Maximum likelihood estimation of a spatial autoregressive model for origin-destination flow variables[1]

Hanbat Jeong[2], Yanli Lin[3], and Lung-fei Lee[4]

This version: November 21, 2021

Abstract

We introduce a spatial autoregressive (SAR) model for an origin-destination flow with the maximum likelihood (ML) estimation method. Each flow $y_{n,ij}$ shows a signal from an origin j to a destination i. A linear SAR model for flows quantifies three channels of spatial influences on $y_{n,ij}$: (1) effect from j to a third-party unit, (2) that from a third-party unit to i, and (3) that among third-party units. Motivated by a panel data model, we accommodate two-way fixed effects for innate characteristics of origin and destination. For a frequent data environment of flows, we also design a SAR Tobit model for flows. The ML estimator's asymptotic properties for the SAR Tobit model are investigated by the spatial near-epoch dependence (NED) concept. Using our models, we capture the significant three channels of spatial influences among the U.S. states' migration flows.

Keywords: Origin-destination flow, Spatial dependence, Tobit model, Maximum likelihood estimation, U.S. migration flow

JEL classification: C31, C51

1. Introduction

This paper develops a spatial autoregressive (SAR) model for an origin-destination flow with estimation methods. Each flow $y_{n,ij}$ can be regarded as a directed outcome generated from origin unit j toward destination unit i (hereafter, j denotes an origin while i represents a destination). A famous example of origin-destination flows is the U.S. states' migration flows. A traditional SAR model (e.g., Cliff and Ord, 1973; Ord, 1975; Anselin, 1988; Kelejian and Prucha, 2001; and Lee, 2004, 2007;) captures the spatial dependence among elements in a univariate variable. As an extension of a univariate SAR model, a corresponding research question is how to measure the spatial dependence among flows. A flow variable contains more information relative to a univariate variable in two aspects. First, two units' (or more) characteristics might affect a flow $y_{n,ij}$. Second, a flow $y_{n,ij}$ contains a direction of influence. Hence, a model specification for a flow variable should be more complex than that for a univariate variable. Relevant works can be found in LeSage and Pace (2008), Fischer and LeSage (2020), and Lee and Yu (2020).

[1] We would like to thank Professors Taehyun Ahn, Jason Blevins, Meta Brown, Chi-Young Choi, Robert de Jong, Suyong Song, Hankyoung Sung, and Bruce Weinberg for their valuable comments and suggestions. Hanbat thanks his colleague Gabi Jiang, and seminar participants at Korean American Economic Association Microeconometrics seminar, Sogang University and University of Seoul and the audience from the 2021 China Meeting in ShanghaiTech University, Shanghai, China, the 91st Annual Meeting of Southern Economic Association (sponsored by Korean-American Economic Association), Houston, U.S., for comments.

[2] Corresponding author: Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. Email: jeong.181@osu.edu.

[3] Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. Email: lin.3019@osu.edu.

[4] Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. Email: lee.1777@osu.edu.

Compared to the existing literature, our model has three advances. First, we clarify channels of spatial influences among flows by specifying a role of each cross-section unit and its position in spatial networks. Since a flow $y_{n,ij}$ contains directional information (from j to i), our model needs to show how spatial effects are generated from j toward i through third-party units with spatial network links. The main issue here is to characterize roles of third-party units in generating spatial spillovers. By introducing a relevant economic model, those channels will be justified. Second, we develop our SARF model for a censored flow variable. When a connection between two units is weak, a flow between them is frequently zero. Also, note that an origin-destination flow $y_{n,ij}$ belongs to a gross flow, which is necessarily nonnegative. They motivate us to build a SARF model with the Tobit structure (hereafter, SARF Tobit model). Relevant asymptotic inferences for the SARF Tobit model will be considered. Third, for an extension of SAR models for flows (hereafter, SARF model), we suggest estimation methods, which robustly control unobserved heterogeneities. Motivated by a traditional panel data model, we accommodate a fixed-effect specification for unobserved characteristics of an origin and a destination.[5] The asymptotic properties of the MLE for the linear SARF and the SARF Tobit models are studied when there exist two-way fixed effects.

For the first contribution, we develop a SARF model and its economic foundation. When there exist n cross-section units in a sample, a set of origin-destination flows can be characterized by an $n \times n$ matrix $Y_N = [y_{n,ij}]$ and an $N \times N$ link (network) matrix specifies their relations, where $N = n^2$. We consider directed forces among multiple units: a flow from j to i can be affected by flows involving third-party units g, h, p, and q (i.e., a flow from j to g, a flow from h to p, or a flow from q to i). Relative to the existing literature (e.g., LeSage and Pace (2008)), we have a general model specification by allowing two $n \times n$ spatial weighting matrices $W_n = [w_{n,ij}]$ and $M_n = [m_{n,ij}]$, which characterize network relationships among n cross-section units.[6] Each element of the ith row of W_n characterizes a relative spatial influence describing an influx into a destination i while an entry of the jth column of M_n shows a directed influence for an outflow from an origin j. In the aspect of agents' decision-making, W_n gives weights for agent j's decisions $y_{n,1j}, \cdots, y_{n,nj}$ for $j = 1, \cdots n$ while M_n contains weights for agents' decisions toward i ($y_{n,i1}, \cdots, y_{n,in}$) for $i = 1, \cdots, n$. We will give a specification example for W_n and M_n in the empirical application.

In consequence, W_n and M_n can generate three-type spatial influences among $y_{n,ij}$s via $N \times N$ matrices: (1) $I_n \otimes W_n$, (2) $M_n' \otimes I_n$, and (3) $M_n' \otimes W_n$. First, an $N \times N$ matrix $I_n \otimes W_n$ characterizes the spatial effect from an origin: a flow from an origin j to a third-party unit g. Second, $M_n' \otimes I_n$ specifies the spatial influence towards a destination: that from other unit q to a destination i. Last, $M_n' \otimes W_n$ characterizes the effect among third-party units. Also, univariate exogenous characteristics of i and j ($x_{n,i}$ and $x_{n,j}$) and exogenous distance or flow variables ($z_{n,ij,1}, \cdots, z_{n,ij,L}$) can affect $y_{n,ij}$. Our resulting model specification can nest the main equation of LeSage and Pace (2008) (equation (20) in LeSage and Pace (2008)), but we clarify channels of spatial influences via directed

[5] Note that a panel data set involves two indexes: (1) cross-section unit i and (2) time-series unit t. This data structure leads to having individual and time fixed effects.

[6] As a special case, we allow $M_n = W_n$. This case can represent the channels (directions) of spatial influences when W_n is a directed network. Another case is the LeSage and Pace's (2008) specification, i.e., $M_n = W_n'$ with a row-normalized W_n. Their specification focuses on having weighted averages of flows instead of considering directional spatial influences.

spatial network links. The first spatial effect arises via $w_{n,ig}y_{n,gj}$. It shows the effect from the flow from j to a third-party unit g through a network link $w_{n,ig}$. A chain $j \mapsto g \mapsto i$ can represent the first channel. The second-type spatial influence is represented by $y_{n,ih}m_{n,hj}$, and a chain $j \mapsto h \mapsto i$ describes this channel. The third spatial effect $w_{n,ig}y_{n,gh}m_{n,hj}$ characterizes the influence from $y_{n,gh}$ via directed network links $w_{n,ig}$ and $m_{n,hj}$. This effect arises when there exist connections (1) from an origin j to h, and (2) from g to a destination i. It can be represented by a chain:

$$j \text{ (origin) } \underbrace{\mapsto}_{m_{n,hj}} \underbrace{h \text{ (3rd party I)} \mapsto g \text{ (3rd party II)}}_{=flow\ y_{n,gh}} \underbrace{\mapsto}_{w_{n,ig}} i \text{ (destination)}.$$

Note that $y_{n,ij}$ would not have self-influence as the diagonal elements of W_n and M_n are zeros (i.e., $w_{n,ii}y_{n,ij} = y_{n,ij}m_{n,jj} = w_{n,ii}y_{n,ij}m_{n,jj} = 0$).[7]

The designed model can be related to an extended gravity equation and a weighted network formation model. When there exist n local representative agents, each flow $y_{n,ij}$ can be considered as an agent j's decision on signal's intensity toward agent i. If there is no spatial spillover effect, the agent j's optimal decision toward i can be represented by the conventional gravity model, which is a function of characteristics of i and j. On the other hand, our model allows that third-party regions' characteristics can affect $y_{n,ij}$ when there is a significant spatial interaction effect. Due to the existence of spatial spillovers, the impacts of characteristics of i and j can be amplified (i.e., multiplier effect). By identifying the model's parameters, the multiplier effects can be quantified as subsequent sections have shown.

Second, in addition to the linear SARF model, we consider the SARF model with a Tobit structure. It is motivated by a specific data environment for flow variables. In some empirical applications, we frequently detect zero observations in flows. It is likely to have a zero value for $y_{n,ij}$ even if there is a connection between units i and j. We can observe many zero values when a level of cross-section units is small (e.g., commuting flows among U.S. counties/cities). This is because a flow outcome between two small units can less occur due to some budgetary reasons: for example, a flow of two counties is more likely to be zero compared to that of two states or two countries. Hence, we consider a case that the range of a flow $y_{n,ij}$ is constrained in some way with a modification of the Tobit model, which is a tool for censored or truncated data. The resulting model is an extension of Qu and Lee (2012), Xu and Lee (2015b), which concern about outcomes of states but not for flow variables with more complex network structures. Thomas-Agnan and LeSage (2014) address this issue for flow variables with focusing on Bayesian estimation procedures (see Section 83.4 in Handbook of Regional Science (2014)).

Third, we suggest methods and their statistical properties for robustly controlling unobserved characteristics in estimating the SARF model. Instead of identifying the effects of univariate characteristics ($x_{n,i}$ and $x_{n,j}$) on $y_{n,ij}$, we specify the effects of origin's characteristics and those of destination's characteristics as fixed effects (i.e., two-way fixed-effect specification). The two-way fixed effects are specified by $2n$ individual parameters (n parameters for origins and n parameters for destinations). We can directly estimate the main parameters and fixed effects for both linear SARF and SARF Tobit specifications. With the linear SARF specification, a concentrated log-likelihood for the main

[7] We note that this assumption would exclude trade problems as internal trade is usually allowed in a trading issue. To apply our model to a trade issue, we then need to allow nonzero diagonal elements of W_n and M_n. In this case, the three-type spatial influences explained above are not able to be separated.

parameters can be established by using linear parameter properties of fixed effects.

For the two cases, we derive the log-likelihood functions. And then, asymptotic properties of the maximum likelihood (ML) estimators are studied. For the spatial Tobit flow model, due to its nonlinear structure, we build a topological structure for asymptotic analysis. On a geographic space, there exist N pairs of flows. A spatial unit is then a pair (i, j) instead of a single cross-section unit i. Each statistic $q_{n,ij}$ is originated from a location of flow (i, j), which is generated by two units i and j. Hence, $\{q_{n,ij}\}$ will construct a random field on a product space of cross-section units. The spatial near-epoch dependence (NED) concept (Jenish and Prucha, 2012) is employed to derive the MLE's asymptotic properties when the SARF Tobit model is considered.[8] Using this device, the law of large numbers (LLN) and the central limit theorem (CLT) are applied to the main statistics.

We also examine the asymptotic distributions of the MLE for the linear SARF and SARF Tobit models if there exist the two-way fixed effects. The existence of fixed effects leads to the incidental parameters (Neyman and Scott, 1948).[9] For both models, we provide the analytical bias corrections. In the linear SARF model, we apply the same strategy of deriving the asymptotic distribution of the MLE and its bias correction as those introduced by Lee and Yu (2010). To derive the MLE's asymptotic distribution for the SARF Tobit model, we employ the idea of Fernandez-Val and Weidner (2016): the second-order Taylor expansion of the concentrated log-likelihood evaluated at the true finite-dimensional parameters. The source of the asymptotic bias is the usage of estimated fixed effects, whose components have slower convergence rates than those of the main parameters' estimates.

By Monte Carlo simulations, we study performance of the MLE and the bias corrected MLE (if we consider the two-way fixed effects). When one disregards a censoring feature of flows, our simulation experiments indicate that (1) estimates of the main spatial interaction parameters are biased and (2) their coverage probabilities are distorted. Under the presence of fixed effects, downward biases in the MLEs of the spatial interaction parameters and the variance parameter are detected. The analytic bias correction procedures for the linear SARF and the SARF Tobit models significantly reduce the magnitudes of downward biases. In selecting a proper specification of spatial weighting matrices (W_n, M_n), our simulations indicate that the Akaike weight based on candidate models' sample log-likelihoods is a reasonable measure.

This paper provides an empirical application: migration flows of the U.S. states in the year 2010. We consider $(W_n, M_n) = (W_n^I, M_n^O)$, where W_n^I contains the shares of historical migrations toward destinations (forces to destinations) and M_n^O consists of those from origins (forces from origins). W_n^I and M_n^O are directed networks and can show different roles of origins and destinations in propagating spatial spillover effects. By the Akaike weights, the chosen specification provides a better fit than other

[8] In the event, there were no censoring, for the asymptotic properties of QMLE for the SARF model, we can apply an extension of the martingale difference central limit theorem (CLT) for a linear quadratic form (Kelejian and Prucha, 2001). And then, consistency and asymptotic normality of the QMLE will be provided. For a linear SAR model, the scenario of spatial unit allocation would not be needed but for the proper rate of convergence of estimator, expanded regions asymptotic is used. It is for nonlinear spatial models under the spatial mixing or spatial NED frameworks, the location setting is needed as in Jenish and Prucha (2009, 2012), so expanded regions asymptotic will be used.

[9] In a panel data setting with large cross-section and time-series observations, the asymptotic bias exists and can be corrected. Our setting is similar with the large panel data setting since there exist n origin units and n units for destinations.

specifications including that of LeSage and Pace (2008), i.e., $(W_n, M_n) = (W_{R,n}^a, W_{R,n}^{a\prime})$, where $W_{R,n}^a$ is the row-normalized states' adjacency matrix. We observe significant spatial influences by the three channels. The estimated average multiplier effect from the linear SARF model is 1.0607, which means the effects of i and j's characteristics are amplified by 106.07% in the equilibrium (from the SARF Tobit model, it is 1.0817). When we control unobservables via fixed effects, the estimates of the spatial interaction parameters from both linear SARF and SARF Tobit specifications are also significant. However, their magnitudes become smaller in absolute values (1.0024 from the linear SARF model and 1.0164 from the SARF Tobit model).

2. Model equations

Assume that there exist n cross-section units (e.g., regions) in a sample. That is, one has $N = n^2$ observations in an origin-destination flow variable. Let $Y_N = [y_{n,ij}]$ be an $n \times n$ matrix of flows, and $W_n = [w_{n,ij}]$ and $M_n = [m_{n,ij}]$ be $n \times n$ spatial weighting matrices characterizing relations among cross-section units $i = 1, \cdots, n$. As a traditional spatial econometric model, we assume that every spatial network link is nonnegative; and each diagonal element is zero. Each $y_{n,ij}$ can be considered as a directed outcome from j (origin) to i (destination). For example, on migration data, $y_{n,ij}$ is a migration flow from state j to state i. To explain Y_N, one can employ K univariate exogenous variables $X_n = [X_{n,1}, \cdots, X_{n,K}] = [x_{n,1}, \cdots, x_{n,n}]'$ with $X_{n,k} = (x_{n,1,k}, \cdots, x_{n,n,k})'$ for $k = 1, \cdots, K$ and $x_{n,i} = (x_{n,i,1}, \cdots, x_{n,i,K})'$ for $i = 1, \cdots, n$; and L exogenous distance or flow variables $z_{n,ij,l}$ for $l = 1, \cdots, L$.

Two spatial networks W_n and M_n allow us to have different sources of spatial influences. The first network matrix W_n characterizes directed spatial influences describing influxes into destinations in terms of columns while the second network M_n specifies directed influences for outflows from origin units in terms of rows. An advantage of this specification is that one can clarify the directions of spatial influences. A SAR model with a flow variable (hereafter, SARF) can be specified as

$$y_{n,ij} = \alpha_0 + \lambda_0 \sum_{g=1}^n w_{n,ig} y_{n,gj} + \gamma_0 \sum_{h=1}^n y_{n,ih} m_{n,hj} + \rho_0 \sum_{g=1}^n \sum_{h=1}^n w_{n,ig} y_{n,gh} m_{n,hj} + \sum_{l=1}^L \beta_{l,0} z_{n,ij,l} + \sum_{k=1}^K \left(b_{k,0} x_{n,i,k} + c_{k,0} x_{n,j,k}\right) + \epsilon_{n,ij}. \quad (1)$$

The specification of explanatory variables is similar to that of LeSage and Pace (2008). They will be introduced later. By aggregation, equation (1) can be consistent with a traditional SAR model for a univariate variable in some special cases.[10] The orders of $w_{n,ig} y_{n,gj}$, $y_{n,ih} m_{n,hj}$, and $w_{n,ig} y_{n,gh} m_{n,hj}$ in (1) are introduced to highlight the directions of spatial effects. The three-type spatial effects characterize the different roles of third-party units on a flow $y_{n,ij}$. The figure below illustrates an example of four regions.

[10] For example, if $\gamma_0 = \rho_0 = 0$, equation (1) can be represented by

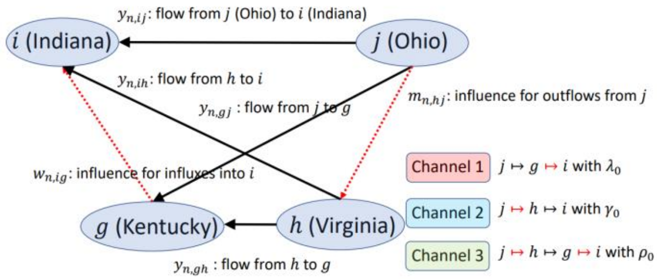
$$y_{n,i.} = \alpha_0 + \sum_{k=1}^K c_{k.0} \bar{x}_{n,k} + \lambda_0 \sum_{g=1}^n w_{n,ig} y_{n,g.} + \sum_{l=1}^L \beta_{l,0} z_{n,i.,l} + \sum_{k=1}^K b_{k,0} x_{n,i,k} + \epsilon_{n,i.},$$

where $y_{n,i.} = \frac{1}{n} \sum_{j=1}^n y_{n,ij}$ (average inflows toward i), $z_{n,i.,l}$s and $\epsilon_{n,i.}$ can be similarly defined, and $\bar{x}_{n,k} = \frac{1}{n} \sum_{j=1}^n x_{n,j,k}$. If $\lambda_0 = \rho_0 = 0$, we have another SAR representation from (1):

$$y_{n,.j} = \alpha_0 + \sum_{k=1}^K b_{k.0} \bar{x}_{n,k} + \gamma_0 \sum_{h=1}^n m_{n,hj} y_{n,.h} + \sum_{l=1}^L \beta_{l,0} z_{n,.j,l} + \sum_{k=1}^K c_{k,0} x_{n,j,k} + \epsilon_{n,.j},$$

where $y_{n,.j} = \frac{1}{n} \sum_{i=1}^n y_{n,ij}$ (average outflows from j), $z_{n,.j,l}$s and $\epsilon_{n,.j}$ can be similarly defined.

Figure 1. Three spatial effects on a flow $y_{n,ij}$



As the first case, we consider $M_n = W_n$, which is a directed regional network, i.e., $w_{n,ij} = 1$ if region j has an influence in region i's economy; and $w_{n,ij} = 0$ otherwise. For the first spatial effect, $w_{n,ig}y_{n,gj}$ shows the influences by a flow from j to a third-party g if there exists a connection from g to i (i.e., an influx into i). Then, this effect can be represented by a chain $j \mapsto g \mapsto i$. Second, if a region h (third-party) and j are linked (i.e., an outflow from j) and $\gamma_0 \neq 0$, the second spatial effect channel exists (i.e., the effect of a flow from h to i). Then, a chain $j \mapsto h \mapsto i$ can illustrate the second spatial effect. Similarly, $w_{n,ig}y_{n,gh}w_{n,hj}$ shows the effect of flows among third-party units when there exist geographic connections (1) between j and a third-party h and (2) between a third-party g and i. A chain $j \mapsto h \mapsto g \mapsto i$ can show the third spatial influence.[11] Note that this effect is distinguished from the first two effects since $w_{n,ii} = w_{n,jj} = 0$.

As a general case, one can specify a different M_n from W_n to present different sources of spatial effects for economic reasonings. Observe that W_n characterizes the effect from column sums of Y_N while M_n is for the effect from row sums of Y_N. Suppose that $y_{n,ij}$ is a decision of agent j toward i (i.e., a signal). Note that W_n provides weights for agent j's decisions $y_{n,1j}, \cdots, y_{n,nj}$ for $j = 1, \cdots, n$. Then, the weighted column sum $\sum_{g=1}^{n} w_{n,ig}y_{n,gj}$ can represent the aggregated signals from j to i. On the other hand, M_n gives weights for agents' decisions toward i, i.e., $y_{n,i1}, \cdots, y_{n,in}$ for $i = 1, \cdots, n$. The weighted row sum $\sum_{h=1}^{n} m_{n,hj}y_{n,ih}$ of the ith row of Y_N shows the aggregated signals toward i via the connections of j. Since W_n and M_n can play different roles in weighting agents' decisions, a practitioner can specify proper settings of W_n and M_n, which are consistent with his/her purpose. We will introduce a theoretical model framework for this setting in Section 2.1, and empirical examples for W_n and M_n in Section 6.

By allowing two spatial weighting matrices (W_n, M_n), our framework also generalizes the LeSage and Pace's (2008) specification. The LeSage and Pace's (2008) specification can be written with a scalar notation:

$$y_{n,ij} = \alpha_0 + \lambda_0 \sum_{g=1}^{n} w_{n,ig}y_{n,gj} + \gamma_0 \sum_{h=1}^{n} w_{n,jh}y_{n,ih} + \rho_0 \sum_{g=1}^{n}\sum_{h=1}^{n} w_{n,ig}w_{n,jh}y_{n,gh} + \sum_{l=1}^{L} \beta_{l,0}z_{n,ij,l} + \sum_{k=1}^{K}\left(b_{k,0}x_{n,i,k} + c_{k,0}x_{n,j,k}\right) + \epsilon_{n,ij}.$$

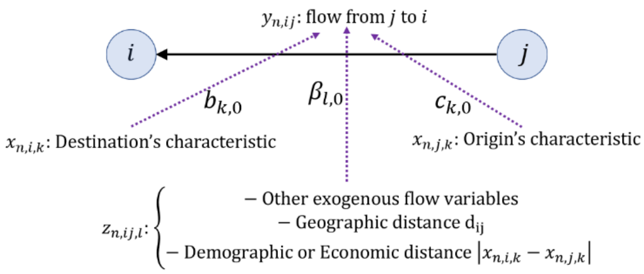
The LeSage and Pace's (2008) specification implies $M_n = W_n'$ with a row-normalized W_n. Under this

[11] Then, $\sum_{g=1}^{n} w_{n,ig}y_{n,gj}$, $\sum_{h=1}^{n} y_{n,ih}m_{n,hj}$, and $\sum_{g=1}^{n}\sum_{h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}$ are local aggregates with specifying directions of influences. In contrast to a univariate SAR model, there is no good rationale of considering a row-normalized W_n when $M_n = W_n$. This is because the first and third channels of spatial effects involve the column sums of W_n.

specification, a flow $y_{n,ij}$ is a weighted average of other flows (i.e., local averages). Moreover, each spatial network link might not contain directional information of a network.

Figure 2 shows potential channels that flow/univariate characteristics affect a flow $y_{n,ij}$. There exist two effects from univariate characteristics X_n. Since a flow $y_{n,ij}$ involves two units i and j, characteristics of both i and j can affect $y_{n,ij}$ with different sensitivities. When $y_{n,ij}$ is a migration flow from state j to state i, population levels and personal incomes of states i and j can be considered as components of X_n. The coefficient $b_{k,0}$ captures the effect of a destination's characteristic $x_{n,i,k}$ on $y_{n,ij}$, while $c_{k,0}$ measures the influence of an origin's characteristic $x_{n,j,k}$ on $y_{n,ij}$.

Figure 2. Potential characteristics affecting a flow y_{ij}



One can test the homogenous effect hypothesis, i.e., $b_{k,0} = c_{k,0}$ for some k. Under the hypothesis $b_{k,0} = c_{k,0}$ for all $k = 1, \cdots, K$, for example, each regressor becomes $x_{n,i,k} + x_{n,j,k}$. In this case, the explanatory variable part of our model can be simplified to $\sum_{l=1}^{L+K} \beta_{l,0} z_{n,ij,l}$, where $z_{n,ij,L+k} = x_{n,i,k} + x_{n,j,k}$ and $\beta_{L+k,0} = b_{k,0}$ for $k = 1, \cdots, K$. As the second-type explanatory variables, additional flows or distance variables $z_{n,ij,1}, \cdots, z_{n,ij,L}$ can be employed, and the parameters $\beta_{1,0}, \cdots, \beta_{L,0}$ capture the linear effects of them. For example, one can utilize geographic distances $\{d_{ij}\}$.

Using a specific data environment of our model, extra exogenous variations explaining $y_{n,ij}$ can be made. First, one can generate a z-variable using a univariate characteristic $\{x_{n,i}\}$. For example, an economic distance can be generated by $|income_i - income_j|$ (or an economic proximity $\frac{1}{|income_i - income_j|}$) where $income_i$ denotes the region i's personal income level. Another example is an income ratio $\frac{income_j}{income_i}$, which captures a relative volume of regions i and j. Compared to a distance variable, a ratio-type variable can capture a variation in $y_{n,ij}$ from asymmetric relations. On the other hand, a z-variable can also generate a x-variable by summation. If $W_n^a = [w_{n,ij}^a]$ denotes an adjacency network matrix, one can generate degrees $deg_j = \sum_{k=1}^{n} w_{n,jk}^a$ to examine the effect of j's network connectivity on $y_{n,ij}$. Other network statistics can also be generated from W_n^a.

For statistical analysis, a stacked vector notation is useful. In a matrix form with $Y_N = [y_{n,ij}]$ as an $n \times n$ matrix of flows,

$$Y_N = \alpha_0 l_n l_n' + \lambda_0 W_n Y_N + \gamma_0 Y_N M_n + \rho_0 W_n Y_N M_n + \sum_{l=1}^{L} \beta_{l,0} Z_{N,l} + \sum_{k=1}^{K} \left(b_{k,0} X_{n,k} l_n' + c_{k,0} l_n X_{n,k}' \right) + \epsilon_N, \quad (2)$$

where each $Z_{N,l} = [z_{n,ij,l}]$ for $l = 1,\cdots,L$ is an $n \times n$ matrix of an explanatory distance/flow variable, and $\epsilon_N = [\epsilon_{n,ij}]$ stands for an $n \times n$ disturbance matrix. A univariate regressor matrix should not contain a constant vector l_n for identification. Then, model (1) can be rewritten as

$$vec(Y_N) = \alpha_0 l_N + \big(\lambda_0(I_n \otimes W_n) + \gamma_0(M_n' \otimes I_n) + \rho_0(M_n' \otimes W_n)\big)vec(Y_N) + \\ + \textstyle\sum_{l=1}^{L} \beta_{l,0} vec(Z_{N,l}) + \sum_{k=1}^{K}\Big(b_{k,0}(l_n \otimes I_n) + c_{k,0}(I_n \otimes l_n)\Big)X_{n,k} + vec(\epsilon_N). \quad (3)$$

In consequence, equation (3) represents the spatial influences among flow units. Observe that three $N \times N$ matrices $I_n \otimes W_n$, $M_n' \otimes I_n$, and $M_n' \otimes W_n$ characterize network relationships among flows, and they correspond to spatial weighting matrices for a univariate SAR model. Let $\boldsymbol{W}_N = I_n \otimes W_n$, $\boldsymbol{M}_N = M_n' \otimes I_n$, $\boldsymbol{R}_N = M_n' \otimes W_n$ and $\boldsymbol{A}_N = \lambda_0 \boldsymbol{W}_N + \gamma_0 \boldsymbol{M}_N + \rho_0 \boldsymbol{R}_N$. Observe that $\boldsymbol{R}_N = \boldsymbol{W}_N \boldsymbol{M}_N$. For each pair (i,j) for a flow, we can find a unique index $f = (j-1)n + i \in \{1,\cdots,N\}$. Hence, the spatial influence between two flows (i,j) and (g,h) (i.e., the (f,f')-element of $\boldsymbol{A}_N$, where $f = (j-1)n + i$ and $f' = (h-1)n + g$) can be characterized by

$$(e_{n,j}' \otimes e_{n,i}')\boldsymbol{A}_N(e_{n,h} \otimes e_{n,g}) = \lambda_0 1(j=h)w_{n,ig} + \gamma_0 1(i=g)m_{n,hj} + \rho_0 w_{n,ig} m_{n,hj},$$

where $1(\cdot)$ denotes the indicator function, and $e_{n,j} = (0,\cdots,1,\cdots,0)'$ with only the j component being one and other entries being zero (which is an n-dimensional unit vector). The below provides a discussion on the three network structures $\boldsymbol{W}_N$, $\boldsymbol{M}_N$, and $\boldsymbol{R}_N$.

Remark. For interpretations, suppose that W_n and M_n are dichotomous networks. Let $c_{ind,w,j} = \sum_{i=1}^{n} w_{n,ij}$ and $c_{ind,m,j} = \sum_{i=1}^{n} m_{n,ij}$ for each j (indegrees); and $c_{outd,w,i} = \sum_{j=1}^{n} w_{n,ij}$ and $c_{outd,m,i} = \sum_{j=1}^{n} m_{n,ij}$ for each i (outdegrees). By the uniform boundedness assumption for row and column sum norms of W_n and M_n, we have $\sum_{i=1}^{n} c_{ind,w,i} = O(n^{\varrho_w})$ and $\sum_{i=1}^{n} c_{ind,m,i} = O(n^{\varrho_m})$ where $0 \le \min\{\varrho_w, \varrho_m\} \le \max\{\varrho_w, \varrho_m\} \le 1$. Then, the densities of W_n and M_n are respectively $\frac{\sum_{i=1}^{n} c_{outd,w,i}}{n(n-1)} = O(n^{\varrho_w - 2})$ and $\frac{\sum_{i=1}^{n} c_{ind,m,i}}{n(n-1)} = O(n^{\varrho_m - 2})$.[12]

Note that the three networks $\boldsymbol{W}_N$, $\boldsymbol{M}_N$, and $\boldsymbol{R}_N$ characterize relations among flows. Then, the vectors of row sums of $\boldsymbol{W}_N$, $\boldsymbol{M}_N$, and $\boldsymbol{R}_N$ are respectively $l_n \otimes (c_{outd,w,1},\cdots,c_{outd,w,n})'$, $(c_{ind,m,1},\cdots,c_{ind,m,n})' \otimes l_n$, and $(c_{ind,m,1},\cdots,c_{ind,m,n})' \otimes (c_{outd,w,1},\cdots,c_{outd,w,n})'$; while their column sums are $l_n' \otimes (c_{ind,w,1},\cdots,c_{ind,w,n})$, $(c_{outd,m,1},\cdots,c_{outd,m,n}) \otimes l_n'$, $(c_{outd,m,1},\cdots,c_{outd,m,n}) \otimes (c_{ind,w,1},\cdots,c_{ind,w,n})$, respectively. The densities of $\boldsymbol{W}_N$, $\boldsymbol{M}_N$, and $\boldsymbol{R}_N$ are respectively $\frac{\sum_{i=1}^{n} c_{outd,w,i}}{n(n^2-1)} = O(n^{\varrho_w - 3})$, $\frac{\sum_{i=1}^{n} c_{ind,m,i}}{n(n^2-1)} = O(n^{\varrho_m - 3})$, and $\frac{(\sum_{i=1}^{n} c_{outd,w,i})(\sum_{i=1}^{n} c_{ind,m,i})}{n^2(n^2-1)} = O(n^{\varrho_w + \varrho_m - 4})$. We observe that the three networks for flows are sparser than W_n and M_n. For $\boldsymbol{R}_N$, moreover, we have $\|\boldsymbol{R}_N\|_\infty \le \|\boldsymbol{W}_N\|_\infty \|\boldsymbol{M}_N\|_\infty$ and $\|\boldsymbol{R}_N\|_1 \le \|\boldsymbol{W}_N\|_1 \|\boldsymbol{M}_N\|_1$ by $\boldsymbol{R}_N = \boldsymbol{W}_N \boldsymbol{M}_N$ and the submultiplicative property of a norm. For asymptotic analysis, the row and column sum norms' magnitudes of $\boldsymbol{W}_N$, $\boldsymbol{M}_N$, and $\boldsymbol{R}_N$ are regulated when W_n and M_n are regulated. ■

Let $S_N = I_N - \boldsymbol{A}_N$ be the spatial filter matrix. If S_N is invertible, the unique reduced form of (3) is

[12] Note that $\sum_{j=1}^{n} c_{ind,m,j} = \sum_{i=1}^{n} c_{outd,m,i}$.

$$vec(Y_N) = S_N^{-1}\left[\alpha_0 l_N + \sum_{l=1}^{L}\beta_{l,0}vec(Z_{N,l}) + \sum_{k=1}^{K}\left(b_{k,0}(l_n \otimes I_n) + c_{k,0}(I_n \otimes l_n)\right)X_{n,k} + vec(\epsilon_N)\right]. \quad (4)$$

The model implies that $\frac{\partial y_f}{\partial z_{f',l}} = \beta_{l,0}[S_N^{-1}]_{ff'}$ for $l = 1,\cdots,L$, and $\frac{\partial y_f}{\partial x_{j,k}} = e'_{N,f}S_N^{-1}\left(b_{k,0}(l_n \otimes e_{n,j}) + c_{k,0}(e_{n,j} \otimes l_n)\right)$ for $j = 1,\cdots,n$ and $k = 1,\cdots,K$, where $e_{N,f} = (0,\ldots,0,1,0,\ldots,0)'$ denotes the N-dimensional unit vector with the unit at the f component and zero elsewhere. Hence, the main part of interpreting our model is understanding the structure of S_N^{-1}. We discuss this issue in Section 2.1.

Note that the spatial filter matrix S_N not only characterizes the equilibrium effects, but it also determines the correlation structure of $vec(Y_N)$. From (4), the variance matrix of $vec(Y_N)$ is $Var(vec(Y_N)) = \sigma_0^2 S_N^{-1}S_N'^{-1}$ where σ_0^2 is the variance of $\epsilon_{n,ij}$. The variance of $y_{n,ij}$ for each pair (i,j) is $\sigma_0^2 e'_{N,f}S_N^{-1}S_N'^{-1}e_{N,f}$, where $f = (j-1)n + i$. As n increases, in order for $Var(vec(Y_N))$ to be bounded, regularity conditions are needed so that S_N^{-1} will not be explosive. A sufficient condition of spatial stability is $\|\boldsymbol{A}_N\|_\infty < 1$. If W_n and M_n are diagonalizable, an eigenvalue of $\boldsymbol{A}_N$ is $\lambda_0\varpi_{1n,i} + \gamma_0\varpi_{2n,j} + \rho_0\varpi_{1,ni}\varpi_{2,nj}$ where $\varpi_{1,ni}$ and $\varpi_{2,ni}$ are respectively eigenvalues of W_n and M_n for $i = 1,\cdots,n$.[13] Then, the parameter space of the stable model can be $\{(\lambda,\gamma,\rho): |\lambda\bar{\varpi}_{1n,i} + \gamma\bar{\varpi}_{2n,j} + \rho\bar{\varpi}_{2n,j}\bar{\varpi}_{1n,i}| < 1, \text{for all } i,j = 1,\ldots,n\}$.

The parameter vector of our interest is $\theta_0 = (\alpha_0,\lambda_0,\gamma_0,\rho_0,\beta_0',b_0',c_0',\sigma_0^2)'$ where $\beta_0 = (\beta_{1,0},\cdots,\beta_{L,0})'$, $b_0 = (b_{1,0},\cdots,b_{K,0})'$, and $c_0 = (c_{1,0},\cdots,c_{K,0})'$. To estimate θ_0, we employ the maximum likelihood (ML) estimation method. Let $\theta = (\alpha,\lambda,\gamma,\rho,\beta',b',c',\sigma^2)'$ with $\beta = (\beta_1,\cdots,\beta_L)'$, $b = (b_1,\cdots,b_K)'$, and $c = (c_1,\cdots,c_K)'$, be a parameter vector in a parameter space. If $\epsilon_{n,ij} \sim i.i.d.N(0,\sigma_0^2)$, the log-likelihood function of the observed continuous dependent variables vector $vec(Y_N)$ is

$$\ln L_N(\theta) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 + \ln\det(S_N(\lambda,\gamma,\rho)) - \frac{1}{2\sigma^2}vec(\epsilon_N(\theta))'vec(\epsilon_N(\theta)), \quad (5)$$

where $S_N(\lambda,\gamma,\rho) = I_N - \boldsymbol{A}_N(\lambda,\gamma,\rho)$ with $\boldsymbol{A}_N(\lambda,\gamma,\rho) = \lambda\boldsymbol{W}_N + \gamma\boldsymbol{M}_N + \rho\boldsymbol{R}_N$, and

$$vec(\epsilon_N(\theta)) = S_N(\lambda,\gamma,\rho)vec(Y_N) - \alpha l_N - \sum_{l=1}^{L}\beta_l vec(Z_{N,l}) - \sum_{k=1}^{K}(b_k(l_n \otimes I_n) + c_k(I_n \otimes l_n))X_{n,k}.$$

The maximum likelihood (ML) estimator $\hat{\theta}_N$ can be obtained by maximizing $\ln L_N(\theta)$. In computing $\hat{\theta}_N$, evaluating $\ln\det(S_N(\lambda,\gamma,\rho))$ might be demanding when n is large. For this issue, we recommend an approximation method using Chebyshev polynomials (see Pace and LeSage, 2004).[14] When $\epsilon_{n,ij}$ is not normally distributed, the log-likelihood function (5) will be a quasi log-likelihood function. We will study the quasi-maximum likelihood (QML) estimator's asymptotic properties in Section 4.

---

[13] More details on the preceding eigenvalues can be found in Appendix Claim A.1. Due to the Kronecker product structure in $\boldsymbol{A}_N$, we note that the above eigenvalues of $\boldsymbol{A}_N$ can be derived even though W_n and M_n are not necessarily simultaneously diagonalizable.

[14] Compared to a univariate SAR model, $S_N(\lambda,\gamma,\rho)$ is an n^2-dimensional square matrix. Hence, a computation cost for $\ln\det(S_N(\lambda,\gamma,\rho))$ exponentially increases when cross-section observations increase. Also, utilizing Chebyshev polynomials is better than considering the Taylor series expansion. Their simulation results have shown that Chebyshev polynomials demonstrate robust performance across a range of spatial interaction parameters.

Since α, β, b, and c are linear parameters of exogenous regressors and σ^2 is the variance parameter of disturbances, we can establish the concentrated log-likelihood function solely relying on λ, γ, and ρ. For notational convenience, let $\delta = (\lambda, \gamma, \rho)'$, $\kappa = (\alpha, \beta', b', c')'$, and $\mathbf{X}_N = [l_N, vec(Z_{N,1}), \cdots, vec(Z_{N,L}), (l_n \otimes I_n)X_n, (I_n \otimes l_n)X_n]$. By the first-order conditions, we have $\hat{\kappa}_N(\delta) = (\mathbf{X}_N'\mathbf{X}_N)^{-1}\mathbf{X}_N' S_N(\delta) vec(Y_N)$ for each δ. By putting $\hat{\kappa}_N(\delta)$ back into $\ln L_N(\theta)$, we obtain

$$\ln L_N(\delta, \sigma^2) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln \sigma^2 + \ln \det(S_N(\delta)) - \frac{1}{2\sigma^2} vec(Y_N)' S_N'(\delta) M_{\mathbf{X}_N} S_N(\delta) vec(Y_N),$$

where $M_{\mathbf{X}_N} = I_N - \mathbf{X}_N(\mathbf{X}_N'\mathbf{X}_N)^{-1}\mathbf{X}_N'$. Using $\hat{\sigma}_N^2(\delta) = \frac{1}{N} vec(Y_N)' S_N'(\delta) M_{\mathbf{X}_N} S_N(\delta) vec(Y_N)$, we obtain the concentrated log-likelihood function:

$$\ln L_N(\delta) = -\frac{N}{2}(\ln 2\pi + 1) - \frac{N}{2}\ln \hat{\sigma}_N^2(\delta) + \ln \det(S_N(\delta)),$$

which can be used for the optimization search on δ for estimation.

