.3877	0.1480	0.1757	95.0
ρ_5	0.4	0.3651	0.1540	0.3837	0.1531	0.1488	91.5	0.3876	0.1561	0.1875	94.2
ρ_6	0.3	0.2683	0.1675	0.2744	0.1626	0.1570	91.7	0.2774	0.1640	0.1605	92.2
σ_v^2	1.0	0.9409	0.0574	0.9717	0.0594	0.0588	90.6	—	—	—	—
		Normal Errors, Homoskedasticity; $N_1 = 20,\ N_2 = 10,\ T = 6$									
β_0	1.0	0.9987	0.4812	1.0032	0.4843	0.4995	95.1	0.9988	0.5008	0.5807	96.2
β_1	1.0	0.9995	0.0155	0.9993	0.0155	0.0155	94.6	0.9994	0.0155	0.0158	95.2
β_2	0.5	0.4914	0.1750	0.4914	0.1750	0.1749	94.5	0.4917	0.1750	0.1746	94.6
β_3	0.5	0.5035	0.4930	0.4996	0.4961	0.5097	94.8	0.5047	0.5129	0.5528	95.4
λ_1	0.7	0.7017	0.0575	0.7007	0.0583	0.0640	94.4	0.7020	0.0628	0.0856	98.0
λ_2	0.5	0.5027	0.0651	0.5003	0.0664	0.0662	93.1	0.5003	0.0683	0.0827	97.4
λ_3	0.4	0.3997	0.0655	0.3989	0.0664	0.0648	92.2	0.3994	0.0679	0.0803	97.4
λ_4	0.3	0.2988	0.0592	0.2982	0.0596	0.0589	92.8	0.2983	0.0595	0.0698	96.3
λ_5	0.2	0.2020	0.0581	0.2012	0.0585	0.0574	94.2	0.2009	0.0586	0.0682	97.5
λ_6	0.2	0.2000	0.0647	0.1996	0.0649	0.0654	94.7	0.1995	0.0648	0.0708	96.3
ρ_1	0.8	0.7812	0.0689	0.7771	0.0702	0.0702	94.3	0.7760	0.0750	0.0925	97.3
ρ_2	0.7	0.6802	0.0791	0.6854	0.0800	0.0760	94.1	0.6862	0.0848	0.0924	95.3
ρ_3	0.5	0.4845	0.1051	0.4896	0.1050	0.0963	92.0	0.4912	0.1071	0.1154	94.7
ρ_4	0.4	0.3867	0.1002	0.3926	0.0996	0.0996	94.5	0.3949	0.1008	0.1170	96.5
ρ_5	0.4	0.3822	0.1001	0.3882	0.0993	0.0977	94.6	0.3907	0.1014	0.1151	96.6
ρ_6	0.3	0.2890	0.1072	0.2908	0.1050	0.1086	94.9	0.2939	0.1065	0.1120	94.8
σ_v^2	1.0	0.9616	0.0403	0.9866	0.0414	0.0422	93.6	—	—	—	—
		Normal Mixture, Heteroskedasticity; $N_1 = 20,\ N_2 = 10,\ T = 6$									
β_0	1.0	0.9736	0.4666	0.9778	0.4696	0.4931	94.7	0.9858	0.5095	0.5903	95.9
β_1	1.0	0.9989	0.0170	0.9987	0.0170	0.0155	92.5	0.9988	0.0171	0.0171	94.6
β_2	0.5	0.4968	0.1641	0.4968	0.1641	0.1748	96.3	0.4965	0.1641	0.1608	94.9
β_3	0.5	0.5336	0.4895	0.5301	0.4926	0.5032	93.9	0.5236	0.5348	0.5738	94.2
λ_1	0.7	0.7022	0.0614	0.7013	0.0623	0.0638	93.3	0.7024	0.0651	0.0862	97.2
λ_2	0.5	0.5047	0.0651	0.5023	0.0664	0.0666	93.1	0.5025	0.0687	0.0838	97.2
λ_3	0.4	0.4016	0.0666	0.4008	0.0674	0.0650	92.0	0.4010	0.0680	0.0817	96.9
λ_4	0.3	0.2994	0.0601	0.2988	0.0605	0.0590	92.7	0.2988	0.0592	0.0696	97.0
λ_5	0.2	0.2023	0.0569	0.2015	0.0572	0.0579	95.0	0.2016	0.0569	0.0664	97.9
λ_6	0.2	0.1996	0.0635	0.1992	0.0638	0.0655	95.0	0.1989	0.0627	0.0688	96.4
ρ_1	0.8	0.7773	0.0730	0.7731	0.0745	0.0737	94.5	0.7745	0.0758	0.0936	97.1
ρ_2	0.7	0.6776	0.0825	0.6828	0.0835	0.0808	93.7	0.6857	0.0847	0.0930	97.0
ρ_3	0.5	0.4808	0.1001	0.4858	0.1001	0.1004	94.2	0.4914	0.1012	0.1167	96.7
ρ_4	0.4	0.3847	0.1055	0.3906	0.1047	0.1025	94.3	0.3951	0.1028	0.1176	96.0
ρ_5	0.4	0.3766	0.0992	0.3827	0.0984	0.1014	94.9	0.3875	0.0982	0.1137	96.6
ρ_6	0.3	0.2851	0.1121	0.2870	0.1098	0.1108	94.7	0.2922	0.1091	0.1111	94.2
σ_v^2	1.0	0.9600	0.0667	0.9850	0.0684	0.0712	93.3	—	—	—	—

Table 6: Monte Carlo Results: FE-3, Time-Variant Spatial Parameters
$W_t = M_t =$ Rook Contiguity with $k = 5$

Parameter		QMLE Mean	QMLE sd	M-Est Mean	M-Est sd	M-Est se	CP95	RM-Est Mean	RM-Est sd	RM-Est se	CP95r
		Normal Errors, Homoskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$									
β_0	1.0	1.0070	0.5988	1.0145	0.5978	0.5918	95.0	1.0123	0.5966	0.5731	93.2
β_1	1.0	0.9972	0.0495	0.9966	0.0494	0.0381	94.6	0.9967	0.0495	0.0388	94.8
λ_1	0.7	0.7093	0.0724	0.6974	0.0826	0.0805	91.4	0.6988	0.0884	0.0949	93.8
λ_2	0.5	0.5082	0.0759	0.4986	0.0829	0.0813	92.7	0.4992	0.0842	0.0893	94.3
λ_3	0.4	0.3991	0.0844	0.3976	0.0879	0.0850	92.6	0.3997	0.0893	0.0927	93.2
λ_4	0.3	0.2976	0.0671	0.2980	0.0679	0.0658	93.9	0.2986	0.0678	0.0691	94.4
λ_5	0.2	0.2009	0.1004	0.1994	0.1021	0.0948	90.5	0.1997	0.1026	0.1002	92.7
ρ_1	0.8	0.8194	0.0858	0.7679	0.1110	0.1007	93.6	0.7656	0.1160	0.1169	93.6
ρ_2	0.7	0.7412	0.1170	0.6717	0.1331	0.1141	93.1	0.6760	0.1377	0.1305	93.5
ρ_3	0.5	0.5454	0.1871	0.4597	0.1725	0.1601	92.6	0.4657	0.1811	0.1776	91.4
ρ_4	0.4	0.4601	0.2040	0.3708	0.1738	0.1630	91.8	0.3775	0.1788	0.1808	91.8
ρ_5	0.4	0.4508	0.2102	0.3582	0.1808	0.1759	91.6	0.3679	0.1864	0.1900	92.1
σ_v^2	1.0	0.6654	0.0568	0.9626	0.0812	0.0762	89.8	—	—	—	—
		Normal Mixture Errors, Heteroskedasticity; $N_1 = 10,\ N_2 = 10,\ T = 5$									
β_0	1.0	0.9895	0.5054	0.9965	0.5023	0.5943	98.5	0.9917	0.5027	0.5099	93.5
β_1	1.0	0.9997	0.0377	0.9991	0.0377	0.0382	95.1	0.9992	0.0377	0.0382	94.8
λ_1	0.7	0.7112	0.0696	0.7007	0.0825	0.0792	90.0	0.7006	0.0854	0.0938	94.4
λ_2	0.5	0.5029	0.0837	0.4940	0.0918	0.0808	89.3	0.4954	0.0936	0.0944	93.1
λ_2	0.4	0.4016	0.0844	0.3993	0.0885	0.0859	91.4	0.4032	0.0917	0.0979	94.4
λ_4	0.3	0.2963	0.0708	0.2964	0.0719	0.0662	91.4	0.2980	0.0711	0.0700	94.0
λ_5	0.2	0.2014	0.0978	0.1999	0.1004	0.0960	92.9	0.2012	0.1003	0.0991	92.8
ρ_1	0.8	0.8160	0.0869	0.7616	0.1185	0.1072	93.2	0.7622	0.1200	0.1180	93.9
ρ_2	0.7	0.7366	0.1115	0.6648	0.1332	0.1233	92.1	0.6751	0.1324	0.1296	92.4
ρ_3	0.5	0.5546	0.1714	0.4683	0.1648	0.1669	94.1	0.4776	0.1678	0.1775	92.6
ρ_4	0.4	0.4683	0.2043	0.3782	0.1776	0.1683	92.4	0.3816	0.1788	0.1789	91.1
ρ_5	0.4	0.4479	0.2199	0.3561	0.1883	0.1831	90.9	0.3654	0.1926	0.1869	91.7
σ_v^2	1.0	0.6678	0.0769	0.9660	0.1101	0.1123	90.5	—	—	—	—
		Normal Errors, Homoskedasticity; $N_1 = 20,\ N_2 = 20,\ T = 5$									
β_0	1.0	1.0036	0.5057	0.9988	0.5055	0.5009	95.00	0.9982	0.5062	0.4573	90.9
β_1	1.0	1.0000	0.0189	0.9994	0.0189	0.0190	95.00	0.9994	0.0189	0.0191	95.2
λ_1	0.7	0.7080	0.0379	0.6973	0.0444	0.0446	93.80	0.6976	0.0479	0.0486	92.4
λ_2	0.5	0.5088	0.0417	0.5016	0.0448	0.0443	94.00	0.5025	0.0464	0.0462	94.3
λ_3	0.4	0.3967	0.0355	0.3980	0.0368	0.0382	95.30	0.3984	0.0369	0.0393	94.8
λ_4	0.3	0.2978	0.0378	0.3011	0.0379	0.0377	94.40	0.3009	0.0382	0.0387	94.4
λ_5	0.2	0.1973	0.0463	0.2007	0.0478	0.0466	94.60	0.2005	0.0486	0.0480	94.4
ρ_1	0.8	0.8424	0.0334	0.7942	0.0476	0.0481	93.70	0.7939	0.0511	0.0519	93.2
ρ_2	0.7	0.7621	0.0423	0.6913	0.0548	0.0553	95.20	0.6909	0.0602	0.0600	95.1
ρ_3	0.5	0.5963	0.0668	0.4957	0.0727	0.0729	94.70	0.4972	0.0761	0.0779	95.1
ρ_4	0.4	0.4908	0.0875	0.3859	0.0838	0.0814	94.20	0.3887	0.0872	0.0859	93.8
ρ_5	0.4	0.4941	0.0896	0.3902	0.0871	0.0847	94.00	0.3927	0.0913	0.0898	94.0
σ_v^2	1.0	0.7220	0.0287	0.9938	0.0388	0.0386	93.90	—	—	—	—
		Normal Mixture Errors, Heteroskedasticity; $N_1 = 20,\ N_2 = 20,\ T = 5$									
β_0	1.0	1.0165	0.3772	1.0097	0.3706	0.4998	99.60	1.0092	0.3700	0.3368	91.1
β_1	1.0	1.0000	0.0190	0.9995	0.0190	0.0190	94.20	0.9996	0.0190	0.0185	93.6
λ_1	0.7	0.7088	0.0372	0.6983	0.0436	0.0444	93.40	0.6991	0.0462	0.0477	94.0
λ_2	0.5	0.5076	0.0409	0.5003	0.0445	0.0444	95.10	0.5017	0.0457	0.0475	95.7
λ_3	0.4	0.3981	0.0365	0.3996	0.0382	0.0382	94.10	0.4003	0.0378	0.0390	94.8
λ_4	0.3	0.2974	0.0371	0.3007	0.0372	0.0377	95.30	0.3011	0.0373	0.0382	95.1
λ_5	0.2	0.1980	0.0441	0.2017	0.0455	0.0467	95.00	0.2014	0.0455	0.0468	95.3
ρ_1	0.8	0.8401	0.0379	0.7912	0.0535	0.0518	93.10	0.7919	0.0523	0.0515	92.8
ρ_2	0.7	0.7607	0.0467	0.6893	0.0598	0.0607	94.80	0.6910	0.0594	0.0597	94.0
ρ_3	0.5	0.5945	0.0724	0.4937	0.0788	0.0785	93.60	0.4950	0.0786	0.0773	94.0
ρ_4	0.4	0.4916	0.0895	0.3867	0.0857	0.0858	95.00	0.3891	0.0855	0.0842	94.3
ρ_5	0.4	0.4885	0.0911	0.3849	0.0879	0.0900	95.50	0.3895	0.0893	0.0893	94.9
σ_v^2	1.0	0.7201	0.0430	0.9910	0.0584	0.0625	95.90	—	—	—	—

