Maximum likelihood estimation of a spatial autoregressive model for origin–destination flow variables

Hanbat Jeong a, *, Lung-fei Lee b

a Department of Economics, Macquarie Business School, Macquarie University, Australia
b Shanghai University of Finance and Economics, Shanghai, China

ABSTRACT

We introduce a spatial autoregressive hurdle model for nonnegative origin–destination flows $y_{N,ij}$. The model incorporates a hurdle formulation to elucidate the different data-generating processes for zero and positive flows. Our model specifies three types of spatial influences on flow $y_{N,ij}$ that quantify the impact of third-party characteristics on the flow $y_{N,ij}$: (i) the effect of outflows from origin $j$, (ii) the effect of inflows to destination $i$, and (iii) the effect of flows among third-party units. We account for two-way fixed effects in the model to capture the inherent characteristics of both origins and destinations. We employ maximum likelihood estimation to estimate the model parameters. To address statistical inference issues, we analyze the asymptotic properties of the ML estimator using the spatial near-epoch dependence concept.

We confirm the presence of an asymptotic bias that arises from the fixed effects, whose dimensions grow with the sample size. Applying our model to migration flows among U.S. states, we estimate significant spatial influences, particularly from inflows to destinations and outflows from origins. Our findings support the notion that zero and positive flow formations are distinct. Consequently, our proposed model outperforms the spatial autoregressive Tobit specification for origin–destination flows, thus providing a better fit to the data.

1. Introduction

This paper aims to develop models for nonnegative origin–destination flows within the spatial autoregressive (SAR) framework. We introduce a maximum likelihood (ML) estimation method for our model, which we refer to as the spatial autoregressive flow (SARF) model. In this specification, each flow $y_{N,ij}$ is a directed outcome originating from $j$ and destined for $i$, where $i,j = 1, \ldots, n$ and $N = n^2$ (with $n$ representing the number of cross-sectional units, such as cities or states). The migration among U.S. states serves as an illustrative example of such flows. Our primary objective is to adapt traditional SAR models to better accommodate the specific data environments often encountered for origin–destination flows. Previous studies on SARF models, such as those by LeSage and Pace (2008), Behrens et al. (2010), Fischer and LeSage (2020), Dargel (2021), and Jin et al. (2023), have aimed to extend linear SAR models to include flow variables and specify the influence of a third-party unit $k$ ($k \neq i, k \neq j$) on $y_{N,ij}$. Furthermore, these models

---

We thank Xiaohong Chen (Editor), the Associate Editor, and three anonymous referees for their great suggestions that substantially improved our paper. We want to thank Taehyun Ahn, Jason Blevis, Meta Brown, Chi-Young Choi, Yanli Lin, Gabi Jiang, Robert de Jong, Su Yong Song, Hankyoung Sung, and Bruce Weinberg for their valuable comments and suggestions. Hanbat thanks 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., and Advances in Econometrics (AiE) Conference and Festschrift in Honor of Joon Y. Park, Bloomington, U.S., for comments.

* Corresponding author.
E-mail addresses: hanbat.jeong@mq.edu.au (H. Jeong), lflee@mail.shufe.edu.cn (L.-f. Lee).
generally assume full observability of all relevant characteristics of the origins and destinations, thereby neglecting unobservable individual effects that may influence the flow.\footnote{Jeong et al. (2023) also explored the SARF model with two-way fixed effects, but their approach falls within the category of linear specifications.}

As a primary contribution, this paper develops a SARF model that accounts for censored flow variables, thus better accommodating the specific data environments often encountered in origin–destination flows. Typically, flow outcomes are gross flows; they are inherently nonnegative and can include zero values. These zero flows might occur due to budgetary constraints or significant geographic distances, making linear models unsuitable. A semicontinuous model with a point mass at zero handles data that has some proportion of zero observations while the rest of the observations are continuous. One approach to modeling zeros through censoring is by extending Tobit models, as described in Tobin’s seminal work (Tobin, 1958). Integrations of Tobit models into the spatial autoregressive (SAR) framework are detailed in works by Qu and Lee (2012), Thomas-Agnan and LeSage (2014), Xu and Lee (2015, 2018). However, Tobit models have limitations in representing the mechanisms underlying the generation of zero and positive flows because they consider both as arising from a single latent variable.