2.1. Economic foundation

The purpose of this section is to provide economic reasonings of model (1). With this point, we will investigate the structure of S_N^{-1}. An economic foundation of the process (4) can be a set of optimal outcomes of representative regional agents. We suppose that there exist n representative agents and each agent j can choose n actions, $y_{n,1j}, \cdots, y_{n,nj}$. Since this problem is choosing the signals' intensities from j, it can be related to weighted network formation. That is, a matrix of flows Y_N can be considered as a weighted network. A recent theoretical work on weighted network formation is Baumann (2021).[15] In the Baumann's (2021) concept, $y_{n,jj}$ is j's self-investment while $y_{n,kj}$ for $k \neq j$ is an amount of investment from j to k. We can relate this signal choice problem to an optimal resource flow model.

For illustrative purposes, let $y_{n,ij}$ be the logged resource flow from j to i (denoted by $\ln(rflow_{ij})$), $z_{n,ij} = \ln(1 + d_{ij})$ and $x_{n,i} = \ln(pop_i)$, where d_{ij} denotes the geographic distance between i and j and pop_i denotes the region i's population level (mass). That is, $L = 1$ and $K = 1$. For interpretations, we consider the log-transformed variables to have the same framework with McCallum's (1995) gravity model: $rflow_{ij} = \overline{rflow}\exp(\epsilon_{ij})(1 + d_{ij})^{\beta_0} pop_i^{b_0} pop_j^{c_0}$, where $\overline{rflow}$ denotes the baseline of $rflow_{ij}$. Then, the coefficients β_0, b_0, and c_0 represent elasticities and their expected signs are $\beta_0 < 0$, $b_0 > 0$, and $c_0 > 0$.[16] For the case of $i = j$, $rflow_{ii} = \overline{rflow}\exp(\epsilon_{ii}) pop_i^{b_0 + c_0}$, which indicates that the i's self-investment $rflow_{ii}$ is only affected by the i's characteristics. The pop_i-elasticity of $rflow_{ii}$ is $b_0 + c_0$.

---

[15] Baumann (2021) establishes an incentive structure of forming $\{y_{n,ij}\}$. Her model has a restriction on the agent's utility function, e.g., $y_{n,ij} = 0$ if and only if $y_{n,ji} = 0$ for $j \neq i$. Even though our model does not rely on her theoretical assumptions, her interpretations on a weighted network link can be applied to a flow $y_{n,ij}$.

[16] A similar structure can be found in estimating the Cobb-Douglas production function. Refer to Section 1.3 in Wooldridge (2010).

Let $\boldsymbol{\eta}_N = (\eta_{n,11}, \cdots, \eta_{n,n1}, \cdots, \eta_{n,1n}, \cdots, \eta_{n,nn})'$ where $\eta_{n,ij} = \beta_0(z_{n,ij} - \bar{z}_n) + b_0(x_{n,i} - \bar{x}_n) + c_0(x_{n,j} - \bar{x}_n) + \epsilon_{n,ij}$ be a vector of exogenous characteristics. Note that $\bar{z}_n$ and $\bar{x}_n$ denote respectively the averages of $\{z_{n,ij}\}$ and $\{x_{n,i}\}$. To justify equation (4), a utility of a representative agent for region j from his/her relation with that for region i is

$$U_j(i) = e'_{n,i} S_{inv.j} \boldsymbol{\eta}_N y_{ij} - \frac{1}{2}\left(y_{n,ij} - \bar{y}_n e'_{n,i} S_{inv.j} l_N\right)^2,$$

where $S_N^{-1} = [S'_{inv.1}, S'_{inv.2}, \cdots, S'_{inv.n}]'$ (i.e., $S_{inv.j}$ is an $n \times N$ submatrix of S_N^{-1}, so $e'_{n,i} S_{inv.j}$ is the $(j-1)n + i$-th row of S_N^{-1}) and $\bar{y}_n$ denotes the social norm/guideline in selecting $y_{n,ij}$. Each $S_{inv.j}$ shows the externalities from spatial influences; and $-\frac{1}{2}\left(y_{n,ij} - \bar{y}_n e'_{n,i} S_{inv.j} l_N\right)^2$ represents a quadratic cost of sending a signal to i.

If $\lambda_0 = \gamma_0 = \rho_0 = 0$, there is no spatial influence (no externality). In this case, $U_j(i) = \eta_{n,ij} y_{n,ij} - \frac{1}{2}\left(y_{n,ij} - \bar{y}_n\right)^2$ with $\bar{y}_n = \beta_0 \bar{z}_n + (b_0 + c_0)\bar{x}_n$, which implies that there is no incentive to consider the effects of a third-party unit. On the other hand, characteristics of a third-party unit can affect a signal $y_{n,ij}$ if some of λ_0, γ_0, and ρ_0 are nonzero. Let $e'_{n,i} S_{inv.j} = (s^{ij}_{inv,11}, s^{ij}_{inv,21}, \cdots, s^{ij}_{inv,nn})$. Then, as $y_{n,ij} = e'_{n,i} S_{inv.j} (\bar{y}_n l_N + \boldsymbol{\eta}_N)$, the optimal resource flow can be characterized by

$$rflow_{ij} = \overline{rflow}^{\sum_{g=1}^n \sum_{h=1}^n s^{ij}_{inv,gh}} \prod_{g=1}^n \prod_{h=1}^n \left(\exp(\epsilon_{gh})(1 + d_{gh})^{\beta_0} pop_g^{b_0} pop_h^{c_0}\right)^{s^{ij}_{inv,gh}}.$$

Then, the optimal resource flow $rflow_{ij}$ is also affected by characteristics of third-party units through $s^{ij}_{inv,gh}$. Also, the effects of i and j's characteristics on $rflow_{ij}$ (represented by the coefficients β_0, b_0, and c_0) are amplified by $\beta_0 s^{ij}_{inv,ij}$, $b_0 s^{ij}_{inv,ij}$, and $c_0 s^{ij}_{inv,ij}$. The remark below describes the structure of $s^{ij}_{inv,gh}$.

Remark (Structure of S_N^{-1}). First, consider the case of $\rho_0 = -\lambda_0\gamma_0$ (LeSage and Pace's (2008) special case) to provide intuitive explanations.[17] By having separable spatial filters (destination-based and origin-based), this case can highlight the roles of the two spatial networks W_n and M_n.

Consider the case of $(i,j) = (g,h)$ to study the spatial multiplier effect. If $\rho_0 = -\lambda_0\gamma_0$ with spatial stability, we have $S_N^{-1} = \left(I_N + \sum_{p=1}^{\infty} \lambda_0^p (I_n \otimes W_n^p)\right)\left(I_N + \sum_{q=1}^{\infty} \gamma_0^q (M_n'^q \otimes I_n)\right)$ and $s^{ij}_{inv,ij}$ is a diagonal element of S_N^{-1}. Then, the (f,f)-element of S_N^{-1} with $f = (j-1)n + i$ is

$$[S_N^{-1}]_{ff} = (e'_{n,j} \otimes e'_{n,i})\left(I_N + \sum_{p=1}^{\infty} \lambda_0^p (I_n \otimes W_n^p)\right)\left(I_N + \sum_{q=1}^{\infty} \gamma_0^q (M_n'^q \otimes I_n)\right)(e_{n,j} \otimes e_{n,i})$$

[17] A similar parameter restriction can be found in a dynamic spatial panel data model (see Section 12.2.2 in Lee and Yu (2015)). In a spatial dynamic panel data model, a similar restriction leads to separable space and time filters. Similarly, the parameter restriction $\rho_0 = -\lambda_0\gamma_0$ separates the spatial dependence among flows into (1) origin-based dependence $M_n' \otimes I_n$ and (2) destination-based dependence $I_n \otimes W_n$ (see Section pages 952-954 of LeSage and Pace (2008)). Even though LeSage and Pace (2008) discuss the special case ($M_n = W_n'$ with a row-normalized W_n), the same idea of separable space filters can be applied.

$$= 1 + \sum_{p=2}^{\infty} \left(\lambda_0^p [W_n^p]_{ii} + \gamma_0^p [M_n^p]_{jj} \right) + \sum_{p=2}^{\infty} \sum_{q=2}^{\infty} \lambda_0^p \gamma_0^q [W_n^p]_{ii} [M_n^q]_{jj},$$

where $w_{n,ii} = m_{n,ii} = 0$.

For the above, note that $[M_n'^q]_{jj} = e_{n,j}' M_n'^q e_{n,j} = e_{n,j}' M_n^q e_{n,j} = [M_n^q]_{jj}$. Hence, $[S_N^{-1}]_{ff}$ is a combination of the feedback effects (1) from i to i, and (2) from j to j, i.e., $\underbrace{i \mapsto \cdots \mapsto i}_{pth-order}$ and $\underbrace{j \mapsto \cdots \mapsto j}_{qth-order}$ for $p, q = 1,2, \cdots, \infty$. For the feedback effect from i to i, note that the middle links might not include i for some order p, even though some might include i.[18] The same logic can be applied to the feedback effect from j to j.

To illustrate the roles of third-party units, we secondly consider the structure of $s_{inv,gh}^{ij}$ when $(i,j) \neq (g,h)$:

$$\sum_{p=1}^{\infty} \left(\lambda_0^p 1(j=h) [W_n^p]_{ig} + \gamma_0^p 1(i=g) [M_n^p]_{hj} \right) + \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} \lambda_0^p \gamma_0^q [W_n^p]_{ig} [M_n^q]_{hj}.$$

We observe that $s_{inv,gh}^{ij}$ consists of (1) chains from third-party g to destination i (i.e., $g \mapsto \cdots \mapsto i$ via $[W_n^p]_{ig}$) and (2) those from origin j to third-party h (i.e., $j \mapsto \cdots \mapsto h$ via $[M_n^p]_{hj}$). From the structure of $s_{inv,gh}^{ij}$, we verify that the first network matrix W_n characterizes relative spatial influences describing influxes into destination units while the second network M_n specifies directed influences for outflows from origin units.

Second, if there is no restriction on λ_0, γ_0, and ρ_0, $S_N^{-1} = I_N + \sum_{p=1}^{\infty} \left(\lambda_0 (I_n \otimes W_n) + \gamma_0 (M_n' \otimes I_n) + \rho_0 (M_n' \otimes W_n) \right)^p$. Then, the (f, f')-element of S_N^{-1} with $f = (j-1)n + i$ and $f' = (h-1)n + g$ is

$$\sum_{p=0}^{\infty} \sum_{p_1+p_2+p_3=p} \frac{p!}{p_1! p_2! p_3!} \lambda_0^{p_1} \gamma_0^{p_2} \rho_0^{p_3} \left[W_n^{p_1+p_3} \right]_{ig} \left[M_n^{p_2+p_3} \right]_{hj}$$

by the trinomial expansion formula. Then, the p-th order effect contains (1) the $p_1 + p_3$-th order effects from g to i originated from $W_n^{p_1+p_3}$ and (2) the $p_2 + p_3$-th order effects from j to h by $M_n^{p_2+p_3}$ such that $p_1 + p_2 + p_3 = p$. ■

2.2. Fixed-effect specification for unobserved characteristics of both origin and destination

In this subsection, we introduce an alternative linear SARF model specification with a fixed effect for each origin unit, and a fixed effect for each destination unit, which can robustly control unobserved characteristics. Consider an extension of the linear SARF model:

$$y_{n,ij} = \lambda_0 \sum_{g=1}^{n} w_{n,ig} y_{n,gj} + \gamma_0 \sum_{h=1}^{n} y_{n,ih} m_{n,hj} + \rho_0 \sum_{g=1}^{n} \sum_{h=1}^{n} w_{n,ig} y_{n,gh} m_{n,hj}$$

---

[18] For the second and third order effects (i.e., $p = 2,3$), paths are $i \mapsto k \mapsto i$ for some k (when $p = 2$) and $i \mapsto k_1 \mapsto k_2 \mapsto i$ for some k_1 and k_2 (when $p = 3$). Since we exclude self-influence, note that k, k_1, and k_2 do not contain i. From the fourth order effect ($p = 4$), the middle links can contain i since a possible path is $i \mapsto k_1 \mapsto k_2 \mapsto k_3 \mapsto i$ for some k_1, k_2, and k_3. By the same logic, potential k_1 and k_3 do not include i. However, k_2 can contain i if there exist paths $k_1 \mapsto i$ and $i \mapsto k_3$.

$$+\boldsymbol{z}_{n,ij}\beta_0 + \alpha_{i,d,0} + \alpha_{j,o,0} + \epsilon_{ij}, \tag{6}$$

where $\boldsymbol{z}_{n,ij} = \left(z_{n,ij,1}, \cdots, z_{n,ij,L}\right)'$, $\alpha_{i,d,0}$ denotes a destination i's fixed-effect component, and $\alpha_{j,o,0}$ is an unobserved fixed-effect component of origin j due to the presence of unknown fixed effects.[19] Let $\omega_0 = (\delta_0', \beta_0', \sigma_0^2)'$ be the true parameter vector for model (6) and $\omega = (\delta', \beta', \sigma^2)'$ be a vector of possible parameter values. Let $\boldsymbol{\alpha}_{n,o,0} = \left(\alpha_{1,o,0}, \cdots, \alpha_{n,o,0}\right)'$ and $\boldsymbol{\alpha}_{n,d,0} = \left(\alpha_{1,d,0}, \cdots, \alpha_{n,d,0}\right)'$ be vectors of the true fixed-effect components and $\boldsymbol{\alpha}_{n,o} = \left(\alpha_{1,o}, \cdots, \alpha_{n,o}\right)'$ and $\boldsymbol{\alpha}_{n,d} = \left(\alpha_{1,d}, \cdots, \alpha_{n,d}\right)'$ be possible vectors of fixed-effect components. The vector/matrix notation of (6) is

$$vec(Y_N) = (\lambda_0 \boldsymbol{W}_N + \gamma_0 \boldsymbol{M}_N + \rho_0 \boldsymbol{R}_N) vec(Y_N) + \boldsymbol{Z}_N \beta_0 + \boldsymbol{\alpha}_{n,o,0} \otimes l_n + l_n \otimes \boldsymbol{\alpha}_{n,d,0} + vec(\epsilon_N), \tag{7}$$

where $\boldsymbol{Z}_N = \left[vec\left(Z_{N,1}\right), \cdots, vec\left(Z_{N,L}\right)\right]$. The reduced form of (7) is

$$vec(Y_N) = S_N^{-1} \left(\boldsymbol{Z}_N \beta_0 + \boldsymbol{\alpha}_{n,o,0} \otimes l_n + l_n \otimes \boldsymbol{\alpha}_{n,d,0} + vec(\epsilon_N)\right).$$

To effectively estimate ω_0, we need to remove the incidental parameters $\boldsymbol{\alpha}_{n,o,0}$ and $\boldsymbol{\alpha}_{n,d,0}$ from the log-likelihood function. The log-likelihood function for estimating ω_0 is

$$\begin{aligned} \ln L_N\left(\omega, \boldsymbol{\alpha}_{n,o}, \boldsymbol{\alpha}_{n,d}\right) = &-\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 + \ln \det\left(S_N(\delta)\right) \\ &-\frac{1}{2\sigma^2} vec\left(\epsilon_N^+\left(\omega, \boldsymbol{\alpha}_{n,o}, \boldsymbol{\alpha}_{n,d}\right)\right)' vec\left(\epsilon_N^+\left(\omega, \boldsymbol{\alpha}_{n,o}, \boldsymbol{\alpha}_{n,d}\right)\right), \end{aligned}$$

where $vec\left(\epsilon_N^+\left(\omega, \boldsymbol{\alpha}_{n,o}, \boldsymbol{\alpha}_{n,d}\right)\right) = vec\left(\epsilon_N^+(\omega)\right) - \boldsymbol{\alpha}_{n,o} \otimes l_n - l_n \otimes \boldsymbol{\alpha}_{n,d}$ and $vec\left(\epsilon_N^+(\omega)\right) = S_N(\delta) vec(Y_N) - \sum_{l=1}^L \beta_l vec\left(Z_{N,l}\right)$. Observe that $\boldsymbol{\alpha}_{n,o}$ and $\boldsymbol{\alpha}_{n,d}$ are linear parameters. By the first-order conditions, we then obtain

$$\widehat{\boldsymbol{\alpha}}_{n,o}\left(\omega, \boldsymbol{\alpha}_{n,d}\right) = \frac{1}{n}(I_n \otimes l_n')\left(vec\left(\epsilon_N^+(\omega)\right) - l_n \otimes \boldsymbol{\alpha}_{n,d}\right), \text{ and } J_n \widehat{\boldsymbol{\alpha}}_{n,d}(\omega) = \frac{1}{n}(l_n' \otimes J_n) vec\left(\epsilon_N^+(\omega)\right).$$

Note that we need to impose a normalization restriction $J_n \widehat{\boldsymbol{\alpha}}_{n,d}(\omega) = \widehat{\boldsymbol{\alpha}}_{n,d}(\omega)$ for identification.[20] Hence, $vec\left(\epsilon_N^+\left(\omega, \widehat{\boldsymbol{\alpha}}_{n,o}\left(\omega, \widehat{\boldsymbol{\alpha}}_{n,d}(\omega)\right), \widehat{\boldsymbol{\alpha}}_{n,d}(\omega)\right)\right) = (J_n \otimes J_n) vec\left(\epsilon_N^+(\omega)\right)$. Then, the concentrated log-likelihood function for estimating ω is

$$\ln L_N(\omega) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 + \ln \det\left(S_N(\delta)\right) - \frac{1}{2\sigma^2} vec\left(\epsilon_N^+(\omega)\right)' (J_n \otimes J_n) vec\left(\epsilon_N^+(\omega)\right)$$

---

[19] Also, equation (6) is an extension of the gravity equation with the two-way fixed effects and assuming $\lambda_0 = \gamma_0 = \rho_0 = 0$ (see Chapter 5 of Feenstra (2003)).

Under specification (6), identifying the sensitivity effects for $x_{n,i}$ and $x_{n,j}$ will not be possible via the estimation of (6), but however, it can be done by a two-step method (Hausman and Taylor, 1981). We might estimate the coefficients of $x_{n,i}$ by a regression of estimated constant term α_0 on x_n if x_n were exogenous, but by an IV approach in the presence of valid IVs for x_n when x_n is endogenous.

[20] Under the restriction $J_n \widehat{\boldsymbol{\alpha}}_{n,d}(\omega) = \widehat{\boldsymbol{\alpha}}_{n,d}(\omega)$, we have

$$\widehat{\boldsymbol{\alpha}}_{n,o}\left(\omega, \widehat{\boldsymbol{\alpha}}_{n,d}(\omega)\right) = \frac{1}{n}(I_n \otimes l_n')\left(vec\left(\epsilon_N^+(\omega)\right) - l_n \otimes \widehat{\boldsymbol{\alpha}}_{n,d}(\omega)\right) = \frac{1}{n}(I_n \otimes l_n') vec\left(\epsilon_N^+(\omega)\right).$$

with $\widehat{\omega}_N = \underset{\omega \in \Theta_\omega}{\operatorname{argmax}} \ln L_N(\omega)$, where Θ_ω denotes a compact parameter space for ω.

Since $E\left(\frac{1}{N}\frac{\partial \ln L_N(\omega_0)}{\partial \omega}\right) \to \mathbf{0}$ as $n \to \infty$, we have consistency of $\widehat{\omega}_N$. However, the direct estimation approach leads to an asymptotic bias of $\widehat{\omega}_N$ since $E\left(\frac{1}{\sqrt{N}}\frac{\partial \ln L_N(\omega_0)}{\partial \omega}\right) = \Lambda_N \neq \mathbf{0}$ for some $\Lambda_N = O(1)$. Hence, a bias correction for $\widehat{\omega}_N$ is needed. In Section 4.2.1, we introduce the asymptotic properties of $\widehat{\omega}_N$ and a bias correction method by analytically evaluating the form of Λ_N.

3. SARF Tobit models

In some application, a flow variable matrix Y_N contains many zero values. For example, a flow outcome between two regions can less occur due to some budgetary reasons if cross-section units are small. Also, an origin-destination flow $y_{n,ij}$ is a gross flow, which is necessarily nonnegative. This section extends the linear SARF model to a model with the Tobit structure (see Tobin (1958)).

We consider the simultaneous SAR Tobit model for a flow variable (hereafter, SARF Tobit). Refer to Qu and Lee (2012), Xu and Lee (2015b), Xu and Lee (2018) for univariate SAR Tobit models. For an $N \times 1$ real vector $\boldsymbol{x} = (x_1, \cdots . x_N)'$, let $F(\boldsymbol{x}) = \big(\max(0, x_1), \cdots, \max(0, x_N)\big)'$.[21] Observe that $F(\cdot)$ is a non-decreasing, convex, and Lipschitz function (since $|F(x_1) - F(x_2)| \leq |x_1 - x_2|$). The SARF Tobit model equation is

$$vec(Y_N) = F(\boldsymbol{A}_N vec(Y_N) + \mathbf{X}_N \kappa_0 + vec(\epsilon_N)),$$

where $\mathbf{X}_N = [\boldsymbol{x}_1', \cdots, \boldsymbol{x}_N']'$ with $\boldsymbol{x}_{n,ij} = \big(1, z_{n,ij,1}, \cdots, z_{n,ij,L}, x_{n,i,1}, \cdots, x_{n,i,K}, x_{n,j,1}, \cdots, x_{n,j,K}\big)$ for each (i,j) (i.e., $\boldsymbol{x}_{n,ij} = \boldsymbol{x}_{N,f}$ with $f = (j-1)n + i$ is the fth row of $\mathbf{X}_N$), and $\kappa_0 = (\alpha_0, \beta_0', b_0', c_0')'$. Using a scalar notation, the model can be written as

$$y_{n,ij} = F\big(\lambda_0 \textstyle\sum_{g=1}^{n} w_{n,ig} y_{n,gj} + \gamma_0 \sum_{h=1}^{n} y_{n,ih} m_{n,hj} + \rho_0 \sum_{g=1}^{n} \sum_{h=1}^{n} w_{n,ig} y_{n,gh} m_{n,hj} + \boldsymbol{x}_{n,ij} \kappa_0 + \epsilon_{n,ij}\big). \quad (8)$$

Or, $y_{n,ij} = 0$ if $y_{n,ij}^* \leq 0$ and $y_{n,ij} = y_{n,ij}^*$ if $y_{n,ij}^* > 0$, where $y_{n,ij}^*$ is the argument inside the $F(\cdot)$ above. Let $vec(Y_N^*)$ be an $N \times 1$ vector whose elements are inside of $F(\cdot)$ of the right-hand-side of (8). Then, the model can be rewritten as $vec(Y_N) = F\big(vec(Y_N^*)\big)$.

Now we consider conditions for model's stability and coherency. Recall that a model can generate a manageable covariance structure under spatial stability. Since the SARF Tobit model involves a nonlinear transformation $F(\cdot)$, a stability condition might be different from that of the linear SARF model. The model's coherency is required to guarantee a unique solution of a nonlinear equation system $vec(Y_N) = F\big(vec(Y_N^*)\big)$. The below assumption states a sufficient condition for the two issues.

Assumption 3.1. Assume that $S_N = I_N - \boldsymbol{A}_N$ is a strictly dominant diagonal matrix. Let $\zeta = \sup_n \|\boldsymbol{A}_N\|_\infty$, and it is assumed that $\zeta < 1$.

---

[21] As Xu and Lee (2015b) mentioned (see Section 2 in Xu and Lee (2015b)), this framework can be extended to the case that censoring points are known and nonzero.

Assumption 3.1 means that $\sum_{f'=1}^{N} \left| [\boldsymbol{A}_N]_{ff'} \right| < 1$ for all $f = 1, \cdots, N$ (Note that $\boldsymbol{A}_N$ has zero diagonal elements due to excluding self-influence). Then, $\|\boldsymbol{A}_N\|_\infty \leq \zeta < 1$ for some $0 \leq \zeta < 1$. Under Assumption 3.1, the system has a solution to a contraction mapping. The condition in Assumption 3.1 is also employed in asymptotic analysis. The detailed arguments can be found in the Appendix.

To derive the log-likelihood function, we rearrange a set of observations such that $vec(Y_N) = \begin{pmatrix} \boldsymbol{y}_{1,N_1} \\ \boldsymbol{y}_{2,N_2} \end{pmatrix}$, where the first N_1 observations in $\boldsymbol{y}_{1,N}$ are zeros while the remaining $N_2 = N - N_1$ observations are positive. Assume $\epsilon_{n,ij} \sim i.i.d.N(0, \sigma_0^2)$, so we have $\frac{\epsilon_{n,ij}}{\sigma_0} \sim i.i.d.N(0,1)$. For notational convenience, let

$$\epsilon_{n,ij}^*(\theta) = \left(y_{n,ij} - \lambda \sum_{g=1}^{n} w_{n,ig} y_{n,gj} - \gamma \sum_{h=1}^{n} y_{n,ih} m_{n,hj} - \rho \sum_{g=1}^{n} \sum_{h=1}^{n} w_{n,ig} y_{n,gh} m_{n,hj} - \boldsymbol{x}_{n,ij} \kappa \right) / \sigma$$

be the normalized residual evaluated at θ. Then, the log-likelihood function for estimation is

$$\begin{aligned} \ln L_N^*(\theta) = \sum_{i,j=1}^{n} 1\left(y_{n,ij} = 0\right) \ln \Phi\left(\epsilon_{n,ij}^*(\theta)\right) - \tfrac{1}{2} \ln 2\pi\sigma^2 \sum_{i,j=1}^{N} 1\left(y_{n,ij} > 0\right) + \ln\det\left(S_{N_2}^*(\delta)\right) \\ - \tfrac{1}{2} \sum_{i,j=1}^{n} 1\left(y_{n,ij} > 0\right) \left(\epsilon_{n,ij}^*(\theta)\right)^2, \end{aligned}$$

where $S_{N_2}^*(\delta)$ is the submatrix of $S_N(\delta)$ corresponding to $y_{n,ij} > 0$. For notational convenience, let $S_{N_2}^*(\delta) = I_{N_2} - \lambda \boldsymbol{W}_{22,N} - \gamma \boldsymbol{M}_{22,N} - \rho \boldsymbol{R}_{22,N}$ where $\boldsymbol{W}_{22,N}$, $\boldsymbol{M}_{22,N}$, and $\boldsymbol{R}_{22,N}$ are respectively the submatrices of $\boldsymbol{W}_N$, $\boldsymbol{M}_N$, and $\boldsymbol{R}_N$ corresponding to positive flows. Note that the number of positive flows (i.e., elements of $\boldsymbol{y}_{2,N_2}$) and their positions $\boldsymbol{W}_{22,N}$, $\boldsymbol{M}_{22,N}$, and $\boldsymbol{R}_{22,N}$ in the whole spatial relations in a sample are stochastic.

The asymptotic properties of $\hat{\theta}_N$ rely on stochastic properties of $\ln\det\left(S_{N_2}^*(\delta)\right)$ and its derivatives with respect to δ. Define $G_N(Y_N) = diag_{f=1}^{N} 1\left(y_{n,ij} > 0 \text{ with } f = (j-1)n + i\right)$, $\widetilde{\boldsymbol{W}}_N = G_N(Y_N) \boldsymbol{W}_N G_N(Y_N)$, $\widetilde{\boldsymbol{M}}_N = G_N(Y_N) \boldsymbol{M}_N G_N(Y_N)$, and $\widetilde{\boldsymbol{R}}_N = G_N(Y_N) \boldsymbol{R}_N G_N(Y_N)$. For $f = 1, \cdots, N$ and for each δ, let $r_{N,f,\lambda}(\delta) = \left[\widetilde{\boldsymbol{W}}_N \tilde{S}_N^{-1}(\delta)\right]_{ff}$, $r_{N,f,\gamma}(\delta) = \left[\widetilde{\boldsymbol{M}}_N \tilde{S}_N^{-1}(\delta)\right]_{ff}$, and $r_{N,f,\rho}(\delta) = \left[\widetilde{\boldsymbol{R}}_N \tilde{S}_N^{-1}(\delta)\right]_{ff}$, where $\tilde{S}_N(\delta) = I_N - \lambda \widetilde{\boldsymbol{W}}_N - \gamma \widetilde{\boldsymbol{M}}_N - \rho \widetilde{\boldsymbol{R}}_N$. Then, we have

$$\ln\det\left(S_{N_2}^*(\delta)\right) = -\sum_{l=1}^{\infty} \left(\frac{tr\left(\left(G_N(Y_N) \boldsymbol{A}_N(\delta) G_N(Y_N)\right)^l\right)}{l} \right),$$

$$tr\left(\boldsymbol{W}_{22,N} {S_{N_2}^*}^{-1}(\delta)\right) = tr\left(\widetilde{\boldsymbol{W}}_N \tilde{S}_N^{-1}(\delta)\right) = \sum_{f=1}^{N} r_{N,f,\lambda}(\delta) = \sum_{i,j=1}^{n} r_{n,ij,\lambda}(\delta),$$

$$tr\left(\boldsymbol{M}_{22,N} {S_{N_2}^*}^{-1}(\delta)\right) = tr\left(\widetilde{\boldsymbol{M}}_N \tilde{S}_N^{-1}(\delta)\right) = \sum_{f=1}^{N} r_{N,f,\gamma}(\delta) = \sum_{i,j=1}^{n} r_{n,ij,\gamma}(\delta), \text{and}$$

$$tr\left(\boldsymbol{R}_{22,N} {S_{N_2}^*}^{-1}(\delta)\right) = tr\left(\widetilde{\boldsymbol{R}}_N \tilde{S}_N^{-1}(\delta)\right) = \sum_{f=1}^{N} r_{N,f,\rho}(\delta) = \sum_{i,j=1}^{n} r_{n,ij,\rho}(\delta).$$

At δ_0, we define $r_{n,ij,\lambda} = r_{n,ij,\lambda}(\delta_0)$, $r_{n,ij,\gamma} = r_{n,ij,\gamma}(\delta_0)$, and $r_{n,ij,\rho} = r_{n,ij,\rho}(\delta_0)$ for each (i,j).

If one wants to control unobservables via the fixed-effect specification, the following log-likelihood function can be utilized:

$$\ln L_N^*(\omega, \boldsymbol{\alpha}_{n,o}, \boldsymbol{\alpha}_{n,d}) = \sum_{i,j=1}^{n} 1(y_{n,ij} = 0) \ln \Phi\left(\epsilon_{n,ij}^{+,*}(\omega, \alpha_{j,o}, \alpha_{i,d})\right) - \frac{1}{2}\ln 2\pi\sigma^2 \sum_{i,j=1}^{n} 1(y_{n,ij} > 0)$$
$$+ \ln\det\left(S_{N_2}^*(\delta)\right) - \frac{1}{2}\sum_{i,j=1}^{n} 1(y_{n,ij} > 0)\left(\epsilon_{n,ij}^{+,*}(\omega, \alpha_{j,o}, \alpha_{i,d})\right)^2 - \frac{\mu}{2}\left(\sum_{j=1}^{n} \alpha_{j,o} - \sum_{i=1}^{n} \alpha_{i,d}\right)^2,$$

where $\epsilon_{n,ij}^{+,*}(\omega, \alpha_{j,o}, \alpha_{i,d}) = \frac{y_{n,ij} - \lambda \sum_{g=1}^{n} w_{n,ig} y_{n,gj} - \gamma \sum_{h=1}^{n} y_{n,ih} m_{n,hj} - \rho \sum_{g,h=1}^{n} w_{n,ig} y_{n,gh} m_{n,hj} - \mathbf{z}_{n,ij}\beta - \alpha_{j,o} - \alpha_{i,d}}{\sigma}$ and μ is an arbitrary positive constant. The penalty term $-\frac{\mu}{2}\left(\sum_{j=1}^{n} \alpha_{j,o} - \sum_{i=1}^{n} \alpha_{i,d}\right)^2$ imposes $\sum_{j=1}^{n} \alpha_{j,o} = \sum_{i=1}^{n} \alpha_{i,d}$ for identification.[22] This setting is consistent with that of nonlinear panel models with individual and time fixed effects with the large n and T setting (Fernandez-Val and Weidner, 2016).[23] Compared to the linear SARF model, the fixed-effect components $\{\alpha_{i,o}\}$ and $\{\alpha_{j,d}\}$ are no longer linear parameters. We will study the asymptotic properties of the MLE in Subsection 4.3.1.

4. Asymptotic properties

In this section, we provide consistency and asymptotic normality of the MLE. Based on moment properties of $\{\epsilon_{n,ij}\}$, we study the QMLE's asymptotic properties for the linear SARF model. But later on, for the SARF Tobit model, the distribution of $\epsilon_{n,ij}$ will be assumed to be normal.[24]

4.1. Topological specification and regularity conditions

To establish the asymptotic properties of the MLE (QMLE) $\hat{\theta}_N$, we provide the topological specification for a cross-section unit i.

Assumption 4.1. In a sample, there exist n cross-section units. A cross-section unit i is located in a space $D_n \subset D$, which is a subset of $\mathbb{R}^d$ $(d \geq 1)$. We assume $\lim_{n\to\infty} \text{card}(D_n) = \infty$, where $\text{card}(D_n)$ is the cardinality of D_n. Let $d(i,j)$ be a distance between i and j. Assume $\min_{i,j} d(i,j) \geq 1$.

By Jenish and Prucha (2009, 2012), this setting was introduced to establish the stochastic properties of spatial mixing and spatial near-epoch dependent (NED) processes. The set D is an irregular lattice containing all potential locations of cross-section units $\{i\}$.[25] Then, we define the location function, $i \mapsto l(i) \in D$ for any i, and $d(i,j) = \|l(i) - l(j)\|_\infty$. The minimum distance assumption leads to avoiding an extreme influence between two cross-section units. Based on Assumption 4.1, the next step

---

[22] Among possible restrictions, we follow the Fernandez-Val and Weidner's (2016) setting. We solve an unconstrained optimization problem with keeping smoothness of the statistical objective function $\ln L_N^*(\omega, \boldsymbol{\alpha}_N)$. One can impose an alternative restriction on $\{\alpha_{i,o}\}$ and $\{\alpha_{j,d}\}$. In the linear SARF model, for example, recall that we need to impose a restriction on $\boldsymbol{\alpha}_{n,d}$ for identification, i.e., $\sum_{i=1}^{n} \alpha_{i,d} = 0$. The same restriction on the time fixed effects is imposed in Lee and Yu (2010).

[23] The incidental parameter problem with the large-T panel data becomes an asymptotic bias problem since the order of bias is $O\left(\frac{n+T}{nT}\right) = o(1)$ when n and T are both large. In our case, the same logic is applied, i.e., the order of bias $=$ $O\left(\frac{\#\text{ of incidental parameters}=2n}{\#\text{ of observations}=n^2}\right) = O\left(\frac{1}{n}\right)$. Relevant reviews can be found in Arellano and Hahn (2007) and Fernandez-Val and Weidner (2018).

[24] For a parametric model, a specific distribution will provide a proper parameter censoring probability. It might be possible to use a nonparametric sieve approach without the normality assumption as in Xu and Lee (2018). But we shall leave that nonparametric approach in this paper because of its complexity in theory and estimation.

[25] The space D can be a combination of geographic/demographic/economic spaces (i.e., characteristic space for a unit i).

is to characterize a distance measure for the two flow outcomes. In contrast to a traditional spatial econometric model, a flow outcome $y_{n,ij}$ involves the two cross-section units i and j. Then, a flow (i,j) can be located at a product space $D \times D$, which is a subspace of $\mathbb{R}^{2d}$.

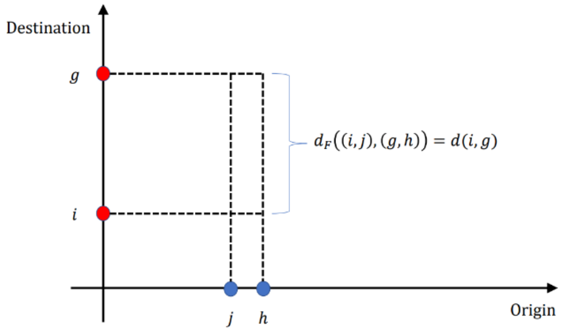
Using $d(i,j)$, we define the distance function for flows (i,j) and (g,h):

$$d_F\big((i,j),(g,h)\big) = \max\{d(i,g), d(j,h)\}.$$

That is, $d_F(\cdot,\cdot)$ takes the maximum value of the distance between origins and that between destinations. This metric satisfies the basic properties: (1) identity of indiscernibles, (2) symmetry, and (3) subadditivity. By the maximum norm's property with Assumption 4.1, $d_F\big((i,j),(g,h)\big) \geq 1$ when $(i,j) \neq (g,h)$. Then, Jenish and Prucha's (2009) Lemma A.1 implies $\operatorname{card}\big(\{(g,h): d_F\big((i,j),(g,h)\big) \leq m\}\big) \leq Cm^{2d}$ for some constant $C > 0$, i.e, there exists an upper bound of the number of flow units around arbitrary (i,j). The purpose of introducing this metric is to generate a device that $Cov\big(q_{n,ij}, q_{n,gh}\big) \to 0$ as $d_F\big((i,j),(g,h)\big) \to \infty$, where $q_{n,ij}$ and $q_{n,gh}$ are respectively components of a random field originated from pairs (i,j) and (g,h). This device will be employed if a statistic $q_{n,ij}$ is a nonlinear function of $\{\epsilon_{n,ij}\}$ (SARF Tobit model case).[26] The remark below illustrates the idea of this metric specification for flows.

Remark. Consider the covariance $Cov\big(y_{n,ij}, y_{n,gh}\big)$ for (i,j), $(g,h) \in D \times D$. The figure below illustrates the topological specification for flows when each cross-section unit is located in $\mathbb{R}$.

Figure 3. Topological specification for flows



If $\rho_0 = -\lambda_0\gamma_0$, we have the separable spatial filter, i.e., $S_N = \big(I_N - \lambda_0(I_n \otimes W_n)\big)\big(I_N - \gamma_0(M_n' \otimes I_n)\big)$. As we mentioned, this case can highlight the roles of W_n and M_n with intuitive manners. Under spatial stability, we have

$$\begin{aligned} Cov\big(y_{n,ij}, y_{n,gh}\big) &= \sigma_0^2\big(e_{n,j}' \otimes e_{n,i}'\big)S_N^{-1}S_N^{-1\prime}\big(e_{n,h} \otimes e_{n,g}\big) \\ &= \sigma_0^2 \underbrace{\Big(\textstyle\sum_{p_1=0}^{\infty}\sum_{p_2=0}^{\infty}\lambda_0^{p_1+p_2}\big[W_n^{p_1}W_n^{p_2\prime}\big]_{ig}\Big)}_{=sum\ A}\underbrace{\Big(\textstyle\sum_{q_1=0}^{\infty}\sum_{q_2=0}^{\infty}\gamma_0^{q_1+q_2}\big[M_n^{q_1}M_n^{q_2\prime}\big]_{hj}\Big)}_{=sum\ B}. \end{aligned}$$

Observe that the two infinite sums are well-defined under the spatial stability condition. Under the NED

[26] Refer to Proposition 1 in Xu and Lee (2019). For the linear SARF model case, $q_{n,ij}$ takes a linear-quadratic form of $\{\epsilon_{n,ij}\}$. Then, a martingale difference central limit theory can be applied. See the supplement file.

framework for a univariate SAR model, the first summation becomes smaller when $d(i,g)$ increases. It implies that $Cov(y_{n,ij}, y_{n,gj})$ decreases when $d(i,g)$ becomes larger. The second summation can be also interpreted similarly. By this setting, we then verify that $Cov(y_{n,ij}, y_{n,gh})$ becomes smaller when $d(i,g)$ or $d(j,h)$ increases. ■

Here are additional regularity assumptions for asymptotic analyses.