Supplemental Materials

"Multi-Dimensional Spatial Panel Data Model with Fixed Effects: Formulation, Estimation and Inference"

The `Supplemental Materials` contain **Appendix B:** Proofs of Theoretical Results, **Appendix C:** A Complete Set of Monte Carlo Results, and **Appendix D:** A Practical Guide.

Appendix B: Proofs of Theoretical Results

We present sketches of the proofs only since once the concise expressions of the key quantities, such as the concentrated AQS functions of δ and their population counterparts, variances of the AQS functions at the true parameter values and Hessian matrices, the proofs of the theorems and corollaries are fairly standard in the spatial econometrics literature. A complication occurs in the case where the spatial parameters vary with time, which makes the estimation and inference problem a high-dimensional one. With the introduction of the finite number of linear contrasts $\mathbf{C}$ and the rate modifier R, we convert the problem back to one of a finite-dimensional nature. Practical applications proceed as usual.

Proof of Theorem 1

Proof of consistency: Substituting $\hat{\beta}_N^*(\delta)$ and $\hat{\sigma}_N^{*2}(\delta)$ given in (3.10) into the δ-component of $S_N^*(\theta)$ given in (3.5), we obtain the concentrated AQS function

$$S_N^{*c}(\delta) = \begin{cases} Y'\mathbf{W}'\mathbb{B}'\hat{V}(\delta)/\hat{\sigma}_N^{*2}(\delta) - \mathtt{tr}[\mathbb{Q}(\rho)\bar{\mathbb{F}}(\delta)], \\ \hat{V}'(\delta)\mathbb{G}(\rho)\hat{V}(\delta)/\hat{\sigma}_N^{*2}(\delta) - \mathtt{tr}[\mathbb{Q}(\rho)\mathbb{G}(\rho)], \end{cases} \tag{B.1}$$

where $\hat{V}(\delta) = \tilde{V}(\hat{\beta}_N^*(\delta), \delta)$. The M-estimator of δ, $\hat{\delta}_{\mathtt{M}}^* = \arg\{S_N^{*c}(\delta) = 0\}$.

By Theorem 5.9 of van der Vaart (1998), $\hat{\delta}_{\mathtt{M}}^*$ is consistent if the identification uniqueness condition (Assumption 7) and the uniform convergence result, $\sup_{\delta\in\delta}\frac{1}{N^*}\left\|S_N^{*c}(\delta) - \bar{S}_N^{*c}(\delta)\right\| \xrightarrow{p} 0$, hold, where $\bar{S}_N^{*c}(\delta)$ is given in (A.3). To show the uniform convergence, we **first** show that $\inf_{\delta\in\Delta}\bar{\sigma}_N^{*2}(\delta) > 0$, where $\bar{\sigma}_N^{*2}(\delta)$ given in (A.2). The second term of $\bar{\sigma}_N^{*2}(\delta)$ is of the form $a'a$ for an $N \times 1$ vector a, and hence is nonnegative. The first term is of the form $\mathtt{tr}(A'A)$ for an $N \times N$ matrix A, and can be shown to be strictly positive.[10] Therefore, $\bar{\sigma}_N^{*2}(\delta) > c > 0$.

[10] We skip the detail as it is lengthy. We refer the reader to Meng and Yang (2021) in a simpler context.

Next, we show that $\sup_{\delta\in\Delta}\left|\hat{\sigma}_N^{*2}(\delta)-\bar{\sigma}_N^{*2}(\delta)\right| = o_p(1)$. Using the notation introduced in Appendix A.2, writing $\hat{V}(\delta) = \mathbf{Q}(\rho)\mathbb{Q}(\rho)\mathbb{Y}(\delta)$, then $\hat{\sigma}_N^{*2}(\delta) = \frac{1}{N^*}Y'\Pi_2(\delta)Y$. We have

$$Y'\Pi_2(\delta)Y = V'(\mathbb{A}'\mathbb{B}')^{-1}\Pi_2(\delta)(\mathbb{B}\mathbb{A})^{-1}V + \zeta_0'\Pi_2(\delta)\zeta_0 + 2\zeta_0'\Pi_2(\delta)(\mathbb{B}\mathbb{A})^{-1}V,$$

by $Y = \zeta_0 + (\mathbb{B}\mathbb{A})^{-1}V$. Compared with $\bar{\sigma}_N^{*2}(\delta)$ given in (A.2), the result follows if

$$\sup_{\delta\in\Delta}\tfrac{1}{N^*}\left|V'(\mathbb{A}'\mathbb{B}')^{-1}\Pi_2(\delta)(\mathbb{B}\mathbb{A})^{-1}V - \sigma_0^2\mathrm{tr}[\Pi_1(\delta)(\mathbb{A}'\mathbb{B}'\mathbb{B}\mathbb{A})^{-1}]\right| = o_p(1), \text{ and}$$
$$\sup_{\delta\in\Delta}\left|\tfrac{1}{N^*}2\zeta_0'\Pi_2(\delta)(\mathbb{B}\mathbb{A})^{-1}V\right| = o_p(1).$$

These results are proved by applying the (weak) uniform law of large numbers (e.g., Andrews, 1992), and Lemma A.2 and Lemma A.3. The detailed proofs are lengthy and are omitted here. They are available from the authors upon request.

With these results and the fact that the second terms in (B.1) and (A.3) are the same and thus cancel out upon taking difference, we are left to show that the differences between the numerators of the first terms of (B.1) and those of (A.3) are uniformly $o_p(1)$.

We have $Y'\mathbf{W}'\mathbb{B}'(\rho)\hat{V}(\delta) = Y'\mathbb{F}'(\lambda)\Pi_2(\delta)Y$ and $\hat{V}'(\delta)\mathbb{G}(\rho)\hat{V}(\delta) = Y'\Pi_4(\delta)Y$, which are

$$V'(\mathbb{A}'\mathbb{B}')^{-1}\mathbb{F}'(\lambda)\Pi_2(\delta)(\mathbb{B}\mathbb{A})^{-1}V + \zeta_0'\mathbb{F}'(\lambda)\Pi_2(\rho)\zeta_0 + 2\zeta_0'\mathbb{F}'(\lambda)\Pi_2(\rho)(\mathbb{B}\mathbb{A})^{-1}V, \text{ and}$$
$$V'(\mathbb{A}'\mathbb{B}')^{-1}\Pi_4(\delta)(\mathbb{B}\mathbb{A})^{-1}V + \zeta_0'\Pi_4(\rho)\zeta_0 + 2\zeta_0'\Pi_4(\rho)(\mathbb{B}\mathbb{A})^{-1}V.$$

Compared with the numerators of the first terms in (A.3), the desired results follow if

$$\sup_{\delta\in\Delta}\tfrac{1}{N^*}\left|V'(\mathbb{A}'\mathbb{B}')^{-1}\mathbb{F}'(\lambda)\Pi_2(\delta)(\mathbb{B}\mathbb{A})^{-1}V - \sigma_0^2\mathtt{tr}[\mathbb{F}'(\lambda)\Pi_1(\delta)(\mathbb{A}'\mathbb{B}'\mathbb{B}\mathbb{A})^{-1}]\right| = o_p(1),$$
$$\sup_{\delta\in\Delta}\tfrac{1}{N^*}\left|\zeta_0'\mathbb{F}'(\lambda)\Pi_2(\rho)(\mathbb{B}\mathbb{A})^{-1}V\right| = o_p(1),$$
$$\sup_{\delta\in\Delta}\tfrac{1}{N^*}\left|V'(\mathbb{A}'\mathbb{B}')^{-1}\Pi_4(\delta)(\mathbb{B}\mathbb{A})^{-1}V - \sigma_0^2\mathtt{tr}[\Pi_3(\delta)(\mathbb{A}'\mathbb{B}'\mathbb{B}\mathbb{A})^{-1}]\right| = o_p(1),$$
$$\sup_{\delta\in\Delta}\tfrac{1}{N^*}\left|\zeta_0'\Pi_4(\rho)(\mathbb{B}\mathbb{A})^{-1}V\right| = o_p(1).$$

The detailed proofs of these results are lengthy but straightforward by applying the (weak) uniform law of large numbers of Andrews (1992) and Lemma A.2 and Lemma A.3. Hence, they are omitted but are available from the authors upon request.