To overcome this limitation, we adopt a hurdle specification, as discussed in Cragg (1971), Deaton and Irish (1984), Lin and Schmidt (1984), Mullahy (1986), and Gurmu and Trivedi (1996).\footnote{Our extension of the SARF Tobit model is suggested by an anonymous referee. We appreciate his/her suggestion.} Our aim is to estimate spatial influences in origin–destination flows, which have a point mass at zero (due to censoring) with a bell-shaped distribution of positive flows. The resulting SARF hurdle model comprises two stages: (i) the participation of a unit \((i,j)\) for \(y_{N,ij}\) in the first stage and (ii) the determination of a latent amount for \(y_{N,ij}\) in the second stage. In our SARF hurdle model, a positive flow is generated only when both stages produce positive outcome values. This approach allows for the generation of zero and positive flows using different latent variables. Specifically, the probability of a zero flow is characterized by a bivariate function derived from these two stages, offering a better representation of zero flows than the SARF Tobit model. The second stage for latent flow amounts incorporates three channels of spatial interactions: (i) the effect of outflows from origin \(j\) (Channel 1), (ii) the effect of inflows to destination \(i\) (Channel 2), and (iii) the effect of flows among third-party units (Channel 3). These enhancements enable our model to effectively elucidate the mechanisms that distinguish zero from positive flows and their spatial influences.

Second, we introduce an estimation method for the SARF hurdle model that robustly controls for unobservables. Motivated by panel data models, we incorporate a fixed-effect specification into the model’s second-stage process (amounts of latent flows). To our knowledge, this paper is the first to adapt fixed-effect specifications for nonlinear SAR models, accommodating panel or origin–destination flows. This specification consists of two components: one capturing unobserved characteristics of the origin, and the other for the destination. These two-way fixed effects are delineated by \(2n\) individual parameters, with \(n\) parameters dedicated to origins and the remaining \(n\) to destinations. Thus, they capture the inherent attributes of both origins and destinations. We derive the log-likelihood function based on the normal distribution of disturbances. Furthermore, we propose a normality test grounded on a generalized normal distribution, which includes the normal distribution as a parametric subset. In the context of a linear model specification, we can utilize a concentrated log-likelihood for main parameters, benefiting from the linear parameter properties of fixed-effect components. However, the SARF hurdle model does not offer closed-form expressions for the fixed-effect estimators. This is because, in this model, fixed effects deviate from the usual linear parameters. As a result, employing a closed-form concentrated log-likelihood function that exclusively relies on the main parameters becomes infeasible.

We investigate the asymptotic properties of the ML estimator (MLE) for statistical inference, focusing on two main features. First, due to the inherent nonlinearity of the model, we employ the spatial near-epoch dependence (NED) concept as discussed in Jenish and Prucha (2012). In doing so, we construct a topological structure for asymptotic analyses. Within the geographic space, there are \(N = n^2\) pairs of flows. In our model, a spatial unit is not a single cross-section unit \(i\) but rather a pair \((i,j)\) positioned within a product space that includes separate categories for origins and destinations. As a result, \(\{y_{N,ij}\}\) constitutes a random field in this product domain. The NED properties of \(\{y_{N,ij}\}\) facilitate the verification of both the law of large numbers (LLN) and the central limit theorem (CLT).

Second, the conventional method, which relies on a closed-form concentrated log-likelihood function for deriving the MLE’s asymptotic distribution, is inapplicable due to the nonlinearity of the model’s fixed-effect parameters. To address this and derive the asymptotic distribution of the MLE, we utilize the approach of Fernandez-Val and Weidner (2016): specifically, the second-order Taylor expansion of the concentrated log-likelihood’s score, evaluated at the true finite-dimensional parameters. Compared to the linear models, we derive an additional expansion to the nonlinear fixed-effect parameter part. Given our framework, where there are \(2n\) incidental parameters with \(N = n^2\) observations, the incidental parameter problem becomes an asymptotic bias issue as outlined in Neyman and Scott (1948). In other words, our approach closely aligns with large panel data models since we have \(n\) origin units and \(n\) destination units. We confirm the existence and quantify the order \(O\left(\frac{1}{\sqrt{n}}\right)\) of the MLE’s asymptotic bias. This bias arises from the estimated fixed effects, which have components that converge at a slower rate of \(\sqrt{n}\), compared to the main parameter estimates that converge at a rate of \(n = \sqrt{N}\). To address this, we introduce an analytical bias correction method, deriving a formula that consistently estimates the asymptotic bias.