Assumption 4.2. (i) Denote $c_{w,c} = \sup_n \|W_n\|_1$, $c_{m,c} = \sup_n \|M_n\|_1$, $c_{w,r} = \sup_n \|W_n\|_\infty$, and $c_{m,r} = \sup_n \|M_n\|_\infty$. The sequences $\{W_n\}$ and $\{M_n\}$ satisfy $\max\{c_{w,c}, c_{m,c}, c_{w,r}, c_{m,r}\} < \infty$, i.e., they are uniformly bounded in both row and column sum norms.

(ii) Θ_δ denotes a compact parameter space for δ. We assume δ_0 belongs to the interior of Θ_δ. The sequence $\{S_N^{-1}(\delta)\}$ satisfies $\max_{\delta\in\Theta_\delta}\{\sup_n \|S_N^{-1}(\delta)\|_\infty, \sup_n \|S_N^{-1}(\delta)\|_1\} < \infty$.

(iii) $w_{n,ij}$ and $m_{n,ij}$ satisfy one of the two conditions:

(iii-1) $w_{n,ij} > 0$ and $m_{n,ij} > 0$ only if $d(i,j) \le \bar{d}$ for some $\bar{d} > 1$; otherwise $w_{n,ij} = 0$.

(iii-2) $w_{n,ij} \le \frac{C_0}{d(i,j)^a}$ and $m_{n,ij} \le \frac{C_0}{d(i,j)^a}$ for some $C_0 > 0$ and $a > 2d$. In this case, we assume $|\lambda_0| c_{w,r} + |\gamma_0| c_{m,r} + |\rho_0| c_{w,r} c_{m,r} \le \zeta$; if $c_{w,c} > c_{w,r}$, there exist at most K_W ($K_W \ge 1$) columns of W_n that the column sum exceeds $c_{w,r}$, where K_W is a constant that does not rely on n.[27]

Assumption 4.3. (i) Elements of $\mathbf{X}_N$ have uniformly bounded constants.

Or, if one wants to assume that $\mathbf{X}_N$ is stochastic, $\max_{k=1,\cdots,2K+L+1} \sup_{n.ij} E|\boldsymbol{x}_{n,ij,k}|^{4+\eta} < \infty$ for some $\eta > 0$; and, $\boldsymbol{X}_N$ and ϵ_N are independent.

(ii) $\lim_{n\to\infty} \frac{1}{N}\mathbf{X}_N'\mathbf{X}_N$ or $\operatorname{plim}_{n\to\infty} \frac{1}{N}\mathbf{X}_N'\mathbf{X}_N$ exists and is nonsingular.

Assumption 4.4. The parameter space Θ of θ is compact. The true value θ_0 belongs to the interior of Θ.

Most assumptions are traditional, but the condition in Assumption 4.2 (iii-2) is introduced to characterize the maximum column sum (and those of its powers) of $\boldsymbol{A}_N$ for the SARF Tobit model if $w_{n,ij}$ and $m_{n,ij}$ are geometric decaying functions of $d(i,j)$. Under this condition, we obtain $\|\boldsymbol{A}_N^l\|_1 \le lK\Gamma\zeta^{l-1}$ for $l \in \mathbb{Z}_+$, where K is a positive integer that does not depend on n (see Lemma C.1).[28] It implies that $\sum_{l=1}^\infty \|\boldsymbol{A}_N^l\|_1 \le K\Gamma \sum_{l=1}^\infty l\zeta^{l-1} < \infty$. Assumption 4.3 gives regularity conditions for exogenous variables $\{\boldsymbol{x}_{n,ij,k}\}$, which provide guidance of generating a z-variable from an x-variable (or generating an x-variable from a z-variable). When a practitioner generates $z_{n,ij} = \frac{1}{|x_{n,i,k} - x_{n,j,k}|}$ for some k, sufficient cross-section variations of $\{x_{n,i,k}\}$ are required to avoid extremely large value of $z_{n,ij}$. If $z_{n,ij} = \frac{x_{n,j,k}}{x_{n,i,k}}$ for some k is defined, $\{x_{n,i,k}\}$ should be bounded away from zero. On the other hand, if one generates an x-variable from a z-variable, it involves summation, i.e., $x_{n,i,l} = \sum_{j=1}^n z_{n,ij,l}$ for some l. To satisfy the regularity conditions, for example, one can assume that $\sum_{j=1}^n z_{n,ij,l}$ is uniformly bounded

---

27 Recall that $\|\boldsymbol{A}_N\|_\infty \le |\lambda_0| c_{w,r} + |\gamma_0| c_{m,c} + |\rho_0| c_{w,r} c_{m,c} \le \zeta < 1$ by Assumption 3.1. Hence, the additional condition implies $\max\{|\lambda_0| c_{w,r} + |\gamma_0| c_{m,c} + |\rho_0| c_{w,r} c_{m,c}, |\lambda_0| c_{w,r} + |\gamma_0| c_{m,r} + |\rho_0| c_{w,r} c_{m,r}\} \le \zeta$.

28 To regulate $\|\boldsymbol{A}_N^l\|_1$ for $l \in \mathbb{Z}_+$, it suffices to introduce the column sum restriction on W_n provided in Assumption 4.2 (iii-2) since an upper bound of $\|\boldsymbol{A}_N^l\|_1$ can be characterized by $\|W_n^{p+r}\|_1$ and $\|M_n\|_\infty^{q+r}$ where $p + q + r = l$. The column sum restriction on W_n is employed to have an upper bound of $\|W_n^{p+r}\|_1$ (see Lemma C.1).

in i and n (if $\{z_{n,ij,l}\}$ are non-stochastic).

4.2. Asymptotic properties of the QMLE of the linear SARF model

In this subsection, we study the asymptotic properties of QMLE for the linear SARF model. Detailed proofs for this subsection can be found in the supplement file. Let $Q_N(\theta) = E\left(\frac{1}{N}\ln L_N(\theta)\right)$ for each $\theta \in \Theta$. For consistency, we establish the uniform convergence of $\frac{1}{N}\ln L_N(\theta) - Q_N(\theta)$ to zero on Θ, and uniform equicontinuity of $\{Q_N(\theta)\}$ on Θ. The two objects can be similarly verified by the traditional techniques for linear SAR models (see Theorem 3.1 in Lee (2004)). The assumption below is a regularity condition for disturbances $\{\epsilon_{n,ij}\}$.

Assumption 4.5. $\epsilon_{n,ij} \sim i.i.d.(0, \sigma_0^2)$ with $\sigma_0^2 > 0$ across pairs (i,j), and $\sup_n E\left|\epsilon_{n,ij}\right|^{4+\eta} < \infty$ for some $\eta > 0$.

The identification uniqueness condition finalizes the argument for consistency. Here are sufficient conditions for identification derived by the information inequality.

Assumption 4.6 (Identification for the linear SARF model). At least, one of the two conditions holds: (i) $S_N^{-1\prime} S_N'(\delta) S_N(\delta) S_N^{-1}$ is not proportional to I_N when $\delta \neq \delta_0$.

(ii) Let $\mathbf{G}_\lambda = \boldsymbol{W}_N S_N^{-1}$, $\mathbf{G}_\gamma = \boldsymbol{M}_N S_N^{-1}$, and $\mathbf{G}_\rho = \boldsymbol{R}_N S_N^{-1}$. Then,

$$\lim_{n\to\infty} \frac{1}{N}\left[\mathbf{G}_\lambda \mathbf{X}_N \kappa_0, \mathbf{G}_\gamma \mathbf{X}_N \kappa_0, \mathbf{G}_\rho \mathbf{X}_N \kappa_0\right]' M_{\mathbf{X}_N}\left[\mathbf{G}_\lambda \mathbf{X}_N \kappa_0, \mathbf{G}_\gamma \mathbf{X}_N \kappa_0, \mathbf{G}_\rho \mathbf{X}_N \kappa_0\right]$$

exists and is nonsingular.

The first identification condition comes from the model's correlation structure. The second identification condition guarantees sufficient variations in the generated regressors $\left[\mathbf{G}_\lambda \mathbf{X}_N \kappa_0, \mathbf{G}_\gamma \mathbf{X}_N \kappa_0, \mathbf{G}_\rho \mathbf{X}_N \kappa_0, \mathbf{X}_N\right]$ (see Assumption 8 in Lee (2004)). Then, we have consistency of $\hat{\theta}_N$.

Theorem 4.1 (Consistency). Under Assumptions 4.1, 4.2 (i), (ii), 4.3 - 4.6, $\hat{\theta}_N \xrightarrow{p} \theta_0$.

The asymptotic distribution of $\hat{\theta}_N$ can be obtained by the Taylor expansion argument: $\sqrt{N}(\hat{\theta}_N - \theta_0) = \left(-\frac{1}{N}\frac{\partial^2 \ln L_N(\tilde{\theta}_N)}{\partial\theta\partial\theta'}\right)^{-1}\frac{1}{\sqrt{N}}\frac{\partial \ln L_N(\theta_0)}{\partial\theta}$, where $\tilde{\theta}_N$ lies between $\hat{\theta}_N$ and θ_0. To have well-definedness of the asymptotic distribution, we introduce the following assumption.

Assumption 4.7. $\Sigma_{\theta_0} = \lim_{n\to\infty} \Sigma_{\theta_0,N}$ is nonsingular where $\Sigma_{\theta_0,N} = E\left(-\frac{1}{N}\frac{\partial^2 \ln L_N(\theta_0)}{\partial\theta\partial\theta'}\right)$.

To finish deriving the asymptotic distribution of $\hat{\theta}_N$, observe that $\frac{1}{\sqrt{N}}\frac{\partial \ln L_N(\theta_0)}{\partial\theta}$ is a summation of martingale differences of a linear-quadratic form. By extending the Kelejian and Prucha's (2001) framework (see Section 1.2 in the supplement file), we derive $\frac{1}{\sqrt{N}}\frac{\partial \ln L_N(\theta_0)}{\partial\theta} \xrightarrow{d} N(\mathbf{0}, \Omega_{\theta_0})$, where $\Omega_{\theta_0} =$

$\lim_{n\to\infty} \Omega_{\theta_0,N}$ and $\Omega_{\theta_0,N} = E\left(\frac{1}{N}\frac{\partial \ln L_N(\theta_0)}{\partial \theta}\frac{\partial \ln L_N(\theta_0)}{\partial \theta'}\right)$. Note that Ω_{θ_0} depends on the 3rd and 4th moments of $\epsilon_{n,ij}$ (see the supplement file). When $\epsilon_{n,ij}$s follow the normal distribution (Assumption 4.3 (ii)), $\Omega_{\theta_0} = \Sigma_{\theta_0}^{-1}$. By applying the Slutsky's lemma, we obtain the following result.

Theorem 4.2 (Asymptotic normality). Under Assumptions 4.1, 4.2 (i), (ii), 4.3 - 4.6, and 4.7, we have $\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(\mathbf{0}, \Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1})$.

4.2.1. Asymptotic distribution of the QMLE for the linear SARF model under the two-way fixed effect specification

In this part, we study the asymptotic properties of $\widehat{\omega}_N$ if one estimates the model (6). First, observe that the first-order conditions at ω_0 are

$$\begin{pmatrix} \frac{\partial \ln L_N(\omega_0)}{\partial \lambda} \\ \frac{\partial \ln L_N(\omega_0)}{\partial \gamma} \\ \frac{\partial \ln L_N(\omega_0)}{\partial \rho} \\ \frac{\partial \ln L_N(\omega_0)}{\partial \beta} \\ \frac{\partial \ln L_N(\omega_0)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} -tr(\boldsymbol{W}_N S_N^{-1}) + \frac{1}{\sigma_0^2}\left(\boldsymbol{W}_N vec(Y_N)\right)'(J_n \otimes J_n) vec(\epsilon_N) \\ -tr(\boldsymbol{M}_N S_N^{-1}) + \frac{1}{\sigma_0^2}\left(\boldsymbol{M}_N vec(Y_N)\right)'(J_n \otimes J_n) vec(\epsilon_N) \\ -tr(\boldsymbol{R}_N S_N^{-1}) + \frac{1}{\sigma_0^2}\left(\boldsymbol{R}_N vec(Y_N)\right)'(J_n \otimes J_n) vec(\epsilon_N) \\ \frac{1}{\sigma_0^2}\boldsymbol{Z}_N'(J_n \otimes J_n) vec(\epsilon_N) \\ -\frac{N}{2\sigma_0^2} + \frac{1}{2\sigma_0^4} vec(\epsilon_N)'(J_n \otimes J_n) vec(\epsilon_N) \end{pmatrix}.$$

Then, a component $\frac{1}{\sqrt{N}}\frac{\partial \ln L_N(\omega_0)}{\partial \omega}$ takes a LQ form: for $l = \lambda,\ \gamma,\ \rho,\ \beta_1,\ \cdots,\ \beta_L,$ and σ^2,

$$\frac{1}{\sqrt{N}} vec(\mathbf{C}_l)'(J_n \otimes J_n) vec(\epsilon_N) + \frac{1}{\sqrt{N}}\frac{1}{\sigma_0^2}\left(vec(\epsilon_N)'\mathbf{G}_l'(J_n \otimes J_n) vec(\epsilon_N) - \sigma_0^2 tr(\mathbf{G}_l)\right),$$

where $\mathbf{G}_{\beta_k} = \mathbf{0}$ for $k = 1, \cdots, L$, $\mathbf{G}_{\sigma^2} = \frac{1}{2\sigma_0^2} I_N$, $vec(\mathbf{C}_l) = \mathbf{G}_l\left(\boldsymbol{Z}_N \beta_0 + \boldsymbol{\alpha}_{n,o,0} \otimes l_n + l_n \otimes \boldsymbol{\gamma}_{n,d,0}\right)$ for $l = \lambda$, γ, and ρ, and $vec(\mathbf{C}_l) = \boldsymbol{Z}_{N,l}$ for β_1, $\cdots$, β_L, and $vec(\mathbf{C}_l) = \mathbf{0}$ for $l = \sigma^2$. Observe that $E\left(\frac{1}{\sqrt{N}} vec(\mathbf{C}_l)'(J_n \otimes J_n) vec(\epsilon_N)\right) = 0$ but $E\left(\frac{1}{\sqrt{N}}\frac{1}{\sigma_0^2}\left(vec(\epsilon_N)'\mathbf{G}_l'(J_n \otimes J_n) vec(\epsilon_N) - \sigma_0^2 tr(\mathbf{G}_l)\right)\right) = \frac{1}{\sqrt{N}}\left(tr\left((J_n \otimes J_n)\mathbf{G}_l\right) - tr(\mathbf{G}_l)\right) \neq 0$. We have $\frac{1}{\sqrt{N}}\frac{\partial \ln L_N(\omega_0)}{\partial \omega} = \frac{1}{\sqrt{N}}\frac{\partial \ln L_N^{(u)}(\omega_0)}{\partial \omega} - \Lambda_N$, where $\frac{1}{\sqrt{N}}\frac{\partial \ln L_N^{(u)}(\omega_0)}{\partial \omega}$ takes a form of
$\frac{1}{\sqrt{N}} vec(\mathbf{C}_l)'(J_n \otimes J_n) vec(\epsilon_N) + \frac{1}{\sqrt{N}}\frac{1}{\sigma_0^2}\left(vec(\epsilon_N)'\mathbf{G}_l'(J_n \otimes J_n) vec(\epsilon_N) - \sigma_0^2 tr\left((J_n \otimes J_n)\mathbf{G}_l\right)\right)$ to be
$E\left(\frac{1}{\sqrt{N}}\frac{\partial \ln L_N^{(u)}(\omega_0)}{\partial \omega}\right) = \mathbf{0}$ and

$$\Lambda_N = \frac{1}{\sqrt{N}}\left(tr\left(\left(I_N - (J_n \otimes J_n)\right)\mathbf{G}_\lambda\right), tr\left(\left(I_N - (J_n \otimes J_n)\right)\mathbf{G}_\gamma\right), tr\left(\left(I_N - (J_n \otimes J_n)\right)\mathbf{G}_\rho\right), \mathbf{0}', \frac{2n-1}{2\sigma_0^2}\right)'.$$

Note that $\Lambda_N = O(1)$ because $N = n^2$, and $tr\left(\left(I_N - (J_n \otimes J_n)\right)\mathbf{G}_l\right) = O(n)$ where $l = \lambda$, γ, and ρ. Since $E\left(\frac{1}{N}\frac{\partial \ln L_N(\omega_0)}{\partial \omega}\right) = O\left(\frac{1}{n}\right) \to 0$ as $n \to \infty$, there is no problem in achieving consistency of $\widehat{\omega}_N$.

The remaining issue is a correction of the asymptotic bias in $\widehat{\omega}_N$. Let $\Lambda_N(\omega)$ be the asymptotic bias term evaluated at $\omega \in \Theta_\omega$. Then, the asymptotic distribution of $\widehat{\omega}_N$ can be characterized by

$$\sqrt{N}(\widehat{\omega}_N - \omega_0) + \Sigma_{\widehat{\omega}_N,N}^{-1}\Lambda_N = \Sigma_{\widehat{\omega}_N,N}^{-1} \frac{1}{\sqrt{N}} \frac{\partial \ln L_N^{(u)}(\omega_0)}{\partial \omega} \xrightarrow{d} N\left(\mathbf{0}, \Sigma_{\omega_0}^{-1}\Omega_{\omega_0}\Sigma_{\omega_0}^{-1}\right),$$

where $\Sigma_{\omega,N} = E\left(-\frac{1}{N}\frac{\partial^2 \ln L_N(\omega)}{\partial\omega\partial\omega'}\right)$, $\Sigma_{\omega_0} = \lim_{n\to\infty} \Sigma_{\omega_0,N}$, $\Omega_{\omega_0} = \lim_{n\to\infty} \Omega_{\omega_0,N}$ and $\Omega_{\omega,N} = E\left(\frac{1}{N}\frac{\partial \ln L_N(\omega)}{\partial\omega}\frac{\partial \ln L_N(\omega)}{\partial\omega'}\right)$. The bias corrected QMLE can then be specified by

$$\widehat{\omega}_N^c = \widehat{\omega}_N + \frac{1}{n}\Sigma_{\widehat{\omega}_N,N}^{-1}\Lambda_N(\widehat{\omega}_N),$$

so we have $\sqrt{N}(\widehat{\omega}_N^c - \omega_0) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_{\omega_0}^{-1}\Omega_{\omega_0}\Sigma_{\omega_0}^{-1}\right)$.

Using $\widehat{\omega}_N$, one can obtain $\widehat{\boldsymbol{\alpha}}_{n,d} = \widehat{\boldsymbol{\alpha}}_{n,d}(\widehat{\omega}_N) = J_n\widehat{\boldsymbol{\alpha}}_{n,d}(\widehat{\omega}_N)$ due to the identification restriction (i.e., $\sum_{i=1}^n \alpha_{i,d} = 0$) and $\widehat{\boldsymbol{\alpha}}_{n,o} = \widehat{\boldsymbol{\alpha}}_{n,o}\left(\widehat{\omega}_N, \widehat{\boldsymbol{\alpha}}_{n,d}\right) = \frac{1}{n}(I_n \otimes l_n')\left(vec\left(\epsilon_N^+(\widehat{\omega}_N)\right) - l_n \otimes \widehat{\boldsymbol{\alpha}}_{n,d}\right)$. For each i, the asymptotic distribution of the ith element of $\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_{n,d} - \boldsymbol{\alpha}_{n,d,0}\right)$ is $\frac{1}{\sqrt{n}}\sum_{j=1}^n \epsilon_{n,ij} + o_p(1) \xrightarrow{d} N(0, \sigma_0^2)$ as $n \to \infty$. For $j = 1, \cdots, n$, the asymptotic distribution of the jth element of $\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_{n,o} - \boldsymbol{\alpha}_{n,o,0}\right)$ is $\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_{n,ij} + o_p(1) \xrightarrow{d} N(0, \sigma_0^2)$ as $n \to \infty$. Details can be found in the supplement file.

4.3. Asymptotic properties of the MLE of the SARF Tobit model

Note that the asymptotic analysis for the MLE $\widehat{\theta}_N$ will be based on normally distributed $\epsilon_{n,ij}$s since the SARF Tobit model comes from the distributional specification. Since $\widehat{\theta}_N$ is a highly nonlinear function of $\{\epsilon_{n,ij}\}$, we will employ the spatial near epoch dependence (NED) concept introduced by Jenish and Prucha (2012). Let $\|x\|_{L_p}$ be the L_p-norm of a random variable x.

Note that the NED concept relates two random fields. Let $q = \{q_{n,ij}: (i,j) \in D_n \times D_n, n \geq 1\}$ and $\epsilon = \{\epsilon_{n,ij}: (i,j) \in D_n \times D_n, n \geq 1\}$ be two random fields.[29] A random field q is L_p-NED on ϵ if $\sup_{n,i,j}\|q_{n,ij}\|_{L_p} < \infty$ and $\left\|q_{n,ij} - E\left(q_{n,ij}|\mathcal{F}_{n,ij}(s)\right)\right\|_{L_p} \leq c_{n,ij}v(s)$, where $p \geq 1$, $\mathcal{F}_{n,ij}(s) = \sigma\left(\epsilon_{n,gh}: d_F\left((i,j),(g,h)\right) \leq s\right)$, $\{c_{n,ij}: n \geq 1\}$ is an array of finite positive constants (NED scaling factor), and $v(s)$ is a sequence such that $v(s) \downarrow 0$ as $s \uparrow \infty$ (NED coefficient). Note that q is a uniform NED random field if $\sup_n \sup_{(i,j)\in D_n\times D_n} c_{n,ij} < \infty$; and q is a geometric random field if $v(s) = O(\tau^s)$ for some $0 < \tau < 1$.

To show the NED properties of $\{1(y_{n,ij} = 0)\}$, the normality assumption is used by showing the uniform boundedness of the essential supremum of $y_{n,ij}^*$'s densities (see Proposition 2 in Xu and Lee (2015b)), where $y_{n,ij}^* = \lambda_0 \sum_{g=1}^n w_{n,ig}y_{n,gj} + \gamma_0 \sum_{h=1}^n y_{n,ih}m_{n,hj} + \rho_0 \sum_{g=1}^n \sum_{h=1}^n w_{n,ig}y_{n,gh}m_{n,hj} + \boldsymbol{x}_{n,ij}\kappa_0 + \epsilon_{n,ij}$.

Assumption 4.8 (Normal distribution assumption on the disturbances). $\epsilon_{n,ij} \sim i.i.d. N(0, \sigma_0^2)$ across

[29] If $\{\boldsymbol{x}_{n,ij}\}$ is stochastic, $\epsilon = \{(\boldsymbol{x}_{n,ij}, \epsilon_{n,ij}): (i,j) \in D_n \times D_n, n \geq 1\}$.

pairs (i,j).

The main point of showing $\hat{\theta}_N \xrightarrow{p} \theta_0$ is uniform convergence of the sample average log-likelihood function, i.e., $\sup_{\theta\in\Theta} \frac{1}{N}[\ln L_N^*(\theta) - E(\ln L_N^*(\theta))] \xrightarrow{p} 0$. Observe that $\ln L_N^*(\theta)$ consists of $\{y_{n,ij}\}$ and its transformations on ϵ. Propositions C.1 and C.2 show the L_2-NED properties of them. Then, we apply the law of large numbers (LLN) for each $\theta \in \Theta$ (Theorem 1 in Jenish and Prucha (2012)) and the compact parameter space assumption (Assumption 4.4) finalizes the proof. The conditions below provide sufficient conditions of identification uniqueness based on Rothenberg (1971). The derivation can be found in the last step of consistency proof.

Assumption 4.9 (Identification for the SARF Tobit model). Assume Assumptions 3.1 and 4.3 (ii) hold.

(a) $I_n \otimes (W_n + W_n')$, $(M_n + M_n') \otimes I_n$, $M_n' \otimes W_n + M_n \otimes W_n'$, $M_n \otimes W_n + M_n \otimes W_n$, $(M_n + M_n') \otimes W_n' W_n$, and $M_n M_n' \otimes (W_n + W_n')$ are linearly independent.

(b) For all $g = 1, \cdots, N$, a set of vectors $\{\mathbb{w}_g^s, \mathbb{m}_g^s, \mathbb{r}_g^s\}$ is linearly independent, where $\mathbb{w}_g^s$, $\mathbb{m}_g^s$, and $\mathbb{r}_g^s$ are respectively $(N-1)\times 1$ vectors consisting of $\sum_{f=1}^N([\boldsymbol{W}_N]_{fg}^2 - [\boldsymbol{W}_N]_{fh}^2)$, $\sum_{f=1}^N([\boldsymbol{M}_N]_{fg}^2 - [\boldsymbol{M}_N]_{fh}^2)$, and $\sum_{f=1}^N([\boldsymbol{R}_N]_{fg}^2 - [\boldsymbol{R}_N]_{fh}^2)$ for $h \neq g$.

(c) $\mathbf{X}_N'\mathbf{X}_N$ is invertible with probability 1. Then, θ_0 is identified.

Condition (a) is for identifying δ_0, (b) is for σ_0^2, and κ_0 can be identified via (c). If needed, this identification condition can be replaced by a high-level assumption such as $\limsup_{n\to\infty}[Q_N^*(\theta) - Q_N^*(\theta_0)] < 0$, where $Q_N^*(\theta) = \frac{1}{N} E(\ln L_N^*(\theta))$ for each $\theta \neq \theta_0$.[30] Then, we have consistency.

Theorem 4.3 (Consistency). Under Assumptions 3.1, 4.1 – 4.4, 4.8 and 4.9, $\hat{\theta}_N \xrightarrow{p} \theta_0$.

Next, we consider the asymptotic distribution of $\hat{\theta}_N$. Note that the asymptotic distribution of $\hat{\theta}_N$ is mainly characterized by the score at θ_0. For each (i,j) and each θ, where we have $\frac{\partial \ln L_N^*(\theta)}{\partial\theta} = \sum_{i=1}^n \sum_{j=1}^n q_{n,ij}(\theta)$ with

$$q_{n,ij}(\theta) = \begin{pmatrix} -1(y_{n,ij}=0)\sigma^{-1}\frac{\phi\left(\epsilon_{n,ij}^*(\theta)\right)\left(\sum_{g=1}^n w_{n,ig}y_{n,gj}\right)}{\Phi\left(\epsilon_{N,f}^*(\theta)\right)} - r_{n,ij,\lambda}(\delta) + 1(y_{n,ij}>0)\sigma^{-1}\epsilon_{n,ij}^*(\theta)\left(\sum_{g=1}^n w_{n,ig}y_{n,gj}\right) \\ -1(y_{n,ij}=0)\sigma^{-1}\frac{\phi\left(\epsilon_{n,ij}^*(\theta)\right)\left(\sum_{h=1}^n y_{n,ih}m_{n,hj}\right)}{\Phi\left(\epsilon_{n,ij}^*(\theta)\right)} - r_{n,ij,\gamma}(\delta) + 1(y_{n,ij}>0)\sigma^{-1}\epsilon_{n,ij}^*(\theta)\left(\sum_{h=1}^n y_{n,ih}m_{n,hj}\right) \\ -1(y_{n,ij}=0)\sigma^{-1}\frac{\phi\left(\epsilon_{n,ij}^*(\theta)\right)\left(\sum_{g,h=1}^n w_{n,ig}y_{n,gh}m_{n,hj}\right)}{\Phi\left(\epsilon_{n,ij}^*(\theta)\right)} - r_{n,ij,\rho}(\delta) + 1(y_{n,ij}>0)\sigma^{-1}\epsilon_{n,ij}^*(\theta)\left(\sum_{g,h=1}^n w_{n,ig}y_{n,gh}m_{n,hj}\right) \\ -1(y_{n,ij}=0)\sigma^{-1}\frac{\phi\left(\epsilon_{n,ij}^*(\theta)\right)\boldsymbol{x}_{n,ij}}{\Phi\left(\epsilon_{n,ij}^*(\theta)\right)} + 1(y_{n,ij}>0)\sigma^{-1}\epsilon_{n,ij}^*(\theta)\boldsymbol{x}_{n,ij} \\ -\frac{1}{2\sigma^2}1(y_{n,ij}=0)\frac{\phi\left(\epsilon_{n,ij}^*(\theta)\right)\epsilon_{n,ij}^*(\theta)}{\Phi\left(\epsilon_{n,ij}^*(\theta)\right)} - \frac{1}{2\sigma^2}1(y_{n,ij}>0) + \frac{1}{2\sigma^2}1(y_{n,ij}>0)\epsilon_{n,ij}^*(\theta)^2 \end{pmatrix}.$$

The main issue is to check the NED properties of a random field $\{\|q_{n,ij}(\theta_0)\|\}$ to apply the CLT

[30] If both W_n and M_n are symmetric, $M_n' \otimes W_n + M_n \otimes W_n'$ and $M_n \otimes W_n + M_n \otimes W_n$ would not be linearly independent (Condition (a) in Assumption 4.9 is violated). In this case, we might need to introduce $\limsup_{n\to\infty}[Q_N^*(\theta) - Q_N^*(\theta_0)] < 0$ for $\theta \neq \theta_0$.

(Corollary 1 in Jenish and Prucha (2012)). The conditions below are introduced for this issue.

Assumption 4.10. (i) $\max_{k=1,\cdots,2K+L+1} \sup_{i,j,n} \|\boldsymbol{x}_{n,ij,k}\|_{L_{8+\eta}} < \infty$ for some $\eta > 0$.

(ii) $\{\boldsymbol{x}_{n,ij,k}\}$ is an α-mixing random field with spatial α-mixing coefficients $\alpha(u,v,r) \leq (u+v)^{\tau}\hat{\alpha}(r)$ for some $\tau \geq 0$; and for some $0 < \tilde{\eta} < 2 + \frac{\eta}{2}$, $\hat{\alpha}(r)$ satisfies $\sum_{r=1}^{\infty} r^{2d(\tau_*+1)-1}\,\hat{\alpha}(r)^{\frac{\tilde{\eta}}{4+2\tilde{\eta}}} < \infty$, where $\tau_* = \frac{\tilde{\eta}\tau}{2+\tilde{\eta}}$.

Assumption 4.11. $\Sigma_{\theta_0}^{*} = \lim_{n\to\infty} \Sigma_{\theta_0,N}^{*}$ is nonsingular and positive definite, where $\Sigma_{\theta_0,N}^{*} = \frac{1}{N} Var\left(\sum_{i=1}^{n}\sum_{j=1}^{n} q_{n,ij}(\theta_0)\right)$.

Assumption 4.12. $a > d \cdot \max(14 + 48\eta^{-1}, 10 + 64\eta^{-1} + 128\eta^{-2})$, where a is in (iii-2) of Assumption 4.2.

Assumption 4.10 are the same as Assumption 3 in Jenish and Prucha (2012). Here, the key is to have the uniform $L_{2+\tilde{\eta}}$-integrability. Assumption 4.11 is for well-definedness of the asymptotic variance. Assumption 4.12 is introduced for the specification in Assumption 4.2 (iii-2). To apply the CLT to an NED random field, we need to check the summability condition for the NED coefficient of $\{\|q_{n,ij}(\theta_0)\|\}$ (see Assumption 4(c) in Jenish and Prucha (2012)).[31] That is, $\sum_{s=1}^{\infty} s^{2d-1}v(s) < \infty$. Assumption 4.12 yields $v(s) = O(s^{-b})$ where $b > 2d$ satisfying $\sum_{s=1}^{\infty} s^{2d-1}v(s) \leq \sum_{s=1}^{\infty} s^{-1-\tau} < \infty$ where $\tau > 0$. Details for the CLT can be found in Propositions C.3 and C.4 in the Appendix.

Theorem 4.4 (Asymptotic normality). Suppose Assumptions 3.1, 4.1 – 4.4, 4.8, 4.9 and 4.11 hold. Under the specification provided in Assumption 4.2 (iii-2), Assumptions 4.10 and 4.12 are additionally required. Then, we have $\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_{\theta_0}^{*-1}\right)$ as $n \to \infty$.

4.3.1. Asymptotic distribution of the MLE for the SARF Tobit model under the two-way fixed effect specification

With a direct estimation approach, the MLE under the two-way fixed-effect specification is defined by

$$(\hat{\omega}_N, \hat{\boldsymbol{\alpha}}_N) = \text{argmax}_{\omega\in\Theta_\omega, \boldsymbol{\alpha}_N} \ln L_N^{*}(\omega, \boldsymbol{\alpha}_N),$$

where $\boldsymbol{\alpha}_N$ is a vector containing identifiable elements of $\boldsymbol{\alpha}_{n,o}$ and $\boldsymbol{\alpha}_{n,d}$. That is, $\boldsymbol{\alpha}_N = \left(\boldsymbol{\alpha}_{n,o}', \boldsymbol{\alpha}_{n,d}'\right)'$. Let $\boldsymbol{\alpha}_N^{0} = \left(\boldsymbol{\alpha}_{n,o,0}', \boldsymbol{\alpha}_{n,d,0}'\right)'$ be the true parameter vector. The main purpose of this subsection is to examine the asymptotic properties of $\hat{\omega}_N$. For this issue, we define $\hat{\boldsymbol{\alpha}}_N(\omega) = \text{argmax}_{\boldsymbol{\alpha}_N} \ln L_N^{*}(\omega, \boldsymbol{\alpha}_N)$ for each ω. Then, we have $\hat{\omega}_N = \text{argmax}_{\omega\in\Theta_\omega} \ln L_N^{*}(\omega)$, where $\ln L_N^{*}(\omega) = \ln L_N^{*}\left(\omega, \hat{\boldsymbol{\alpha}}_N(\omega)\right)$ denotes the concentrated log-likelihood function. For the discussion in this subsection, we will use the notations based on double indexes. For example, we denote $\ln L_N^{*}(\omega, \boldsymbol{\alpha}_N) = \sum_{i=1}^{n}\sum_{j=1}^{n} \ell_{n,ij}^{*}\left(\omega, \alpha_{j,o}, \alpha_{i,d}\right)$ for each $\left(\omega, \alpha_{j,o}, \alpha_{i,d}\right)$, where $\ell_{n,ij}^{*}\left(\omega, \alpha_{j,o}, \alpha_{i,d}\right)$ denotes the (i,j)-component of the log-likelihood.[32]

---

[31] For the LLN, it suffices to show $v(s) \downarrow 0$ as $s \uparrow \infty$.

[32] Note that

$\ell_{n,ij}^{*}\left(\omega, \alpha_{j,o}, \alpha_{i,d}\right) = 1\left(y_{n,ij} = 0\right) \ln \Phi\left(\epsilon_{n,ij}^{+,*}\left(\omega, \alpha_{j,o}, \alpha_{i,d}\right)\right)$

Our goals are to verify (i) $\widehat{\omega}_N \xrightarrow{p} \omega_0$ and (ii) $\sqrt{N}(\widehat{\omega}_N - \omega_0) - \Sigma^{*,-1}_{\omega_0,N}\Lambda^*_N \xrightarrow{d} N\big(\mathbf{0}, \Sigma^{*,-1}_{\omega_0}\big)$ as $n \to \infty$, where $\Sigma^*_{\omega_0}$ denotes the limiting variance of $\sqrt{N}(\widehat{\omega}_N - \omega_0)$, $\Sigma^*_{\omega_0,N}$ is a matrix satisfying $\Sigma^*_{\omega_0} = \lim_{n\to\infty} \Sigma^*_{\omega_0,N}$, and $\Lambda^*_N(\omega, \boldsymbol{\alpha}_N)$ with $\Lambda^*_N = \Lambda^*_N(\omega_0, \boldsymbol{\alpha}^0_N)$ is an asymptotic bias term originated from the existence of fixed effects. We will derive $\Sigma^*_{\omega_0,N}$, $\Sigma^*_{\omega_0}$, and Λ^*_N later. Due to the presence of possible asymptotic bias of $\widehat{\omega}_N$ because of the many individual effects, in a subsequent section, we consider an asymptotic bias adjustment procedure and a bias-adjusted estimator for ω_0.