Proof of asymptotic normality: Applying the mean value theorem (MVT) to each element of $S_N^*(\hat{\theta}_{\mathtt{M}}^*)$, we have

$$0 = \tfrac{1}{\sqrt{N^*}}S_N^*(\hat{\theta}_{\mathtt{M}}^*) = \tfrac{1}{\sqrt{N^*}}S_N^*(\theta_0) + \left[\tfrac{1}{N^*}\tfrac{\partial}{\partial\theta'}S_N^*(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}\right]\sqrt{N^*}(\hat{\theta}_{\mathtt{M}}^* - \theta_0), \tag{B.2}$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}_N^*$ and θ_0. The result of the theorem follows if

(a) $\frac{1}{\sqrt{N^*}} S_N^*(\theta_0) \xrightarrow{D} N[0, \lim_{N^* \to \infty} \Omega_*(\theta_0)]$,

(b) $\frac{1}{N^*}[\frac{\partial}{\partial \theta'} S_N^*(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}} - \frac{\partial}{\partial \theta'} S_N^*(\theta_0)] = o_p(1)$, and

(c) $\frac{1}{N^*}[\frac{\partial}{\partial \theta'} S_N^*(\theta_0) - \mathrm{E}(\frac{\partial}{\partial \theta'} S_N^*(\theta_0))] = o_p(1)$.

The key step in the **proof of (a)** is to write $S_{nT}^*(\theta_0)$ as linear-quadratic (lq) forms in V so that the central limit theorem (CLT) for LQ form of Kelejian and Prucha (2001) can be applied to prove its asymptotic normality for each LQ element. The joint asymptotic normality of $\frac{1}{\sqrt{N^*}} S_N^*(\theta_0)$ therefore follows by applying Cramér-Wold device.

The proofs of (b) and (c) are done by invoking certain types of (weak) law of large numbers, which are tedious but straightforward, and therefore are omitted. The interested readers can refer to Meng and Yang (2021) or directly write to the authors. ■

Proof of Theorem 2. When T is fixed, the proof is similar to that of Theorem 2 since the number of parameters involved is still finite. When T increases with the sample size N, the dimension of θ increases with N. Therefore, the standard methods for studying the asymptotic properties of the M-estimator cannot be applied. With the introduction of finite number of linear contrasts $\mathbf{C}$ and the convergence rate modifier R above Theorem 2, this 'high-dimensional problem' is converted to a fixed-dimensional one. The rest of the proof is similar to the proof of Theorem 1. ■

Proof of Theorem 3. The standard tools, the CLT for LQ forms of Kelejian and Prucha (2001) and the weak ULLN of Andrews (1992), used in studying the asymptotic properties of the M-estimators remain valid when the errors are independent and heteroskedastic. Therefore, the theorem is proved along a similar line as that of Theorem 1. ■

Proof of Theorem 4. Similar to Theorem 2, the problem is non-standard as the dimension of δ vector increases with T. Again, with the introduction of the finite number of linear contrasts $\mathbf{C}$ and the convergence rate modifier R above Theorem 4, the high-dimensional problem is reduced to a finite dimensional one. ■

Proof of Corollary 1. The proof is carried out along a similar line of Meng and Yang (2021) for genuinely unbalanced 2D spatial panel data models. ■

Proof of Corollary 2. The proof is an extension of that of Corollary 1 allowing spatial parameters to vary with time, by using linear contrast $\mathbf{C}$ and rate modifier R. ■

Proof of Corollary 3. The proof is an extension of that of Corollary 1 allowing errors to be heteroskedastic. ∎

Proof of Corollary 4. The proof extends that of Corollary 2 by further allowing the errors to be heteroskedastic. ∎

Appendix C: A Complete Set of Monte Carlo Results

A complete set of Monte Carlo results is given, including those reported in the main text. The results are summarized in six long tables that parallel Tables 1-6 in the main text. They are labeled by whether the spatial parameters change with time:

Case 1. Spatial parameters are time invariant,

Case 2. Spatial parameters vary with time;

by the FE specifications:

`FE-1`: The (i, j, t) specification given in Table A of the main text, without STIC effect,

`FE-2`: The (i, j, t) specification given in Table A of the main text, with STIC,

`FE-3`: The (ij, jt) specification given in Table A of the main text, without STIC effect;

by errors distributions (DGP = 1, 2, 3, for normal, mixture normal, and gamma), and by whether the errors are homoskedastic (Het = 1) and heteroskedastic (Het = 2).

The observations and conclusions drawn from the complete set of Monte Carlo results are consistent with those presented in the main text.

Table 1: Monte Carlo Results: Case 1, FE-1

Wt = Mt = Circular World with r = 10

	QMLE		M-Est				RM-Est			
par	Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
			N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 1							
1	1.0073	0.4606	1.0053	0.4677	0.4715	94.2	1.0057	0.4681	0.4816	94.6
1	0.9993	0.0292	0.9992	0.0292	0.0302	95.7	0.9992	0.0292	0.0300	95.5
1	0.9433	0.0604	0.9891	0.0627	0.0638	94.2	0.0000	0.0000	0.0000	0.0
0.2	0.1976	0.0924	0.1980	0.0949	0.0941	94.1	0.1979	0.0950	0.0974	93.6
0.3	0.0639	0.1941	0.2523	0.1693	0.1659	94.3	0.2525	0.1697	0.1693	94.4
			N1 = 10; N2 = 10; T = 10; DGP = 1; Het = 1							
1	0.9942	0.2278	0.9933	0.2297	0.2255	95.0	0.9933	0.2297	0.2260	94.8
1	0.9996	0.0197	0.9996	0.0197	0.0199	94.5	0.9996	0.0197	0.0199	94.4
1	0.9711	0.0454	0.9980	0.0466	0.0451	93.1	0.0000	0.0000	0.0000	0.0
0.2	0.1985	0.0617	0.1992	0.0636	0.0615	94.7	0.1991	0.0637	0.0630	94.7
0.3	0.1359	0.1261	0.2738	0.1160	0.1096	94.0	0.2739	0.1161	0.1114	94.0
			N1 = 20; N2 = 20; T = 5; DGP = 1; Het = 1							
1	1.0031	0.2043	1.0046	0.2047	0.2084	95.8	1.0047	0.2048	0.2091	95.2
1	1.0003	0.0127	1.0003	0.0127	0.0130	95.6	1.0003	0.0127	0.0130	95.5
1	0.9775	0.0324	0.9974	0.0331	0.0319	94.7	0.0000	0.0000	0.0000	0.0
0.2	0.2000	0.0324	0.1995	0.0326	0.0320	94.9	0.1995	0.0326	0.0327	94.9
0.3	0.2115	0.0774	0.2911	0.0714	0.0681	94.6	0.2911	0.0714	0.0694	94.6
			N1 = 20; N2 = 20; T = 10; DGP = 1; Het = 1							
1	1.0039	0.1681	1.0060	0.1688	0.1732	95.0	1.0060	0.1689	0.1754	95.0
1	1.0001	0.0090	1.0000	0.0090	0.0094	96.4	1.0000	0.0090	0.0094	96.5
1	0.9878	0.0219	0.9985	0.0222	0.0225	95.7	0.0000	0.0000	0.0000	0.0
0.2	0.1994	0.0302	0.1988	0.0305	0.0302	94.5	0.1988	0.0305	0.0311	95.6
0.3	0.2488	0.0516	0.2973	0.0494	0.0498	95.6	0.2973	0.0495	0.0511	96.2
			N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 2							
1	1.0128	0.5310	1.0091	0.5464	0.4709	89.9	1.0092	0.5465	0.5566	94.4
1	1.0005	0.0305	1.0005	0.0306	0.0302	93.9	1.0005	0.0306	0.0306	94.3
1	0.9434	0.0810	0.9892	0.0845	0.0941	95.0	0.0000	0.0000	0.0000	0.0
0.2	0.1960	0.1077	0.1969	0.1117	0.0937	89.4	0.1969	0.1115	0.1136	93.4
0.3	0.0581	0.2092	0.2467	0.1827	0.1662	92.9	0.2442	0.1826	0.1785	94.5
			N1 = 10; N2 = 10; T = 10; DGP = 1; Het = 2							
1	0.9973	0.3769	0.9930	0.3881	0.3389	90.1	0.9926	0.3887	0.3994	95.1
1	0.9992	0.0220	0.9993	0.0221	0.0208	93.8	0.9993	0.0221	0.0218	94.1
1	0.9714	0.0588	0.9982	0.0603	0.0661	95.7	0.0000	0.0000	0.0000	0.0
0.2	0.1977	0.0787	0.1988	0.0820	0.0687	89.2	0.1989	0.0820	0.0838	94.6
0.3	0.1360	0.1265	0.2742	0.1172	0.1131	93.6	0.2731	0.1172	0.1232	95.9

			N1 = 20; N2 = 20; T = 5; DGP = 1; Het = 2							
1	0.9997	0.1803	1.0011	0.1803	0.1917	96.4	1.0005	0.1804	0.1857	96.2
1	0.9998	0.0155	0.9998	0.0155	0.0134	90.8	0.9998	0.0155	0.0154	94.3
1	0.9767	0.0432	0.9965	0.0441	0.0489	96.2	0.0000	0.0000	0.0000	0.0
0.2	0.1993	0.0385	0.1985	0.0390	0.0385	94.8	0.1988	0.0390	0.0394	95.0
0.3	0.2106	0.0833	0.2905	0.0770	0.0707	93.6	0.2893	0.0773	0.0768	95.0
			N1 = 20; N2 = 20; T = 10; DGP = 1; Het = 2							
1	1.0044	0.2062	1.0054	0.2064	0.1774	90.2	1.0054	0.2064	0.2109	94.5
1	1.0001	0.0101	1.0001	0.0101	0.0094	92.9	1.0001	0.0101	0.0099	94.6
1	0.9886	0.0328	0.9993	0.0332	0.0363	96.3	0.0000	0.0000	0.0000	0.0
0.2	0.2010	0.0212	0.2008	0.0212	0.0197	93.0	0.2008	0.0212	0.0215	95.5
0.3	0.2475	0.0550	0.2957	0.0525	0.0454	90.4	0.2943	0.0524	0.0524	94.6

			N1 = 10; N2 = 10; T = 5; DGP = 3; Het = 1							
1	1.0100	0.4875	1.0099	0.4905	0.4724	93.3	1.0105	0.4905	0.4816	93.8
1	1.0004	0.0293	1.0004	0.0293	0.0301	95.7	1.0004	0.0293	0.0300	95.7
1	0.9419	0.1048	0.9876	0.1098	0.1055	91.0	0.0000	0.0000	0.0000	0.0
0.2	0.1953	0.0981	0.1954	0.0993	0.0943	93.5	0.1952	0.0994	0.0974	93.1
0.3	0.0660	0.1956	0.2536	0.1696	0.1659	94.5	0.2538	0.1701	0.1683	94.1
			N1 = 10; N2 = 10; T = 10; DGP = 3; Het = 1							
1	0.9863	0.2283	0.9856	0.2294	0.2252	93.7	0.9858	0.2294	0.2242	93.0
1	1.0006	0.0204	1.0006	0.0203	0.0198	94.2	1.0006	0.0203	0.0198	94.5
1	0.9668	0.0781	0.9936	0.0801	0.0757	92.0	0.0000	0.0000	0.0000	0.0
0.2	0.2007	0.0602	0.2012	0.0616	0.0615	94.8	0.2012	0.0616	0.0629	95.0
0.3	0.1369	0.1232	0.2748	0.1130	0.1097	95.2	0.2749	0.1131	0.1114	94.6
			N1 = 20; N2 = 20; T = 5; DGP = 3; Het = 1							
1	0.9998	0.2139	1.0021	0.2140	0.2083	94.3	1.0021	0.2138	0.2088	94.5
1	0.9996	0.0131	0.9996	0.0131	0.0130	95.3	0.9996	0.0131	0.0130	95.5
1	0.9787	0.0530	0.9987	0.0541	0.0542	94.6	0.0000	0.0000	0.0000	0.0
0.2	0.1998	0.0331	0.1992	0.0333	0.0320	93.3	0.1992	0.0333	0.0327	93.4
0.3	0.2088	0.0762	0.2887	0.0704	0.0685	94.6	0.2887	0.0703	0.0697	95.5
			N1 = 20; N2 = 20; T = 10; DGP = 3; Het = 1							
1	0.9959	0.1739	0.9977	0.1746	0.1730	94.4	0.9978	0.1745	0.1749	94.3
1	0.9995	0.0098	0.9995	0.0098	0.0094	94.6	0.9995	0.0098	0.0094	94.7
1	0.9885	0.0372	0.9992	0.0376	0.0386	94.5	0.0000	0.0000	0.0000	0.0
0.2	0.2010	0.0303	0.2005	0.0306	0.0301	94.0	0.2005	0.0306	0.0311	94.7
0.3	0.2468	0.0521	0.2955	0.0499	0.0499	94.9	0.2955	0.0499	0.0512	95.7

Note: The first column of each panel are values of β_0, β_1, σ_v^2, λ, and ρ.

Table 2: Monte Carlo Results: Case 1, FE-2
Wt = Mt = Circular World with r = 10

	QMLE		M-Est				RM-Est			
par	Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
			N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 1							
1	1.0158	0.4293	1.0136	0.4334	0.4248	93.9	1.0139	0.4342	0.4319	93.5
1	0.9994	0.0290	0.9993	0.0290	0.0301	95.8	0.9993	0.0290	0.0299	95.3
0.5	0.4933	0.1954	0.4933	0.1954	0.1984	95.4	0.4933	0.1954	0.1961	95.2
0.5	0.5041	0.2106	0.5036	0.2104	0.2028	92.1	0.5036	0.2105	0.2029	91.8
1	0.9435	0.0602	0.9849	0.0624	0.0634	93.8	0.0000	0.0000	0.0000	0.0
0.2	0.1962	0.0961	0.1968	0.0977	0.0966	93.8	0.1967	0.0979	0.0999	93.2
0.3	0.0643	0.1958	0.2256	0.1729	0.1687	94.3	0.2261	0.1733	0.1711	93.8
			N1 = 10; N2 = 10; T = 10; DGP = 1; Het = 1							
1	0.9836	0.3002	0.9802	0.3050	0.3085	95.3	0.9801	0.3051	0.3132	95.0
1	1.0005	0.0188	1.0005	0.0188	0.0199	96.1	1.0005	0.0188	0.0198	96.0
0.5	0.5015	0.1414	0.5015	0.1414	0.1413	95.6	0.5015	0.1414	0.1407	95.2
0.5	0.4982	0.1998	0.4979	0.1994	0.1966	94.2	0.4979	0.1993	0.1962	94.2
1	0.9710	0.0455	0.9959	0.0466	0.0450	93.1	0.0000	0.0000	0.0000	0.0
0.2	0.2009	0.0687	0.2019	0.0707	0.0707	93.2	0.2019	0.0707	0.0728	93.6
0.3	0.1313	0.1239	0.2575	0.1151	0.1150	94.7	0.2576	0.1152	0.1171	94.2
			N1 = 20; N2 = 20; T = 5; DGP = 1; Het = 1							
1	0.9965	0.1919	0.9974	0.1917	0.1937	95.5	0.9974	0.1918	0.1935	95.1
1	0.9994	0.0133	0.9994	0.0133	0.0134	95.3	0.9994	0.0133	0.0133	95.8
0.5	0.5032	0.1352	0.5033	0.1352	0.1393	95.8	0.5033	0.1352	0.1387	95.4
0.5	0.5022	0.1011	0.5024	0.1012	0.1015	94.6	0.5024	0.1012	0.1013	94.8
1	0.9776	0.0321	0.9966	0.0327	0.0318	94.2	0.0000	0.0000	0.0000	0.0
0.2	0.1994	0.0265	0.1991	0.0266	0.0262	94.6	0.1992	0.0266	0.0266	94.6
0.3	0.2114	0.0766	0.2852	0.0710	0.0662	93.3	0.2853	0.0709	0.0670	93.2
			N1 = 20; N2 = 20; T = 10; DGP = 1; Het = 1							
1	1.0122	0.1455	1.0131	0.1458	0.1536	96.2	1.0131	0.1458	0.1540	96.2
1	0.9997	0.0092	0.9996	0.0092	0.0093	95.7	0.9996	0.0092	0.0093	95.7
0.5	0.4999	0.0946	0.4999	0.0946	0.0985	95.9	0.4999	0.0946	0.0984	95.8
0.5	0.4968	0.0994	0.4969	0.0994	0.1015	95.5	0.4969	0.0994	0.1014	95.9
1	0.9878	0.0217	0.9980	0.0220	0.0225	95.9	0.0000	0.0000	0.0000	0.0
0.2	0.1991	0.0237	0.1988	0.0238	0.0237	94.3	0.1988	0.0238	0.0242	95.0
0.3	0.2491	0.0487	0.2949	0.0466	0.0468	94.5	0.2949	0.0466	0.0476	95.1

			N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 2							
1	1.0283	0.4967	1.0243	0.5095	0.4241	88.5	1.0249	0.5100	0.5045	93.0
1	1.0001	0.0301	1.0001	0.0301	0.0301	94.2	1.0001	0.0301	0.0302	94.4
0.5	0.4971	0.2032	0.4971	0.2032	0.1982	93.5	0.4971	0.2032	0.2001	94.3

0.5	0.5057	0.2161	0.5048	0.2168	0.2038	92.1	0.5049	0.2169	0.2107	93.2
1	0.9426	0.0813	0.9840	0.0844	0.0933	94.3	0.0000	0.0000	0.0000	0.0
0.2	0.1918	0.1207	0.1930	0.1252	0.0960	85.5	0.1928	0.1251	0.1270	92.8
0.3	0.0613	0.2151	0.2220	0.1909	0.1676	90.8	0.2188	0.1911	0.1854	92.2
N1 = 10; N2 = 10; T = 10; DGP = 1; Het = 2										
1	0.9892	0.2853	0.9862	0.2864	0.3094	95.9	0.9869	0.2873	0.2921	94.7
1	0.9994	**0.0199**	**0.9994**	0.0199	0.0199	95.7	0.9994	0.0199	0.0205	96.6
0.5	0.5031	0.1461	0.5031	0.1461	0.1413	94.4	0.5031	0.1461	0.1449	94.8
0.5	0.4972	0.2027	0.4970	0.2025	0.1963	94.7	0.4971	0.2026	0.1960	94.9
1	0.9713	0.0537	0.9963	0.0551	0.0582	95.4	0.0000	0.0000	0.0000	0.0
0.2	0.2016	0.0688	0.2025	0.0697	0.0709	95.0	0.2023	0.0698	0.0702	94.6
0.3	0.1300	0.1243	0.2567	0.1148	0.1157	95.2	0.2585	0.1149	0.1175	95.3
N1 = 20; N2 = 20; T = 5; DGP = 1; Het = 2										
1	0.9901	0.1854	0.9907	0.1861	0.1930	95.7	0.9914	0.1862	0.1854	94.5
1	0.9996	0.0142	0.9996	0.0142	0.0134	93.3	0.9996	0.0142	0.0147	95.6
0.5	0.5038	0.1165	0.5038	0.1165	0.1393	97.6	0.5038	0.1165	0.1194	95.1
0.5	0.4962	0.1052	0.4963	0.1053	0.1008	93.8	0.4965	0.1053	0.1012	93.5
1	0.9781	0.0545	0.9971	0.0556	0.0627	96.3	0.0000	0.0000	0.0000	0.0
0.2	0.2015	0.0284	0.2012	0.0286	0.0262	92.6	0.2010	0.0286	0.0275	93.1
0.3	0.2022	0.0954	0.2766	0.0885	0.0668	87.0	0.2797	0.0883	0.0895	94.7
N1 = 20; N2 = 20; T = 10; DGP = 1; Het = 2										
1	1.0108	0.1655	1.0117	0.1657	0.1537	92.2	1.0117	0.1658	0.1710	95.0
1	0.9997	0.0102	0.9997	0.0102	0.0093	93.1	0.9997	0.0102	0.0101	94.7
0.5	0.4996	0.0984	0.4996	0.0984	0.0985	94.2	0.4996	0.0984	0.1020	95.3
0.5	0.4981	0.1033	0.4982	0.1033	0.1015	94.0	0.4982	0.1033	0.1014	94.2
1	0.9876	0.0302	0.9978	0.0305	0.0343	97.0	0.0000	0.0000	0.0000	0.0
0.2	0.1992	0.0237	0.1990	0.0238	0.0237	95.5	0.1990	0.0238	0.0238	95.6
0.3	0.2495	0.0503	0.2953	0.0482	0.0468	94.4	0.2943	0.0481	0.0507	95.8