Here, we will provide basic ideas of showing the two objects. Detailed discussions and proofs can be found in Appendix D and the supplement file. By Proposition D.2 (ii), the first-order conditions around ω_0 give

$$\mathbf{0} = \frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\widehat{\omega}_N, \widehat{\boldsymbol{\alpha}}_N)}{\partial \omega} = \frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\omega_0, \widehat{\boldsymbol{\alpha}}_N(\omega_0))}{\partial \omega} - \Sigma^*_{\omega_0,N}\sqrt{N}(\widehat{\omega}_N - \omega_0) + o_p(1),$$

where
$\Sigma^*_{\omega_0,N} = E\left(-\frac{1}{N}\frac{\partial^2 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega\partial\omega'}\right) - \frac{1}{n}\left\{E\left(\frac{1}{n}\frac{\partial^2 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega\partial\boldsymbol{\alpha}'_N}\right)E\left(-\frac{1}{n}\frac{\partial^2 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}'_N}\right)^{-1}E\left(\frac{1}{n}\frac{\partial^2 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\boldsymbol{\alpha}_N\partial\omega'}\right)\right\}$, and $\widehat{\boldsymbol{\alpha}}_N = \widehat{\boldsymbol{\alpha}}_N(\widehat{\omega}_N)$. Let $\widehat{\boldsymbol{\alpha}}^0_N = \widehat{\boldsymbol{\alpha}}_N(\omega_0)$, $\widehat{\boldsymbol{\alpha}}^0_{n,0} = \widehat{\boldsymbol{\alpha}}_{n,0}(\omega_0)$, and $\widehat{\boldsymbol{\alpha}}^0_{n,d} = \widehat{\boldsymbol{\alpha}}^0_{n,d}(\omega_0)$. Elements of those vectors are similarly defined. To study the asymptotic distribution of $\sqrt{N}(\widehat{\omega}_N - \omega_0)$, therefore, the main issue is to examine $\frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\omega_0,\widehat{\boldsymbol{\alpha}}_N(\omega_0))}{\partial\omega}$. By the second-order Taylor expansion of $\frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\omega_0,\widehat{\boldsymbol{\alpha}}^0_N)}{\partial\omega}$ around the true parameters $\boldsymbol{\alpha}^0_N$, Proposition D.2 (ii) yields

$$\begin{aligned}
\frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\omega_0,\widehat{\boldsymbol{\alpha}}^0_N)}{\partial\omega} = {} & \frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega} \\
& + \frac{1}{n}\sum_{j=1}^n \frac{\partial^2 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega\partial\alpha_{j,o}}\left(\hat{\alpha}^0_{j,o} - \alpha_{j,o,0}\right) + \frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega\partial\alpha_{i,d}}\left(\hat{\alpha}^0_{i,d} - \alpha_{i,d,0}\right) \\
& + \frac{1}{2n}\sum_{k=1}^n\sum_{j=1}^n \frac{\partial^3 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega\partial\alpha_{j,o}\partial\alpha_{k,o}}\left(\hat{\alpha}^0_{j,o} - \alpha_{j,o,0}\right)\left(\hat{\alpha}^0_{k,o} - \alpha_{k,o,0}\right) \\
& + \frac{1}{2n}\sum_{k=1}^n\sum_{i=1}^n \frac{\partial^3 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega\partial\alpha_{i,d}\partial\alpha_{k,o}}\left(\hat{\alpha}^0_{i,d} - \alpha_{i,d,0}\right)\left(\hat{\alpha}^0_{k,o} - \alpha_{k,o,0}\right) \\
& + \frac{1}{2n}\sum_{l=1}^n\sum_{j=1}^n \frac{\partial^3 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega\partial\alpha_{j,o}\partial\alpha_{l,d}}\left(\hat{\alpha}^0_{j,o} - \alpha_{j,o,0}\right)\left(\hat{\alpha}^0_{l,d} - \alpha_{l,d,0}\right) \\
& + \frac{1}{2n}\sum_{l=1}^n\sum_{i=1}^n \frac{\partial^3 \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega\partial\alpha_{i,d}\partial\alpha_{l,d}}\left(\hat{\alpha}^0_{i,d} - \alpha_{i,d,0}\right)\left(\hat{\alpha}^0_{l,d} - \alpha_{l,d,0}\right) + o_p(1).
\end{aligned} \tag{9}$$

Note that the first term of the right-hand-side above has zero mean and characterizes the asymptotic variance $\Sigma^*_{\omega_0}$. That is, $\frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega} \xrightarrow{d} N\big(\mathbf{0}, \Sigma^*_{\omega_0}\big)$ as $n \to \infty$. By the 2nd ~ 7th terms above, $\frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\omega_0,\widehat{\boldsymbol{\alpha}}^0_N)}{\partial\omega}$ is not centered at zero even for a large n. Those components would give some possible asymptotic bias terms. First, it comes from the usage of $\widehat{\boldsymbol{\alpha}}^0_N$ instead of $\boldsymbol{\alpha}^0_N$, whose components have slower convergence rates than $\sqrt{N} = n$ that is the convergence rate or $\widehat{\omega}_N$.[33] Second, the correlation

---

$$-1(y_{n,ij} > 0)\left\{\frac{1}{2}\ln 2\pi\sigma^2 + \left[\sum_{l=1}^\infty\left(\frac{\left(G_N(Y_N)A_N(\delta)G_N(Y_N)\right)^l}{l}\right)\right]_{(j-1)n+i,(j-1)n+i} - 1(y_{n,ij} > 0)\left(\epsilon^{+,*}_{n,ij}(\omega, \alpha_{j,o}, \alpha_{i,d})\right)^2\right\}$$
$$-\frac{\mu}{2N}\left(\sum_{j=1}^n \alpha_{j,o} - \sum_{i=1}^n \alpha_{i,d}\right)^2 .$$

[33] Indeed, we have $\sqrt{n}\left(\hat{\alpha}^0_{j,o} - \alpha_{j,o,0}\right) = O_p(1)$ for $j = 1,\cdots,n$ and $\sqrt{n}\left(\hat{\alpha}^0_{i,d} - \alpha_{i,d,0}\right) = O_p(1)$ for $i = 1,\cdots,n$ by Lemma D.2 in the supplement file. This $\sqrt{n}$-convergence rate is the same as that of the fixed-effect estimates in the linear SARF model.

between $\hat{\alpha}_{j,o}^{0}$ (and $\hat{\alpha}_{i,d}^{0}$) and the second and third order derivatives of the log-likelihood, which are related to $\widehat{\omega}_N$. Third, the variances of $\hat{\alpha}_{j,o}^{0}$ and $\hat{\alpha}_{i,d}^{0}$ form the asymptotic bias term.[34]

To represent the asymptotic bias terms, the following notations are employed:

Let $\bar{\mathcal{H}}_n = E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial \boldsymbol{\alpha}_N \partial \boldsymbol{\alpha}_N'}\right)^{-1} = \begin{bmatrix} \bar{\mathcal{H}}_{(\alpha_o\alpha_o),n} & \bar{\mathcal{H}}_{(\alpha_o\alpha_d),n} \\ \bar{\mathcal{H}}'_{(\alpha_o\alpha_d),n} & \bar{\mathcal{H}}_{(\alpha_d\alpha_d),n} \end{bmatrix}$,
$a_{n,ij} = \left[\bar{\mathcal{H}}_{(\alpha_o\alpha_o),n}\right]_{ij}$, $b_{n,ij} = \left[\bar{\mathcal{H}}_{(\alpha_o\alpha_d),n}\right]_{ij}$, and $c_{n,ij} = \left[\bar{\mathcal{H}}_{(\alpha_d\alpha_d),n}\right]_{ij}$;
$\frac{\partial \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial \alpha_{j,o}} = \sum_{i=1}^{n} q_{n,ij}^{\alpha_o}$ for $j = 1,\cdots,n$; $\frac{\partial \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial \alpha_{i,d}} = \sum_{j=1}^{n} q_{n,ij}^{\alpha_d}$ for $i = 1,\cdots,n$;
$\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial \omega \partial \alpha_{j,o}} = \sum_{i=1}^{n} h_{n,ij}^{\omega\alpha_o}$ for $j = 1,\cdots,n$; $\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial \omega \partial \alpha_{i,d}} = \sum_{j=1}^{n} h_{n,ij}^{\omega\alpha_d}$ for $i = 1,\cdots,n$;
$\frac{\partial^3 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial \omega \partial \alpha_{j,o}^2} = \sum_{i=1}^{n} t_{n,ij}^{\omega\alpha_o}$ for $j = 1,\cdots,n$; $\frac{\partial^3 \ln L_N^*(\omega\omega_0,\boldsymbol{\alpha}_N^0)}{\partial \omega \partial \alpha_{i,d}^2} = \sum_{j=1}^{n} t_{n,ij}^{\omega\alpha_d}$ for $i = 1,\cdots,n$;
$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial \alpha_{j,o}^3} = \sum_{i=1}^{n} t_{n,ij}^{\alpha_o}$ for $j = 1,\cdots,n$; and $\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial \alpha_{i,d}^3} = \sum_{j=1}^{n} t_{n,ij}^{\alpha_d}$ for $i = 1,\cdots,n$,
where $q_{n,ij}^{\alpha_o} = \frac{\partial \ell_{n,ij}^*(\omega_0,\alpha_{j,o,0},\alpha_{i,d,0})}{\partial \alpha_{j,o}}$, $q_{n,ij}^{\alpha_d} = \frac{\partial \ell_{n,ij}^*(\omega_0,\alpha_{j,o,0},\alpha_{i,d,0})}{\partial \alpha_{i,d}}$,
$h_{n,ij}^{\omega\alpha_o} = \frac{\partial^2 \ell_{n,ij}^*(\omega_0,\alpha_{j,o,0},\alpha_{i,d,0})}{\partial \omega \partial \alpha_{j,o}} = \frac{\partial^2 \ell_{n,ij}^*(\omega_0,\alpha_{j,o,0},\alpha_{i,d,0})}{\partial \omega \partial \alpha_{i,d}} = h_{n,ij}^{\omega\alpha_d}$,
$t_{n,ij}^{\omega\alpha_o} = \frac{\partial^3 \ell_{n,ij}^*(\omega_0,\alpha_{j,o,0},\alpha_{i,d,0})}{\partial \omega \partial \alpha_{j,o}^2} = \frac{\partial^3 \ell_{n,ij}^*(\omega_0,\alpha_{j,o,0},\alpha_{i,d,0})}{\partial \omega \partial \alpha_{i,d}^2} = t_{n,ij}^{\omega\alpha_d}$, and
$t_{n,ij}^{\alpha_o} = \frac{\partial^3 \ell_{n,ij}^*(\omega_0,\alpha_{j,o,0},\alpha_{i,d,0})}{\partial \alpha_{j,o}^3} = \frac{\partial^3 \ell_{n,ij}^*(\omega_0,\alpha_{j,o,0},\alpha_{i,d,0})}{\partial \alpha_{i,d}^3} = t_{n,ij}^{\alpha_d}$. We provide the forms of them in Appendix D.

Using the notations above, we represent components characterizing the asymptotic bias of $\widehat{\omega}_N$:

$\Lambda_{1,N}^* = \frac{1}{n}\sum_{j=1}^{n} a_{n,jj} \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} E\left(q_{n,kj}^{\alpha_o} h_{n,ij}^{\omega\alpha_o}\right)$,
$\Lambda_{2,N}^* = \frac{1}{n}\sum_{i=1}^{n} c_{n,ii} \frac{1}{n}\sum_{l=1}^{n}\sum_{j=1}^{n} E(q_{n,il}^{\alpha_d} h_{n,ij}^{\omega\alpha_d})$,
$\Lambda_{3,N}^* = \frac{1}{n}\sum_{j=1}^{n} a_{n,jj}\left(\frac{1}{n}\sum_{k=1}^{n} E\left(h_{n,kj}^{\omega\alpha_o}\right)\right)\sum_{i=1}^{n} E\left(h_{n,ij}^{\alpha_o} v_{\alpha_o,n,j}\right)$,
$\Lambda_{4,N}^* = \frac{1}{n}\sum_{i=1}^{n} c_{n,ii}\left(\frac{1}{n}\sum_{l=1}^{n} E\left(h_{n,il}^{\omega\alpha_d}\right)\right)\sum_{j=1}^{n} E\left(h_{n,ij}^{\alpha_d} v_{\alpha_d,n,i}\right)$,
$\Lambda_{5,N}^* = \frac{1}{2n}\sum_{j=1}^{n} \widetilde{\omega}_{n,(\alpha_o\alpha_o),jj}\, a_{n,jj}^2 \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n} E\left(q_{n,kj}^{\alpha_o} q_{n,lj}^{\alpha_o}\right)$, and
$\Lambda_{6,N}^* = \frac{1}{2n}\sum_{i=1}^{n} \widetilde{\omega}_{n,(\alpha_d\alpha_d),ii} c_{n,ii}^2 \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n} E(q_{n,ik}^{\alpha_d} q_{n,il}^{\alpha_d})$,

where $v_{\alpha_o,n,j} = \frac{1}{n}\sum_{k=1}^{n}\sum_{p=1}^{n} a_{n,jk} q_{n,pk}^{\alpha_o} + \frac{1}{n}\sum_{l=1}^{n}\sum_{q=1}^{n} b_{n,jl} q_{n,lq}^{\alpha_d}$,
$\widetilde{\omega}_{n,(\alpha_o\alpha_o),jj} = \frac{1}{n}\sum_{i=1}^{n} E\left(t_{n,ij}^{\omega\alpha_o}\right) + \frac{1}{n}\pi_{\alpha_o,n,j}\sum_{i=1}^{n} E\left(t_{n,ij}^{\alpha_o}\right) + \frac{1}{n}\sum_{i=1}^{n} \pi_{\alpha_d,n,i} E\left(t_{n,ij}^{\alpha_d}\right)$,
$\pi_{\alpha_o,n,j} = \frac{1}{n}\sum_{k=1}^{n}\sum_{p=1}^{n} a_{n,jk} E\left(h_{n,pk}^{\omega\alpha_o}\right) + \frac{1}{n}\sum_{l=1}^{n}\sum_{q=1}^{n} b_{n,jl} E\left(h_{n,lq}^{\omega\alpha_d}\right)$ for $j = 1,\cdots,n$,
$v_{\alpha_d,n,i} = \frac{1}{n}\sum_{k=1}^{n}\sum_{p=1}^{n} b_{n,ki} q_{n,pk}^{\alpha_o} + \frac{1}{n}\sum_{l=1}^{n}\sum_{q=1}^{n} c_{n,il} q_{n,lq}^{\alpha_d}$,
$\widetilde{\omega}_{n,(\alpha_d\alpha_d),ii} = \frac{1}{n}\sum_{j=1}^{n} E\left(t_{n,ij}^{\omega\alpha_d}\right) + \frac{1}{n}\sum_{j=1}^{n} \pi_{\alpha_o,n,j} E\left(t_{n,ij}^{\alpha_o}\right) + \frac{1}{n}\pi_{\alpha_d,n,i}\sum_{j=1}^{n} E\left(t_{n,ij}^{\alpha_d}\right)$, and
$\pi_{\alpha_d,n,i} = \frac{1}{n}\sum_{k=1}^{n}\sum_{p=1}^{n} b_{n,ki} E\left(h_{n,pk}^{\omega\alpha_o}\right) + \frac{1}{n}\sum_{l=1}^{n}\sum_{q=1}^{n} c_{n,il} E\left(h_{n,lq}^{\omega\alpha_d}\right)$ for $i = 1,\cdots n$.

By Proposition D.1, $E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial \boldsymbol{\alpha}_N \partial \boldsymbol{\alpha}_N'}\right)^{-1}$ can be approximated by a diagonal matrix and its off-

[34] Those three points are raised by Hahn and Newey (2004). Refer to Section 2 in Hahn and Newey (2004).

diagonal components of are of $O\left(\frac{1}{n}\right)$. Then, the second term in (9) can be approximated by $\Lambda_{1,N}^{*}+\Lambda_{3,N}^{*}$; the third term's approximation in (9) is $\Lambda_{2,N}^{*}+\Lambda_{4,N}^{*}$; the fourth term is approximated by $\Lambda_{5,N}^{*}$; the seventh term's approximation is $\Lambda_{6,N}^{*}$; and the fifth and sixth terms are stochastically negligible (see Proposition D.3). Note that $\Lambda_{1,N}^{*}$, $\Lambda_{3,N}^{*}$ and $\Lambda_{5,N}^{*}$ are originated from the estimators of the origins' fixed effects $\widehat{\boldsymbol{\alpha}}_{n,0}^{0}$ while $\Lambda_{2,N}^{*}$, $\Lambda_{4,N}^{*}$, and $\Lambda_{6,N}^{*}$ comes from $\widehat{\boldsymbol{\alpha}}_{n,d}^{0}$. This additive separation is originated from the additive separability of $\alpha_{j,o}$ and $\alpha_{i,d}$. For the following result, we define

$$\mathbb{S}_{2,(\omega,\omega),N}(\delta)=\begin{bmatrix} s_{\lambda\lambda,N}(\delta) & * & * & * & * \\ s_{\lambda\gamma,N}(\delta) & s_{\gamma\gamma,N}(\delta) & * & * & * \\ s_{\lambda\rho,N}(\delta) & s_{\gamma\rho,N}(\delta) & s_{\rho\rho,N}(\delta) & * & * \\ 0 & 0 & 0 & 0 & * \\ s_{\lambda\sigma^2,N}(\delta) & s_{\gamma\sigma^2,N}(\delta) & s_{\rho\sigma^2,N}(\delta) & 0 & s_{\sigma^2\sigma^2,N}(\omega\delta) \end{bmatrix}, \text{ for } \delta\in\Theta_{\delta},$$

where $s_{\lambda\lambda,N}(\delta)=tr\left(\boldsymbol{W}_{22,N}^{2}S_{N_2}^{*-2}(\delta)\right)$, $s_{\lambda\gamma,N}(\delta)=tr\left(\boldsymbol{W}_{22,N}\boldsymbol{M}_{22,N}S_{N_2}^{*-2}(\delta)\right)$, $s_{\lambda\rho,N}(\delta)=tr\left(\boldsymbol{W}_{22,N}\boldsymbol{R}_{22,N}S_{N_2}^{*-2}(\delta)\right)$, $s_{\lambda\sigma^2,N}(\delta)=-tr\left(\boldsymbol{W}_{22,N}S_{N_2}^{*-1}(\delta)\right)-tr\left(\boldsymbol{W}_{22,N}S_{N_2}^{*-2}(\delta)\boldsymbol{A}_{N}(\delta)\right)$, $s_{\gamma\gamma,N}(\delta)=tr\left(\boldsymbol{M}_{22,N}^{2}S_{N_2}^{*-2}(\delta)\right)$, $s_{\gamma\rho,N}(\delta)=tr\left(\boldsymbol{M}_{22,N}\boldsymbol{R}_{22,N}S_{N_2}^{*-2}(\delta)\right)$, $s_{\gamma\sigma^2,N}(\delta)=-tr\left(\boldsymbol{M}_{22,N}S_{N_2}^{*-1}(\delta)\right)-tr\left(\boldsymbol{M}_{22,N}S_{N_2}^{*-2}(\delta)\boldsymbol{A}_{N}(\delta)\right)$, $s_{\rho\rho,N}(\delta)=tr\left(\boldsymbol{R}_{22,N}^{2}S_{N_2}^{*-2}(\delta)\right)$, $s_{\rho\sigma^2,N}(\delta)=-tr\left(\boldsymbol{R}_{22,N}S_{N_2}^{*-1}(\delta)\right)-tr\left(\boldsymbol{R}_{22,N}S_{N_2}^{*-2}(\delta)\boldsymbol{A}_{N}(\delta)\right)$, and $s_{\sigma^2\sigma^2,N}(\delta)=2tr\left(S_{N_2}^{*-1}(\delta)\boldsymbol{A}_{N}(\delta)\right)+tr\left(S_{N_2}^{*-2}(\delta)\boldsymbol{A}_{N}^{2}(\delta)\right)$. The theorem below states the asymptotic properties of $\widehat{\omega}_{N}$.

Theorem 4.5. Suppose Assumptions 3.1, 4.1 – 4.4, 4.8, 4.10 and 4.12 hold. In addition, we assume that (i) each element of $\boldsymbol{\alpha}_{N}$ is a bounded constant in $\mathbb{R}$ for all N, (ii) $\Sigma_{\omega_0}^{*}=\lim_{n\to\infty}\Sigma_{\omega_0,N}^{*}$ is nonsingular and $\Sigma_{\omega_0}^{*}>0$, (iii) $\mathbb{S}_{2,(\omega,\omega),N}(\delta)\geq 0$ for $\delta\in\Theta_{\delta}$, and (iv) $E\left(-\frac{1}{n}\frac{\partial^2\ln L_N^{*}(\omega_0,\boldsymbol{\alpha}_N^{0})}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)$ is nonsingular and $E\left(-\frac{1}{n}\frac{\partial^2\ln L_N^{*}(\omega_0,\boldsymbol{\alpha}_N^{0})}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)>0$ under a large n. Then, we have (i) $\widehat{\omega}_N\xrightarrow{p}\omega_0$, and (ii) $\sqrt{N}(\widehat{\omega}_N-\omega_0)\xrightarrow{d}N\left(\Sigma_{\omega_0}^{*-1}\Lambda_{\infty}^{*},\Sigma_{\omega_0}^{*-1}\right)$ as $n\to\infty$, where $\Lambda_{\infty}^{*}=\lim_{n\to\infty}\Lambda_{N}^{*}$ with $\Lambda_N^{*}=\Lambda_{1,N}^{*}+\Lambda_{2,N}^{*}+\Lambda_{3,N}^{*}+\Lambda_{4,N}^{*}+\Lambda_{5,N}^{*}+\Lambda_{6,N}^{*}$.

Showing Theorem 4.5 is based on deriving the asymptotic expansion of $\frac{1}{\sqrt{N}}\frac{\partial\ln L_N^{*}(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)}{\partial\omega}$ under regularity conditions for the Taylor approximation of $\frac{1}{\sqrt{N}}\frac{\partial\ln L_N^{*}(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)}{\partial\omega}$. Condition (i) Theorem 4.5 is for consistency of $\widehat{\omega}_N$ and $\widehat{\boldsymbol{\alpha}}_N$. Then, all components of $\ln L_N^{*}(\omega)$ (i.e., $\{\ell_{n,ij}^{*}(\omega)\}$) satisfy the NED properties using the same arguments in Theorem 4.3. For well-definedness of the asymptotic variance of $\widehat{\omega}_N$, Condition (ii) is introduced. Condition (iii) leads to strict concavity of $\ln L_N^{*}(\omega,\boldsymbol{\alpha}_N)$.[35] Under a large n, Condition (iv) guarantees for invertibility of $E\left(-\frac{1}{n}\frac{\partial^2\ln L_N^{*}(\omega_0,\boldsymbol{\alpha}_N^{0})}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)$ whose inverse is a component of the Taylor approximation of $\frac{1}{\sqrt{N}}\frac{\partial\ln L_N^{*}(\omega,\widehat{\boldsymbol{\alpha}}_N(\omega))}{\partial\omega}$.[36] With Conditions (i), (ii), and (iii), we can establish consistency of $\widehat{\boldsymbol{\alpha}}_N$ and the asymptotic expansion of $\frac{1}{\sqrt{N}}\frac{\partial\ln L_N^{*}(\omega_0,\widehat{\boldsymbol{\alpha}}_N^{0})}{\partial\omega}$. It implies

[35] Under this condition, the log-likelihood function under the Olsen's (1978) reparameterization becomes strictly concave in the transformed parameters. The reason for having this condition is that Jacobian term under the transformation relies on the spatial interaction parameters as well as variance parameter. This condition can be removed when all eigenvalues of W_n and M_n are real-valued. For this issue, refer to Lemma 2 in Liu, Xu, and Lee (2021) and our supplement file.

[36] In detail, the diagonal components $\bar{\mathcal{H}}_n$ are one of the main components of $\{\hat{\alpha}_{j,o}^{0}-\alpha_{j,o,0}\}$ and $\{\hat{\alpha}_{i,d}^{0}-\alpha_{i,d,0}\}$ in (9).

$$\sqrt{N}(\widehat{\omega}_N-\omega_0)-\Sigma^{*-1}_{\omega_0,N}\Lambda^*_N=\Sigma^{*-1}_{\omega_0,N}\frac{1}{\sqrt{N}}\frac{\partial \ln L^*_N(\omega_0,\boldsymbol{\alpha}^0_N)}{\partial\omega}\xrightarrow{d}N\left(\mathbf{0},\Sigma^{*-1}_{\omega_0}\right)\ \text{as}\ n\to\infty.$$

Based on the results of Theorem 4.5, we can define a bias corrected MLE

$$\widehat{\omega}^c_N=\widehat{\omega}_N-\frac{1}{n}\Sigma^{*-1}_{\omega_0,N}(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)\widehat{\Lambda}^*_N,$$

where $\Sigma^*_{\omega_0,N}(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)$ denotes the asymptotic variance matrix evaluated at $(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)$, and $\widehat{\Lambda}^*_N$, is a consistent estimator of Λ^*_∞ with employing $(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)$. To obtain $\widehat{\Lambda}^*_N$, we can apply the similar idea of getting a truncated sum of sample covariances in time series literature.[37] Note that $\Lambda^*_{1,N}$, $\Lambda^*_{3,N}$, and $\Lambda^*_{5,N}$ take a form of $\Lambda^*_{o,N}=\frac{1}{n}\sum_{j=1}^n d_{n,j}\frac{1}{n}\sum_{i=1}^n\sum_{k=1}^n E\left(A_{n,ij}B_{n,kj}\right)$ while $\Lambda^*_{2,N}$, $\Lambda^*_{4,N}$, and $\Lambda^*_{6,N}$ take $\Lambda^*_{d,N}=\frac{1}{n}\sum_{i=1}^n d_{n,i}\frac{1}{n}\sum_{j=1}^n\sum_{k=1}^n E\left(A_{n,ij}B_{n,ik}\right)$ where $\{d_{n,i}\}$ are non-stochastic bounded weights, and $\{A_{n,ij}\}$ and $\{B_{n,ij}\}$ are random components. For consistent estimators of $\Lambda^*_{o,\infty}=\lim_{n\to\infty}\Lambda^*_{o,N}$ and $\Lambda^*_{d,\infty}=\lim_{n\to\infty}\Lambda^*_{d,N}$, we design

$$\widehat{\Lambda}^*_{o,N}=\frac{1}{n}\sum_{j=1}^n\hat{d}_{n,j}\frac{1}{n}\sum_{i=1}^n\sum_{k\in nbd(i,s_n)}\hat{A}_{n,ij}\hat{B}_{n,kj}\ \text{and}\ \widehat{\Lambda}^*_{d,N}=\frac{1}{n}\sum_{i=1}^n\hat{d}_{n,i}\frac{1}{n}\sum_{i=1}^n\sum_{k\in nbd(i,s_n)}\hat{A}_{n,ij}\hat{B}_{n,ik},$$

where $nbd(i,s_n)$ denotes the i's s_n-th order neighboring units induced by spatial weighting matrices, and $\hat{A}_{n,ij}$, $\hat{B}_{n,ij}$ and $\hat{d}_{n,i}$ are respectively $A_{n,ij}$, $B_{n,ij}$, and $d_{n,i}$ evaluated at $(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)$.[38] If $s_n\to\infty$ and $\frac{\sup_{n,i}\text{card}(\{k:k\in nbd(i,s_n)\})}{n}\to 0$, we have $\widehat{\Lambda}^*_{o,N}\xrightarrow{p}\Lambda^*_{o,\infty}$ and $\widehat{\Lambda}^*_{d,N}\xrightarrow{p}\Lambda^*_{d,\infty}$ as $n\to\infty$.[39]

The asymptotic property of the bias corrected estimator $\widehat{\omega}^c_N$ is stated in Theorem 4.6.

Theorem 4.6. Assume that the conditions of Theorem 4.5 hold. If $\widehat{\Lambda}^*_N\xrightarrow{p}\Lambda^*_\infty$ as $n\to\infty$, we have

$$\sqrt{N}(\widehat{\omega}^c_N-\omega_0)\xrightarrow{d}N\left(\mathbf{0},\Sigma^{*-1}_{\omega_0}\right)\ \text{as}\ n\to\infty.$$

Theorem 4.6 can be verified by showing $\Sigma^*_{\omega_0,N}(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)\xrightarrow{p}\Sigma^*_{\omega_0}$ as $n\to\infty$. The result follows by the continuous mapping theorem with $\|\widehat{\omega}_N-\omega_0\|\xrightarrow{p}0$ and $\|\widehat{\boldsymbol{\alpha}}_N-\boldsymbol{\alpha}^0_N\|_\infty\xrightarrow{p}0$ as $n\to\infty$. Compared to the bias correction for the linear SARF model with fixed effects, we need to employ the estimates $\widehat{\boldsymbol{\alpha}}_N$ in evaluating $\Sigma^*_{\omega_0,N}(\widehat{\omega}_N,\widehat{\boldsymbol{\alpha}}_N)$ and $\widehat{\Lambda}^*_N$. Since our model implies $\frac{\text{\# of origin units}}{\text{\# of destination units}}=\frac{n}{n}=1$, there is no restriction on a sample size n as Lee and Yu (2010).[40]

5. Monte Carlo simulations

[37] In a nonlinear panel setting with time-dependent but cross-sectionally independent observations, Hahn and Kuersteiner (2011) and Fernandez-Val and Weidner (2016) apply the truncation idea.

[38] For example, consider W_n is a sparse adjacency matrix and $M_n=W_n$. Then, $k\in nbd(i,s_n)$ if $[W_n^l]_{ik}\neq 0$ for some $l\in\{0,1,\cdots,s_n\}$.

[39] Note that $\text{card}(\{k:k\in nbd(i,s_n)\})$ corresponds to the trimming parameter for the time dimension in Fernandez-Val and Weidner (2016). To show $\widehat{\Lambda}^*_{o,N}\xrightarrow{p}\Lambda^*_{o,\infty}$ and $\widehat{\Lambda}^*_{d,N}\xrightarrow{p}\Lambda^*_{d,\infty}$ as $n\to\infty$, hence, we can apply the similar strategy of the proof of Theorem 4.3 in Fernandez-Val and Weidner (2016) (see Part II of the proof of Theorem 4.3 in Fernandez-Val and Weidner (2016)).

[40] To have the asymptotic normality of the bias corrected MLE for spatial dynamic panel data models, Lee and Yu (2010) verify that $\frac{n}{T^3}\to 0$ and $\frac{T}{n^3}\to 0$ are required. Those conditions are introduced to achieve the asymptotic equivalence of the infeasible bias corrected MLE and the feasible bias corrected MLE.

5.1 Finite sample performance

In this subsection, we conduct Monte Carlo simulations to study the finite sample performance of the MLE $\hat{\theta}_N$. Also, we investigate misspecification errors when one uses the linear SARF model but the true data generating process (DGP) is the SARF Tobit model. Two DGPs are considered:

DGP 1. Linear SARF model: equation (3); and **DGP 2**. SARF Tobit model: equation (6)

In generating the data, we utilize the same $X_{n,1}$ and $Z_{N,1}$ in Section 6. That is, $n = 48$, $K = 1$, and $L = 1$. We consider an adjacency matrix of states' borders (denoted by W_n^a) and suppose $M_n = W_n^a$ in this subsection: $w_{n,ij}^a = 1$ if states i and j with $j \neq i$ are bordering states; and $w_{n,ij}^a = 0$ otherwise. Both W_n^a and M_n are symmetric spatial weighting matrices. For the first experiment, we consider $\theta_0 = (1,\ 0.02,\ 0.02,\ 0.01,\ -4,\ 1,\ 1,\ 1)'$, which satisfies the spatial stability and model coherency. Simulation results for additional parameter sets are provided in the supplement file. The disturbances $\epsilon_{n,ij}$s are independently drawn from the standard normal distribution.[41] For the SARF Tobit model (**DGP 2**), we generate $vec(Y_N)$ by the contraction mapping with a tolerance 10^{-6}.

In order to evaluate finite sample performance of $\hat{\theta}_N$, we consider three criteria: (i) empirical bias, (ii) empirical standard deviation (STD), and (iii) 95% coverage probability. The number of sample repetitions is 1,000 for each experiment.

Table 1. Simulation results for the linear SARF and SARF Tobit models

DGP 1 (Linear SARF). $\theta_0 = (\alpha_0, \lambda_0, \gamma_0, \rho_0, \beta_0, b_0, c_0, \sigma_0^2)' = (1,\ 0.02,\ 0.02,\ 0.01,\ -4,\ 1,\ 1,\ 1)'$

	α_0	λ_0	γ_0	ρ_0	β_0	b_0	c_0	σ_0^2
Bias	0.0001	-0.0003	-0.0004	0.0000	0.0001	0.0014	0.0014	0.0005
STD	0.0001	0.0014	0.0014	0.0004	0.0008	0.0015	0.0015	0.0001
95% CP	1.0000	0.9980	1.0000	0.9620	1.0000	1.0000	1.0000	1.0000

DGP 2 (SARF Tobit). $\theta_0 = (\alpha_0, \lambda_0, \gamma_0, \rho_0, \beta_0, b_0, c_0, \sigma_0^2)' = (1,\ 0.02,\ 0.02,\ 0.01,\ -4,\ 1,\ 1,\ 1)'$
% of nonzero observations (average): 83.38%

	α_0	λ_0	γ_0	ρ_0	β_0	b_0	c_0	σ_0^2
Bias	0.0002	-0.0014	-0.0015	0.0002	0.0011	0.0032	0.0032	0.0005
STD	0.0001	0.0015	0.0014	0.0004	0.0007	0.0013	0.0013	0.0001
95% CP	1.0000	0.9960	0.9990	0.9510	1.0000	1.0000	1.0000	1.0000
Misspecification from assuming the linear SARF model								
Bias	0.0011	-0.0081	-0.0081	0.0013	0.0090	0.0158	0.0159	0.0015
STD	0.0007	0.0011	0.0010	0.0004	0.0093	0.0033	0.0022	0.0091
95% CP	1.0000	0.0200	0.0130	0.1560	0.9970	0.9990	1.0000	0.9970

For both cases, we observe reasonable performance of the MLE in terms of biases and CPs. Under the larger spatial influence case, biases become slightly larger (in absolute values) while there is no significant change in STDs. In $\hat{\lambda}_N$ and $\hat{\gamma}_N$, we detect downward biases. On the other hand, upward biases are observed in $\hat{b}_N$, $\hat{c}_N$, and $\hat{\sigma}_N^2$.

[41] In the supplement file, we provide additional simulation results for non-normally distributed $\epsilon_{n,ij}$s: (1) uniform, (2) Logistic, (3) Gamma, (4) Beta, and (5) mixed normal distributions. We observe that the MLE performs well except for estimating the variance parameter σ_0^2 with the case of the mixed normal distribution.

Table 1 also shows the simulation results under DGP 2 for the SARF Tobit model. By adjusting β_0, we can control a proportion of nonzero observations. We provide average percentages of nonzero observations. For misspecification analyses, we also report the estimation results using the SARF model.

When the model is correctly specified (i.e., SARF Tobit model), the MLE performs well and overall performance is similar with that provided, which is for the linear SARF model in Table 1. For all cases, biases increase when we do not consider the Tobit structure. By observing the estimates' low levels of CPs for λ_0, γ_0, and ρ_0, we observe that statistical inference for the spatial interaction parameters would be invalid under the misspecification. Under the model misspecification, we detect large biases in estimates of the linear sensitivity parameters β_0, b_0, and c_0. Misspecification biases increase when a percentage of zero observations increases. Under larger spatial influences, misspecification biases tend to be larger (except for estimates of b_0 and c_0).

As the second issue, we estimate the parameter $\omega_0 = (\lambda_0, \gamma_0, \rho_0, \beta_0)'$ when there exist two-way fixed effects. We investigate the finite sample performance of the MLE and the bias corrected MLE. We consider $n = 25$ in the main draft for computational tractability.[42] We consider the LeSage and Pace's (2008) specification with a row-normalized rook matrix as for a chess board, i.e., $(W_n, M_n) = (W_{R,n}^r, W_{R,n}^{r\prime})$, where $W_{R,n}^r = [w_{R,n,ij}^r]$ with $w_{R,n,ij}^r = \frac{w_{n,ij}^r}{\sum_{k=1}^n w_{n,ik}^r}$. We utilize the first 25 states' geographic locations for constructing $Z_{N,1}$ while $X_{n,1}$ is excluded. For this experiment, $\omega_0 = (0.1, 0.1, 0.05, -0.15, 0.8)'$ and $\omega_0 = (-0.1, -0.1, -0.05, -0.15, 0.8)'$ are considered. For the fixed effects, we draw $\alpha_{j,o}$ from $N(1, 0.01^2)$ and set $\alpha_{i,d} = \alpha_{i,o}$ for $i = 1, \cdots, n$. The trimming spatial order s_n for all $i = 1, \cdots, n$ is defined by $k \in nbd(i, s_n)$ if $\left[\left(W_{R,n}^r\right)^l\right]_{ik} \neq 0$ for some $l \in \{0, 1, \cdots, s_n\}$. We report the bias corrected MLE for the SARF Tobit model with $s_n = 1$ and 2.[43]

Table 2. Simulation results for the linear SARF and SARF Tobit models with fixed effects

Case 1:

DGP1 (Linear SARF). $n = 25$, $\omega_0 = (\lambda_0, \gamma_0, \rho_0, \beta_0, \sigma_0^2)' = (0.1, 0.1, 0.05, -0.15, 0.8)'$

	λ_0	γ_0	ρ_0	β_0	σ_0^2
MLE					
Bias	-0.0660	-0.0664	0.0071	0.0009	-0.0700
STD	0.0515	0.0491	0.0880	0.0244	0.0456
95% CP	0.7480	0.7620	0.9500	0.9420	0.5880
Bias corrected MLE					
Bias	-0.0070	-0.0072	-0.0009	0.0011	-0.0148
STD	0.0534	0.0510	0.0948	0.0243	0.0490
95% CP	0.9320	0.9340	0.9360	0.9380	0.8660

DGP2 (SARF Tobit). $n = 25$, $\omega_0 = (\lambda_0, \gamma_0, \rho_0, \beta_0, \sigma_0^2)' = (0.1, 0.1, 0.05, -0.15, 0.8)'$

% of nonzero observations (average): 92.22%

	λ_0	γ_0	ρ_0	β_0	σ_0^2
MLE					
Bias	-0.0704	-0.0706	0.0066	0.0009	-0.0704
STD	0.0553	0.0528	0.0956	0.0245	0.0481

[42] Since we employ the direct estimation method for the SARF Tobit model with the fixed effects, a large n leads to a linearly increasing parameter space. We observe that the case of $n = 25$ not only provides a sufficient sample size (i.e., $N = n^2 = 625$), but it also takes a reasonable time for sample repetitions. In the supplement file, we provide the simulation results for a smaller $n = 16$ and a larger $n = 36$.

[43] The average value of $\mathrm{card}(\{k : k \in nbd(i, s_n = 1)\})$ is 4.2 while that of $\mathrm{card}(\{k : k \in nbd(i, s_n = 2)\})$ is 9.16.

95% CP	0.7440	0.7560	0.9460	0.9440	0.6160
Bias corrected MLE with $s_n = 1$					
Bias	-0.0118	-0.0120	0.0120	0.0024	0.0193
STD	0.0543	0.0519	0.0936	0.0246	0.0546
95% CP	0.9360	0.9520	0.9540	0.9340	0.8820
Bias corrected MLE with $s_n = 2$					
Bias	-0.0204	-0.0203	0.0136	0.0021	-0.0027
STD	0.0569	0.0541	0.1009	0.0249	0.0537
95% CP	0.9040	0.9400	0.9340	0.9360	0.8840

Case 2:
DGP1 (Linear SARF). $n = 25$, $\omega_0 = (\lambda_0, \gamma_0, \rho_0, \beta_0, \sigma_0^2)' = (-0.1, -0.1, -0.05, -0.15, 0.8)'$

	λ_0	γ_0	ρ_0	β_0	σ_0^2
MLE					
Bias	-0.0554	-0.0553	-0.0054	0.0022	-0.0793
STD	0.0499	0.0482	0.0851	0.0242	0.0456
95% CP	0.7880	0.7960	0.9560	0.9340	0.5180
Bias corrected MLE					
Bias	-0.0036	-0.0033	0.0026	0.0013	-0.0155
STD	0.0522	0.0505	0.0924	0.0243	0.0495
95% CP	0.9420	0.9240	0.9340	0.9340	0.8680

DGP2 (SARF Tobit). $n = 25$, $\omega_0 = (\lambda_0, \gamma_0, \rho_0, \beta_0, \sigma_0^2)' = (-0.1, -0.1, -0.05, -0.15, 0.8)'$
% of nonzero observations (average): 79.36%

	λ_0	γ_0	ρ_0	β_0	σ_0^2
MLE					
Bias	-0.0691	-0.0696	-0.0056	0.0023	-0.0836
STD	0.0619	0.0601	0.1060	0.0245	0.0521
95% CP	0.7840	0.8040	0.9560	0.9380	0.5380
Bias corrected MLE with $s_n = 1$					
Bias	0.0052	0.0044	0.0386	0.0059	0.0246
STD	0.0623	0.0604	0.1066	0.0247	0.0609
95% CP	0.9460	0.9460	0.9460	0.9240	0.8680
Bias corrected MLE with $s_n = 2$					
Bias	-0.0135	-0.0138	0.0240	0.0050	-0.0029
STD	0.0650	0.0624	0.1136	0.0250	0.0594
95% CP	0.9280	0.9360	0.9420	0.9320	0.8720

Downward biases in the MLEs of λ_0, γ_0, and σ_0^2 are detected for both models. Those downward biases for the SARF Tobit model tend to be larger than those for the linear SARF model. Magnitudes of those biases decrease when n increases. These results are consistent with estimating the contemporaneous spatial effect, dynamic effect, and the variance parameters in a spatial dynamic panel data model (Lee and Yu, 2010). We do not capture a significant bias in estimating ρ_0 and β_0. After correcting the asymptotic biases, the magnitudes of the biases in the MLEs of λ_0, γ_0, and σ_0^2 are reduced and the CPs of them increase to be more adequate. For the bias correction of $\hat{\lambda}_N$ and $\hat{\gamma}_N$ in case of the SARF Tobit model, choosing $s_n = 1$ significantly reduces the downward biases in $\hat{\lambda}_N$ and $\hat{\gamma}_N$; and choosing $s_n = 2$ performs well for the bias correction of $\hat{\sigma}_N^2$.

5.2 Selecting spatial weighting matrices $(\boldsymbol{W_n}, \boldsymbol{M_n})$

In this subsection, we consider the model comparison issue in selecting W_n and/or M_n. If a practitioner wants to select a proper (W_n, M_n) among various possible spatial weighting matrices, a statistical criterion can be a reasonable guidance for this issue. To generate asymmetric spatial influences, we construct $W_n^e = \left[w_{n,ij}^e\right]$, $w_{n,ij}^e = \frac{1}{|income_i - income_j|}\left(\frac{\ln income_j}{\ln income_i}\right)$ and its row-normalized version, and $income_i$ denotes the state i's real personal income level in the year 2010. We will use W_n^e in the application part, and provide a justification on its specification.