N1 = 10; N2 = 10; T = 5; DGP = 3; Het = 1										
1	1.0013	0.4537	0.9990	0.4545	0.4246	92.5	1.0000	0.4543	0.4295	91.8
1	1.0005	0.0293	1.0005	0.0293	0.0300	95.2	1.0005	0.0293	0.0298	95.3
0.5	0.5078	0.1906	0.5078	0.1906	0.1980	95.5	0.5078	0.1906	0.1953	95.9
0.5	0.4997	0.2083	0.4994	0.2089	0.2028	94.0	0.4996	0.2090	0.2029	94.2
1	0.9418	0.1048	0.9831	0.1093	0.1047	90.4	0.0000	0.0000	0.0000	0.0
0.2	0.1956	0.1008	0.1963	0.1021	0.0966	92.9	0.1960	0.1022	0.0993	92.6
0.3	0.0655	0.1985	0.2260	0.1748	0.1682	93.9	0.2266	0.1756	0.1694	92.3
N1 = 10; N2 = 10; T = 10; DGP = 3; Het = 1										
1	0.9886	0.3158	0.9868	0.3196	0.3092	93.9	0.9870	0.3199	0.3122	94.0
1	1.0005	0.0201	1.0004	0.0201	0.0198	94.6	1.0004	0.0201	0.0198	94.8
0.5	0.4984	0.1389	0.4984	0.1389	0.1409	95.2	0.4984	0.1389	0.1396	95.1

0.5	0.5088	0.2061	0.5087	0.2063	0.1978	93.9	0.5087	0.2063	0.1962	93.8
1	0.9662	0.0762	0.9909	0.0781	0.0751	93.0	0.0000	0.0000	0.0000	0.0
0.2	0.1976	0.0707	0.1981	0.0725	0.0709	94.6	0.1981	0.0726	0.0728	94.1
0.3	0.1368	0.1260	0.2628	0.1166	0.1147	94.8	0.2632	0.1168	0.1167	94.3
			N1 = 20; N2 = 20; T = 5; DGP = 3; Het = 1							
1	1.0013	0.1994	1.0025	0.1997	0.1939	94.8	1.0026	0.1997	0.1934	94.5
1	1.0004	0.0130	1.0004	0.0130	0.0134	96.1	1.0004	0.0130	0.0133	96.4
0.5	0.5041	0.1423	0.5041	0.1423	0.1393	95.1	0.5041	0.1423	0.1390	95.6
0.5	0.4979	0.1028	0.4981	0.1030	0.1014	93.9	0.4981	0.1030	0.1013	94.1
1	0.9784	0.0523	0.9974	0.0533	0.0542	95.2	0.0000	0.0000	0.0000	0.0
0.2	0.1988	0.0270	0.1986	0.0271	0.0262	94.2	0.1986	0.0271	0.0266	94.2
0.3	0.2115	0.0740	0.2854	0.0684	0.0662	94.0	0.2855	0.0684	0.0669	94.4
			N1 = 20; N2 = 20; T = 10; DGP = 3; Het = 1							
1	1.0060	0.1554	1.0069	0.1558	0.1536	95.2	1.0069	0.1558	0.1540	95.1
1	1.0002	0.0093	1.0002	0.0093	0.0093	94.0	1.0002	0.0093	0.0093	94.1
0.5	0.4977	0.0998	0.4977	0.0998	0.0985	95.2	0.4977	0.0998	0.0982	94.8
0.5	0.4926	0.1033	0.4927	0.1034	0.1015	94.3	0.4927	0.1034	0.1016	94.1
1	0.9877	0.0377	0.9980	0.0381	0.0385	94.9	0.0000	0.0000	0.0000	0.0
0.2	0.1999	0.0245	0.1996	0.0246	0.0237	94.5	0.1996	0.0246	0.0242	94.8
0.3	0.2491	0.0488	0.2950	0.0468	0.0469	95.4	0.2950	0.0468	0.0476	96.0

Note: The first column of each panel are values of β_0, β_1, β_2, β_3, σ_v^2, λ, and ρ.

Table 3: Monte Carlo Results: Case 1, FE-3
Wt = Mt = Queen Contiguity

	QMLE		M-Est				RM-Est			
par	Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
			N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 1							
5	5.0233	0.8390	5.0530	0.8583	0.7863	92.9	5.0535	0.8598	0.8269	93.1
1	0.9985	0.0466	0.9982	0.0466	0.0463	95.7	0.9982	0.0466	0.0459	95.0
1	0.7100	0.0523	0.9842	0.0726	0.0735	93.2	0.0000	0.0000	0.0000	0.0
0.2	0.1980	0.1022	0.1933	0.1061	0.0900	89.6	0.1933	0.1066	0.1045	92.3
0.3	0.2637	0.1451	0.2904	0.1364	0.1149	89.2	0.2905	0.1376	0.1303	91.7
			N1 = 10; N2 = 10; T = 10; DGP = 1; Het = 1							
5	4.9399	0.6677	4.9660	0.6839	0.6293	92.9	4.9676	0.6852	0.6673	93.4
1	1.0006	0.0290	1.0004	0.0290	0.0293	95.3	1.0004	0.0290	0.0292	95.5
1	0.8091	0.0410	0.9972	0.0505	0.0496	94.8	0.0000	0.0000	0.0000	0.0
0.2	0.2058	0.0669	0.2022	0.0698	0.0600	89.7	0.2020	0.0700	0.0694	92.8
0.3	0.2645	0.0961	0.2870	0.0920	0.0768	89.3	0.2874	0.0920	0.0873	91.8
			N1 = 20; N2 = 20; T = 5; DGP = 1; Het = 1							
5	4.9967	0.5900	5.0018	0.5926	0.5618	93.4	5.0023	0.5929	0.5270	89.5
1	0.9987	0.0214	0.9987	0.0214	0.0213	94.9	0.9987	0.0214	0.0213	95.2
1	0.7565	0.0286	0.9956	0.0376	0.0363	94.2	0.0000	0.0000	0.0000	0.0
0.2	0.1977	0.0530	0.1968	0.0540	0.0477	91.4	0.1967	0.0541	0.0530	94.2
0.3	0.3019	0.0684	0.3004	0.0673	0.0596	90.9	0.3006	0.0674	0.0655	94.2
			N1 = 20; N2 = 20; T = 10; DGP = 1; Het = 1							
5	5.0079	0.4587	5.0141	0.4610	0.4544	93.9	5.0143	0.4611	0.4559	93.9
1	1.0006	0.0145	1.0005	0.0145	0.0144	96.8	1.0005	0.0145	0.0144	96.6
1	0.8534	0.0211	0.9982	0.0247	0.0243	95.2	0.0000	0.0000	0.0000	0.0
0.2	0.1991	0.0346	0.1982	0.0351	0.0327	93.0	0.1982	0.0352	0.0365	95.8
0.3	0.2985	0.0440	0.2988	0.0432	0.0404	93.6	0.2989	0.0432	0.0448	95.8
			N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 2							
5	5.0410	0.8331	5.0717	0.8532	0.7896	92.8	5.0707	0.8519	0.8251	92.8
1	1.0004	0.0454	1.0002	0.0454	0.0464	95.4	1.0002	0.0454	0.0459	95.1
1	0.7132	0.0542	0.9888	0.0751	0.0758	93.7	0.0000	0.0000	0.0000	0.0
0.2	0.1969	0.1014	0.1921	0.1054	0.0905	89.7	0.1922	0.1054	0.1054	92.1
0.3	0.2608	0.1482	0.2878	0.1391	0.1154	89.6	0.2870	0.1391	0.1312	91.8
			N1 = 10; N2 = 10; T = 10; DGP = 1; Het = 2							
5	4.9678	0.6929	4.9960	0.7110	0.6318	91.7	4.9935	0.7100	0.6711	92.9
1	0.9999	0.0281	0.9996	0.0281	0.0293	95.4	0.9997	0.0281	0.0293	96.0
1	0.8072	0.0410	0.9948	0.0506	0.0508	94.6	0.0000	0.0000	0.0000	0.0
0.2	0.2019	0.0661	0.1980	0.0693	0.0604	91.1	0.1983	0.0694	0.0705	94.5
0.3	0.2711	0.0947	0.2934	0.0909	0.0768	89.3	0.2925	0.0912	0.0882	93.6

			N1 = 20; N2 = 20; T = 5 DGP = 1; Het = 2							
5	5.0038	0.5429	5.0084	0.5455	0.5618	94.6	5.0095	0.5456	0.4994	90.7
1	0.9999	0.0215	0.9999	0.0215	0.0213	94.4	0.9999	0.0215	0.0214	94.2
1	0.7582	0.0278	0.9978	0.0365	0.0372	95.5	0.0000	0.0000	0.0000	0.0
0.2	0.2014	0.0510	0.2006	0.0519	0.0475	92.2	0.2004	0.0520	0.0525	94.1
0.3	0.2972	0.0669	0.2958	0.0658	0.0596	91.6	0.2963	0.0659	0.0653	94.3
			N1 = 20; N2 = 20; T = 10 DGP = 1; Het = 2							
5	4.9808	0.4710	4.9888	0.4745	0.4853	95.4	4.9893	0.4746	0.4956	95.6
1	1.0003	0.0144	1.0002	0.0144	0.0141	94.9	1.0002	0.0144	0.0141	95.0
1	0.8528	0.0211	0.9978	0.0246	0.0249	95.1	0.0000	0.0000	0.0000	0.0
0.2	0.1997	0.0333	0.1988	0.0340	0.0319	93.3	0.1988	0.0340	0.0354	95.5
0.3	0.3040	0.0434	0.3010	0.0430	0.0397	92.7	0.3010	0.0430	0.0437	95.3

			N1 = 10; N2 = 10; T = 5; DGP = 3; Het = 1							
5	4.9774	0.8324	5.0047	0.8513	0.7812	92.9	5.0084	0.8527	0.8092	92.1
1	0.9988	0.0463	0.9986	0.0464	0.0463	95.0	0.9986	0.0464	0.0457	94.6
1	0.7115	0.0825	0.9865	0.1143	0.1111	91.4	0.0000	0.0000	0.0000	0.0
0.2	0.2011	0.1003	0.1967	0.1044	0.0891	89.6	0.1961	0.1049	0.1028	91.3
0.3	0.2551	0.1491	0.2823	0.1399	0.1147	88.9	0.2832	0.1411	0.1288	91.2
			N1 = 10; N2 = 10; T = 10; DGP = 3; Het = 1							
5	4.9539	0.6647	4.9825	0.6837	0.6316	92.6	4.9816	0.6837	0.6650	93.0
1	1.0000	0.0297	0.9997	0.0298	0.0292	95.2	0.9997	0.0298	0.0290	94.6
1	0.8054	0.0645	0.9925	0.0795	0.0779	93.8	0.0000	0.0000	0.0000	0.0
0.2	0.2035	0.0684	0.1995	0.0716	0.0605	90.4	0.1997	0.0715	0.0701	92.1
0.3	0.2733	0.0944	0.2954	0.0907	0.0770	90.0	0.2952	0.0909	0.0875	93.4
			N1 = 20; N2 = 20; T = 5; DGP = 3; Het = 1							
5	5.0209	0.5597	5.0258	0.5621	0.5621	94.6	5.0264	0.5618	0.5231	90.9
1	0.9990	0.0222	0.9989	0.0222	0.0213	94.3	0.9989	0.0222	0.0213	94.2
1	0.7568	0.0434	0.9960	0.0571	0.0567	94.3	0.0000	0.0000	0.0000	0.0
0.2	0.1983	0.0524	0.1974	0.0533	0.0477	92.1	0.1973	0.0534	0.0529	94.6
0.3	0.2999	0.0657	0.2984	0.0647	0.0596	93.3	0.2986	0.0646	0.0654	95.5
			N1 = 20; N2 = 20; T = 10; DGP = 3; Het = 1							
5	4.9945	0.4915	5.0022	0.4952	0.4863	95.1	5.0022	0.4952	0.4859	94.1
1	0.9994	0.0146	0.9994	0.0146	0.0141	94.1	0.9994	0.0146	0.0141	93.8
1	0.8530	0.0328	0.9980	0.0384	0.0397	95.2	0.0000	0.0000	0.0000	0.0
0.2	0.1994	0.0339	0.1986	0.0346	0.0320	93.2	0.1986	0.0346	0.0354	95.7
0.3	0.3023	0.0433	0.2993	0.0429	0.0400	92.8	0.2993	0.0429	0.0439	95.5

Note: The first column of each panel are values of β_0, β_1, σ_v^2, λ, and ρ.

Table 4: Monte Carlo Results: Case 2, FE-1
Wt = Mt = Rook Contiguity

	QMLE		M-Est				RM-Est			
par	Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
				N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 1						
0.7	0.7024	0.0840	0.6982	0.0912	0.1004	90.3	0.6975	0.0968	0.1119	92.6
0.5	0.5049	0.0861	0.5010	0.0888	0.0854	91.9	0.5002	0.0883	0.0962	94.5
0.4	0.3975	0.0898	0.3956	0.0924	0.0994	92.9	0.3956	0.0941	0.1096	95.9
0.3	0.3027	0.0802	0.3015	0.0815	0.0859	95.6	0.3017	0.0821	0.0889	95.3
0.2	0.2078	0.0980	0.2063	0.0991	0.0967	90.9	0.2056	0.0986	0.1005	92.6
0.8	0.7558	0.1046	0.7665	0.1108	0.1090	92.5	0.7684	0.1125	0.1208	91.9
0.7	0.6640	0.1167	0.6761	0.1185	0.1062	92.0	0.6807	0.1185	0.1195	92.8
0.5	0.4669	0.1437	0.4792	0.1445	0.1427	93.3	0.4851	0.1470	0.1539	94.1
0.4	0.3702	0.1461	0.3833	0.1452	0.1439	92.9	0.3875	0.1479	0.1510	92.9
0.4	0.3566	0.1534	0.3585	0.1496	0.1499	93.6	0.3631	0.1505	0.1570	92.0
1.0	0.9757	0.3633	0.9811	0.3656	0.3547	92.7	0.9828	0.3649	0.3717	93.8
1.0	1.0004	0.0275	0.9997	0.0276	0.0272	94.4	0.9996	0.0275	0.0283	94.9
1.0	0.9364	0.0607	0.9753	0.0633	0.0651	92.9	0.0000	0.0000	0.0000	0.0
				N1 = 10; N2 = 10; T = 5; DGP = 2; Het = 2						
0.7	0.704	0.0930	0.6998	0.1001	0.0985	86.8	0.6993	0.1008	0.1117	92.1
0.5	0.4997	0.0883	0.4958	0.0910	0.0851	90.0	0.4959	0.0916	0.0974	93.7
0.4	0.4023	0.1001	0.4007	0.1029	0.0993	89.6	0.4019	0.1030	0.1137	94.1
0.3	0.2999	0.0952	0.2987	0.0967	0.0852	90.9	0.2990	0.0942	0.0937	94.3
0.2	0.2044	0.0976	0.2029	0.0987	0.0978	93.0	0.2023	0.0970	0.0981	93.5
0.8	0.7504	0.1159	0.7605	0.1227	0.1114	89.5	0.7642	0.1178	0.1243	91.2
0.7	0.6623	0.1182	0.6745	0.1199	0.1104	92.6	0.6798	0.1182	0.1203	93.7
0.5	0.4614	0.1496	0.4734	0.1507	0.1467	92.8	0.4823	0.1505	0.1532	92.7
0.4	0.3675	0.1598	0.3806	0.1591	0.1460	91.8	0.3838	0.1566	0.1538	92.5
0.4	0.3568	0.1589	0.3588	0.1550	0.1534	92.7	0.3667	0.1514	0.1525	92.7
1.0	0.9855	0.3465	0.9910	0.3489	0.3566	94.2	0.9933	0.3499	0.3695	94.2
1.0	0.9982	0.0411	0.9975	0.0411	0.0271	94.7	0.9977	0.0411	0.0282	95.3
1.0	0.9312	0.1018	0.9699	0.1061	0.1001	88.7	0.0000	0.0000	0.0000	0.0
				N1 = 20; N2 = 20; T = 5; DGP = 1; Het = 1						
0.7	0.7008	0.0473	0.6988	0.0489	0.0490	93.3	0.6997	0.0524	0.0575	95.7
0.5	0.5038	0.0450	0.5022	0.0456	0.0460	94.5	0.5026	0.0469	0.0503	95.5
0.4	0.3997	0.0427	0.3992	0.0431	0.0439	94.1	0.3995	0.0432	0.0465	95.8
0.3	0.3005	0.0438	0.3001	0.0440	0.0429	93.5	0.2997	0.0443	0.0461	95.1
0.2	0.2016	0.0444	0.2011	0.0446	0.0447	93.6	0.2007	0.0449	0.0445	94.5
0.8	0.7915	0.0485	0.7927	0.0501	0.0494	93.1	0.7920	0.0536	0.0571	94.0
0.7	0.6924	0.0507	0.6928	0.0514	0.0524	94.7	0.6924	0.0554	0.0576	95.5
0.5	0.4962	0.0668	0.4957	0.0669	0.0666	94.2	0.4963	0.0687	0.0702	95.0

0.4	0.3905	0.0721	0.3903	0.0718	0.0712	94.6	0.3921	0.0727	0.0755	95.0
0.4	0.3948	0.0708	0.3933	0.0701	0.0708	94.7	0.3954	0.0727	0.0717	94.1
1.0	0.9969	0.2023	0.9983	0.2026	0.2044	94.5	0.9994	0.2030	0.2041	94.3
1.0	0.9997	0.0136	0.9994	0.0136	0.0137	96.1	0.9995	0.0137	0.0138	95.7
1.0	0.9735	0.0323	0.9946	0.0330	0.0330	95.0	0.0000	0.0000	0.0000	0.0
			N1 = 20; N2 = 20; T = 5; DGP = 2; Het = 2							
0.7	0.7001	0.0465	0.6980	0.0481	0.0489	92.2	0.6994	0.0519	0.0571	94.8
0.5	0.5039	0.0438	0.5023	0.0444	0.0463	94.8	0.5034	0.0467	0.0513	95.9
0.4	0.4007	0.0444	0.4002	0.0448	0.0439	93.6	0.4005	0.0446	0.0468	95.4
0.3	0.3009	0.0431	0.3005	0.0434	0.0431	93.9	0.3006	0.0436	0.0459	96.3
0.2	0.2024	0.0428	0.2019	0.0430	0.0447	95.7	0.2019	0.0430	0.0439	95.8
0.8	0.7911	0.0517	0.7924	0.0533	0.0513	91.1	0.7922	0.0544	0.0561	93.4
0.7	0.6898	0.0538	0.6901	0.0546	0.0558	95.1	0.6909	0.0557	0.0579	96.2
0.5	0.4956	0.0707	0.4951	0.0708	0.0691	93.3	0.4971	0.0701	0.0697	94.2
0.4	0.3902	0.0732	0.3899	0.0729	0.0732	94.4	0.3917	0.0721	0.0743	94.9
0.4	0.3899	0.0699	0.3884	0.0693	0.0733	95.4	0.3908	0.0694	0.0715	94.7
1.0	0.9947	0.1926	0.9961	0.1928	0.2035	96.2	0.9964	0.1944	0.1943	95.1
1.0	0.9992	0.0146	0.9990	0.0147	0.0137	93.4	0.9991	0.0147	0.0144	95.1
1.0	0.9706	0.0522	0.9917	0.0533	0.0556	93.4	0.0000	0.0000	0.0000	0.0

Note: The first column of each panel are values of λ_1 ... λ_6, ρ_1 ... ρ_6, β_0, β_1, σ_v^2.