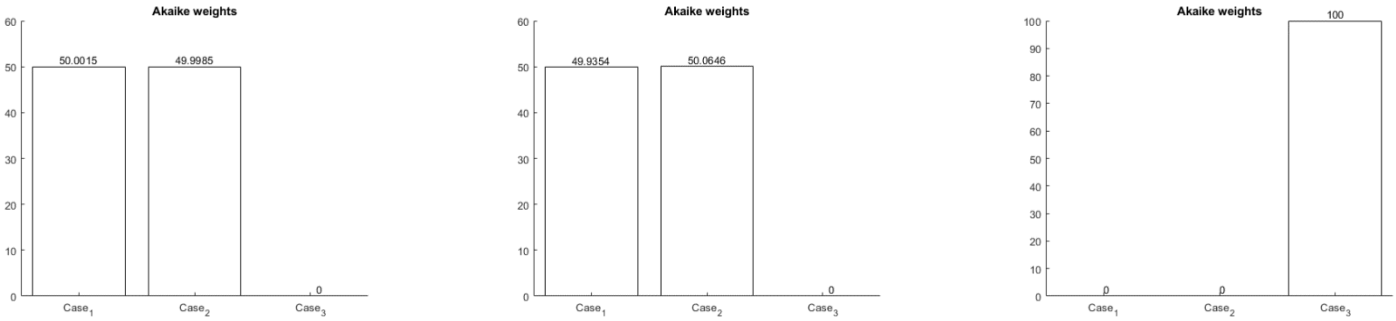
From W_n^e, three specifications are considered: (1) Case 1: $(W_n, M_n) = (W_n^e, W_n^e)$, (2) Case 2: LeSage and Pace's (2008) specification: $(W_n, M_n) = \left(W_n^e, W_n^{e\prime}\right)$, and (3) Case 3: LeSage and Pace's (2008) specification with row-normalization: $(W_n, M_n) = \left(W_{R,n}^e, W_{R,n}^{e\prime}\right)$, where $W_{R,n}^e = \left[w_{R,n,ij}^e\right]$ with $w_{R,n,ij}^e = \frac{w_{n,ij}^e}{\sum_{k=1}^n w_{n,ik}^e}$. By considering an asymmetric spatial weighting matrix, we make a difference between the case of $M_n = W_n^e$ and the LeSage and Pace's (2008) specification (i.e., $M_n = W_n^{e\prime}$). We measure the degree of matrix's asymmetry using $\frac{\|A-A'\|_2}{2\|A\|_2}$, where A is a square matrix. Using a matrix norm, this measure evaluates the normalized distance between A and A' and is located in $[0,1]$. We observe $\frac{\left\|W_n^e - W_n^{e\prime}\right\|_2}{2\|W_n^e\|_2} = 0.0001$ and $\frac{\|W_{R,n}^e - W_{R,n}^{e\prime}\|_2}{2\|W_{R,n}^e\|_2} = 0.2349$.[44] It implies that the row normalization for W_n^e generates a relatively high level of asymmetry. We generate the data for the three cases by considering $\theta_0 = (1, 0.02, 0.02, 0.01, -4, 1, 1, 1)'$ since this parameter vector provides the representative simulation results in Section 5.1 by generating sufficient zero values (about 17%). We estimate the model using the three candidate specifications.

For model comparison, we consider a measure based on the sample log-likelihood. The theoretical foundation of this model framework is the Akaike information criterion (AIC) (see Akaike, 1973). AIC measures the Kullback-Leibler divergence, which captures the distance between the (unknown) true model's distribution and that of a candidate model. For each DGP, we consider the quantity $\Delta_l = \exp\left((AIC_{true} - AIC_l)/2\right)$, where AIC_{true} denotes the AIC evaluated at the true model and AIC_l is the AIC evaluated at Case $l = 1,2$, and 3. This measure is the relative likelihood of the model l capturing the information loss from using the model l. The Akaike weight defined by $\bar{\Delta}_l = \frac{\Delta_l}{\sum_{r=1}^R \Delta_r}$, where R denotes the number of candidate models, represents the probability that model l minimizes the information loss among candidate models.[45] The Akaike weights are valuable indices in empirical applications as the true unknown AIC_{true} will be canceled from the numerator and the denominator. The figures below show how this measure is reasonable in determining a proper specification. We only report the results when the SARF Tobit model is taken. The simulation results for the linear SARF model are similar with the below.

[44] A baseline matrix norm can be replaced by other norms. For example, $\frac{\left\|W_n^e - W_n^{e\prime}\right\|_1}{2\|W_n^e\|_1} = 0.0001$, $\frac{\left\|W_n^e - W_n^{e\prime}\right\|_\infty}{2\|W_n^e\|_\infty} = 0.0001$, $\frac{\|W_{R,n}^e - W_{R,n}^{e\prime}\|_1}{2\|W_{R,n}^e\|_1} = 0.2674$ and $\frac{\|W_{R,n}^e - W_{R,n}^{e\prime}\|_\infty}{2\|W_{R,n}^e\|_\infty} = 0.5292$.

[45] For details, refer to Section 2.9 of Burnham and Anderson (2002).

Figure 3. Akaike weights



True DGP: Case 1 True DGP: Case 2 True DGP: Case 3

The numbers in the bar graphs show the averages of Akaike weights (multiplied by 100). In terms of average, the Akaike weight takes the highest value at each true model. When Case 1 or Case 2 is true, the Akaike weights from Cases 1 and 2 are similar while those of Case 3 are close to zeros. If the true DGP follows Case 3, the Akaike weights for Cases 1 and 2 are almost zeros. Those results imply that the Akaike weight is a reasonable measure in selecting a proper specification on (W_n, M_n) when each candidate specification generates distinct spatial influences.

6. Application: States' migration flow

In this section, we consider the migration flows among the 48 U.S. contiguous states (excluding Alaska and Hawaii) as an application. For each pair (i,j), $y_{n,ij}$ denotes the logged migration flows (added 1) from state j in year 2010 to state i in year 2011 (i.e., $y_{n,ij} = \ln(mflow_{ij} + 1)$). As univariate explanatory variables $\{x_{n,i}\}$, we consider the states' (1) logged population levels $\{x_{n,i,1}\}$, (2) percentage growth rates of per capita real personal incomes $\{x_{n,i,2}\}$ in year 2010, (3) insured unemployment rate $\{x_{n,i,3}\}$ in year 2010, (4) 5-year average housing burden ratios $\{x_{n,i,4}\}$ from year 2006 to year 2010, and (5) the logged degrees of nodes in the states' adjacency network $\{x_{n,i,5}\}$.[46] For $z_{n,ij,1}$, the logged (kilometer-based) geographic distances (d_{ij}) added 1 is employed. In addition to d_{ij}, demographic and/or economic distances are considered as z-variables. For this, we consider the income growth differential $|x_{n,i,2} - x_{n,j,2}|$ as $z_{n,ij,2}$, insured unemployment rate differential $|x_{n,i,3} - x_{n,j,3}|$ as $z_{n,ij,3}$, and housing burden ratio differential $|x_{n,i,4} - x_{n,j,4}|$ as $z_{n,ij,4}$.[47] All variables are collected from the U.S Census. For details of variable specifications, refer to the supplement file.

For combinations of spatial weighting matrices, six specifications are considered. First, we consider the shares of historical migration influxes and outflows (from 2009 to 2010). An $n \times n$ matrix $W_n^I = [w_{n,ij}^I]$ is designed to present forces toward destinations. Each entry $w_{n,ij}^I$ is the share of migration flow from j to i among migration influxes to i. To represent forces from origins, we consider an n-

[46] Since the states' adjacency matrix is symmetric, its outdegree is the same as its indegree. Hence, it is a degree.

[47] Taking the absolute value on $x_{n,i,k} - x_{n,j,k}$ generates additional variations relative to $x_{n,i,k}$ and $x_{n,j,k}$. Other types of z-variables can be considered to show an incentive of migrations (e.g., $x_{n,i,k} - x_{n,j,k}$ or $\frac{x_{n,i,k}}{x_{n,j,k}}$). In our estimation results, however, it is difficult to identify the relevant sensitivity parameter due to small variations in an alternative z-variable.

dimensional square matrix $M_n^O = [m_{n,ij}^O]$ whose (i,j)-element $m_{n,ij}^O$ is the share of migration flow from j to i among migration outflows from j. Those two matrices are directed networks and can show different roles of origins and destinations in spreading spatial spillover effects. For alternative specifications, we consider the states' adjacency matrix W_n^a and a matrix constructed by their economic relations W_n^e. We construct W_n^e by the economic distance $|income_i - income_j|$ and its product with $\frac{\ln(income_j)}{\ln(income_i)}$. The weight $\frac{\ln(income_j)}{\ln(income_i)}$ generates asymmetric influences. If $income_j > income_i$, a signal from j to i ($w_{n,ij}^e$) is larger than that from i to j (i.e., a larger effect from a higher income region). From W_n^a, W_n^e, and their row normalized versions, e.g., $W_{R,n}^a = [w_{R,n,ij}^a]$ with $w_{R,n,ij}^a = \frac{w_{n,ij}^a}{\sum_{k=1}^n w_{n,ik}^a}$, we have six specifications: (1) $(W_n, M_n) = (W_n^I, M_n^O)$, (2) $(W_n, M_n) = (W_n^a, W_n^a)$, (3) $(W_n, M_n) = (W_{R,n}^a, W_{R,n}^{a\prime})$ (i.e., LeSage and Pace's (2008) specification), (4) $(W_n, M_n) = (W_n^e, W_n^e)$, (5) $(W_n, M_n) = (W_n^e, W_n^{e\prime})$, and (6) $(W_n, M_n) = (W_{R,n}^e, W_{R,n}^{e\prime})$. In the supplement file, we report the estimation results when $(W_n, M_n) = (W_n^e, W_n^e)$ and $(W_n, M_n) = (W_n^e, W_n^{e\prime})$, i.e., the cases (4) and (5).

Then, we estimate the resource flow model discussed in Section 2.1. When $\lambda_0 \neq 0$ or $\gamma_0 \neq 0$ or $\rho_0 \neq 0$, there exist effects from third-party's characteristics. Then, $s_{inv,gh}^{ij}$ is the $((j-1)n + i, (h-1)n + g)$-element of S_N^{-1} which characterizes the spatial spillover effect. The weight $s_{inv,ij}^{ij}$ usually takes a number greater than one and shows how the spatial spillovers work to amplify the characteristics' effects (i.e., intensity of the multiplier effect). If there is no spatial spillover effect, $s_{inv,ij}^{ij} = 1$ for all (i,j) and $s_{inv,kl}^{ij} = 0$ for $(g,h) \neq (i,j)$ (i.e., the conventional gravity model).[48] For example, note that $\beta_{1,0}$ presents the elasticity of geographic distance d_{ij}. Then, $\beta_{1,0} s_{inv,ij}^{ij} > \beta_{1,0}$, where $\beta_{1,0} s_{inv,ij}^{ij}$ is the amplified elasticity of d_{ij}, if $s_{inv,ij}^{ij} > 1$ since $s_{inv,ij}^{ij}$ is a diagonal element of S_N^{-1}. If $(g,h) \neq (i,j)$, $s_{inv,gh}^{ij}$ represents how much the third parties (g,h)'s characteristics affect $mflow_{ij}$. We will report $\{s_{inv,gh}^{ij}\}$ after providing the estimation results.

Table 3. Descriptive statistics: States' migration flows

Variables	Mean	Std. dev.	Minimum	Maximum
Migration flows ($\times 10^4$)	1.9253	17.4002	0.0000	527.1168
Geographic dist. (km)	1662.0418	957.1254	60.9591	4283.9987
Population ($\times 10^6$)	6.3887	6.9237	0.5644	37.3277
Personal income growth (%)	1.4545	1.5102	-1.2748	7.2790
Insured unemployment rate (%)	3.3831	0.8768	1.2346	5.2431
Housing burden ratio (%)	33.8697	4.8457	26.3536	49.1117
Logged degree of W_n^a	1.4204	0.4151	0	2.0794

48 In the McCallum's (1995) gravity model framework with $L = 1$ and $K = 1$ for simplicity, we have
$$mflow_{ij} = \overline{mflow}(1 + d_{ij})^{\beta_{1,0}} pop_i^{b_{1,0}} pop_j^{c_{1,0}} \exp(\epsilon_{ij}),$$
where $\overline{mflow}$ is absorbed in a constant term for estimation.

Figure 4. Heatmap for the U.S. state level migration flows

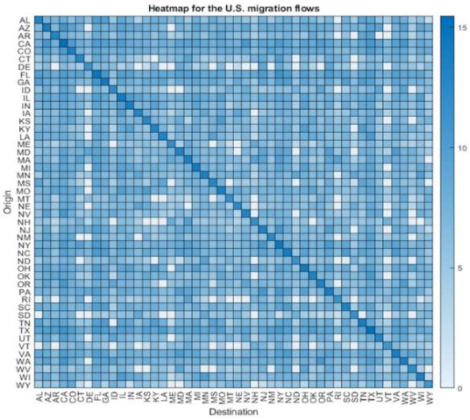


Table 3 shows the descriptive statistics. From the heatmap of the U.S. migration flows (Figure 4), we observe two features: (1) intrastate migration flows $\{y_{n,ii}\}$ are dominant over interstate ones $\{y_{n,ij}\}$ ($i \neq j$); (2) 93.88% observations are nonzero, there exist unignorable zeros of the states' migration flows. For the first feature, the traditional logistic specification (e.g., Sasser (2010)) is weak in explaining the dominating intrastate migration. It assumes that economic conditions in a state other than a given origin and destination pair have no impact on the migration choice. On the contrary, our model suggests that the impact of the characteristics of the origin and destination pair can be amplified by the third-party's characteristics via spatial spillover effects. We propose that local moves are more intense for three possible reasons: (1) the relocation costs, including the moving cost (measured by d_{ij}) and the information cost (by the demographic and economic distances), (2) the larger impact of a state i's kth characteristics $x_{n,i,k}$ on $y_{n,ii}$ (i.e., $b_{k,0} + c_{k,0}$); and (3) the amplified effects of (1) and (2) by the spatial multiplier effect ($s_{inv,gh}^{ii}$) characterized by the feedback effects $i \mapsto \cdots \mapsto i$ via spatial networks. For the second feature, a migration flow $mflow_{ij}$ is the aggregation of individuals' relocation decisions. An individual chooses to migrate to places only in the case that her utility is maximized by balancing the benefit to move and the cost to migrate. As a result, zero values of migration flows (about 6%) can occur when we aggregate individuals' choices. It motivates us to consider the SARF Tobit model.

Table 4. Estimation results I: States' migration flows

	Linear SARF				SARF Tobit			
Parameters\Specification	(W_n^I, M_n^O)	(W_n^a, W_n^a)	$(W_{R,n}^a, W_{R,n}^{a'})$	$(W_{R,n}^e, W_{R,n}^{e'})$	(W_n^I, M_n^O)	(W_n^a, W_n^a)	$(W_{R,n}^a, W_{R,n}^{a'})$	$(W_{R,n}^e, W_{R,n}^{e'})$
Constant	-9.0100***	-15.3024***	-16.5112***	-9.1665***	-9.5311***	-15.7306***	-7.9918***	-19.2525***
	[1.2075]	[0.8087]	[0.8475]	[0.9905]	[1.2565]	[0.8622]	[0.9046]	[1.0438]
λ_0	0.5725***	0.0545***	0.2516***	0.1961***	0.6044***	0.0556***	0.3532***	0.1177***
	[0.0400]	[0.0045]	[0.0228]	[0.0249]	[0.0410]	[0.0048]	[0.0236]	[0.0268]
γ_0	0.6221***	0.0520***	0.2910***	0.1436***	0.6250***	0.0498***	0.3875***	0.0538**
	[0.0385]	[0.0046]	[0.0228]	[0.0253]	[0.0406]	[0.0050]	[0.0234]	[0.0271]
ρ_0	-0.3283***	-0.0028**	-0.1004***	-0.4127***	-0.3194***	-0.0018	-0.2677***	-0.2487***
	[0.0848]	[0.0011]	[0.0326]	[0.0361]	[0.0833]	[0.0012]	[0.0340]	[0.0386]
$\beta_{1,0}$ $(d_{ij} \mapsto y_{n,ij})$	-0.6910***	-0.7214***	-0.6876***	-1.0709***	-0.6926***	-0.7140***	-0.7441***	-1.0093***
	[0.0249]	[0.0273]	[0.0293]	[0.0267]	[0.0261]	[0.0290]	[0.0313]	[0.0282]
$\beta_{2,0}$ $(\lvert x_{n,i,2} - x_{n,j,2}\rvert \mapsto y_{n,ij})$	0.0113	0.0302	0.0135	0.0100	0.0124	0.0169	-0.0235	0.0486*
	[0.0209]	[0.0222]	[0.0220]	[0.0240]	[0.0223]	[0.0236]	[0.0236]	[0.0254]
$\beta_{3,0}$ $(\lvert x_{n,i,3} - x_{n,j,3}\rvert \mapsto y_{n,ij})$	0.0577	0.0800**	0.0573	0.0234	0.0524	0.0484	0.0473	0.0514
	[0.0377]	[0.0401]	[0.0396]	[0.0440]	[0.0401]	[0.0427]	[0.0424]	[0.0465]
$\beta_{4,0}$ $(\lvert x_{n,i,4} - x_{n,j,4}\rvert \mapsto y_{n,ij})$	-0.0234***	-0.0409***	-0.0326***	-0.0185***	-0.0231***	-0.0361***	-0.0113	-0.0373***

	[0.0073]	[0.0077]	[0.0077]	[0.0084]	[0.0078]	[0.0082]	[0.0082]	[0.0089]
$b_{1,0}$ $(x_{n,i,1} \mapsto y_{n,ij})$	0.4325***	0.8414***	0.7805***	0.8344***	0.4587***	0.8926***	0.5355***	1.1341***
	[0.0550]	[0.0422]	[0.0424]	[0.0502]	[0.0584]	[0.0451]	[0.0449]	[0.0535]
$b_{2,0}$ $(x_{n,i,2} \mapsto y_{n,ij})$	0.0419**	0.0509**	0.0445**	0.0020	0.0448**	0.0396*	-0.0038	0.0321
	[0.0206]	[0.0218]	[0.0217]	[0.0234]	[0.0219]	[0.0232]	[0.0233]	[0.0248]
$b_{3,0}$ $(x_{n,i,3} \mapsto y_{n,ij})$	-0.0152	-0.0384	-0.0557	-0.0262	-0.0130	-0.0407	-0.0606	-0.0148
	[0.0377]	[0.0389]	[0.0389]	[0.0425]	[0.0400]	[0.0415]	[0.0416]	[0.0450]
$b_{4,0}$ $(x_{n,i,4} \mapsto y_{n,ij})$	0.0029	0.0200	0.0099	-0.0030	0.0007	0.0104	0.0054	-0.0029
	[0.0083]	[0.0089]	[0.0088]	[0.0097]	[0.0088]	[0.0095]	[0.0094]	[0.0103]
$b_{5,0}$ $(x_{n,i,5} \mapsto y_{n,ij})$	-0.0108	-0.9941***	-0.0118	-0.1415	-0.0134	-1.1996***	-0.1532	-0.0091
	[0.0740]	[0.1375]	[0.0777]	[0.0853]	[0.0786]	[0.1465]	[0.0833]	[0.0901]
$c_{1,0}$ $(x_{n,j,1} \mapsto y_{n,ij})$	0.4484***	0.7771***	0.7762***	0.6990***	0.4374***	0.7836***	0.5332***	0.9865***
	[0.0550]	[0.0417]	[0.0420]	[0.0491]	[0.0572]	[0.0444]	[0.0446]	[0.0522]
$c_{2,0}$ $(x_{n,j,2} \mapsto y_{n,ij})$	0.0537**	0.0525**	0.0561***	0.0029	0.0556**	0.0391*	0.0162	0.0424*
	[0.0209]	[0.0218]	[0.0217]	[0.0234]	[0.0222]	[0.0232]	[0.0233]	[0.0248]
$c_{3,0}$ $(x_{n,j,3} \mapsto y_{n,ij})$	0.0233	0.0139	-0.0107	0.0500	0.0314	0.0366	-0.0160	0.0695
	[0.0371]	[0.0389]	[0.0389]	[0.0423]	[0.0394]	[0.0414]	[0.0416]	[0.0448]
$c_{4,0}$ $(x_{n,j,4} \mapsto y_{n,ij})$	0.0137	0.0399***	0.0273***	0.0200**	0.0121	0.0406***	0.0203**	0.0216**
	[0.0084]	[0.0089]	[0.0088]	[0.0096]	[0.0089]	[0.0095]	[0.0094]	[0.0102]
$c_{5,0}$ $(x_{n,j,5} \mapsto y_{n,ij})$	0.2114**	-0.6951***	0.2125***	0.1451*	0.2334***	-0.7289***	0.0708	0.3125***
	[0.0754]	[0.1357]	[0.0783]	[0.0855]	[0.0801]	[0.1446]	[0.0838]	[0.0906]
σ_0^2	1.5960***	1.7990***	1.7622***	2.1139***	1.7963***	2.0268***	2.0097***	2.3533***
	[0.0488]	[0.0533]	[0.0529]	[0.0627]	[0.0582]	[0.0635]	[0.0643]	[0.0739]
Log-likelihood	-3867.5132	-3977.4477	-3964.3566	-4135.4614	-3969.7881	-4078.7266	-4114.9061	-4218.8884
Akaike weight× 100	100.0000	0.0000	0.0000	0.0000	100.0000	0.0000	0.0000	0.0000

Note: Theoretical standard deviations are in parenthesis. Estimates that are significant at the 10%, 5%, and 1% levels are respectively marked by "*", "**", and "***".

Table 4 reports the estimation results for the linear SARF and SARF Tobit models. The Akaike weights suggest that the first setting $(W_n, M_n) = (W_n^I, M_n^O)$ is the best for both models. We observe that the estimates from the SARF Tobit model tend to be larger than those from the linear SARF model (in absolute values). However, the two models (linear SARF and SARF Tobit) yield the same sign and similar significance of estimates. For interpretations, we focus on the estimates from the SARF Tobit model. First, by the estimates for λ_0 and γ_0, we detect significant positive spatial effects from flows $y_{n,gj}$ (for $g \neq i$) and $y_{n,ih}$ (for $h \neq j$) on $y_{n,ij}$. It implies that the overall inflow (to destination i) or outflow (from origin j) tendency has a positive influence on $y_{n,ij}$. For example, the migration flow from Ohio to Indiana increases if that from Ohio to a third-party connected state (e.g., Kentucky) or that from a third-party state to Indiana increases. Second, for the parameter ρ_0, a negative effect of migration flows among third-party states on $y_{n,ij}$ is captured.[49] Third, the geographic distance has significantly negative effect on a state's migration flow $y_{n,ij}$. The geographic-distance-elasticity of the migration flow is -0.7030. The effects of the income growth difference $(z_{n,ij,2})$ and relative labor market condition $(z_{n,ij,3})$ on $y_{n,ij}$ are not significantly captured. The housing burden ratio differential $(z_{n,ij,4})$ negatively affects $y_{n,ij}$.

Fourth, the estimates for $b_{1,0}$, $b_{2,0}$ $c_{1,0}$, and $c_{2,0}$ show that the logged population levels and the personal income growths of both origin and destination significantly positively affect $y_{n,ij}$. The effect of origin's population is larger than that of destination's population. The importance of the origin state' network connections on a migration flow is significantly identified (the estimate of $c_{5,0}$). Last. we reject

49 In the supplement file (Section 3.2.1), we report the estimation results when the averages of inflows and outflows ($y_{n,i.}$ and $y_{n.j}$) are considered as dependent variables (with the conventional SAR model). We do not capture significant spatial influences among those averages of flows. It seems that the aggregation leads to losing information in spatial influences among the origin-destination flows.

the hypothesis $H_0\colon b_{k,0} = c_{k,0}$ for $k = 5$ (logged degrees of W_n^a).

Table 5. Estimation results II (with the fixed-effect specification): States' migration flows

	Linear SARF				SARF Tobit			
Parameters\Specification	(W_n^I, M_n^O)	(W_n^a, W_n^a)	$(W_{R,n}^a, W_{R,n}^{a\prime})$	$(W_{R,n}^e, W_{R,n}^{e\prime})$	(W_n^I, M_n^O)	(W_n^a, W_n^a)	$(W_{R,n}^a, W_{R,n}^{a\prime})$	$(W_{R,n}^e, W_{R,n}^{e\prime})$
λ_0	0.1505***	0.0377***	0.1299***	0.0685***	0.2198***	0.1237***	0.1377***	0.0484*
	[0.0567]	[0.0052]	[0.0262]	[0.0267]	[0.0599]	[0.0121]	[0.0278]	[0.0282]
γ_0	0.1448***	0.0417***	0.2157***	-0.0084	0.5059***	0.1317***	0.2249***	-0.0417
	[0.0552]	[0.0053]	[0.0254]	[0.0269]	[0.0554]	[0.0113]	[0.0271]	[0.0283]
ρ_0	-0.2287***	0.0026**	0.1338***	-0.1920***	0.8670***	-0.0102***	0.1506***	-0.1712***
	[0.0733]	[0.0012]	[0.0401]	[0.0546]	[0.1312]	[0.0003]	[0.0427]	[0.0571]
$\beta_{1,0}$ $(d_{ij} \mapsto y_{n,ij})$	-0.7870***	-0.8051***	-0.7458***	-1.1089***	-0.7693***	-0.0091	-0.7315***	-1.1708***
	[0.0253]	[0.0292]	[0.0313]	[0.0259]	[0.0260]	[0.0205]	[0.0332]	[0.0275]
$\beta_{2,0}$ $(\lvert x_{n,i,2} - x_{n,j,2}\rvert \mapsto y_{n,ij})$	0.0299	0.0327	0.0201	0.0650**	0.0229	0.2282***	0.0150	0.0761**
	[0.0299]	[0.0306]	[0.0303]	[0.0321]	[0.0316]	[0.0370]	[0.0323]	[0.0342]
$\beta_{3,0}$ $(\lvert x_{n,i,3} - x_{n,j,3}\rvert \mapsto y_{n,ij})$	-0.0185	-0.0188	-0.0257	0.0014	-0.0247	0.0426	-0.0397	0.0056
	[0.0464]	[0.0475]	[0.0470]	[0.0498]	[0.0488]	[0.0563]	[0.0499]	[0.0529]
$\beta_{4,0}$ $(\lvert x_{n,i,4} - x_{n,j,4}\rvert \mapsto y_{n,ij})$	-0.0191**	-0.0215**	-0.0189**	-0.0218**	-0.0139	-0.0421***	-0.0223**	-0.0271***
	[0.0091]	[0.0094]	[0.0092]	[0.0098]	[0.0096]	[0.0110]	[0.0098]	[0.0104]
σ_0^2	1.5315***	1.5454***	1.5103***	1.7084***	1.6674***	2.0201***	1.7429***	2.0501***
	[0.0422]	[0.0440]	[0.0430]	[0.0484]	[0.0489]	[0.0574]	[0.0509]	[0.0575]
Log-likelihood	-3697.0327	-3741.8366	-3717.7547	-3842.8731	-3790.7025	-5270.1536	-3828.8817	-3950.8906
Akaike weight× 100	100.0000	0.0000	0.0000	0.0000	100.0000	0.0000	0.0000	0.0000

Note: Theoretical standard deviations are in parenthesis. Estimates that are significant at the 10%, 5%, and 1% levels are respectively marked by "*", "**", and "***".

Also, we consider the SARF model with the fixed-effect specification. The bias corrected ML estimates are provided in Table 5. Even for the specifications with fixed effects, we observe that (W_n^I, M_n^O) is the best. After controlling for some time-invariant state characteristics, the estimates of λ_0, γ_0 and ρ_0 tend to become smaller in absolute values. Significant negative impact of the bilateral distance on $y_{n,ij}$ is found. When the difference between the housing burden ratio of the origin state (j) and the destination state (i) increases, the linear SARF model with fixed effects captures that the outflow of migrants from j to i decreases. We do not detect significant effects of other characteristics.

Table 6. Equilibrium effects

	Linear SARF $s_{inv,ij}^{ij}$	Linear SARF $s_{inv,gh}^{ij}$ with $(g,h) \neq (i,j)$	Linear SARF + Fixed-effect $s_{inv,ij}^{ij}$	Linear SARF + Fixed-effect $s_{inv,gh}^{ij}$ with $(g,h) \neq (i,j)$
Mean	1.0607	0.0028	1.0024	0.0002
25th-percentile	1.0416	0.0003	1.0017	0.0000
Median	1.0556	0.0008	1.0023	0.0000
75th-percentile	1.0739	0.0019	1.0030	0.0000
	SARF Tobit $s_{inv,ij}^{ij}$	SARF Tobit $s_{inv,gh}^{ij}$ with $(g,h) \neq (i,j)$	SARF Tobit + Fixed-effect $s_{inv,ij}^{ij}$	SARF Tobit + Fixed-effect $s_{inv,gh}^{ij}$ with $(g,h) \neq (i,j)$
Mean	1.0675	0.0044	1.0164	0.0000
25th-percentile	1.0454	0.0007	1.0098	0.0000
Median	1.0615	0.0016	1.0161	0.0000
75th-percentile	1.0828	0.0039	1.0245	0.0068

Table 6 reports the statistics of $\{s_{inv,gh}^{ij}\}$ showing the equilibrium effect under $(W_n, M_n) = (W_n^I, M_n^O)$. The multiplier effects from the SARF Tobit models ($\{s_{inv,ij}^{ij}\}$) are greater than those from the linear specifications. When we include fixed effects in the models, the multiplier effects become smaller. The off-diagonal elements of S_N^{-1} (i.e., $\{s_{inv,gh}^{ij}\}_{(g,h)\neq(i,j)}$) are much less than the diagonal elements of S_N^{-1}

($\{s_{inv,ij}^{ij}\}$). It implies that only a small number of third-party units have significant influences on $mflow_{ij}$.

7. Conclusion

We develop a spatial autoregressive model for an origin-destination flow dependent variable (SARF) with estimation methods. Using a similar structure of flow data with a panel data set, our model accommodates the two-way fixed effect specification. For a specific data environment of a flow variable with possible zero flow, we also consider the SARF Tobit model. We study the asymptotic properties of the maximum likelihood (ML) estimator (quasi-ML estimator for the linear SARF model). To establish the asymptotic properties of the MLE for the SARF Tobit model, the near-epoch dependence concept developed by Jenish and Prucha (2012) is employed. Monte Carlo simulation results are provided to deliver finite sample properties of the ML and QML estimators. Last, we apply our models to the U.S. states' migration flows. We detect significant spatial influences from neighboring migration flows.

Appendix. Mathematical proofs

Throughout this section, we will use the following notations. Recall that $c_{w,c,j} = \sum_{i=1}^{n} w_{n,ij}$ and $c_{m,c,j} = \sum_{i=1}^{n} m_{n,ij}$ for each j; and $c_{w,r,i} = \sum_{j=1}^{n} w_{n,ij}$ and $c_{m,r,i} = \sum_{j=1}^{n} m_{n,ij}$ for each i. Note that all elements $w_{n,ij}$ and $m_{n,ij}$ are nonnegative by construction. Then, $c_{w,c} = \sup_n \max_{j=1,\cdots,n} c_{w,c,j} = \sup_n \|W_n\|_1$, $c_{m,c} = \sup_n \max_{j=1,\cdots,n} c_{m,c,j} = \sup_n \|M_n\|_1$, $c_{w,r} = \sup_n \max_{i=1,\cdots,n} c_{w,r,i} = \sup_n \|W_n\|_\infty$, and $c_{m,r} = \sup_n \max_{i=1,\cdots,n} c_{m,r,i} = \sup_n \|M_n\|_\infty$.

A. Spatial stability

Here we introduce more details about spatial stability for the SARF model. Note that the cross product $(I_n \otimes W_n)(M_n' \otimes I_n) = M_n' \otimes W_n$ by Kronecker mixed product rule. We assume that W_n and M_n are diagonalizable, i.e., $W_n = \Gamma_{1n}\Lambda_{1n}\Gamma_{1n}^{-1}$ and $M_n = \Gamma_{2n}\Lambda_{2n}\Gamma_{2n}^{-1}$, where $\Lambda_{j,n} = \mathrm{diag}(\varpi_{jn,1}, \dots, \varpi_{jn,n}), j = 1,2$ are diagonal matrices of eigenvalues and corresponding Γ_{jn} are eigenvector matrices. As $M_n' = (\Gamma_{2n}\Lambda_{2n}\Gamma_{2n}^{-1})' = \Gamma_{2n}'^{-1}\Lambda_{2n}\Gamma_{2n}'$, so M_n' has the same eigenvalue matrix but its eigenvector matrix is $\Gamma_{2n}'^{-1}$. The eigenvalues of $\lambda(I_n \otimes W_n) + \gamma(M_n' \otimes I_n) + \rho(M_n' \otimes W_n)$ are in the subsequent Claim A.1.

Claim A.1. An eigenvalue of $\lambda(I_n \otimes W_n) + \gamma(M_n' \otimes I_n) + \rho(M_n' \otimes W_n)$ is $\lambda\varpi_{1n,i} + \gamma\varpi_{2n,j} + \rho\varpi_{2n,j}\varpi_{1n,i}$ for $i, j = 1, \dots, n$.

Proof of Claim A.1. Let $x_{1n,i}$ be the ith eigenvector corresponding to $\varpi_{1n,i}$ of W_n, and $x_{2n,j}$ be the jth eigenvector corresponding to $\varpi_{2n,j}$ of M_n'. For arbitrary i and j,

$$
\begin{aligned}
&[\lambda(I_n \otimes W_n) + \gamma(M_n' \otimes I_n) + \rho(M_n' \otimes W_n)](x_{2j} \otimes x_{1i}) \\
&= \lambda(x_{2n,j} \otimes W_n x_{1n,i}) + \gamma(M_n' x_{2n,j} \otimes x_{1n,i}) + \rho(M_n' x_{2n,j} \otimes W_n x_{1n,i}) \\
&= (\lambda\varpi_{1n,i} + \gamma\varpi_{2n,j} + \rho\varpi_{2n,j}\varpi_{1n,i})(x_{2n,j} \otimes x_{1n,i})
\end{aligned}
$$

Thus, we have the claimed result. ■

For spatial stability, the parameter space of the stable model can be

$$\{(\lambda, \gamma, \rho) : |\lambda \varpi_{1n,i} + \gamma \varpi_{2n,j} + \rho \varpi_{2n,j} \varpi_{1n,i}| < 1, \text{for all } i, j = 1, \dots, n\}.$$

This implies that, with $\delta = (\lambda, \gamma, \rho)'$ in this parameter space,

$$\det(S_N(\delta)) = \prod_{i,j=1}^{n} \left(1 - (\lambda \varpi_{1n,i} + \gamma \varpi_{2n,j} + \rho \varpi_{2n,j} \varpi_{1n,i})\right) > 0.$$

When $\rho = -\lambda\gamma$, $S_N(\delta) = I_N - \lambda(I_n \otimes W_n) - \gamma(M_n' \otimes I_n) - \rho(M_n' \otimes W_n) = (I_N - \lambda(I_n \otimes W_n))(I_N - \gamma(M_n' \otimes I_n))$. This would be a separable spatial filter case (LeSage and Pace, 2008). Then the eigenvalues of $S_N(\delta)$ can be factorized into $(1 - \lambda\varpi_{1n,i})(1 - \varpi_{2n,j}), i, j = 1, \dots n$ and $\det(S_N(\delta)) = \prod_{i,j=1}^{n} (1 - \lambda\varpi_{1n,i})(1 - \gamma\varpi_{2n,j})$.

B. Model's coherency

Now we consider model's coherency for the SARF Tobit model. Note that $\boldsymbol{A}_N$ has zero diagonal elements due to excluding self-influence, i.e., $w_{n,ii} = 0$ and $m_{n,jj} = 0$ for $i, j = 1, \dots, n$, and under Assumption 3.1, $\|\boldsymbol{A}_N\|_\infty = \max_f \sum_{f'=1}^{N} \boldsymbol{a}_{ff'} \leq \zeta < 1$, where $\boldsymbol{a}_{ff'}$ denotes the (f, f')-element of $|\boldsymbol{A}_N|$. By spectral radius theorem, for any $r \times r$ principal submatrix $\boldsymbol{A}_{N,r}$ of $\boldsymbol{A}_N$, we have $\max_i |\varphi_i(\boldsymbol{A}_{N,r})| \leq \|\boldsymbol{A}_{N,r}\|_\infty \leq \|\boldsymbol{A}_N\|_\infty \leq \zeta$, where $\varphi_i(\boldsymbol{A}_{N,r})$ is the ith characteristic root of $\boldsymbol{A}_{N,r}$. For each r, let $\varrho_{i,r} = \varphi_i(I_r - \boldsymbol{A}_{N,r})$ denote the ith eigenvalue of $I_r - \boldsymbol{A}_{N,r}$. Then, $\varrho_{i,r} = 1 - \varphi_i(\boldsymbol{A}_{N,r})$. If $\varphi_i(\boldsymbol{A}_{N,r})$ is real, $\varrho_{i,r} = 1 - \varphi_i(\boldsymbol{A}_{N,r}) \geq 1 - \zeta$. If $\varrho_{i,r}$ is complex (i.e., $\varphi_i(\boldsymbol{A}_{N,r}) = c_{R,r,i} + \mathbf{i} \cdot c_{I,r,i}$ where $c_{R,r,i}, c_{I,r,i} \in \mathbb{R}$, and $\mathbf{i} = \sqrt{-1}$), its conjugate $\bar{\varrho}_{i,r}$ is also an eigenvalue of $I_r - \boldsymbol{A}_{N,r}$. Then,

$$\begin{aligned}
\varrho_i \bar{\varrho}_i = (1 - c_{R,r,i} - \mathbf{i} \cdot c_{I,r,i})(1 - c_{R,r,i} + \mathbf{i} \cdot c_{I,r,i}) &= 1 - 2c_{R,r,i} + c_{R,r,i}^2 + c_{I,r,i}^2 \\
&\geq 1 - 2\sqrt{c_{R,r,i}^2 + c_{I,r,i}^2} + c_{R,r,i}^2 + c_{I,r,i}^2 \\
&\geq \left(1 - \max_i |\varphi_i(\boldsymbol{A}_{N,r})|\right)^2 \geq (1 - \zeta)^2
\end{aligned}$$

since $|\varphi_i(\boldsymbol{A}_{N,r})| = \sqrt{c_{R,r,i}^2 + c_{I,r,i}^2} \leq \zeta$. Thus, we have the corresponding principal minor $|I_r - \boldsymbol{A}_{N,r}| \geq (1 - \zeta)^r > 0$.