Table 5: Monte Carlo Results: Case 2, FE-2
Wt = Mt = Rook Contiguity

	QMLE		M-Est				RM-Est			
par	Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
			N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 1							
0.7	0.6986	0.0834	0.6946	0.0893	0.0915	90.4	0.6959	0.0944	0.1323	95.0
0.5	0.5003	0.0902	0.4986	0.0913	0.0992	93.4	0.4977	0.0924	0.1195	97.5
0.4	0.3946	0.0977	0.3928	0.1003	0.1022	91.4	0.3942	0.1024	0.1355	96.4
0.3	0.2986	0.0770	0.2977	0.0781	0.0766	92.3	0.2976	0.0781	0.1013	97.5
0.2	0.2041	0.0971	0.2019	0.0989	0.1004	93.1	0.2021	0.0993	0.1315	97.2
0.2	0.1949	0.0988	0.1946	0.0995	0.0986	91.1	0.1945	0.0994	0.1075	93.5
0.8	0.7572	0.1017	0.7706	0.1053	0.1012	89.2	0.7734	0.1081	0.1327	93.3
0.7	0.6686	0.1164	0.6637	0.1163	0.1132	92.4	0.6664	0.1174	0.1341	95.6
0.5	0.4686	0.1402	0.4851	0.1406	0.1437	93.4	0.4882	0.1442	0.1801	95.1
0.4	0.3638	0.1464	0.3816	0.1452	0.1387	91.9	0.3877	0.1480	0.1757	95.0
0.4	0.3651	0.1540	0.3837	0.1531	0.1488	91.5	0.3876	0.1561	0.1875	94.2
0.3	0.2683	0.1675	0.2744	0.1626	0.1570	91.7	0.2774	0.1640	0.1605	92.2
1.0	0.9849	0.5414	0.9925	0.5454	0.5750	93.0	0.9962	0.5510	0.6531	96.1
1.0	0.9974	0.0254	0.9969	0.0254	0.0254	94.7	0.9968	0.0255	0.0264	95.3
0.5	0.5104	0.1690	0.5114	0.1689	0.1727	95.0	0.5115	0.1692	0.1752	95.2
0.5	0.5272	0.6381	0.5215	0.6431	0.6647	93.4	0.5177	0.6486	0.6771	93.8
1.0	0.9409	0.0574	0.9717	0.0594	0.0588	90.6	0.0000	0.0000	0.0000	0.0
			N1 = 10; N2 = 10; T = 6; DGP = 2; Het = 2							
0.7	0.6957	0.0800	0.6915	0.0858	0.0894	91.5	0.6940	0.0899	0.1427	95.7
0.5	0.5057	0.0971	0.5041	0.0983	0.0986	91.5	0.5060	0.1016	0.1308	97.7
0.4	0.3966	0.0948	0.3949	0.0970	0.1017	92.4	0.3963	0.0973	0.1344	97.7
0.3	0.2981	0.0799	0.2969	0.0807	0.0765	90.4	0.2974	0.0776	0.1039	98.2
0.2	0.2018	0.0969	0.1996	0.0987	0.1007	93.5	0.2016	0.0966	0.1330	97.3
0.2	0.2012	0.0983	0.2010	0.0990	0.0989	91.4	0.2026	0.0963	0.1109	95.4
0.8	0.7566	0.1043	0.7702	0.1077	0.1021	90.1	0.7736	0.1082	0.1415	93.1
0.7	0.6598	0.1288	0.6550	0.1287	0.1192	92.0	0.6567	0.1273	0.1475	95.8
0.5	0.4624	0.1488	0.4788	0.1489	0.1473	92.3	0.4853	0.1484	0.1791	94.9
0.4	0.3588	0.1507	0.3769	0.1495	0.1418	92.3	0.3813	0.1446	0.1804	95.6
0.4	0.3627	0.1610	0.3814	0.1600	0.1524	91.9	0.3852	0.1558	0.1909	94.5
0.3	0.2566	0.1677	0.2630	0.1629	0.1606	91.8	0.2625	0.1613	0.1648	91.5
1.0	0.9839	0.5723	0.9913	0.5762	0.5725	92.2	0.9890	0.6069	0.7247	95.9
1.0	0.9977	0.0256	0.9972	0.0257	0.0253	93.9	0.9976	0.0257	0.0269	95.4
0.5	0.5089	0.1698	0.5098	0.1699	0.1727	95.4	0.5086	0.1699	0.1737	95.6
0.5	0.5025	0.6268	0.4970	0.6316	0.6620	96.0	0.4924	0.6652	0.7469	95.5
1.0	0.9416	0.0862	0.9725	0.0891	0.0927	90.4	0.0000	0.0000	0.0000	0.0

			N1 = 20; N2 = 10; T = 6; DGP = 1; Het = 1							
0.7	0.7017	0.0575	0.7007	0.0583	0.0640	94.4	0.7020	0.0628	0.0856	98.0
0.5	0.5027	0.0651	0.5003	0.0664	0.0662	93.1	0.5003	0.0683	0.0827	97.4
0.4	0.3997	0.0655	0.3989	0.0664	0.0648	92.2	0.3994	0.0679	0.0803	97.4
0.3	0.2988	0.0592	0.2982	0.0596	0.0589	92.8	0.2983	0.0595	0.0698	96.3
0.2	0.2020	0.0581	0.2012	0.0585	0.0574	94.2	0.2009	0.0586	0.0682	97.5
0.2	0.2000	0.0647	0.1996	0.0649	0.0654	94.7	0.1995	0.0648	0.0708	96.3
0.8	0.7812	0.0689	0.7771	0.0702	0.0702	94.3	0.7760	0.0750	0.0925	97.3
0.7	0.6802	0.0791	0.6854	0.0800	0.0760	94.1	0.6862	0.0848	0.0924	95.3
0.5	0.4845	0.1051	0.4896	0.1050	0.0963	92.0	0.4912	0.1071	0.1154	94.7
0.4	0.3867	0.1002	0.3926	0.0996	0.0996	94.5	0.3949	0.1008	0.1170	96.5
0.4	0.3822	0.1001	0.3882	0.0993	0.0977	94.6	0.3907	0.1014	0.1151	96.6
0.3	0.2890	0.1072	0.2908	0.1050	0.1086	94.9	0.2939	0.1065	0.1120	94.8
1.0	0.9987	0.4812	1.0032	0.4843	0.4995	95.1	0.9988	0.5008	0.5807	96.2
1.0	0.9995	0.0155	0.9993	0.0155	0.0155	94.6	0.9994	0.0155	0.0158	95.2
0.5	0.4914	0.1750	0.4914	0.1750	0.1749	94.5	0.4917	0.1750	0.1746	94.6
0.5	0.5035	0.4930	0.4996	0.4961	0.5097	94.8	0.5047	0.5129	0.5528	95.4
1.0	0.9616	0.0403	0.9866	0.0414	0.0422	93.6	0.0000	0.0000	0.0000	0.0
			N1 = 20; N2 = 10; T = 6; DGP = 2; Het = 2							
0.7	0.7022	0.0614	0.7013	0.0623	0.0638	93.3	0.7024	0.0651	0.0862	97.2
0.5	0.5047	0.0651	0.5023	0.0664	0.0666	93.1	0.5025	0.0687	0.0838	97.2
0.4	0.4016	0.0666	0.4008	0.0674	0.0650	92.0	0.4010	0.0680	0.0817	96.9
0.3	0.2994	0.0601	0.2988	0.0605	0.0590	92.7	0.2988	0.0592	0.0696	97.0
0.2	0.2023	0.0569	0.2015	0.0572	0.0579	95.0	0.2016	0.0569	0.0664	97.9
0.2	0.1996	0.0635	0.1992	0.0638	0.0655	95.0	0.1989	0.0627	0.0688	96.4
0.8	0.7773	0.0730	0.7731	0.0745	0.0737	94.5	0.7745	0.0758	0.0936	97.1
0.7	0.6776	0.0825	0.6828	0.0835	0.0808	93.7	0.6857	0.0847	0.0930	97.0
0.5	0.4808	0.1001	0.4858	0.1001	0.1004	94.2	0.4914	0.1012	0.1167	96.7
0.4	0.3847	0.1055	0.3906	0.1047	0.1025	94.3	0.3951	0.1028	0.1176	96.0
0.4	0.3766	0.0992	0.3827	0.0984	0.1014	94.9	0.3875	0.0982	0.1137	96.6
0.3	0.2851	0.1121	0.2870	0.1098	0.1108	94.7	0.2922	0.1091	0.1111	94.2
1.0	0.9736	0.4666	0.9778	0.4696	0.4931	94.7	0.9858	0.5095	0.5903	95.9
1.0	0.9989	0.0170	0.9987	0.0170	0.0155	92.5	0.9988	0.0171	0.0171	94.6
0.5	0.4968	0.1641	0.4968	0.1641	0.1748	96.3	0.4965	0.1641	0.1608	94.9
0.5	0.5336	0.4895	0.5301	0.4926	0.5032	93.9	0.5236	0.5348	0.5738	94.2
1.0	0.9600	0.0667	0.9850	0.0684	0.0712	93.3	0.0000	0.0000	0.0000	0.0

Note: The first column of each panel are values of λ_1 ... λ_6, ρ_1 ... ρ_6, β_0, β_1, β_2, β_3, σ_v^2.