C. Asymptotic properties of the MLE for the SARF Tobit model

C.1. Consistency of the MLE

To prove consistency, the following propositions will be employed. First, consider the sequences $\{y_{n,ij}\}$, $\{\sum_{g=1}^{n} w_{n,ig} y_{n,gj}\}$, $\{\sum_{h=1}^{n} y_{n,ih} m_{n,hj}\}$, $\{\sum_{g=1}^{n} \sum_{h=1}^{n} w_{n,ig} y_{n,gh} m_{n,hj}\}$, $\{\epsilon_{n,ij}^*(\theta)\}$, and $\{y_{n,ij}^*\}$, where

$y^*_{n,ij} = \lambda_0 \sum_{g=1}^n w_{n,ig} y_{n,gj} + \gamma_0 \sum_{h=1}^n y_{n,ih} m_{n,hj} + \rho_0 \sum_{g=1}^n \sum_{h=1}^n w_{n,ig} y_{n,gh} m_{n,hj} + \boldsymbol{x}_{n,ij}\kappa_0 + \epsilon_{n,ij}$ as in (6) inside $F(\cdot)$, $\epsilon^*_{n,ij}(\theta) = \left(y_{n,ij} - \lambda \sum_{g=1}^n w_{n,ig} y_{n,gj} - \gamma \sum_{h=1}^n y_{n,ih} m_{n,hj} - \rho \sum_{g=1}^n \sum_{h=1}^n w_{n,ig} y_{n,gh} m_{n,hj} - \boldsymbol{x}_{n,ij}\kappa\right)/\sigma$, and $\boldsymbol{x}_{n,ij} = \left(1, z_{n,ij,1}, \cdots, z_{n,ij,L}, x_{n,i,1}, \cdots, x_{n,i,K}, x_{n,j,1}, \cdots, x_{n,j,K}\right)$ (i.e., $\boldsymbol{x}_{n,ij} = \boldsymbol{x}_{N,f}$ with $f = (j-1)n + i$ is the fth row of $\mathbf{X}_N$). For the propositions below, the notation $A \leq^* B$, where $A = \left[a_{ff'}\right]$ and $B = \left[b_{ff'}\right]$ means $\left|a_{ff'}\right| \leq \left|b_{ff'}\right|$ for all $f, f' = 1, \cdots, N$, and the notation $|A|$ for a matrix $A = \left[a_{ff'}\right]$ means $|A| = \left[\left|a_{ff'}\right|\right]$.

Before establishing Proposition C.1, we introduce the following lemma.

Lemma C.1. Let $\boldsymbol{a}_{n,(i,j),(g,h)}$ be the (f, f')-element of $|\boldsymbol{A}_N|$, where $f = (j-1)n + i$ and $f' = (h-1)n + g$. Assume that the model's spatial stability and coherency hold. Under Assumption 4.2 (iii-1), $\boldsymbol{a}_{ff'} > 0$ can be only if $d_F\left((i,j),(g,h)\right) \leq \bar{d}$ and $\boldsymbol{a}_{n,(i,j),(g,h)} = 0$ otherwise. Then, $\sum_{f'=1}^N \sum_{l=1}^\infty [|\boldsymbol{A}_N|^l]_{ff'} \leq \sum_{l=\left[s/\bar{d}\right]+1}^\infty \zeta^l \to 0$ as $s \to \infty$, where $\left[s/\bar{d}\right]$ is the biggest integer that is less or equal than $s/\bar{d}$. Under Assumption 4.2 (iii-2), $\boldsymbol{a}_{n,(i,j),(g,h)} \leq \tilde{C}_0 d_F\left((i,j),(g,h)\right)^{-a}$ for some $\tilde{C}_0 > 0$. Moreover, $\sum_{l=1}^\infty [|\boldsymbol{A}_N|^l]_{ff'} \leq C_1 d_F\left((i,j),(g,h)\right)^{-a}$ for some $C_1 > 0$.

Proof of Lemma C.1. Note that $\left(e'_{n,j} \otimes e'_{n,i}\right)(I_n \otimes W_n)\left(e_{n,h} \otimes e_{n,g}\right) = e'_{n,j} e_{n,j} w_{n,ig}$, $\left(e'_{n,j} \otimes e'_{n,i}\right)(M'_n \otimes I_n)\left(e_{n,h} \otimes e_{n,g}\right) = m_{n,hj} e'_{n,i} e_{n,g}$, and $\left(e'_{n,j} \otimes e'_{n,i}\right)(M'_n \otimes W_n)\left(e_{n,h} \otimes e_{n,g}\right) = w_{n,ig} m_{n,hj}$. Then,

$$\boldsymbol{a}_{n,(i,j),(g,h)} = |\lambda_0| 1(j = h) w_{n,ig} + |\gamma_0| 1(i = g) m_{n,hj} + |\rho_0| w_{n,ig} m_{n,hj},$$

so there exist four cases for $\boldsymbol{a}_{n,(i,j),(g,h)}$. The column sum vector of $\boldsymbol{A}_N$ is

$$\lambda_0 l'_n \otimes \left(c_{w,c,1}, \cdots, c_{w,c,n}\right) + \gamma_0 \left(c_{m,r,1}, \cdots, c_{m,r,n}\right) \otimes l'_n + \rho_0 \left(c_{m,r,1}, \cdots, c_{m,r,n}\right) \otimes \left(c_{w,c,1}, \cdots, c_{w,c,n}\right),$$

so $\|\boldsymbol{A}_N\|_1 \leq |\lambda_0| c_{w,c} + |\gamma_0| c_{m,r} + |\rho_0| c_{w,c} c_{m,r} = \Gamma < \infty$. Note that there exist the n same column sum components in the first part.

Case 1, Assumption 4.2 (iii-1): Suppose Assumption 4.2 (iii-1) holds. First, if $i = g$ and $j = h$, we have $\boldsymbol{a}_{n,(i,j),(g,h)} = 0$. Second, when $i \neq g$ and $j = h$, $\boldsymbol{a}_{n,(i,j),(g,h)} = |\lambda_0| w_{n,ig} > 0$ only if $d_F\left((i,j),(g,h)\right) \leq \bar{d}$ since $d_F\left((i,j),(g,h)\right) = \max\{d(i,j), d(g,h)\} = d(i,g) \leq \bar{d}$. Third, when $i = g$ and $j \neq h$, $\boldsymbol{a}_{n,(i,j),(g,h)} = |\gamma_0| m_{n,hj} > 0$ only if $d_F\left((i,j),(g,h)\right) \leq \bar{d}$ since $d_F\left((i,j),(g,h)\right) = \max\{d(i,g), d(j,h)\} = d(j,h) \leq \bar{d}$. Fourth, when $i \neq g$ and $j \neq h$, $\boldsymbol{a}_{n,(i,j),(g,h)} = |\rho_0| w_{n,ig} m_{n,hj} > 0$ only if $d_F\left((i,j),(g,h)\right) \leq \bar{d}$, because both $w_{n,ig} > 0$ and $m_{n,hj} > 0$ only if both $d(i,g) \leq \bar{d}$ and $d(j,h) \leq \bar{d}$, then $d_F\left((i,j),(g,h)\right) = \max\{d(i,g), d(j,h)\} \leq \bar{d}$. Hence, we can show that $\boldsymbol{a}_{n,(i,j),(g,h)} > 0$ only if $d_F\left((i,j),(g,h)\right) \leq \bar{d}$ and $\boldsymbol{a}_{n,(i,j),(g,h)} = 0$ otherwise.

For some large $s > 0$, we observe

$\sum_{f'=1}^N \sum_{l=1}^\infty [|\boldsymbol{A}_N|^l]_{ff'}$

$$= \sum_{l=\left[s/_{\bar{d}}\right]+1}^{\infty} \sum_{g,h=1}^{n} \sum_{i_1,j_1=1}^{n} \sum_{i_2,j_2=1}^{n} \cdots \sum_{i_{l-1},j_{l-1}=1}^{n} \boldsymbol{a}_{n,(i,j),(i_1,j_1)} \boldsymbol{a}_{n,(i_1,j_1),(i_2,j_2)} \cdots \boldsymbol{a}_{n,(i_{l-1},j_{l-1}),(g,h)}$$
$$\leq \sum_{l=\left[s/_{\bar{d}}\right]+1}^{\infty} \zeta^l.$$

The first equality comes from Claim C.2.3 of Qu and Lee (2015). By Assumption 3.1, the following inequality holds because

$$\begin{aligned}
&\sum_{g,h=1}^{n} \sum_{i_1,j_1=1}^{n} \sum_{i_2,j_2=1}^{n} \cdots \sum_{i_{l-1},j_{l-1}=1}^{n} \boldsymbol{a}_{n,(i,j),(i_1,j_1)} \boldsymbol{a}_{n,(i_1,j_1),(i_2,j_2)} \cdots \boldsymbol{a}_{n,(i_{l-1},j_{l-1}),(g,h)} \\
&\quad = \sum_{i_1,j_1=1}^{n} \sum_{i_2,j_2}^{n} \cdots \sum_{i_{l-1},j_{l-1}=1}^{n} \boldsymbol{a}_{n,(i,j),(i_1,j_1)} \cdots \boldsymbol{a}_{n,(i_{l-2},j_{l-2}),(i_{l-1},j_{l-1})} \underbrace{\sum_{g,h=1}^{n} \boldsymbol{a}_{n,(i_{l-1},j_{l-1}),(g,h)}}_{\leq\zeta} \\
&\quad \leq \zeta \sum_{i_1,j_1=1}^{n} \cdots \sum_{i_{l-2},j_{l-2}=1}^{n} \boldsymbol{a}_{n,(i,j),(i_1,j_1)} \cdots \boldsymbol{a}_{n,(i_{l-3},j_{l-3}),(i_{l-2},j_{l-2})} \underbrace{\sum_{i_{l-1},j_{l-1}=1}^{n} \boldsymbol{a}_{n,(i_{l-2},j_{l-2}),(i_{l-1},j_{l-1})}}_{\leq\zeta} \\
&\quad \leq \cdots \leq \zeta^l.
\end{aligned}$$

Then, $\sum_{f'=1}^{N} \sum_{l=1}^{\infty} [|\boldsymbol{A}_N|^l]_{ff'} \leq \sum_{l=\left[s/_{\bar{d}}\right]+1}^{\infty} \zeta^l \leq (1-\zeta)^{-1} \zeta^{s/_{\bar{d}}} \to 0$ as $s \to \infty$.

Case 2, Assumption 4.2 (iii-2): Consider the case of Assumption 4.2 (iii-2). First, $\boldsymbol{a}_{n,(i,j),(g,h)} = 0$ if $i = g$ and $j = h$. Second, $\boldsymbol{a}_{n,(i,j),(g,h)} = |\lambda_0| w_{n,ig}$ if $i \neq g$ and $j = h$. Then, $\boldsymbol{a}_{n,(i,j),(g,h)} = |\lambda_0| w_{n,ig} \leq |\lambda_0| C_0 d_F\big((i,j),(g,h)\big)^{-a}$ since $d_F\big((i,j),(g,h)\big) = \max\{d(i,g), d(j,h)\} = d(i,g)$. Third, $\boldsymbol{a}_{n,(i,j),(g,h)} = |\gamma_0| m_{n,hj} \leq |\gamma_0| C_0 d_F\big((i,j),(g,h)\big)^{-a}$ if $i = g$ and $j \neq h$. Last, if $i \neq g$ and $j \neq h$, $\boldsymbol{a}_{n,(i,j),(g,h)} \leq |\rho_0| w_{n,ig} m_{n,hj} \leq |\rho_0| C_0^2 d(i,g)^{-a} d(j,h)^{-a} \leq |\rho_0| C_0^2 d_F\big((i,j),(g,h)\big)^{-a}$ since $d(i,g) \geq 1$ and $d(j,h) \geq 1$ with $a > 1$. Hence, we have $\boldsymbol{a}_{n,(i,j),(g,h)} \leq \tilde{C}_0 d_F\big((i,j),(g,h)\big)^{-a}$ for some $\tilde{C}_0 > 0$.

As the next step, we will show $\left\|\boldsymbol{A}_N^l\right\|_1 \leq lK\Gamma\zeta^{l-1}$ for $l \in \mathbb{Z}_+$, where K is a positive integer that does not depend on n. First, if $c_{w,c} \leq c_{w,r}$, $\left\|\boldsymbol{A}_N^l\right\|_1 \leq \big(|\lambda_0| c_{w,r} + |\gamma_0| c_{m,r} + |\rho_0| c_{w,r} c_{m,r}\big)^l \leq \zeta^l$. Consider the case of $c_{w,c} > c_{w,r}$. Then, we have $\left\|W_n^p\right\|_1 \leq p c_{w,c} K_W c_{w,r}^{p-1}$ for $p \in \mathbb{Z}_+$ by Claim C.1.2 of Qu and Lee (2015). For $l = 2,3,4,\cdots$, by the triangle inequality, we have

$$\begin{aligned}
\left\|\boldsymbol{A}_N^l\right\|_1 &\leq \sum_{p+q+r=l} \frac{l!}{p!q!r!} |\lambda_0|^p |\gamma_0|^q |\rho_0|^r \left\|W_n^{p+r}\right\|_1 \|M_n\|_\infty^{q+r} \\
&\leq \sum_{p+q+r=l} \frac{l!}{p!q!r!} |\lambda_0|^p |\gamma_0|^q |\rho_0|^r (p+r) K_W c_{w,c} c_{w,r}^{p+r-1} c_{m,r}^{q+r} \\
&\leq lK_W \bar{c} \left(\underbrace{\sum_{p+q+r=l} \frac{l!}{p!q!r!} |\lambda_0|^p |\gamma_0|^q |\rho_0|^r c_{w,r}^{p+r} c_{m,r}^{q+r}}_{=\left(|\lambda_0| c_{w,r} + |\gamma_0| c_{m,r} + |\rho_0| c_{w,r} c_{m,r}\right)^l} \right) \leq lK\Gamma\zeta^{l-1},
\end{aligned}$$

where K is a positive constant satisfying $K_W \bar{c} \zeta \leq K\Gamma$ and $\bar{c} > 1$ such that $c_{w,c} = \bar{c} c_{w,r}$. The second and third inequalities hold since $\left\|W_n^{p+r}\right\|_1 \leq (p+r) K_W c_{w,c} c_{w,r}^{p+r-1} = (p+r) K_W \bar{c} c_{w,r}^{p+r}$, and $\frac{l!}{p!q!r!}(p + r) \leq l \frac{l!}{p!q!r!}$ for $p, q, r \in \mathbb{Z}_+$ such that $p + q + r = l \in \mathbb{Z}_+$.

For any $l \in \mathbb{Z}_+$, we construct two matrices $\boldsymbol{A}_{1N} = [\boldsymbol{a}_{1n,(i,j),(g,h)}]$ and $\boldsymbol{A}_{2N} = [\boldsymbol{a}_{2n,(i,j),(g,h)}]$ as follows: $\boldsymbol{a}_{1n,(i,j),(g,h)} = \boldsymbol{a}_{n,(i,j),(g,h)} \cdot 1(\boldsymbol{a}_{n,(i,j),(g,h)} \leq \tilde{C}_0 \left(\frac{d_F((i,j),(g,h))}{l}\right)^{-a})$ and $\boldsymbol{a}_{2n,(i,j),(g,h)} = \boldsymbol{a}_{n,(i,j),(g,h)} \cdot 1(\boldsymbol{a}_{n,(i,j),(g,h)} > \tilde{C}_0 \left(\frac{d_F((i,j),(g,h))}{l}\right)^{-a})$, then $|\boldsymbol{A}_N| = \boldsymbol{A}_{1N} + \boldsymbol{A}_{2N}$ and $\boldsymbol{a}_{1n,(i,j),(g,h)}\boldsymbol{a}_{2n,(i,j),(g,h)} = 0$. At least one of the items $\boldsymbol{a}_{n,(i,j),(i_1,j_1)}$, $\boldsymbol{a}_{n,(i_1,j_1),(i_2,j_2)}$, $\cdots$, and $\boldsymbol{a}_{n,(i_{l-1},j_{l-1}),(g,h)}$ would be less or equal to $\tilde{C}_0 \left(\frac{d_F((i,j),(g,h))}{l}\right)^{-a}$, because there exists at least two neighboring points in the chain $(i,j) \to (i_1,j_1) \to \cdots (i_{l-1},j_{l-1}) \to (g,h)$ such that their distance is at least $\frac{d_F((i,j),(g,h))}{l}$. Hence,

$$[\boldsymbol{A}_{2N}^l]_{ff'} = \sum_{i_1,j_1=1}^n \cdots \sum_{i_{l-1},j_{l-1}=1}^n \boldsymbol{a}_{n,(i,j),(i_1,j_1)} \cdots \boldsymbol{a}_{n,(i_{l-1},j_{l-1}),(g,h)} \cdot 1\left(\text{all } \boldsymbol{a}'s > \tilde{C}_0 \left(\frac{d_F((i,j),(g,h))}{l}\right)^{-a}\right),$$

and we have

$$\begin{aligned}
[|\boldsymbol{A}_N|^l]_{ff'} = [|\boldsymbol{A}_N|^l - \boldsymbol{A}_{2N}^l]_{ff'} &\leq |\boldsymbol{A}_{1N}|_{max} \sum_{m=0}^{l-1} \|\boldsymbol{A}_{2N}\|_\infty^m \||\boldsymbol{A}_N|^{l-m-1}\|_1 \\
&\leq \tilde{C}_0 \left(\frac{d_F((i,j),(g,h))}{l}\right)^{-a} \zeta^{l-1} K\Gamma \sum_{m=0}^{l-1}(l-m-1) \\
&\leq \tilde{C}_0^* d_F \left(\frac{d_F((i,j),(g,h))}{l}\right)^{-\alpha} l^{2+a}\zeta^{l-1},
\end{aligned}$$

where $\tilde{C}_0^* = \tilde{C}_0 K\Gamma$. The first inequality follows by Lemma A.3. in Xu and Lee (2015b) and $\|\boldsymbol{A}_N^l\|_1 \leq lK\Gamma\zeta^{l-1}$ for $l \in \mathbb{Z}_+$ by Assumption 4.2, all elements in $\boldsymbol{A}_{1N}$ are less or equal to $\tilde{C}_0 \left(\frac{d_F((i,j),(g,h))}{l}\right)^{-a}$, and $\sum_{m=0}^{l-1}(l-m-1) = \sum_{m=1}^{l-1} m = \frac{l(l-1)}{2} \leq l^2$. Then,

$$\begin{aligned}
\sum_{l=1}^\infty [|\boldsymbol{A}_N|^l]_{ff'} &= \sum_{l=1}^\infty \sum_{i_1,j_1=1}^n \cdots \sum_{i_{l-1},j_{l-1}=1}^n \boldsymbol{a}_{n,(i,j),(i_1,j_1)} \cdots \boldsymbol{a}_{n,(i_{l-1},j_{l-1}),(g,h)} \\
&\leq \tilde{C}_0^* d_F \left(d_F((i,j),(g,h))\right)^{-a} \zeta^{-1} \sum_{l=1}^\infty l^{2+a}\zeta^l \\
&\leq C_2 d_F \left(d_F((i,j),(g,h))\right)^{-a},
\end{aligned}$$

where $C_2 = \tilde{C}_0^* \zeta^{-1} \sum_{l=1}^\infty l^{2+a}\zeta^l < \infty$. ■

Proposition C.1. Assume that the model's spatial stability and coherency hold.

(i) If $\sup_{n,i,j} E|\epsilon_{n,ij}|^p < \infty$ for some $p \geq 1$, we have uniform L_p-boundedness of $\{y_{n,ij}\}$, $\{\sum_{g=1}^n w_{n,ig} y_{n,gj}\}$, $\{\sum_{h=1}^n y_{n,ih} m_{n,hj}\}$, $\{\sum_{g=1}^n \sum_{h=1}^n w_{n,ig} y_{n,gh} m_{n,hj}\}$, $\{\epsilon_{n,ij}^*(\theta)\}$, and $\{y_{n,ij}^*\}$.

(ii) Under Assumptions 4.1, 4.2 (iii-1), 4.3, and 4.4, $\{y_{n,ij}\}$, $\{\sum_{g=1}^n w_{n,ig} y_{n,gj}\}$, $\{\sum_{h=1}^n y_{n,ih} m_{n,hj}\}$, $\{\sum_{g=1}^n \sum_{h=1}^n w_{n,ig} y_{n,gh} m_{n,hj}\}$, $\{\epsilon_{n,ij}^*(\theta)\}$, and $\{y_{n,ij}^*\}$ are geometrically L_2-NED on ϵ. For example, $\left\|y_{n,ij} - E\left(y_{n,ij}|\mathcal{F}_{n,ij}(s)\right)\right\|_{L_2} \leq C\zeta^{s/\bar{d}}$ where C is a constant, and $\bar{d}$ is a constant defined in Assumption 4.2 (iii-1).

(iii) Under Assumptions 4.1, 4.2 (iii-2), 4.3, and 4.4, $\{y_{n,ij}\}$, $\{\sum_{g=1}^n w_{n,ig} y_{n,gj}\}$, $\{\sum_{h=1}^n y_{n,ih} m_{n,hj}\}$, $\{\sum_{g=1}^n \sum_{h=1}^n w_{n,ig} y_{n,gh} m_{n,hj}\}$, $\{\epsilon_{n,ij}^*(\theta)\}$, and $\{y_{n,ij}^*\}$ are uniformly L_2-NED on ϵ. For example,

$\left\| y_{n,ij} - E\left(y_{n,ij}|\mathcal{F}_{n,ij}(s)\right)\right\|_{L_2} \leq Cs^{2d-a}$ where C is a constant, and both d and a are constants such that $a > 2d$ in Assumption 4.2 (iii-2).

Proof of C.1 (i). Recall that

$$vec(Y_N) = F\big(vec(Y_N^*)\big) = F\big(\boldsymbol{A}_N vec(Y_N) + \mathbf{X}_N\kappa_0 + vec(\epsilon_N)\big).$$

Under the model's coherency (Assumption 3.1), $vec(Y_N)$ can be represented by a unique explicit function of $\mathbf{X}_N\kappa_0 + vec(\epsilon_N)$. Denote the unique solution of $vec(Y_N) = F\big(\boldsymbol{A}_N vec(Y_N) + \mathbf{X}_N\kappa_0 + vec(\epsilon_N)\big)$ as $\mathbf{y}_N(\mathbf{X}_N\kappa_0 + vec(\epsilon_N))$ with fth element $\mathbf{y}_{N,f} = e_{N,f}'\mathbf{y}_N(\mathbf{X}_N\kappa_0 + vec(\epsilon_N))$. By the mean value theorem for a convex function (see Wegge (1974)), we have

$$\mathbf{y}_N(\mathbf{X}_N\kappa_0 + vec(\epsilon_N)) - \mathbf{y}_N(\mathbf{0}) = \overline{\nabla F} \times [\boldsymbol{A}_N\mathbf{y}_N(\mathbf{X}_N\kappa_0 + vec(\epsilon_N)) + \mathbf{X}_N\kappa_0 + vec(\epsilon_N) - (\boldsymbol{A}_N\mathbf{y}_N(\mathbf{0}) + 0)]$$

where $\overline{\nabla F} = diag(\overline{\nabla F_1}, \cdots, \overline{\nabla F_N})$, $\overline{\nabla F_f}$ $(f = 1, \cdots, N)$ is a subgradient of $F(\cdot)$ at a point lying between $e_{N,f}'\boldsymbol{A}_N\mathbf{y}_N(\mathbf{X}_N\kappa_0 + vec(\epsilon_N))$ and 0.[50] Note that for the Tobit model, $\mathbf{y}_N(\mathbf{0}) = \mathbf{0}$ and the sub-gradients of $F(\cdot)$ lie between 0 and 1. Under spatial stability, we have a Neumann series expansion of $\mathbf{y}_N(\mathbf{X}_N\kappa_0 + vec(\epsilon_N))$, i.e., $\mathbf{y}_N(\mathbf{X}_N\kappa_0 + vec(\epsilon_N)) = (I_N - \overline{\nabla F}\boldsymbol{A}_N)^{-1}\overline{\nabla F}(\mathbf{X}_N\kappa_0 + vec(\epsilon_N))$. Since

$$(I_N - \overline{\nabla F}\boldsymbol{A}_N)^{-1}\overline{\nabla F} = \textstyle\sum_{l=0}^{\infty}(\overline{\nabla F}\boldsymbol{A}_N)^l\overline{\nabla F} \leq^* \sum_{l=0}^{\infty}|\boldsymbol{A}_N|^l \equiv \mathbb{M}_N = \left[\mathbf{m}_{N,ff'}\right],$$

where $A = \left[a_{ff'}\right] \leq^* B = \left[b_{ff'}\right]$ indicates $\left|a_{ff'}\right| \leq \left|b_{ff'}\right|$ for all f and f'. Then, we have $\left|\mathbf{y}_{N,f}(\mathbf{X}_N\kappa_0 + vec(\epsilon_N))\right| \leq \sum_{f'=1}^{N}\mathbf{m}_{N,ff'}\left|\boldsymbol{x}_{N,f'}\kappa_0 + \epsilon_{N,f'}\right|$. By the Minkowski's inequality, we have $\left\|\mathbf{y}_{N,f}(\mathbf{X}_N\kappa_0 + vec(\epsilon_N))\right\|_{L_p} \leq \sum_{f'=1}^{N}\mathbf{m}_{N,ff'}\left\|\boldsymbol{x}_{N,f'}\kappa_0 + \epsilon_{N,f'}\right\|_{L_p} < \infty$ uniformly in n. Hence, $\{y_{n,ij}\}$ is uniformly L_p-bounded. Using the same strategy, we can show that $\left\{\sum_{g=1}^{n} w_{n,ig}y_{n,gj}\right\}$, $\left\{\sum_{h=1}^{n} y_{n,ih}m_{n,hj}\right\}$, $\left\{\sum_{g=1}^{n}\sum_{h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}\right\}$, and $\left\{y_{n,ij}^*\right\}$ are uniformly L_p-bounded. For $\left\{\epsilon_{n,ij}^*(\theta)\right\}$, we can employ that θ belongs to a compact parameter space Θ.

Proofs of (ii) and (iii). Consider the NED properties of $\{y_{n,ij}\}$. Choose two possible bases $\epsilon_N^{(1)}$ and $\epsilon_N^{(2)}$, which generate $Y_N^{(1)}$ and $Y_N^{(2)}$, respectively. That is, $vec\left(Y_N^{(j)}\right) = F\left(\boldsymbol{A}_N vec\left(Y_N^{(j)}\right) + \mathbf{X}_N^{(j)}\kappa_0 + vec\left(\epsilon_N^{(j)}\right)\right)$ for $j = 1,2$. Using the same way in the proof of (i), we obtain $vec\left(Y_N^{(1)}\right) - vec\left(Y_N^{(2)}\right) = \left(I_N - \widetilde{\nabla F_y}\boldsymbol{A}_N\right)^{-1}\widetilde{\nabla F_y} \times \left[\left(\mathbf{X}_N^{(1)} - \mathbf{X}_N^{(2)}\right)\kappa_0 + \left(vec\left(\epsilon_N^{(1)}\right) - vec\left(\epsilon_N^{(2)}\right)\right)\right]$ and $\left(I_N - \widetilde{\nabla F_y}\boldsymbol{A}_N\right)^{-1}\widetilde{\nabla F_y} \leq^* \mathbb{M}_N$, where $\widetilde{\nabla F_y}$ is a diagonal matrix containing the sub-gradients of $F(\cdot)$ evaluated between the two points. Note that $E\left(y_{n,ij}|\mathcal{F}_{n,ij}(s)\right)$ is an approximation of $y_{n,ij}$, which is a function of $\left\{\left(\boldsymbol{x}_{n,gh}, \epsilon_{n,gh}\right): d_F\left((i,j),(g,h)\right) \leq s\right\}$. Then, we have

[50] The mean value theorem is applied to each element of $F\big(\boldsymbol{A}_N vec(Y_N) + \mathbf{X}_N\kappa_0 + vec(\epsilon_N)\big)$.

$$\left\|y_{n,ij} - E\left(y_{n,ij}|\mathcal{F}_{n,ij}(s)\right)\right\|_{L_2} \leq \sum_{g,h:d_F((i,j),(g,h))>s}^{n} \mathbf{m}_{n,(i,j),(g,h)} \cdot \left\|\boldsymbol{x}_{n,gh}\kappa_0 + \epsilon_{n,gh}\right\|_{L_2}$$
$$\leq \sup_{n,g,h}\left\|\boldsymbol{x}_{n,gh}\kappa_0 + \epsilon_{n,gh}\right\|_{L_2} \cdot \sup_{n,g,h} \sum_{g,h:d_F((i,j),(g,h))>s} \mathbf{m}_{n,(i,j),(g,h)},$$

where $\mathbf{m}_{n,(i,j),(g,h)} = \mathbf{m}_{N,ff'}$ (i.e., $f = (j-1)n + i$ and $f' = (h-1)n + g$).

Note that $\sup_{n,g,h}\left\|\boldsymbol{x}_{n,gh}\kappa_0 + \epsilon_{n,gh}\right\|_{L_2} < \infty$ by Assumptions 4.3 and 4.5. To show the NED properties of $\{y_{n,ij}\}$, we need to show $\sup_{n,g,h} \sum_{g,h:d_F((i,j),(g,h))>s} \mathbf{m}_{n,(i,j),(g,h)} \to 0$ as $s \to \infty$. Using the results from Lemma C.1 with the similar argument of Proposition 1 in Xu and Lee (2015), we finish the proof. The details can be found in the supplement file. ■

Next, we consider the NED properties of $1(y_{n,ij} > 0)$, which is a component of $\ln L_N^*(\theta)$. Before discussing this issue, an additional condition is needed. The normality assumption (Assumption 4.8) helps to restrict an upper bound of probability densities of $\{y_{n,ij}^*\}$.[51] Here are relevant lemmas and proposition. Ideas of the proofs are the same as Xu and Lee's (2015) Lemma 2 and Proposition 2. Modified proofs for our framework can be found in the supplement file.

Lemma C.2. When M_n is an n-dimensional symmetric matrix, $x_n' M_n x_n \geq \min_{i=1,\dots,n} \varphi_i(M_n)\, x_n' x_n$, where x_n is a nonzero n-dimensional vector.

Lemma C.3. Assume that the model's spatial stability and coherency hold (Assumption 3.1). Under Assumption 4.8, the essential supremums of densities of $\{y_{n,ij}^*\}$ are uniformly bounded in i, j, and n.

Proposition C.2. Assume that the model's spatial stability and coherency hold.

(i) Under Assumptions 4.1, 4.2 (iii-1), 4.3, and 4.8, $\{1(y_{n,ij} > 0)\}$ is uniformly and geometrically L_2-NED on ϵ. That is, $\left\|1(y_{n,ij} > 0) - E\left(1(y_{n,ij} > 0)|\mathcal{F}_{n,ij}(s)\right)\right\|_{L_2} \leq C\zeta^{s/3\bar{d}}$ where C is a constant, and $\bar{d}$ is a constant defined in Assumption 4.2 (iii-1).

(ii) Under Assumptions 4.1, 4.2 (iii-2), 4.3, and 4.8, $\{1(y_{n,ij} > 0)\}$ is uniformly L_2-NED on ϵ. That is, $\left\|1(y_{n,ij} > 0) - E\left(1(y_{n,ij} > 0)|\mathcal{F}_{n,ij}(s)\right)\right\|_{L_2} \leq Cs^{(2d-a)/3}$ where C is a constant, and both d and a are constants such that $a > 2d$ in Assumption 4.2 (iii-2).

Here is the proof of consistency.

Proof of consistency. Under the unique identification condition, for consistency, it suffices to show (1) uniform convergence $\sup_{\theta\in\Theta} \frac{1}{N}[\ln L_N^*(\theta) - E\ln L_N^*(\theta)] \xrightarrow{p} 0$ and (2) the uniform equicontinuity of $\left\{\frac{1}{N} E \ln L_N^*(\theta)\right\}$.

Step 1 (**Uniform convergence**): Note that

[51] Note that the event of $\{y_{n,ij} > 0\}$ is the same as that of $\{y_{n,ij}^* > 0\}$.

$$
\begin{aligned}
\frac{1}{N}[\ln L_N^*(\theta) - E(\ln L_N^*(\theta))] = &\frac{1}{N}\sum_{i.j=1}^{n}\left[1\left(y_{n,ij}=0\right)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right) - E\left(1\left(y_{n,ij}=0\right)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right)\right)\right] \\
&-\frac{1}{2N}\ln 2\pi\sigma^2\sum_{i,j=1}^{n}\left[1\left(y_{n,ij}>0\right) - E1\left(y_{n,ij}>0\right)\right] \\
&+\frac{1}{N}\left[\ln\left|S_{N_2}^*(\delta)\right| - E\ln\left|S_{N_2}^*(\delta)\right|\right] \\
&-\frac{1}{2N}\sum_{i,j=1}^{n}\left[1\left(y_{n,ij}>0\right)\left(\epsilon_{n,ij}^*(\theta)\right)^2 - E1\left(y_{n,ij}>0\right)\left(\epsilon_{n,ij}^*(\theta)\right)^2\right]. \qquad (\mathrm{C}.1)
\end{aligned}
$$

First, note that $\frac{1}{2N}\ln 2\pi\sigma^2\sum_{i,j=1}^{n}\left[1\left(y_{n,ij}>0\right) - E1\left(y_{n,ij}>0\right)\right] \xrightarrow{p} 0$ uniformly in Θ since each σ^2 is a positive constant due to the compact parameter space assumption. Second, consider the last term of (C.1). Observe the components of $\left\{\left(\epsilon_{n,ij}^*(\theta)\right)^2\right\}$:

$$
\begin{aligned}
\left(\epsilon_{n,ij}^*(\theta)\right)^2 = &\frac{1}{\sigma^2}y_{n,ij}^2 + \frac{\lambda^2}{\sigma^2}\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gj}\right)^2 + \frac{\gamma^2}{\sigma^2}\left(\sum_{h=1}^{n} y_{n,ih}m_{n,hj}\right)^2 + \frac{\rho^2}{\sigma^2}\left(\sum_{g,h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}\right)^2 \\
&+\frac{1}{\sigma^2}\left(\boldsymbol{x}_{n,ij}\kappa\right)^2 - \frac{2\lambda}{\sigma^2}y_{n,ij}\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gj}\right) - \frac{2\gamma}{\sigma^2}y_{n,ij}\left(\sum_{h=1}^{n} y_{n,ih}m_{n,hj}\right) \\
&-\frac{2\rho}{\sigma^2}y_{n,ij}\left(\sum_{g,h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}\right) - \frac{2}{\sigma^2}y_{n,ij}\left(\boldsymbol{x}_{n,ij}\kappa\right) + \frac{2\lambda\gamma}{\sigma^2}\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gj}\right)\left(\sum_{h=1}^{n} y_{n,ih}m_{n,hj}\right) \\
&+\frac{2\lambda\rho}{\sigma^2}\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gj}\right)\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}\right) \\
&+\frac{2\gamma\rho}{\sigma^2}\left(\sum_{h=1}^{n} y_{n,ih}m_{n,hj}\right)\left(\sum_{g,h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}\right) \\
&+\frac{2\lambda}{\sigma^2}\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gj}\right)\left(\boldsymbol{x}_{n,ij}\kappa\right) + \frac{2\gamma}{\sigma^2}\left(\sum_{h=1}^{n} y_{n,ih}m_{n,hj}\right)\left(\boldsymbol{x}_{n,ij}\kappa\right) \\
&+\frac{2\rho}{\sigma^2}\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}\right)\left(\boldsymbol{x}_{n,ij}\kappa\right).
\end{aligned}
$$

By Proposition C.1 and using the compact parameter space assumption, $\left\{\left(\epsilon_{n,ij}^*(\theta)\right)^2\right\}$ for each θ is uniformly L_2-NED on ϵ, and uniformly $L_{2+\eta}$-bounded for some $\eta > 0$. Since $\left\{1\left(y_{n.ij}>0\right)\right\}$ is also uniformly L_2-NED on ϵ by Proposition C.2, $\left\{1\left(y_{n,ij}>0\right)\left(\epsilon_{n,ij}^*(\theta)\right)^2\right\}$ is uniformly L_1-NED on ϵ. Hence, it satisfies the conditions for the WLLN: $\sup_{\theta\in\Theta}\frac{1}{N}\sum_{i,j=1}^{n}\left[1\left(y_{n,ij}>0\right)\left(\epsilon_{n,ij}^*(\theta)\right)^2 - E1\left(y_{n,ij}>0\right)\left(\epsilon_{n,ij}^*(\theta)\right)^2\right] \xrightarrow{p} 0$.

Third, we will consider $\sup_{\theta\in\Theta}\frac{1}{N}\sum_{i,j=1}^{n}\left[1\left(y_{n,ij}=0\right)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right) - E\left(1\left(y_{n,ij}=0\right)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right)\right)\right] \xrightarrow{p} 0$. Let $\ell_{1,N}(\theta) = \frac{1}{N}\sum_{i,j=1}^{n}\left[1\left(y_{n,ij}=0\right)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right) - E\left(1\left(y_{n,ij}=0\right)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right)\right)\right]$ for each θ. Observe that Θ is compact by Assumption 4.4 and $\ell_{1,N}(\theta) \xrightarrow{p} 0$ for each $\theta \in \Theta$. By Theorem 1 in Andrews (1992), it suffices to check the stochastic equicontinuity of $\left\{\ell_{1,N}(\theta)\right\}$ as Xu and Lee's (2015) proof of Theorem 1. Observe that

$$
\begin{aligned}
\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right) &= \ln\left[1 - \Phi\left(\tilde{\lambda}\sum_{g=1}^{n} w_{n,ig}y_{n,gj} + \tilde{\gamma}\sum_{h=1}^{n} y_{n,ih}m_{n,hj} + \tilde{\rho}\sum_{g,h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj} + \boldsymbol{x}_{n,ij}\tilde{\kappa}\right)\right] \\
&= \ln\Phi\left(-\tilde{\lambda}\sum_{g=1}^{n} w_{n,ig}y_{n,gj} - \tilde{\gamma}\sum_{h=1}^{n} y_{n,ih}m_{n,hj} - \tilde{\rho}\sum_{g,h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj} - \boldsymbol{x}_{n,ij}\tilde{\kappa}\right),
\end{aligned}
$$

where $\tilde{\lambda}=\lambda/\sigma$, $\tilde{\gamma}=\gamma/\sigma$, $\tilde{\rho}=\rho/\sigma$, and $\tilde{\kappa}=\kappa/\sigma$ located at a close and bounded subset of $\mathbb{R}^{3+L+2K}$ (due to the compact parameter space assumption). By Lemma A.9 of Xu and Lee (2015b), $|\ln\Phi(x_1)-\ln\Phi(x_2)|\le(2|x_1|+2|x_2|+C_2)\cdot|x_1-x_2|$ for some $C_2>0$. For two sets of parameters $(\tilde{\lambda}_1,\tilde{\gamma}_1,\tilde{\rho}_1,\tilde{\kappa}_1')'$ and $(\tilde{\lambda}_2,\tilde{\gamma}_2,\tilde{\rho}_2,\tilde{\kappa}_2')'$ and for some $C_2>0$, then,

$$
\begin{aligned}
&\frac{1}{N}\sum_{i,j=1}^{n}1(y_{n,ij}=0)\begin{bmatrix}\ln\Phi\left(-\tilde{\lambda}_1\sum_{g=1}^{n}w_{n,ig}y_{n,gj}-\tilde{\gamma}_1\sum_{h=1}^{n}y_{n,ih}m_{n,hj}-\tilde{\rho}_1\sum_{g,h=1}^{n}w_{n,ig}y_{n,gh}m_{n,hj}-\boldsymbol{x}_{n,ij}\tilde{\kappa}_1\right)\\-\ln\Phi\left(-\tilde{\lambda}_2\sum_{g=1}^{n}w_{n,ig}y_{n,gj}-\tilde{\gamma}_2\sum_{h=1}^{n}y_{n,ih}m_{n,hj}-\tilde{\rho}_2\sum_{g,h=1}^{n}w_{n,ig}y_{n,gh}m_{n,hj}-\boldsymbol{x}_{n,ij}\tilde{\kappa}_2\right)\end{bmatrix}\\
&\le\frac{1}{N}\sum_{i,j=1}^{n}\begin{bmatrix}2\left|\tilde{\lambda}_1\sum_{g=1}^{n}w_{n,ig}y_{n,gj}+\tilde{\gamma}_1\sum_{h=1}^{n}y_{n,ih}m_{n,hj}+\tilde{\rho}_1\sum_{g,h=1}^{n}w_{n,ig}y_{n,gh}m_{n,hj}+\boldsymbol{x}_{n,ij}\tilde{\kappa}_1\right|\\+2\left|\tilde{\lambda}_2\sum_{g=1}^{n}w_{n,ig}y_{n,gj}+\tilde{\gamma}_2\sum_{h=1}^{n}y_{n,ih}m_{n,hj}+\tilde{\rho}_2\sum_{g,h=1}^{n}w_{n,ig}y_{n,gh}m_{n,hj}+\boldsymbol{x}_{n,ij}\tilde{\kappa}_2\right|+C_2\end{bmatrix}\\
&\quad\times\left|(\tilde{\lambda}_1-\tilde{\lambda}_2)\sum_{g=1}^{n}w_{n,ig}y_{n,gj}+(\tilde{\gamma}_1-\tilde{\gamma}_2)\sum_{h=1}^{n}y_{n,ih}m_{n,hj}+(\tilde{\rho}_1-\tilde{\rho}_2)\sum_{g,h=1}^{n}w_{n,ig}y_{n,gh}m_{n,hj}+\boldsymbol{x}_{n,ij}(\tilde{\kappa}_1-\tilde{\kappa}_2)\right|\\
&\le\frac{1}{N}\sum_{i,j=1}^{n}\underbrace{\left[4\tilde{\lambda}_m\left|\sum_{g=1}^{n}w_{n,ig}y_{n,gj}\right|+4\tilde{\gamma}_m\left|\sum_{h=1}^{n}y_{n,ih}m_{n,hj}\right|+4\tilde{\rho}_m\left|\sum_{g,h=1}^{n}w_{ig}y_{gh}m_{hj}\right|+4\left|\boldsymbol{x}_{n,ij}\right|\tilde{\kappa}_m+C_2\right]}_{\equiv term1}\\
&\quad\times\underbrace{\left(\left|\sum_{g=1}^{n}w_{n,ig}y_{n,gj}\right|+\left|\sum_{h=1}^{n}y_{n,ih}m_{n,hj}\right|+\left|\sum_{g,h=1}^{n}w_{n,ig}y_{n,gh}m_{n,hj}\right|+\left|\boldsymbol{x}_{n,ij}\right|l_{L+2K}\right)}_{\equiv term2}\\
&\quad\times\left(\left|\tilde{\lambda}_1-\tilde{\lambda}_2\right|+|\tilde{\gamma}_1-\tilde{\gamma}_2|+|\tilde{\rho}_1-\tilde{\rho}_2|+\|\tilde{\kappa}_1-\tilde{\kappa}_2\|\right),
\end{aligned}
$$

where $\tilde{\lambda}_m$, $\tilde{\gamma}_m$, $\tilde{\rho}_m$, and $\tilde{\kappa}_m$ are respectively the supremums of $\tilde{\lambda}$, $\tilde{\gamma}$, $\tilde{\rho}$, and $\tilde{\kappa}$, and $\|\cdot\|$ denotes the Euclidean vector norm. By the compact parameter space assumption, they are finite. By Proposition C.1, the components in *term1* and *term2* above are uniformly $L_{4+\eta}$-bounded for some $\eta>0$. Then, $\|term1\cdot term2\|_{L_{2+\frac{\eta}{2}}}\le\|term1\|_{L_{4+\eta}}\cdot\|term2\|_{L_{4+\eta}}$ by the generalized Hölder's inequality. By applying Lemma 1(a) in Andrews (1992), $\{\ell_{1,N}(\theta)\}$ is stochastic equicontinuous and $\left\{\frac{1}{N}\sum_{i,j=1}^{n}E\left(1(y_{n,ij}=0)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right)\right)\right\}$ is equicontinuous.

Last, we will show $\sup_{\theta\in\Theta}\frac{1}{N}\left[\ln\det\left(S_{N_2}^*(\delta)\right)-E\ln\det\left(S_{N_2}^*(\delta)\right)\right]\xrightarrow{p}0$. Let $\boldsymbol{A}_N(\delta)=\lambda\boldsymbol{W}_N+\gamma\boldsymbol{M}_N+\rho\boldsymbol{R}_N$ for each δ and its (f,f')-element be $\boldsymbol{a}_{n,(i,j),(g,h)}(\delta)$, where $f=(j-1)n+i$ and $f'=(h-1)n+g$. Then, by the Taylor expansion,

$$
\begin{aligned}
\ln\det\left(S_{N_2}^*(\delta)\right)&=-\sum_{l=1}^{\infty}\frac{1}{l}tr\left(\left(G_N(Y_N)\boldsymbol{A}_N(\delta)G_N(Y_N)\right)^l\right)\\
&=-\sum_{i,j=1}^{n}1(y_{n,ij}>0)\left[\sum_{l=1}^{\infty}\frac{1}{l}tr\left(\left(G_N(Y_N)\boldsymbol{A}_N(\delta)G_N(Y_N)\right)^l\right)\right]_{(j-1)n+i,(j-1)n+i}\\
&=-\sum_{i,j=1}^{n}1(y_{n,ij}>0)\sum_{l=1}^{\infty}\frac{1}{l}\sum_{i_1,j_1}\cdots\sum_{i_{l-1},j_{l-1}}\boldsymbol{a}_{n,(i,j),(i_1,j_1)}(\delta)\cdots\boldsymbol{a}_{n,(i_{l-1},j_{l-1})(i,j)}(\delta)\\
&\qquad\times1\left(\cap_{h=1}^{l-1}\{y_{n,i_hj_h}>0\}\right)
\end{aligned}
$$

For each (i,j) and $q\in\mathbb{Z}_+$, define $\boldsymbol{B}_{1,n,ij}^q=\sum_{l=1}^{q}\frac{1}{l}b_{l,n,ij}$ with $b_{l,n,ij}=1(y_{n,ij}>0)\sum_{i_1,j_1}\cdots\sum_{i_{l-1},j_{l-1}}\boldsymbol{a}_{n,(i,j),(i_1,j_1)}(\delta)\cdots\boldsymbol{a}_{n,(i_{l-1},j_{l-1})(i,j)}(\delta)\cdot1\left(\cap_{h=1}^{l-1}\{y_{n,i_hj_h}>0\}\right)$ and $\boldsymbol{B}_{2,n,ij}^q=\sum_{l=q+1}^{\infty}\frac{1}{l}b_{l.n,ij}$. By using Lemma A.8 in Xu and Lee (2015b), $\{b_{l,n,ij}\}$ is uniformly L_2-NED on ϵ for $l\in\mathbb{Z}_+$. For each l, we have $\frac{1}{N}\sum_{i,j=1}^{n}(b_{l,n,ij}-Eb_{l,n,ij})\xrightarrow{p}0$ by applying Theorem 1 in Jenish and Prucha (2012). By the compact parameter space assumption, we obtain $\sup_{\delta\in\Theta_\delta}\left|\frac{1}{N}\sum_{i,j=1}^{n}(\boldsymbol{B}_{1,n,ij}^q-E\boldsymbol{B}_{1,n,ij}^q)\right|\xrightarrow{p}0$ for $q\in\mathbb{Z}_+$. Using the same expansion technique in Lemma C.1.,

$$\sup_{\delta\in\Theta_\delta}\left|\frac{1}{N}\sum_{i,j=1}^{n}\boldsymbol{B}_{2,n,ij}^{q}\right| \le \sum_{l=q+1}^{\infty}\frac{\zeta^l}{l} \le \frac{\zeta^{q+1}}{(q+1)(1-\zeta)}.$$

Take a positive integer q_ε such that $\frac{\zeta^{q_\varepsilon+1}}{(q_\varepsilon+1)(1-\zeta)} < \frac{\varepsilon}{2}$ for an arbitrary small $\varepsilon > 0$. It implies $\sup_{\delta\in\Theta_\delta}\left|\frac{1}{N}\sum_{i,j=1}^{n}\boldsymbol{B}_{2,n,ij}^{q_\varepsilon}\right| < \frac{\varepsilon}{2}$ and $\sup_{\delta\in\Theta_\delta}\left|\frac{1}{N}\sum_{i,j=1}^{n}E\boldsymbol{B}_{2,n,ij}^{q_\varepsilon}\right| < \frac{\varepsilon}{2}$. Then, $\sup_{\delta\in\Theta_\delta}\left|\frac{1}{N}\sum_{i,j=1}^{n}\left(\boldsymbol{B}_{2,n,ij}^{q_\varepsilon} - E\boldsymbol{B}_{2,n,ij}^{q_\varepsilon}\right)\right| < \varepsilon$. By combining (1) $\sup_{\delta\in\Theta_\delta}\left|\frac{1}{N}\sum_{i,j=1}^{n}\left(\boldsymbol{B}_{1,n,ij}^{q_\varepsilon} - E\boldsymbol{B}_{1,n,ij}^{q_\varepsilon}\right)\right| \xrightarrow{p} 0$ and (2) $\sup_{\delta\in\Theta_\delta}\left|\frac{1}{N}\sum_{i,j=1}^{n}\left(\boldsymbol{B}_{2,n,ij}^{q_\varepsilon} - E\boldsymbol{B}_{2,n,ij}^{q_\varepsilon}\right)\right| < \varepsilon$, we obtain the desired result.

Step 2 (**Equicontinuity of** $\left\{\frac{1}{N}E(\ln L_N^*(\theta))\right\}$): Recall that

$$\frac{1}{N}E(\ln L_N^*(\theta)) = \frac{1}{N}\sum_{i,j=1}^{N}E\left(1\left(y_{n,ij}=0\right)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right)\right) - \frac{1}{2N}\ln 2\pi\sigma^2\sum_{i,j=1}^{n}E1\left(y_{n,ij}>0\right)$$
$$+\frac{1}{N}E\ln\det\left(S_{N_2}^*(\delta)\right) - \frac{1}{2N}\sum_{i,j=1}^{n}E1\left(y_{n,ij}>0\right)\left(\epsilon_{n,ij}^*(\theta)\right)^2.$$

By Step 1, we have verified that a family of functions $\left\{\frac{1}{N}\sum_{i,j=1}^{n}E\left(1\left(y_{n,ij}=0\right)\ln\Phi\left(\epsilon_{n,ij}^*(\theta)\right)\right)\right\}$ is equicontinuous. By the compact parameter space assumption, $\left\{\frac{1}{2N}\ln 2\pi\sigma^2\sum_{i,j=1}^{n}E1\left(y_{n,ij}>0\right)\right\}$ is equicontinuous. Consider $\frac{1}{N}E\ln\det\left(S_{N_2}^*(\delta)\right)$. Since

$$\sup_n\sup_{\delta\in\Theta_\delta}\left|\frac{\partial}{\partial\delta}\frac{1}{N}E\ln\det\left(S_{N_2}^*(\delta)\right)\right| = \sup_n\sup_{\delta\in\Theta_\delta}\begin{pmatrix}\left|\frac{1}{N}tr\left(\widetilde{\boldsymbol{W}}_N\tilde{S}_N^{-1}(\delta)\right)\right|\\ \left|\frac{1}{N}tr\left(\widetilde{\boldsymbol{M}}_N\tilde{S}_N^{-1}(\delta)\right)\right|\\ \left|\frac{1}{N}tr\left(\widetilde{\boldsymbol{R}}_N\tilde{S}_N^{-1}(\delta)\right)\right|\end{pmatrix} \le^* \frac{1}{1-\zeta}\begin{pmatrix}c_{w,r}\\ c_{m.c}\\ c_{w,r}c_{m,c}\end{pmatrix} < \infty,$$

$\left\{\frac{1}{N}E\ln\det\left(S_{N_2}^*(\delta)\right)\right\}$ is equicontinuous. For the last component, observe that $\left\{E1\left(y_{n,ij}>0\right)\left(\epsilon_{n,ij}^*(\theta)\right)^2\right\}$ is a sequence of uniformly $L_{2+\eta}$-bounded components for some $\eta > 0$ by Step 1. With the compact parameter space assumption, the last component is also equicontinuous.

Step 3 (**Identification uniqueness**): In this part, we will derive the identification conditions provided in Assumption 4.9. Those also come from Rothenberg (1971): θ_0 is uniquely identified if and only if there is no observationally equivalent $\theta \in \Theta$. Suppose $\ln L_N^*(\theta_0) = \ln L_N^*(\theta_1)$ for some $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$. Consider a specific event $\bigcap_{i,j=1}^{n}\left\{y_{n,ij}>0\right\}$, and note that $P\left(\bigcap_{i,j=1}^{n}\left\{y_{n,ij}>0\right\}\right) > 0$. In this case, $\ln L_N^*(\theta_0) = \ln L_N^*(\theta_1)$ is equivalent to

$$\frac{N}{2}\ln\sigma_0^2 - \ln\det\left(S_{N_2}^*(\delta_0)\right) + \frac{1}{2}\sum_{i,j=1}^{n}\left(\epsilon_{n,ij}^*(\theta_0)\right)^2 = \frac{N}{2}\ln\sigma_1^2 - \ln\det\left(S_{N_2}^*(\delta_1)\right) + \frac{1}{2}\sum_{i,j=1}^{n}\left(\epsilon_{n,ij}^*(\theta_1)\right)^2.$$

Since $y_{n,ij} > 0$ for all $i,j = 1,\cdots,n$, we differentiate the above with respect to $y_{n,gh}$ (for some $g,h \in \{1,\cdots,n\}$):

$$-\frac{1}{\sigma_0^2}\sum_{i,j=1}^{n}\big(y_{n,ij}-(e_{n,j}'\otimes e_{n,i}')\boldsymbol{A}_N(\delta_0)vec(Y_N)-\boldsymbol{x}_{n,ij}\kappa_0\big)(e_{n,j}'\otimes e_{n,i}')\boldsymbol{A}_N(\delta_0)(e_{n,h}\otimes e_{n,g})$$
$$+\frac{1}{\sigma_0^2}\big(y_{n,gh}-(e_{n,h}'\otimes e_{n,g}')\boldsymbol{A}_N(\delta_0)vec(Y_N)-\boldsymbol{x}_{n,gh}\kappa_0\big)$$
$$=-\frac{1}{\sigma_1^2}\sum_{i,j=1}^{n}\big(y_{N,f}-(e_{n,j}'\otimes e_{n,i}')\boldsymbol{A}_N(\delta_1)vec(Y_N)-\boldsymbol{x}_{N,f}\kappa_1\big)(e_{n,j}'\otimes e_{n,i}')\boldsymbol{A}_N(\delta_1)(e_{n,h}\otimes e_{n,g})$$
$$+\frac{1}{\sigma_0^2}\big(y_{n,gh}-(e_{n,h}'\otimes e_{n,g}')\boldsymbol{A}_N(\delta_1)vec(Y_N)-\boldsymbol{x}_{n,gh}\kappa_1\big). \tag{C.2}$$

Let $f=(j-1)n+i$ and $f'=(h-1)n+g$. By differentiating both sides of (C.2) with respect to $y_{n,gh}$, we have
$\frac{1}{\sigma_0^2}\sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_0)]_{ff'}^2+\frac{1}{\sigma_0^2}=\frac{1}{\sigma_1^2}\sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_1)]_{ff'}^2+\frac{1}{\sigma_1^2}$, which is equivalent that

$$\begin{aligned}\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}&=\sum_{f=1}^{N}\left(\frac{1}{\sigma_1^2}[\boldsymbol{A}_N(\delta_1)]_{ff'}^2-\frac{1}{\sigma_0^2}[\boldsymbol{A}_N(\delta_0)]_{ff'}^2\right)\\
&=\left(\frac{\lambda_1^2}{\sigma_1^2}-\frac{\lambda_0^2}{\sigma_0^2}\right)\sum_{f=1}^{N}[\boldsymbol{W}_N]_{ff'}^2+\left(\frac{\gamma_1^2}{\sigma_1^2}-\frac{\gamma_0^2}{\sigma_0^2}\right)\sum_{f=1}^{N}[\boldsymbol{M}_N]_{ff'}^2+\left(\frac{\rho_1^2}{\sigma_1^2}-\frac{\rho_0^2}{\sigma_0^2}\right)\sum_{f=1}^{N}[\boldsymbol{R}_N]_{ff'}^2\\
&\quad+\left(\frac{\lambda_1\gamma_1}{\sigma_1^2}-\frac{\lambda_0\gamma_0}{\sigma_0^2}\right)\sum_{f=1}^{N}2[\boldsymbol{W}_N]_{ff'}[\boldsymbol{M}_N]_{ff'}+\left(\frac{\lambda_1\rho_1}{\sigma_1^2}-\frac{\lambda_0\rho_0}{\sigma_0^2}\right)\sum_{f=1}^{N}2[\boldsymbol{W}_N]_{ff'}[\boldsymbol{R}_N]_{ff'}\\
&\quad+\left(\frac{\gamma_1\rho_1}{\sigma_1^2}-\frac{\gamma_0\rho_0}{\sigma_0^2}\right)\sum_{f=1}^{N}2[\boldsymbol{M}_N]_{ff'}[\boldsymbol{R}_N]_{ff'}\\
&=\left(\frac{\lambda_1^2}{\sigma_1^2}-\frac{\lambda_0^2}{\sigma_0^2}\right)\sum_{f=1}^{N}[\boldsymbol{W}_N]_{ff'}^2+\left(\frac{\gamma_1^2}{\sigma_1^2}-\frac{\gamma_0^2}{\sigma_0^2}\right)\sum_{f=1}^{N}[\boldsymbol{M}_N]_{ff'}^2+\left(\frac{\rho_1^2}{\sigma_1^2}-\frac{\rho_0^2}{\sigma_0^2}\right)\sum_{f=1}^{N}[\boldsymbol{R}_N]_{ff'}^2\end{aligned}$$

for all $f'=1,\cdots,N$. The above holds since $[\boldsymbol{W}_N]_{ff'}^2=1(j=h)w_{n,ig}^2$, $[\boldsymbol{M}_N]_{ff'}^2=1(i=g)m_{n,hj}^2$, $[\boldsymbol{R}_N]_{ff'}^2=w_{n,ig}^2m_{n,hj}^2$, $[\boldsymbol{W}_N]_{ff'}[\boldsymbol{M}_N]_{ff'}=1(i=g,j=h)w_{n,ig}m_{n,hj}=0$, $[\boldsymbol{W}_N]_{ff'}[\boldsymbol{R}_N]_{ff'}=\underbrace{1(j=h)w_{n,ig}}_{=[\boldsymbol{W}_N]_{ff'}}\underbrace{w_{n,ig}\,m_{n,hj}}_{=[\boldsymbol{R}_N]_{ff'}}=0$, and $[\boldsymbol{M}_N]_{ff'}[\boldsymbol{R}_N]_{ff'}=\underbrace{1(i=g)m_{n,hj}}_{=[\boldsymbol{M}_N]_{ff'}}\underbrace{w_{n,ig}m_{n,hj}}_{=[\boldsymbol{R}_N]_{ff'}}=0$. For all $f'=1,\cdots,N$, note that a set of vectors $\left\{\mathbb{w}_{f'}^s,\mathbb{m}_{f'}^s,\mathbb{r}_{f'}^s\right\}$ is linearly independent, where $\mathbb{w}_{f'}^s$, $\mathbb{m}_{f'}^s$, and $\mathbb{r}_{f'}^s$ are respectively consist of $\sum_{f=1}^{N}\left([\boldsymbol{W}_N]_{ff'}^2-[\boldsymbol{W}_N]_{ff''}^2\right)$, $\sum_{f=1}^{N}\left([\boldsymbol{M}_N]_{ff'}^2-[\boldsymbol{M}_N]_{ff''}^2\right)$, and $\sum_{f=1}^{N}\left([\boldsymbol{R}_N]_{ff'}^2-[\boldsymbol{R}_N]_{ff''}^2\right)$ for $f''\neq f'$. This condition implies $\frac{\lambda_1^2}{\sigma_1^2}=\frac{\lambda_0^2}{\sigma_0^2}$, $\frac{\gamma_1^2}{\sigma_1^2}=\frac{\gamma_0^2}{\sigma_0^2}$, $\frac{\rho_1^2}{\sigma_1^2}=\frac{\rho_0^2}{\sigma_0^2}$, and $\frac{1}{\sigma_1^2}=\frac{1}{\sigma_0^2}$, so we have $\sigma_1^2=\sigma_0^2$, $|\lambda_0|=|\lambda_1|$, $|\gamma_1|=|\gamma_0|$, and $|\rho_1|=|\rho_0|$.

By differentiating both sides of (C.2) with respect to $y_{n,kl}$ with $(k,l)\neq(g,h)$, we have

$$\begin{aligned}&-\big([\boldsymbol{A}_N(\delta_0)]_{f''f'}+[\boldsymbol{A}_N(\delta_0)]_{f'f''}\big)+\sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_0)]_{ff''}[\boldsymbol{A}_N(\delta_0)]_{ff'}\\
&\quad=-\big([\boldsymbol{A}_N(\delta_1)]_{f''f'}+[\boldsymbol{A}_N(\delta_1)]_{f'f''}\big)+\sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_1)]_{ff''}[\boldsymbol{A}_N(\delta_1)]_{ff'},\end{aligned} \tag{C.3}$$

where $f''=(l-1)n+k$, since $\sigma_0^2=\sigma_1^2$. First, we consider the first part of (C.3):

$$\begin{aligned}&[\boldsymbol{A}_N(\delta_0)]_{f''f'}+[\boldsymbol{A}_N(\delta_0)]_{f'f''}-[\boldsymbol{A}_N(\delta_1)]_{f''f'}-[\boldsymbol{A}_N(\delta_1)]_{f'f''}\\
&=(\lambda_0-\lambda_1)\big([\boldsymbol{W}_N]_{f''f'}+[\boldsymbol{W}_N]_{f'f''}\big)+(\gamma_0-\gamma_1)\big([\boldsymbol{M}_N]_{f''f'}+[\boldsymbol{M}_N]_{f'f''}\big)\\
&\quad+(\rho_0-\rho_1)\big([\boldsymbol{R}_N]_{f''f'}+[\boldsymbol{R}_N]_{f'f''}\big)\\
&=(\lambda_0-\lambda_1)1(h=l)\big(w_{n,gk}+w_{n,kg}\big)+(\gamma_0-\gamma_1)1(g=k)\big(m_{n,hl}+m_{n,lh}\big)\\
&\quad+(\rho_0-\rho_1)\big(m_{n,hl}w_{n,kg}+m_{n,lh}w_{n,gk}\big)\end{aligned}$$

$$= \left(e_{n,h}' \otimes e_{n,g}'\right)\begin{pmatrix}(\lambda_0-\lambda_1)I_n \otimes (W_n+W_n') + (\gamma_0-\gamma_1)(M_n+M_n')\otimes I_n \\ +(\rho_0-\rho_1)(M_n'\otimes W_n + M_n \otimes W_n')\end{pmatrix}\left(e_{n,l}\otimes e_{n,k}\right).$$

Consider the second part of (C.3) $\sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_0)]_{ff''}[\boldsymbol{A}_N(\delta_0)]_{ff'} - \sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_1)]_{ff''}[\boldsymbol{A}_N(\delta_1)]_{ff'}$. Then, we have

$$\begin{aligned}
&\sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_0)]_{fh}[\boldsymbol{A}_N(\delta_0)]_{fg} - \sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_1)]_{fh}[\boldsymbol{A}_N(\delta_1)]_{fg} \\
&= \sum_{f=1}^{N}\begin{pmatrix}(\lambda_0^2-\lambda_1^2)[\boldsymbol{W}_N]_{ff''}[\boldsymbol{W}_N]_{ff'} + (\gamma_0^2-\gamma_1^2)[\boldsymbol{M}_N]_{ff''}[\boldsymbol{M}_N]_{ff'} + (\rho_0^2-\rho_1^2)[\boldsymbol{R}_N]_{ff''}[\boldsymbol{R}_N]_{ff'} \\ +(\lambda_0\gamma_0-\lambda_1\gamma_1)\left([\boldsymbol{W}_N]_{ff''}[\boldsymbol{M}_N]_{ff'} + [\boldsymbol{M}_N]_{ff''}[\boldsymbol{W}_N]_{ff'}\right) \\ +(\lambda_0\rho_0-\lambda_1\rho_1)\left([\boldsymbol{W}_N]_{ff''}[\boldsymbol{R}_N]_{ff'} + [\boldsymbol{R}_N]_{ff''}[\boldsymbol{W}_N]_{ff'}\right) \\ +(\gamma_0\rho_0-\gamma_1\rho_1)\left([\boldsymbol{M}_N]_{ff''}[\boldsymbol{R}_N]_{ff'} + [\boldsymbol{R}_N]_{ff''}[\boldsymbol{M}_N]_{ff'}\right)\end{pmatrix} \\
&= (\lambda_0\gamma_0-\lambda_1\gamma_1)\left(w_{n,gk}m_{n,hl} + w_{n,kg}m_{n,lh}\right) + (\lambda_0\rho_0-\lambda_1\rho_1)\left(m_{n,hl}+m_{n,lh}\right)\left(\textstyle\sum_{i=1}^{n} w_{n,ig}w_{n,ik}\right) \\
&\quad +(\gamma_0\rho_0-\gamma_1\rho_1)\left(w_{n,gk}+w_{n,kg}\right)\left(\textstyle\sum_{j=1}^{n} m_{n,hj}m_{n,lj}\right) \\
&= \left(e_{n,h}'\otimes e_{n,g}'\right)\begin{pmatrix}(\lambda_0\gamma_0-\lambda_1\gamma_1)(M_n\otimes W_n + M_n \otimes W_n) \\ +(\lambda_0\rho_0-\lambda_1\rho_1)\left((M_n+M_n')\otimes W_n'W_n\right) \\ +(\gamma_0\rho_0-\gamma_1\rho_1)\left(M_nM_n'\otimes (W_n+W_n')\right)\end{pmatrix}\left(e_{n,l}\otimes e_{n,k}\right)
\end{aligned}$$

since $\lambda_1^2=\lambda_0^2$, $\gamma_1^2=\gamma_0^2$, and $\rho_1^2=\rho_0^2$. Since $I_n\otimes(W_n+W_n')$, $(M_n+M_n')\otimes I_n$, $M_n'\otimes W_n + M_n\otimes W_n'$, $M_n\otimes W_n + M_n \otimes W_n$, $(M_n+M_n')\otimes W_n'W_n$, and $M_nM_n'\otimes(W_n+W_n')$ are linearly independent, relation (C.3) implies $\lambda_0=\lambda_1$, $\gamma_0=\gamma_1$, and $\rho_0=\rho_1$.

By putting $\lambda_0=\lambda_1$, $\gamma_0=\gamma_1$, $\rho_0=\rho_1$, and $\sigma_0^2=\sigma_1^2$, (C.2) becomes

$\sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_0)]_{ff'}\boldsymbol{x}_{N,f}\kappa_0 - \boldsymbol{x}_{N,f'}\kappa_0 = \sum_{f=1}^{N}[\boldsymbol{A}_N(\delta_0)]_{ff'}\boldsymbol{x}_{N,f}\kappa_1 - \boldsymbol{x}_{N,f'}\kappa_1$, which is equivalent to $e_{N,f'}'(\boldsymbol{A}_N'(\delta_0)-I_N)\mathbf{X}_N\kappa_0 = e_{N,f'}'(\boldsymbol{A}_N'(\delta_0)-I_N)\mathbf{X}_N\kappa_1$ for an arbitrary f'. It implies $S_N'\mathbf{X}_N\kappa_0 = S_N'\mathbf{X}_N\kappa_1$, so $\mathbf{X}_N\kappa_0=\mathbf{X}_N\kappa_1$ due to invertibility of S_N'. By multiplying $\mathbf{X}_N'$ on both sides, $\mathbf{X}_N'\mathbf{X}_N\kappa_0 = \mathbf{X}_N'\mathbf{X}_N\kappa_1$. The invertibility assumption for $\mathbf{X}_N'\mathbf{X}_N$ yields $\kappa_0=\kappa_1$. Under the conditions in Assumption 4.8, we cannot have $\theta_1\neq\theta_0$. ■

C.2. Asymptotic distribution of the MLE

To prove the asymptotic normality, we need to show the following properties. Recall that $r_{n,ij,\lambda} = \left[\widetilde{\boldsymbol{W}}_N\tilde{S}_N^{-1}\right]_{ff} = \left[\sum_{l=0}^{\infty}\widetilde{\boldsymbol{W}}_N\widetilde{\boldsymbol{A}}_N^l\right]_{ff}$, $r_{n,ij,\gamma} = \left[\widetilde{\boldsymbol{M}}_N\tilde{S}_N^{-1}\right]_{ff} = \left[\sum_{l=0}^{\infty}\widetilde{\boldsymbol{M}}_N\widetilde{\boldsymbol{A}}_N^l\right]_{ff}$, and $r_{n,ij,\rho} = \left[\widetilde{\boldsymbol{R}}_N\tilde{S}_N^{-1}\right]_{ff} = \left[\sum_{l=0}^{\infty}\widetilde{\boldsymbol{R}}_N\widetilde{\boldsymbol{A}}_N^l\right]_{ff}$. Define $r_{n,ij,\lambda\lambda} = \left[\widetilde{\boldsymbol{W}}_N^2\tilde{S}_N^{-2}\right]_{ff}$, $r_{n,ij,\gamma\gamma} = \left[\widetilde{\boldsymbol{M}}_N^2\tilde{S}_N^{-2}\right]_{ff}$, $r_{n,ij,\rho\rho} = \left[\widetilde{\boldsymbol{R}}_N^2\tilde{S}_N^{-2}\right]_{ff}$, $r_{n,ij,\lambda\gamma} = \left[\widetilde{\boldsymbol{W}}_N\widetilde{\boldsymbol{M}}_N\tilde{S}_N^{-2}\right]_{ff}$, $r_{n,ij,\lambda\rho} = \left[\widetilde{\boldsymbol{W}}_N\widetilde{\boldsymbol{R}}_N\tilde{S}_N^{-2}\right]_{ff}$, and $r_{n,ij,\gamma\rho} = \left[\widetilde{\boldsymbol{M}}_N\widetilde{\boldsymbol{R}}_N\tilde{S}_N^{-2}\right]_{ff}$, where $f=(j-1)n+i$. Note that each term in the above can be represented by a Neuman series expansion and an indicator function.

Proposition C.3. Assume that the model's spatial stability and coherency hold.

(i) Under Assumptions 4.1, 4.2 (iii-1), 4.4, and 4.5, $\{r_{n,ij,\lambda}\}$, $\{r_{n,ij,\gamma}\}$, $\{r_{n,ij,\rho}\}$ are uniformly and geometrically L_2-NED on ϵ with the NED coefficient $s\zeta^{s/3\bar{d}}$, where $\bar{d}$ is a constant defined in Assumption 4.2 (iii-1). Moreover, $\{r_{n,ij,\lambda\lambda}\}$, $\{r_{n,ij,\gamma\gamma}\}$, $\{r_{n,ij,\rho\rho}\}$, $\{r_{n,ij,\lambda\gamma}\}$, $\{r_{n,ij,\lambda\rho}\}$, and $\{r_{n,ij,\gamma\rho}\}$ are uniformly and geometrically L_2-NED on ϵ with the NED coefficient $s^2\zeta^{s/3\bar{d}}$.

(ii) Under Assumptions 4.1, 4.2 (iii-2), 4.5, and 4.5, $\{r_{n,ij,\lambda}\}$, $\{r_{n,ij,\gamma}\}$, $\{r_{n,ij,\rho}\}$, $\{r_{n,ij,\lambda\lambda}\}$, $\{r_{n,ij,\gamma\gamma}\}$, $\{r_{n,ij,\rho\rho}\}$, $\{r_{n,ij,\lambda\gamma}\}$, $\{r_{n,ij,\lambda\rho}\}$, and $\{r_{n,ij,\gamma\rho}\}$ are uniformly L_2-NED on ϵ with the NED coefficient $s^{(2d-a)/3}$, where both d and a are constants such that $a > 2d$ in Assumption 4.2 (iii-2).

The proof of Proposition C.3 can be found in the supplement file. For example, the idea of proving the NED properties of $\{r_{n,ij,\lambda}\}$ is to represent $r_{n,ij,\lambda}$ as a Neumann series expansion, i.e.,

$$r_{n,ij,\lambda} = \sum_{l=0}^{\infty}\sum_{i_1,j_1}\cdots\sum_{i_l,j_l} 1(j = j_1) w_{n,ii_1}\boldsymbol{a}_{n,(i_1,j_1),(i_2,j_2)}\cdots\boldsymbol{a}_{n,(i_l,j_l),(i,j)} \times 1\left(\{y_{n,ij} > 0\} \cap \left(\cap_{h=1}^{l}\{y_{n,i_h j_h} > 0\}\right)\right).$$

Then, we decompose $r_{n,ij,\lambda}$ as a finite member of terms (i.e., $\left[\sum_{l=0}^{m}\widetilde{\boldsymbol{W}}_N\widetilde{\boldsymbol{A}}_N^l\right]_{ff}$ for a finite integer m) and a remaining infinite sum (i.e., $\left[\sum_{l=m+1}^{\infty}\widetilde{\boldsymbol{W}}_N\widetilde{\boldsymbol{A}}_N^l\right]_{ff}$). For the finite summation term, we can show its NED properties by applying Propositions C.1 and C.2. The remaining infinite summation term is small under a large m (i.e., $\left[\sum_{l=m+1}^{\infty}\widetilde{\boldsymbol{W}}_N\widetilde{\boldsymbol{A}}_N^l\right]_{ff} \to 0$ as $m \to \infty$).

Note that deriving the asymptotic distribution of $\hat{\theta}_N$ relies on the Taylor expansion argument: $\sqrt{N}(\hat{\theta}_N - \theta_0) = \left(-\frac{1}{N}\frac{\partial^2 \ln L_N^*(\tilde{\theta}_N)}{\partial\theta\partial\theta'}\right)^{-1}\frac{1}{\sqrt{N}}\frac{\partial \ln L_N^*(\theta_0)}{\partial\theta}$, where $\tilde{\theta}_N$ lies between $\hat{\theta}_N$ and θ_0. After establishing $\hat{\theta}_N \xrightarrow{p} \theta_0$, our direction of proof is to show $\frac{1}{N}\frac{\partial^2 \ln L_N(\tilde{\theta}_N)}{\partial\theta\partial\theta'} - E\left(\frac{1}{N}\frac{\partial^2 \ln L_N(\theta_0)}{\partial\theta\partial\theta'}\right) = o_p(1)$ and $\frac{1}{\sqrt{N}}\sum_{i,j=1}^{n} q_{n,ij}(\theta_0) \xrightarrow{d} N(\boldsymbol{0}, \Sigma_{\theta_0}^*)$, where $q_{n,ij}(\theta)$ denotes the (i,j)-component of the score evaluated at θ (i.e., $\frac{\partial \ln L_N^*(\theta_0)}{\partial\theta} = \sum_{i,j=1}^{n} q_{n,ij}(\theta_0)$) and $\Sigma_{\theta_0}^* = \lim_{n\to\infty}\Sigma_{\theta_0,N}^*$ with $\Sigma_{\theta_0,N}^* = \frac{1}{N}Var\left(\sum_{i,j=1}^{n} q_{n,ij}(\theta_0)\right)$. In the supplement file, we provide the first and second order conditions. The set of first-order conditions can be written as the summation of $q_{n,ij}(\theta)$ for each $\theta \in \Theta$. The next proposition characterizes the asymptotic distribution of $\frac{1}{\sqrt{N}}\sum_{i,j=1}^{n} q_{n,ij}(\theta_0)$.

Proposition C.4. We additionally assume Assumptions 4.10 and 4.11. Under Assumption 4.2. (iii-2), Assumption 4.12 is additionally needed. Then, $\frac{1}{\sqrt{N}}\sum_{i,j=1}^{n} q_{n,ij}(\theta_0) \xrightarrow{d} N(\boldsymbol{0}, \Sigma_{\theta_0}^*)$, where $\Sigma_{\theta_0}^* = \lim_{n\to\infty}\Sigma_{\theta_0,N}^*$ with $\Sigma_{\theta_0,N}^* = \frac{1}{N}Var\left(\sum_{i,j=1}^{n} q_{n,ij}(\theta_0)\right)$.

Proposition C.4 is the application of Corollary 1 of Jenish and Prucha (2012). Assumption 4.10 corresponds to Assumption 3 in Jenish and Prucha (2012). Hence, the remaining point of proving Proposition C.4 is to have the uniform $L_{2+\tilde{\eta}}$-integrability of $\{\|q_{n,ij}(\theta_0)\|\}$ (Assumption 4 in Jenish and Prucha (2012). Proving asymptotic normality is showing $\frac{1}{N}\frac{\partial^2 \ln L_N^*(\tilde{\theta}_N)}{\partial\theta\partial\theta'} - E\left(\frac{1}{N}\frac{\partial^2 \ln L_N^*(\theta_0)}{\partial\theta\partial\theta'}\right) = o_p(1)$ and applying the Slutsky's lemma with Proposition C.4. For having $\frac{1}{N}\frac{\partial^2 \ln L_N^*(\tilde{\theta}_N)}{\partial\theta\partial\theta'} - E\left(\frac{1}{N}\frac{\partial^2 \ln L_N^*(\theta_0)}{\partial\theta\partial\theta'}\right) = o_p(1)$, we then need to check regularity conditions (Assumption 2 in Jenish and Prucha (2012)) to apply Theorem 1 in Jenish and Prucha (2012). The detailed proofs of Proposition C.4 and Theorem 4 (asymptotic normality) can be found in the supplement file.

D. Asymptotic properties of the MLE for the SARF Tobit model with the two-way fixed effects

Fernandez-Val and Weidner (2016) study the asymptotic distribution of parameters in nonlinear panel models with individual and time effects when n and T are large (i.e., large-T version of the incidental parameter problem). To derive the asymptotic distribution of $\widehat{\omega}_N$, we will employ the notions established by Fernandez-Val and Weidner (2016) due to some similarities in terms of the framework. First, they consider the case $0 < \lim_{n,T\to\infty} \frac{n}{T} < \infty$ (Assumption 4.1 (i) in Fernandez-Val and Weidner (2016)). It corresponds to our case since we always have the same number of units (n) for origins and destinations.[52] Second, our fixed-effect specification belongs to the additive separable two-way fixed-effect specification (Assumption 4.1 (iii) in Fernandez-Val and Weidner (2016)). Third, our statistical objective function $\ln L_N^*(\omega, \boldsymbol{\alpha}_N)$ is infinitely differentiable, so it satisfies smoothness conditions (Assumption 4.1 (iv) in Fernandez-Val and Weidner (2016)). The strict concavity of $\ln L_N^*(\omega, \boldsymbol{\alpha}_N)$ in Assumption 4.1 (v) in Fernandez-Val and Weidner (2016) can be achieved by the reparameterization (Olsen, 1978). The difference comes from the dependence concept among observations. Their framework allows a general type of time dependence with cross-sectionally independent samples. On the other hand, our setting considers the weak cross-sectional dependence (characterized by the NED concept) across origins and destinations.

In the main draft, we provide the brief description of the arguments for consistency and asymptotic normality. The supplement file contains the arguments in detail.

The proposition below shows the structure of $E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}$. Note that the diagonal terms of $E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)$ are of $O(1)$, and its off-diagonal terms are of $O\left(\frac{1}{n}\right)$. When $E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)$ is invertible under a large n, $E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}$ can be approximated by a diagonal matrix.