Table 6: Monte Carlo Results: Case 2, FE-3
Wt = Mt = Rook Contiguity

	QMLE		M-Est				RM-Est			
par	Mean	sd	Mean	sd	se	CP95	Mean	sd	se	CP95r
			N1 = 10; N2 = 10; T = 5; DGP = 1; Het = 1							
0.7	0.7093	0.0724	0.6974	0.0826	0.0805	91.4	0.6988	0.0884	0.0949	93.8
0.5	0.5082	0.0759	0.4986	0.0829	0.0813	92.7	0.4992	0.0842	0.0893	94.3
0.4	0.3991	0.0844	0.3976	0.0879	0.0850	92.6	0.3997	0.0893	0.0927	93.2
0.3	0.2976	0.0671	0.2980	0.0679	0.0658	93.9	0.2986	0.0678	0.0691	94.4
0.2	0.2009	0.1004	0.1994	0.1021	0.0948	90.5	0.1997	0.1026	0.1002	92.7
0.8	0.8194	0.0858	0.7679	0.1110	0.1007	93.6	0.7656	0.1160	0.1169	93.6
0.7	0.7412	0.1170	0.6717	0.1331	0.1141	93.1	0.6760	0.1377	0.1305	93.5
0.5	0.5454	0.1871	0.4597	0.1725	0.1601	92.6	0.4657	0.1811	0.1776	91.4
0.4	0.4601	0.2040	0.3708	0.1738	0.1630	91.8	0.3775	0.1788	0.1808	91.8
0.4	0.4508	0.2102	0.3582	0.1808	0.1759	91.6	0.3679	0.1864	0.1900	92.1
1.0	1.0070	0.5988	1.0145	0.5978	0.5918	95.0	1.0123	0.5966	0.5731	93.2
1.0	0.9972	0.0495	0.9966	0.0494	0.0381	94.6	0.9967	0.0495	0.0388	94.8
1.0	0.6654	0.0568	0.9626	0.0812	0.0762	89.8	0.0000	0.0000	0.0000	0.0
			N1 = 10; N2 = 10; T = 5; DGP = 2; Het = 2							
0.7	0.7112	0.0696	0.7007	0.0825	0.0792	90.0	0.7006	0.0854	0.0938	94.4
0.5	0.5029	0.0837	0.4940	0.0918	0.0808	89.3	0.4954	0.0936	0.0944	93.1
0.4	0.4016	0.0844	0.3993	0.0885	0.0859	91.4	0.4032	0.0917	0.0979	94.4
0.3	0.2963	0.0708	0.2964	0.0719	0.0662	91.4	0.2980	0.0711	0.0700	94.0
0.2	0.2014	0.0978	0.1999	0.1004	0.0960	92.9	0.2012	0.1003	0.0991	92.8
0.8	0.8160	0.0869	0.7616	0.1185	0.1072	93.2	0.7622	0.1200	0.1180	93.9
0.7	0.7366	0.1115	0.6648	0.1332	0.1233	92.1	0.6751	0.1324	0.1296	92.4
0.5	0.5546	0.1714	0.4683	0.1648	0.1669	94.1	0.4776	0.1678	0.1775	92.6
0.4	0.4683	0.2043	0.3782	0.1776	0.1683	92.4	0.3816	0.1788	0.1789	91.1
0.4	0.4479	0.2199	0.3561	0.1883	0.1831	90.9	0.3654	0.1926	0.1869	91.7
1.0	0.9895	0.5054	0.9965	0.5023	0.5943	98.5	0.9917	0.5027	0.5099	93.5
1.0	0.9997	0.0377	0.9991	0.0377	0.0382	95.1	0.9992	0.0377	0.0382	94.8
1.0	0.6678	0.0769	0.9660	0.1101	0.1123	90.5	0.0000	0.0000	0.0000	0.0
			N1 = 20; N2 = 20; T = 5; DGP = 1; Het = 1							
0.7	0.7080	0.0379	0.6973	0.0444	0.0446	93.8	0.6976	0.0479	0.0486	92.4
0.5	0.5088	0.0417	0.5016	0.0448	0.0443	94.0	0.5025	0.0464	0.0462	94.3
0.4	0.3967	0.0355	0.3980	0.0368	0.0382	95.3	0.3984	0.0369	0.0393	94.8
0.3	0.2978	0.0378	0.3011	0.0379	0.0377	94.4	0.3009	0.0382	0.0387	94.4
0.2	0.1973	0.0463	0.2007	0.0478	0.0466	94.6	0.2005	0.0486	0.0480	94.4
0.8	0.8424	0.0334	0.7942	0.0476	0.0481	93.7	0.7939	0.0511	0.0519	93.2
0.7	0.7621	0.0423	0.6913	0.0548	0.0553	95.2	0.6909	0.0602	0.0600	95.1
0.5	0.5963	0.0668	0.4957	0.0727	0.0729	94.7	0.4972	0.0761	0.0779	95.1

0.4	0.4908	0.0875	0.3859	0.0838	0.0814	94.2	0.3887	0.0872	0.0859	93.8
0.4	0.4941	0.0896	0.3902	0.0871	0.0847	94.0	0.3927	0.0913	0.0898	94.0
1.0	1.0036	0.5057	0.9988	0.5055	0.5009	95.0	0.9982	0.5062	0.4573	90.9
1.0	1.0000	0.0189	0.9994	0.0189	0.0190	95.0	0.9994	0.0189	0.0191	95.2
1.0	0.7220	0.0287	0.9938	0.0388	0.0386	93.9	0.0000	0.0000	0.0000	0.0
			N1 = 20; N2 = 20; T = 5; DGP = 2; Het = 2							
0.7	0.7088	0.0372	0.6983	0.0436	0.0444	93.4	0.6991	0.0462	0.0477	94.0
0.5	0.5076	0.0409	0.5003	0.0445	0.0444	95.1	0.5017	0.0457	0.0475	95.7
0.4	0.3981	0.0365	0.3996	0.0382	0.0382	94.1	0.4003	0.0378	0.0390	94.8
0.3	0.2974	0.0371	0.3007	0.0372	0.0377	95.3	0.3011	0.0373	0.0382	95.1
0.2	0.1980	0.0441	0.2017	0.0455	0.0467	95.0	0.2014	0.0455	0.0468	95.3
0.8	0.8401	0.0379	0.7912	0.0535	0.0518	93.1	0.7919	0.0523	0.0515	92.8
0.7	0.7607	0.0467	0.6893	0.0598	0.0607	94.8	0.6910	0.0594	0.0597	94.0
0.5	0.5945	0.0724	0.4937	0.0788	0.0785	93.6	0.4950	0.0786	0.0773	94.0
0.4	0.4916	0.0895	0.3867	0.0857	0.0858	95.0	0.3891	0.0855	0.0842	94.3
0.4	0.4885	0.0911	0.3849	0.0879	0.0900	95.5	0.3895	0.0893	0.0893	94.9
1.0	1.0165	0.3772	1.0097	0.3706	0.4998	99.6	1.0092	0.3700	0.3368	91.1
1.0	1.0000	0.0190	0.9995	0.0190	0.0190	94.2	0.9996	0.0190	0.0185	93.6
1.0	0.7201	0.0430	0.9910	0.0584	0.0625	95.9	0.0000	0.0000	0.0000	0.0

Note: The first column of each panel are values of λ_1 ... λ_6, ρ_1 ... ρ_6, β_0, β_1, σ_v^2.

Appendix D: A Practical Guide

We have introduced a comprehensive set of econometric methods for the estimation and inference of multidimensional spatial panel data models with multidimensional fixed effects. To help applied researchers apply these methods to solve their own economic problems, we first make available online at http://www.mysmu.edu/faculty/zlyang/SubPages/research.htm the full set of Matlab codes in two zip-folders `MatlabCode2025TI` and `MatlabCode2025TV`, which produce all the Monte Carlo results. The former contains the codes for the case of time-invariant parameters and the latter the codes for the case of time-varying parameters. A brief description of the main matlab programs and sub-functions is given below.

1. **The Case of Time-Invariant Spatial Parameters**
   - `mD_SPD_FETI_A.m`: main program for `FE-1`
   - `mD_SPD_FETI_A.m`: main program for `FE-2`
   - `mD_SPD_FETI_A.m`: main program for `FE-3`
   - `FnQMLC.m`: sub-function for computing the QML function
   - `FnAQSC.m`: sub-function for computing the AQS function
   - `FnRAQSC.m`: sub-function for computing the robust AQS function
   - `queen.m`: sub-function for generating `queen` spatial weight matrix
   - `rook.m`: sub-function for generating `rook` spatial weight matrix
2. **The Case of Time-Variant Spatial Parameters**
   - `mD_SPD_FETV_A.m`: main program `for FE-1`
   - `mD_SPD_FETV_A.m`: main program `for FE-2`
   - `mD_SPD_FETV_A.m`: main program `for FE-3`
   - `FnQMLtC.m`: sub-function for computing the QML function
   - `FnAQStC.m`: sub-function for computing the AQS function
   - `FnRAQStC.m`: sub-function for computing the robust AQS function

All main programs implement three methods: QML, M and robust M, which can easily be modified to suit practical applications. The sub-functions do not need modification.

Next, we briefly describe how to modify the main program for empirical applications. Suppose that we have the data file and the file for spatial weight matrix in excel format, `Munnell.xls` and `weight_Munnell.xls`. Replace the beginning part of the main program for data generation by the following:

```
data = importdata('Munnell.xls');
WMat = importdata('weight_Munnell.xls');
Y = data(:,4);
X = data(:,5:8);
      ⋮
```

which imports the data and spatial weight matrix, and defines the response vector `Y` and the regressors' matrix `X`.

The researchers then choose an FE specification that suits their problem as the 'best', and define the D matrix in Model (2.1), by referring Table A and the general descriptions given in Sec. 2. Finally, the researchers decide whether there are interesting STIC effects that they want to estimate. If so, the researchers proceed to decompose D and ϕ in Model (2.1).

Alternatively, the STIC effects can be recovered after the model estimation. Given the estimates $\hat{\theta}$ of the common parameters θ, the FE parameters are estimated through (3.2) as $\hat{\phi} = \hat{\phi}(\hat{\beta}, \hat{\delta})$. The STIC effects are linear contrasts, $C\phi$, of the FE parameters ϕ for a suitably defined linear contrast matrix C. Their estimates are $C\hat{\phi}$, and their VC matrix is $C\text{Var}(\hat{\phi})C'$. Therefore, inference about $C\phi$ can be made as long as we have a 'good' estimate $\widehat{\text{Var}}(\hat{\phi})$ of $\text{Var}(\hat{\phi})$, so that $C\widehat{\text{Var}}(\hat{\phi})C'$ consistently estimates $C\text{Var}(\hat{\phi})C'$. This method may be easier to implement and can be numerically more stable, but joint inference with common parameters cannot be directly carried out. See Pirotte and Yang (2024) for details on this method.