Proposition D.1. We denote $E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1} \equiv \bar{\mathcal{H}}_n = \begin{bmatrix} \bar{\mathcal{H}}_{(\alpha_o\alpha_o),n} & \bar{\mathcal{H}}_{(\alpha_o\alpha_d),n} \\ \bar{\mathcal{H}}'_{(\alpha_o\alpha_d),n} & \bar{\mathcal{H}}_{(\alpha_d\alpha_d),n} \end{bmatrix}$, a block matrix. Under the same regularity conditions for Theorem 4.5, (i) $a_{n,jj} = \left[\bar{\mathcal{H}}_{(\alpha_o\alpha_o),n}\right]_{jj}$ for $j = 1,\cdots,n$ and $c_{n,ii} = \left[\bar{\mathcal{H}}_{(\alpha_d\alpha_d),n}\right]_{ii}$ for $i = 1,\cdots,n$ are of $O(1)$; and $\left\|\bar{\mathcal{H}}_n - \widetilde{\mathcal{H}}_n\right\|_{max} = O\left(\frac{1}{n}\right)$, where $\|A_n\|_{max} = \max_{i,j}\left|[A_n]_{ij}\right|$ and $\widetilde{\mathcal{H}}_n = diag\left(\left\{a_{n,jj}\right\}_{j=1}^n, \left\{c_{n,ii}\right\}_{i=1}^n\right)$ is an approximation of $\bar{\mathcal{H}}_n$.

Note that the dimension of $\widehat{\boldsymbol{\alpha}}_N(\omega)$ (and $\boldsymbol{\alpha}_N^0$) is $2n$, which grows as n increases. In order to evaluate an $2n \times 1$ vector, an $2n \times 2n$ matrix (e.g., $-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}$), and so on, we consider the q-norm $\|\cdot\|_q$ for $2 \le q \le \infty$.[53] To characterize parameters near the true ones, we define the closed balls of radius $r \ge 0$: (i) for ω_0, let $\mathcal{B}(\omega_0, r) = \{\omega : \|\omega - \omega_0\| \le r\}$, and (ii) for $\boldsymbol{\alpha}_N^0$, let $\mathcal{B}_q(\boldsymbol{\alpha}_N^0, r) = \{\boldsymbol{\alpha}_N : \|\boldsymbol{\alpha}_N - \boldsymbol{\alpha}_N^0\|_q \le r\}$.

Step 1: As the first step to show the asymptotic distribution of $\widehat{\omega}_N$, we need to have Taylor

[52] In our setting, we always have $\frac{\#\text{ of origin units}}{\#\text{ of destination units}} = 1$.

[53] The q-norm for a matrix and/or a tensor is defined the induced vector norm. Details can be found in the supplement file.

approximations of $\widehat{\boldsymbol{\alpha}}_N(\omega)-\boldsymbol{\alpha}_N^0$ and $\frac{1}{\sqrt{N}}\frac{\partial \ln L_N^*(\omega,\widehat{\boldsymbol{\alpha}}_N(\omega))}{\partial\omega}$ for given ω. To obtain the Taylor expansions of them, some regularity conditions should be checked (see Lemma D.1 in the supplement file). Those conditions are the counterpart of Assumption B.1 in Fernandez-Val and Weidner (2016).

The implication of the proposition below gives bounds of Taylor approximations' ($\widehat{\boldsymbol{\alpha}}_N(\omega)-\boldsymbol{\alpha}_N^0$ and $\frac{1}{\sqrt{N}}\frac{\partial \ln L_N^*(\omega,\widehat{\boldsymbol{\alpha}}_N(\omega))}{\partial\omega}$) the remainder terms if one takes r_α-consistent estimator $\widetilde{\omega}_N$ for $\widehat{\boldsymbol{\alpha}}_N(\widetilde{\omega}_N)-\boldsymbol{\alpha}_N^0$ and $\frac{1}{\sqrt{N}}\frac{\partial \ln L_N^*(\widetilde{\omega}_N,\widehat{\boldsymbol{\alpha}}_N(\widetilde{\omega}_N))}{\partial\omega}$.

Proposition D.2. Assume the results in Lemma D.1 in the supplement hold and $\sup_{\omega\in\mathcal{B}(\omega_0,r_\omega)}\|\widehat{\boldsymbol{\alpha}}_N(\omega)-\boldsymbol{\alpha}_N^0\|_q=o_p(r_\alpha)$. Let $q=4+\eta$ for some $\eta>0$ and $0\le\varepsilon<\frac{1}{4}-\frac{1}{q}$. For $\omega\in\mathcal{B}(\omega_0,r_\omega)$ and $\boldsymbol{\alpha}_N\in\mathcal{B}_q(\boldsymbol{\alpha}_N^0,r_\alpha)$ where $r_\alpha=o(n^{-\varepsilon})$ and $r_\omega=o\left(n^{-\frac{1}{q}-\varepsilon}\right)$, we have the two results below:

(i) For a given $\omega\in\mathcal{B}(r_\omega,\omega_0)$, the Taylor expansion of $\widehat{\boldsymbol{\alpha}}_N(\omega)$ around $\boldsymbol{\alpha}_N^0$ is

$$\widehat{\boldsymbol{\alpha}}_N(\omega)-\boldsymbol{\alpha}_N^0=\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N}+\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\omega'}(\omega-\omega_0)$$

$$+\frac{1}{2}\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\sum_{j=1}^n\left\{u_{\alpha_o,j}\frac{1}{n}\frac{\partial^3\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'\partial\alpha_{j,o}}\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N}\right\}$$

$$+\frac{1}{2}\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\sum_{i=1}^n\left\{u_{\alpha_d,i}\frac{1}{n}\frac{\partial^3\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'\partial\alpha_{i,d}}\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N}\right\}+\mathcal{R}_N^\alpha(\omega),$$

where $u_{\alpha_o,j}$ is the jth element of an $n\times 1$ vector $\bar{\mathcal{H}}_{(\alpha_o\alpha_o),n}\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_{n,o}}+\bar{\mathcal{H}}_{(\alpha_o\alpha_d),n}\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_{n,d}}$, $u_{\alpha_d,i}$ denotes the ith element of $\bar{\mathcal{H}}_{(\alpha_d\alpha_o),n}\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_{n,o}}+\bar{\mathcal{H}}_{(\alpha_d\alpha_d),n}\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_{n,d}}$, $\mathcal{R}_N^\alpha(\omega)$ denotes the remainder term.[54] Note that $\|\mathcal{R}_N^\alpha(\omega)\|_q=o_p\left(n^{-1+\frac{1}{q}}\right)+o_p\left(n^{\frac{1}{q}}\cdot\|\omega-\omega_0\|\right)$ for $\omega\in\mathcal{B}(\omega_0,r_\omega)$.

(ii) For a given $\omega\in\mathcal{B}(r_\omega,\omega_0)$, the Taylor expansion of $\frac{1}{\sqrt{N}}\frac{\partial\ln L_N^*(\omega,\widehat{\boldsymbol{\alpha}}_N(\omega))}{\partial\omega}$ can be represented by

$$\frac{1}{\sqrt{N}}\frac{\partial\ln L_N^*(\omega,\widehat{\boldsymbol{\alpha}}_N(\omega))}{\partial\omega}=-\Sigma_{\omega_0,N}^*\sqrt{N}(\omega-\omega_0)+U_N^{(0)}+U_N^{(1)}+\mathcal{R}_N(\omega),$$

where

$$\Sigma_{\omega_0,N}^*=E\left(-\frac{1}{N}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\omega'}\right)-\frac{1}{n}\left\{E\left(\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'}\right)E\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}E\left(\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\omega'}\right)\right\},$$

$$U_N^{(0)}=\frac{1}{\sqrt{N}}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega}+E\left(\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'}\right)\cdot E\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\cdot\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N},$$

$U_N^{(1)}=U_N^{(1,a)}+U_N^{(1,b)}$ with $U_N^{(1,a)}=U_N^{(1,a,1)}+U_N^{(1,a,2)}$,

$$U_N^{(1,a,1)}=\left\{\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'}-E\left(\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'}\right)\right\}\cdot E\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\cdot\frac{1}{n}\frac{\partial\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N},$$

$$U_N^{(1,a,2)}=-E\left(\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'}\right)\cdot E\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\left\{-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}-E\left(\left(-\frac{1}{n}\frac{\partial^2\ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)\right)\right\}$$

[54] Since $\bar{\mathcal{H}}_n$ is symmetric, note that $\bar{\mathcal{H}}_{(\alpha_d\alpha_o),n}=\bar{\mathcal{H}}_{(\alpha_o\alpha_d),n}'$.

$$\cdot E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\frac{1}{n}\frac{\partial \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N},$$

$$U_N^{(1,b)} = \frac{1}{2}\sum_{g=1}^{\dim(\boldsymbol{\alpha}_N)}\left(E\left(\frac{1}{n}\frac{\partial^3 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'\partial\alpha_g}\right) + E\left(\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'}\right)\cdot E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\cdot\right.$$

$$\left.E\left(\frac{1}{n}\frac{\partial^3 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'\partial\alpha_g}\right)\right)\cdot\left[\overline{\mathcal{H}}_n\frac{1}{n}\frac{\partial \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N}\right]_g E\left(-\frac{1}{n}\frac{\partial^2 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'}\right)^{-1}\cdot\frac{1}{n}\frac{\partial \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N},$$

$\mathcal{R}_N(\omega)$ denotes the remainder term satisfying $\|\mathcal{R}_N(\omega)\| = o_p(1) + o_p(n\cdot\|\omega-\omega_0\|)$ for $\omega \in \mathcal{B}(\omega_0, r_\omega)$, and α_g denote the gth-element of $\boldsymbol{\alpha}_N$ and $\dim(\boldsymbol{\alpha}_N) = 2n$.

In terms of the Fernandez-Val and Weidner's (2016) expressions. the reminder terms satisfy $\sup_{\omega\in\mathcal{B}(\omega_0,r_\omega)}\frac{n^{1-\frac{1}{q}}\|\mathcal{R}_N^\alpha(\omega)\|_q}{1+n\cdot\|\omega-\omega_0\|} = o_p(1)$ and $\sup_{\omega\in\mathcal{B}(\omega_0,r_\omega)}\frac{\|\mathcal{R}_N(\omega)\|}{1+n\cdot\|\omega-\omega_0\|} = o_p(1)$. Since we will finally achieve $\|\widehat{\omega}_N - \omega_0\| = O_p(n^{-1}) = O_p\left(\frac{1}{\sqrt{N}}\right)$, we will have $\|\mathcal{R}_N^\alpha(\widehat{\omega}_N)\|_q = o_p\left(n^{-1+\frac{1}{q}}\right)$ and $\|\mathcal{R}_N(\widehat{\omega}_N)\| = o_p(1)$.

Step 2: One additional condition for Proposition D.2 is $\|\widehat{\boldsymbol{\alpha}}_N(\omega) - \boldsymbol{\alpha}_N^0\|_q = O\left(n^{-\frac{1}{2}+\frac{1}{q}}\right)$ if $\|\omega-\omega_0\| = O\left(n^{-\frac{1}{2}}\right)$. Lemmas D.2 and D.3 in the supplement file shows that $\|\widehat{\omega}_N - \omega_0\| = O_p\left(n^{-\frac{1}{2}}\right)$ and $\|\widehat{\boldsymbol{\alpha}}_N(\widehat{\omega}_N) - \boldsymbol{\alpha}_N^0\|_q = O\left(n^{-\frac{1}{2}+\frac{1}{q}}\right)$ if $\Sigma_{\omega_0}^* = \lim_{n\to\infty}\Sigma_{\omega_0,N}^*$ is nonsingular and $\Sigma_{\omega_0}^* > 0$. Showing Lemmas D.2 and D.3 relies on strict concavity of the log-likelihood function. For this, we consider the reparameterization suggested by Olsen (1978), i.e., $\mathcal{T}: (\omega, \boldsymbol{\alpha}_N) \mapsto (\omega^*, \boldsymbol{\alpha}_N^*)$. By showing the results for the re-parameterized MLEs, we can obtain the desired results using the functional invariance property of the MLE (i.e., $(\widehat{\omega}_N, \widehat{\boldsymbol{\alpha}}_N) = \mathcal{T}^{-1}(\widehat{\omega}_N^*, \widehat{\boldsymbol{\alpha}}_N^*)$). After verifying $U_N^{(0)} + U_N^{(1)} = O_p(1)$, we can achieve $\|\widehat{\omega}_N - \omega_0\| = O_p(n^{-1}) = O_p\left(\frac{1}{\sqrt{N}}\right)$ (a part of Proposition D.3 shows $U_N^{(0)} + U_N^{(1)} = O_p(1)$).

Then, $U_N^{(0)}$ is the main part of the asymptotic distribution of $\widehat{\omega}_N$ while $U_N^{(1)}$ characterizes the asymptotic bias of $\widehat{\omega}_N$. We will show that $U_N^{(0)} \xrightarrow{d} N\left(0, \Sigma_{\omega_0}^*\right)$ where $\Sigma_{\omega_0}^* = \lim_{n\to\infty}\Sigma_{\omega_0,N}^*$ and $U_N^{(1)} - \Lambda_N^* = o_p(1)$ for some Λ_N^*. The lemma below characterizes the form of Λ_N^*.

Step 3: The proposition below characterizes the key terms of the asymptotic expansion $\widehat{\omega}_N$: (i) $U_N^{(0)}$ and (ii) $U_N^{(1)}$.

Proposition D.3. (i) $U_N^{(0)} \xrightarrow{d} N\left(\mathbf{0}, \Sigma_{\omega_0}^*\right)$ as $n\to\infty$; and (ii) $U_N^{(1,a,1)} - \left(\Lambda_{1,N}^* + \Lambda_{2,N}^*\right) \xrightarrow{p} 0$, $U_N^{(1,a,2)} - \left(\Lambda_{3,N}^* + \Lambda_{4,N}^*\right) \xrightarrow{p} 0$, and $U_N^{(1,b)} - (\Lambda_{5,N}^* + \Lambda_{6,N}^*) \xrightarrow{p} 0$ as $n\to\infty$,

where $\Lambda_{1,N}^* = \frac{1}{n}\sum_{j=1}^n a_{n,jj}\frac{1}{n}\sum_{k=1}^n\sum_{i=1}^n E\left(q_{n,kj}^{\alpha_o}h_{n,ij}^{\omega\alpha_o}\right)$,
$\Lambda_{2,N}^* = \frac{1}{n}\sum_{i=1}^n c_{n,ii}\frac{1}{n}\sum_{l=1}^n\sum_{j=1}^n E(q_{n,il}^{\alpha_d}h_{n,ij}^{\omega\alpha_d})$,
$\Lambda_{3,N}^* = \frac{1}{n}\sum_{j=1}^n a_{n,jj}\left(\frac{1}{n}\sum_{k=1}^n E\left(h_{n,kj}^{\omega\alpha_o}\right)\right)\sum_{i=1}^n E\left(h_{n,ij}^{\alpha_o}v_{\alpha_o,n,j}\right)$,
$\Lambda_{4,N}^* = \frac{1}{n}\sum_{i=1}^n c_{n,ii}\left(\frac{1}{n}\sum_{l=1}^n E\left(h_{n,il}^{\omega\alpha_d}\right)\right)\sum_{j=1}^n E\left(h_{n,ij}^{\alpha_d}v_{\alpha_d,n,i}\right)$,
$\Lambda_{5,N}^* = \frac{1}{2n}\sum_{j=1}^n \widetilde{\omega}_{n,(\alpha_o\alpha_o),jj}\,a_{n,jj}^2\frac{1}{n}\sum_{k=1}^n\sum_{l=1}^n E\left(q_{n,kj}^{\alpha_o}q_{n,lj}^{\alpha_o}\right)$, and
$\Lambda_{6,N}^* = \frac{1}{2n}\sum_{i=1}^n \widetilde{\omega}_{n,(\alpha_d\alpha_d),ii}c_{n,ii}^2\frac{1}{n}\sum_{k=1}^n\sum_{l=1}^n E(q_{n,ik}^{\alpha_d}q_{n,il}^{\alpha_d})$,

with $v_{\alpha_o,n,j} = \frac{1}{n}\sum_{k=1}^{n}\sum_{p=1}^{n} a_{n,jk} q_{n,pk}^{\alpha_o} + \frac{1}{n}\sum_{l=1}^{n}\sum_{q=1}^{n} b_{n,jl} q_{n,lq}^{\alpha_d}$ for $j = 1, \cdots, n$,
$v_{\alpha_d,n,i} = \frac{1}{n}\sum_{k=1}^{n}\sum_{p=1}^{n} b_{n,ki} q_{n,pk}^{\alpha_o} + \frac{1}{n}\sum_{l=1}^{n}\sum_{q=1}^{n} c_{n,il} q_{n,lq}^{\alpha_d}$ for $i = 1, \cdots n$,
$\widetilde{\omega}_{n,(\alpha_o\alpha_o),jj} = \frac{1}{n}\sum_{i=1}^{n} E\left(t_{n,ij}^{\omega\alpha_o}\right) + \frac{1}{n}\pi_{\alpha_o,n,j}\sum_{i=1}^{n} E\left(t_{n,ij}^{\alpha_o}\right) + \frac{1}{n}\sum_{i=1}^{n}\pi_{\alpha_d,n,i} E\left(t_{n,ij}^{\alpha_d}\right)$ for $j = 1, \cdots, n$,
$\widetilde{\omega}_{n,(\alpha_d\alpha_d),ii} = \frac{1}{n}\sum_{j=1}^{n} E\left(t_{n,ij}^{\omega\alpha_d}\right) + \frac{1}{n}\sum_{j=1}^{n}\pi_{\alpha_o,n,j} E\left(t_{n,ij}^{\alpha_o}\right) + \frac{1}{n}\pi_{\alpha_d,n,i}\sum_{j=1}^{n} E\left(t_{n,ij}^{\alpha_d}\right)$ for $i = 1, \cdots n$,
where $\pi_{\alpha_o,n,j} = \frac{1}{n}\sum_{k=1}^{n}\sum_{p=1}^{n} a_{n,jk} E\left(h_{n,pk}^{\omega\alpha_o}\right) + \frac{1}{n}\sum_{l=1}^{n}\sum_{q=1}^{n} b_{n,jl} E\left(h_{n,lq}^{\omega\alpha_d}\right)$ for $j = 1, \cdots, n$, and
$\pi_{\alpha_d,n,i} = \frac{1}{n}\sum_{k=1}^{n}\sum_{p=1}^{n} b_{n,ki} E\left(h_{n,pk}^{\omega\alpha_o}\right) + \frac{1}{n}\sum_{l=1}^{n}\sum_{q=1}^{n} c_{n,il} E\left(h_{n,lq}^{\omega\alpha_d}\right)$ for $i = 1, \cdots n$.

Then, $\Lambda_N^* = \Lambda_{1,N}^* + \Lambda_{2,N}^* + \Lambda_{3,N}^* + \Lambda_{4,N}^* + \Lambda_{5,N}^* + \Lambda_{6,N}^*$.

Step 4: By applying Proposition D.2 (ii) and Proposition D.3, we have

$$\sqrt{N}(\widehat{\omega}_N - \omega_0) = \Sigma_{\omega_0}^*\left(U_N^{(0)} + U_N^{(1)}\right) + o_p(1) \xrightarrow{d} N\left(\Sigma_{\omega_0}^{*-1}\Lambda_\infty^*, \Sigma_{\omega_0}^{*-1}\right) \text{ as } n \to \infty,$$

where $\Lambda_\infty^* = \lim_{n\to\infty}\Lambda_N^*$ with $\Lambda_N^* = \Lambda_{1,N}^* + \Lambda_{2,N}^* + \Lambda_{3,N}^* + \Lambda_{4,N}^* + \Lambda_{5,N}^* + \Lambda_{6,N}^*$ (for details, refer to Lemma D.4 in the supplement).

Derivatives: In this part, we provide the detailed forms of the key derivative components. Consider the first-order derivatives. Then, $\frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\omega} = \sum_{i=1}^{n}\sum_{j=1}^{n} q_{n,ij}^{\omega}\left(\omega, \alpha_{j,o}, \alpha_{i,d}\right)$ with

$$q_{n,ij}^{\omega}\left(\omega, \alpha_{j,o}, \alpha_{i,d}\right) = \left(\frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda} \quad \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma} \quad \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho} \quad \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta'} \quad \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma^2}\right)'$$
$$= \begin{pmatrix} -1\left(y_{n,ij}=0\right)\sigma^{-1}\frac{\phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gj}\right)}{\Phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)} - r_{n,ij,\lambda}(\delta) + 1\left(y_{n,ij}>0\right)\sigma^{-1}\epsilon_{n,ij}^{+,*}\left(\omega,\alpha_{j,o},\alpha_{i,d}\right)\left(\sum_{g=1}^{n} w_{n,ig}y_{n,gj}\right) \\ -1\left(y_{n,ij}=0\right)\sigma^{-1}\frac{\phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\sum_{h=1}^{n} y_{n,ih}m_{n,hj}\right)}{\Phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)} - r_{n,ij,\gamma}(\delta) + 1\left(y_{n,ij}>0\right)\sigma^{-1}\epsilon_{n,ij}^{+,*}\left(\omega,\alpha_{j,o},\alpha_{i,d}\right)\left(\sum_{h=1}^{n} y_{n,ih}m_{n,hj}\right) \\ -1\left(y_{n,ij}=0\right)\sigma^{-1}\frac{\phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\sum_{g,h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}\right)}{\Phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)} - r_{n,ij,\rho}(\delta) + 1\left(y_{n,ij}>0\right)\sigma^{-1}\epsilon_{n,ij}^{+,*}\left(\omega,\alpha_{j,o},\alpha_{i,d}\right)\left(\sum_{g,h=1}^{n} w_{n,ig}y_{n,gh}m_{n,hj}\right) \\ -1\left(y_{n,ij}=0\right)\sigma^{-1}\frac{\phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\boldsymbol{z}_{n,ij}}{\Phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)} + 1\left(y_{n,ij}>0\right)\sigma^{-1}\epsilon_{n,ij}^{+,*}\left(\omega,\alpha_{j,o},\alpha_{i,d}\right)\boldsymbol{z}_{n,ij} \\ -\frac{1}{2\sigma^2}1\left(y_{n,ij}=0\right)\frac{\phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})}{\Phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)} - \frac{1}{2\sigma^2}1\left(y_{n,ij}>0\right) + \frac{1}{2\sigma^2}1\left(y_{n,ij}>0\right)\epsilon_{n,ij}^{+,*}\left(\omega,\alpha_{j,o},\alpha_{i,d}\right)^2 \end{pmatrix}.$$

At the true parameter values, let $q_{n,ij}^{\omega} = q_{n,ij}^{\omega}\left(\omega_0, \alpha_{j,o,0}, \alpha_{i,d,0}\right)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n$. Other quantities below are similarly defined.

Consider $\frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\boldsymbol{\alpha}_N} = \left(\frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,o}}, \cdots, \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,o}}, \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,d}}, \cdots, \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,d}}\right)'$. Observe that

$$\begin{pmatrix} \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,o}} \\ \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{2,o}} \\ \vdots \\ \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,o}} \end{pmatrix} = \sum_{i=1}^{n}\begin{pmatrix} q_{n,i1}^{\alpha_o}\left(\omega,\alpha_{1,o},\alpha_{i,d}\right) \\ q_{n,i2}^{\alpha_o}\left(\omega,\alpha_{2,o},\alpha_{i,d}\right) \\ \vdots \\ q_{n,in}^{\alpha_o}\left(\omega,\alpha_{n,o},\alpha_{i,d}\right) \end{pmatrix}, \text{ and } \begin{pmatrix} \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,d}} \\ \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{2,d}} \\ \vdots \\ \frac{\partial \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,d}} \end{pmatrix} = \sum_{j=1}^{n}\begin{pmatrix} q_{n,1j}^{\alpha_d}\left(\omega,\alpha_{j,o},\alpha_{1,d}\right) \\ q_{n,2j}^{\alpha_d}\left(\omega,\alpha_{j,o},\alpha_{2,d}\right) \\ \vdots \\ q_{n,n,j}^{\alpha_d}\left(\omega,\alpha_{j,o},\alpha_{n,d}\right) \end{pmatrix}$$

where $q^{\alpha_o}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) = q_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) - \mu\frac{1}{n}\sum_{j=1}^{n}\alpha_{j,o} + \mu\alpha_{i,d}$ for $j = 1,\cdots,n$, $q^{\alpha_d}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) = q_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) + \mu\alpha_{j,o} - \mu\frac{1}{n}\sum_{i=1}^{n}\alpha_{i,d}$ for $i = 1,\cdots,n$, and $q_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) = -1(y_{n,ij} = 0)\sigma^{-1}\frac{\phi\left(\epsilon^{+,*}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})\right)}{\Phi\left(\epsilon^{+,*}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})\right)} + 1(y_{n,ij} > 0)\sigma^{-1}\epsilon^{+,*}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$ for each (i,j).

Consider the relevant components of the second-order derivatives.

$$\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'} = \begin{bmatrix}
\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,o}\partial\alpha_{1,o}} & 0 & \cdots & 0 & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,o}\partial\alpha_{1,d}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,o}\partial\alpha_{2,d}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,o}\partial\alpha_{n,d}} \\
* & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{2,o}\partial\alpha_{2,o}} & \cdots & 0 & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{2,o}\partial\alpha_{1,d}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{2,o}\partial\alpha_{2,d}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{2,o}\partial\alpha_{n,d}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
* & * & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,o}\partial\alpha_{n,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,o}\partial\alpha_{1,d}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,o}\partial\alpha_{2,d}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,o}\partial\alpha_{n,d}} \\
* & * & \cdots & * & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{1,d}\partial\alpha_{1,d}} & 0 & \cdots & 0 \\
* & * & \cdots & * & * & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{2,d}\partial\alpha_{2,d}} & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
* & * & \cdots & * & * & * & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{n,d}\partial\alpha_{n,d}}
\end{bmatrix},$$

where $\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{j,o}^2} = \sum_{i=1}^{n} h^{\alpha_o}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$ for $j = 1,\dots,n$; $\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{j,o}\partial\alpha_{k,o}} = -\mu$ for $j \neq k$;
$\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{j,o}\partial\alpha_{i,d}} = h_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) + \mu$ for $j = 1,\dots,n$ and $i = 1,\dots,n$;
$\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{i,d}^2} = \sum_{j=1}^{n} h^{\alpha_d}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$ for $i = 1,\dots,n$; and
$\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{i,d}\partial\alpha_{l,d}} = -\mu$ for $i \neq l$ with
$h^{\alpha_o}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) = h_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) - \frac{\mu}{n}$ for $j = 1,\dots,n$,
$h^{\alpha_d}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) = h_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) - \frac{\mu}{n}$ for $i = 1,\dots,n$.
$h_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d}) = 1(y_{n,ij} = 0)\sigma^{-2}\psi\left(\epsilon^{+,*}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})\right) - 1(y_{n,ij} > 0)\sigma^{-2}$.

Consider the elements of $h^{\omega\alpha_o}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$ for $j = 1,\cdots,n$ and $h^{\omega\alpha_d}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$ for $i = 1,\cdots,n$.

$$\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\omega\partial\boldsymbol{\alpha}_N'} = \begin{bmatrix}
\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{1,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{2,o}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{n,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{1,d}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{2,d}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{n,d}} \\
\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{1,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{2,o}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{n,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{1,d}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{2,d}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{n,d}} \\
\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{1,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{2,o}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{n,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{1,d}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{2,d}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{n,d}} \\
\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{1,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{2,o}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{n,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{1,d}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{2,d}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{n,d}} \\
\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma\partial\alpha_{1,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma\partial\alpha_{2,o}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma\partial\alpha_{n,o}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma\partial\alpha_{1,d}} & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma\partial\alpha_{2,d}} & \cdots & \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma\partial\alpha_{n,d}}
\end{bmatrix},$$

where for $j = 1,\dots,n$,

$\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{j,o}} = \sum_{i=1}^{n} h^{\lambda\alpha_o}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$, $\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{j,o}} = \sum_{i=1}^{n} h^{\gamma\alpha_o}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$,
$\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{j,o}} = \sum_{i=1}^{n} h^{\rho\alpha_o}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$, $\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{j,o}} = \sum_{i=1}^{n} h^{\beta\alpha_o}_{n,ij}(\omega,\alpha_{j,o},\alpha_{i,d})$, and

$\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma^2\partial\alpha_{j,o}} = \sum_{i=1}^n h_{n,ij}^{\sigma^2\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d})$ with

$$
\begin{aligned}
h_{n,ij}^{\lambda\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= 1(y_{n,ij}=0)\sigma^{-2}\psi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\textstyle\sum_{g=1}^n w_{n,ig}y_{n,gj}\right)\\
&\quad -1(y_{n,ij}>0)\sigma^{-2}\left(\textstyle\sum_{g=1}^n w_{n,ig}y_{n,gj}\right),\\
h_{n,ij}^{\gamma\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= 1(y_{n,ij}=0)\sigma^{-2}\psi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\textstyle\sum_{h=1}^n y_{n,ih}m_{n,hj}\right)\\
&\quad -1(y_{n,ij}>0)\sigma^{-2}\left(\textstyle\sum_{h=1}^n y_{n,ih}m_{n,hj}\right),\\
h_{n,ij}^{\rho\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= 1(y_{n,ij}=0)\sigma^{-2}\psi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\textstyle\sum_{g,h=1}^n w_{n,ig}y_{n,gh}m_{n,hj}\right)\\
&\quad -1(y_{n,ij}>0)\sigma^{-2}\left(\textstyle\sum_{g,h=1}^n w_{n,ig}y_{n,gh}m_{n,hj}\right),\\
h_{n,ij}^{\beta\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= 1(y_{n,ij}=0)\sigma^{-2}\psi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\boldsymbol{z}_{n,ij} - 1(y_{n,ij}>0)\sigma^{-2}\boldsymbol{z}_{n,ij}, \text{ and}\\
h_{n,ij}^{\sigma^2\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= 1(y_{n,ij}=0)\frac{1}{2\sigma^3}\left[\frac{\phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)}{\Phi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)} + \psi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right]\\
&\quad -1(y_{n,ij}>0)\frac{1}{\sigma^3}\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d}).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{i,d}} = \sum_{j=1}^n h_{n,ij}^{\lambda\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\ \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{i,d}} = \sum_{j=1}^n h_{n,ij}^{\gamma\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\\
&\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{i,d}} = \sum_{j=1}^n h_{n,ij}^{\rho\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\ \frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{i,d}} = \sum_{j=1}^n h_{n,ij}^{\beta\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}), \text{ and}\\
&\frac{\partial^2 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma^2\partial\alpha_{i,d}} = \sum_{j=1}^n h_{n,ij}^{\sigma^2\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) \text{ for all } i=1,\cdots,n.
\end{aligned}
$$

For the third-order derivative, define $\varphi(x) = \frac{d(\psi(x))}{dx}$. For $\frac{\partial^3 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'\partial\alpha_{j,o}}$ and $\frac{\partial^3 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\omega\partial\boldsymbol{\alpha}_N'\partial\alpha_{i,d}}$, the relevant terms are

For $j=1,\dots,n$,

$$
\begin{aligned}
&\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{j,o}^2} = \sum_{i=1}^n t_{n,ij}^{\lambda\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\ \frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{j,o}^2} = \sum_{i=1}^n t_{n,ij}^{\gamma\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\\
&\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{j,o}^2} = \sum_{i=1}^n t_{n,ij}^{\rho\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\ \frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{j,o}^2} = \sum_{i=1}^n t_{n,ij}^{\beta\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}), \text{ and}\\
&\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma^2\partial\alpha_{j,o}^2} = \sum_{i=1}^n t_{n,ij}^{\sigma^2\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),
\end{aligned}
$$

where $t_{n,ij}^{\lambda\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) = -1(y_{n,ij}=0)\sigma^{-3}\varphi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\sum_{g=1}^n w_{n,ig}y_{n,gj}\right)$,

$$
\begin{aligned}
t_{n,ij}^{\gamma\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= -1(y_{n,ij}=0)\sigma^{-3}\varphi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\textstyle\sum_{h=1}^n y_{n,ih}m_{n,hj}\right),\\
t_{n,ij}^{\rho\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= -1(y_{n,ij}=0)\sigma^{-3}\varphi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\left(\textstyle\sum_{g,h=1}^n w_{n,ig}y_{n,gh}m_{n,hj}\right),\\
t_{n,ij}^{\beta\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= -1(y_{n,ij}=0)\sigma^{-3}\varphi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\boldsymbol{z}_{n,ij}, \text{ and}\\
t_{n,ij}^{\sigma^2\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) &= -1(y_{n,ij}=0)\frac{1}{2\sigma^4}\begin{bmatrix} 2\psi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right) \\ +\varphi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\end{bmatrix} + 1(y_{n,ij}>0)\frac{1}{\sigma^4}.
\end{aligned}
$$

For $j=1,\dots,n$ and $i=1,\dots,n$,

$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{j,o}\partial\alpha_{i,d}} = t_{n,ij}^{\lambda\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\ \frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{j,o}\partial\alpha_{i,d}} = t_{n,ij}^{\gamma\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),$$
$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{j,o}\partial\alpha_{i,d}} = t_{n,ij}^{\rho\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\ \frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{j,o}\partial\alpha_{i,d}} = t_{n,ij}^{\beta\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}), \text{ and}$$
$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma^2\partial\alpha_{j,o}\partial\alpha_{i,d}} = t_{n,ij}^{\sigma^2\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}).$$

For $i = 1, \dots, n$, we have

$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\lambda\partial\alpha_{i,d}^2} = \sum_{j=1}^n t_{n,ij}^{\lambda\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\ \frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\gamma\partial\alpha_{i,d}^2} = \sum_{j=1}^n t_{n,ij}^{\gamma\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),$$
$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\rho\partial\alpha_{i,d}^2} = \sum_{j=1}^n t_{n,ij}^{\rho\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}),\ \frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\beta\partial\alpha_{i,d}^2} = \sum_{j=1}^n t_{n,ij}^{\beta\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}), \text{ and}$$
$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\sigma^2\partial\alpha_{i,d}^2} = \sum_{j=1}^n t_{n,ij}^{\sigma^2\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}).$$

For $\frac{\partial^3 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'\partial\alpha_{j,o}}$ and $\frac{\partial^3 \ln L_N^*(\omega_0,\boldsymbol{\alpha}_N^0)}{\partial\boldsymbol{\alpha}_N\partial\boldsymbol{\alpha}_N'\partial\alpha_{i,d}}$, the relevant terms (see the derivation of second order derivatives above) are

$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{j,o}^3} = \sum_{i=1}^n t_{n,ij}^{\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) \text{ for } j = 1, \dots, n;\ \frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{j,o}^2\partial\alpha_{k,o}} = 0 \text{ for } k \neq l;$$
$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{j,o}^2\partial\alpha_{i,d}} = t_{n,ij}^{\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) \text{ for } j = 1, \dots, n \text{ and } i = 1, \dots, n;$$
$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{i,d}^2\partial\alpha_{j,o}} = t_{n,ij}^{\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) \text{ for } j = 1, \dots, n \text{ and } i = 1, \dots, n;$$
$$\frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{i,d}^2\partial\alpha_{l,d}} = 0 \text{ for } i \neq l; \text{ and } \frac{\partial^3 \ln L_N^*(\omega,\boldsymbol{\alpha}_N)}{\partial\alpha_{i,d}^3} = \sum_{j=1}^n t_{n,ij}^{\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) \text{ for } i = 1, \dots, n,$$

where $t_{n,ij}^{\alpha_o}(\omega,\alpha_{j,o},\alpha_{i,d}) = -1(y_{n,ij} = 0)\sigma^{-3}\varphi\left(\epsilon_{n,ij}^{+,*}(\omega,\alpha_{j,o},\alpha_{i,d})\right)$.

References

Andrews, D.W., 1992. Generic uniform convergence. *Econometric Theory* 8, 241-257.

Arellano, M., and Hahn J. 2007. Understanding bias in nonlinear panel models: Some recent development. *Econometric Society Monographs* 43:381.

Amemiya, T. 1984. Tobit models: A survey. *Journal of Econometrics* 24, 3-61.

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

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csáki, F., (eds.), 2nd International Symposium on Information Theory. Budapest: Akadémia Kiadó, 267–281.

Baumann, L., 2021. A model of weighted network formation. *Theoretical Economics* 16, 1-23.

Burnham, K. P., and Anderson, D. R. 2002. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. 2nd ed. New York: Springer-Verlag.

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

Feenstra, R.C. 2003. Advanced International Trade: Theory and Evidence. Princeton University Press. Princeton.

Fernandez-Val, I., and M. Weidner 2016. Individual and time effects in nonlinear panel models with large N, T, *Journal of Econometrics* 192, 291-312.

Fernandez-Val, I., and M. Weidner 2018. Fixed Effects Estimation of Large-T Panel Data Models.

Annual Review of Economics 10, 109-138.
Fischer, M.M, and J.P. LeSage 2020. Network dependence in multi-indexed data on international trade flows. *Journal of Spatial Econometrics* 1, 1-26.
Hahn, J., and W. Newey 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72(4), 1295-1319.
Hahn, J., and G. Kuersteiner 2011. Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric theory* 27, 1152–1191.
Hausman, J.A., and W.E. Taylor 1981. Panel data and unobservable individual effects. *Econometrica* 49 (6), 1377-1398.
Jenish, N., and I. R. Prucha, 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics* 150, 86-98.
Jenish, N., and I. R. Prucha, 2012. On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170, 178-190.
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.
Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica* 72, 1899-1925.
Lee, L.F., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489-514.
Lee, L.F., and J. Yu 2010. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26, 564-597.
Lee, L.F., and J. Yu 2015. Spatial panels. Chapter 5 in The Oxford Handbook of Panel Data, edited by Badi H. Baltagi, Oxford University Press, Oxford.
Lee, L.F., and J.,Yu 2020. Estimating flow data models of international trade: Dual gravity and spatial interactions. Working paper.
LeSage, J.P., and R.K. Pace, 2008. Spatial econometric modeling of origin-destination flows. *Journal of Regional Science* 48, 941-967.
Liu, T., X. Xu, and L.F. Lee 2021. Consistency without compactness of the parameter space in spatial econometrics, Working paper.
McCallum, J. 1995. National borders matter: Canada-U.S. regional trade patterns. *American Economic Review* 85 (3), 615-623.
Neyman, J., and E. Scott 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16 (1), 1-32.
Olsen, R.J. 1978. Note on the uniqueness of the maximum likelihood estimator for the Tobit model. *Econometrica* 46 (5), 1211-1215.
Ord, J.K., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.
Pace, R.K, and J.P. LeSage, 2004. Chebyshev approximation of log-determinants of spatial weight matrices. *Computational Statistics and Data Analysis* 45, 179-196.
Rothenberg, T. J., 1971. Identification in parametric models. *Econometrica* 39 (3), 577-591.
Qu, X., and L.F. Lee, 2012. LM tests for spatial correlation in spatial models with limited dependent variables. *Regional Science and Urban Economics* 42, 430-445.
Qu, X., and L.F. Lee, 2015. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics* 184, 209-232.
Sasser, A.C., 2010. Voting with their feet: Relative economic conditions and state migration patterns.

Regional Science and Urban Economics 40, 122-135.
Thomas-Agnan, C., and J.P. LeSage, 2014. Spatial Econometric OD-Flow Models. In: Handbook of Regional Science Springer. Berlin Heidelberg pp. 1653-1673.
Wegge, L.L, 1974. Mean value theorem for convex functions. *Journal of Mathematical Economics* 1(2), 207-208.
Tobin, J., 1958. Estimation of relationships for limited dependent variables. *Econometrica* 26, 24-36.
Wooldridge, J.M. 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.
Xu, X. and L.F. Lee, 2015a. A spatial autoregressive model with a nonlinear transformation of the dependent variable. *Journal of Econometrics* 186, 1-18.
Xu, X., and L.F. Lee, 2015b. Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics* 188, 264-280.
Xu, X., and L.F. Lee, 2018. Sieve maximum likelihood estimation of the spatial autoregressive Tobit model. *Journal of Econometrics* 203, 96-112.
Xu, X., and L.F. Lee, 2019. Theoretical foundations for spatial econometric research. *Regional Science and Urban Economics* 76, 2-12.