We conduct Monte Carlo simulations to assess the finite sample performance of the MLE and its bias-corrected version. Our analytical bias correction method effectively reduces the biases observed in the MLEs for the parameters in the second-stage process. The presence of fixed effects in the SARF hurdle model does not affect the finite sample performance of the MLE for the first-stage parameters. Unlike existing model specifications, such as LeSage and Pace (2008), our approach permits the utilization of two spatial weighting matrices—\(W_n\) and \(M_n\)—to depict a more intricate spatial dependence among flows. Specifically, \(W_n\) portrays...
destination-based dependence, and \( M_a \) determines origin-based dependence. By distinctly specifying the sources of spatial dependence, our model offers a better fit. This is an improvement over the traditional framework that relies on a single row-normalized spatial weighting matrix. Nevertheless, this introduces potential model selection concerns, especially when multiple prior specifications for \((W_a, M_a)\) exist. Encouragingly, our simulations suggest that the Akaike weight, based on the sample log-likelihoods of candidate models, reliably selects the true model. This is especially true when the data-generating process (DGP) is similar to our empirical application.

In this paper, we undertake an empirical analysis using migration flows across U.S. states for the year 2010. The distribution of the logged migration flows aligns with our model’s assumptions (a point mass at zero with a bell-shaped distribution of positive flows). We compare our model to the SARF Tobit model. Our findings reveal significant spatial influences through the first two channels of the model: (i) the effect of outflows from origins and (ii) the effect of inflows to destinations. The Akaike weights demonstrate a superior performance of two spatial weighting matrices grounded in historical migration flows over the conventional specification presented by LeSage and Pace (2008). When examining model fit, our SARF hurdle model consistently outperforms the SARF Tobit model. To elaborate, the SARF Tobit model produces fewer zero flows (0.62%), a sharp contrast to the actual data’s 8.70% and the 7.71% generated by our SARF hurdle model. All zero flows are attributed to the first stage process, highlighting the distinct nature of the two-stage processes involved in U.S. migrations. Additionally, the positive flow distribution produced by our model aligns more closely with actual data than the SARF Tobit model, which tends to exaggerate variance.

We also delve into the determinants of migration, revealing significant spatial influences, particularly through the channels related to geographic and economic distances. From a policy perspective, these insights can guide strategies to encourage migration, such as supporting long-distance migrants or addressing disparities in states’ unemployment rates and housing burden ratios.

Furthermore, based on the model’s DGP and estimated parameters, we provide a computational method for estimating marginal effects. Our results suggest that a 1% increase in geographic distance between states results in an average decline of 0.6876% in migration flows. Similarly, a 1% increase in the differential of housing burden ratios corresponds to an average decrease of 0.0297% in migration flows. Lastly, a 1% rise in the differential of insured unemployment rates is linked to an average drop of 0.0433% in migration flows.

The paper is organized as follows: Section 2 introduces a model specification and derives a log-likelihood function, setting the foundational methodology. Section 3 delves into the asymptotic properties of the MLE. In Section 4, we report on the Monte Carlo simulation results to evaluate the finite sample performance of the MLE and discuss the model selection issue of selecting spatial weighting matrices. Section 5 applies this framework to consider the migration flows across U.S. states, providing a practical application. The paper concludes in Section 6. Additional detailed derivations and asymptotic analyses are provided in the appendices for further technical depth.

2. Model

2.1. Model specification

A sample has \( n \) cross-section units (regions), so there are \( N = n^2 \) observations of origin–destination flows. Let \( Y_N = \{y_{N,ij}\} \) be an \( n \times n \) matrix of flows, and \( W_a = \{w_{a,ij}\} \) and \( M_a = \{m_{a,ij}\} \) be \( n \times n \) spatial weighting matrices with zero diagonals (i.e., \( w_{a,ii} = m_{a,ii} = 0 \)) characterizing relations among cross-section units. As a traditional spatial econometric model, we assume that every spatial weight embodies a matrix of flows, and \( M_a \) characterizes Channel 3 (origin-to-destination dependence) among regions. The decision to engage in an activity (i.e., whether \( \{y_{N,ij} = 0\} \) or \( \{y_{N,ij} > 0\} \)) and the activity levels (the value of \( y_{N,ij} \) when \( y_{N,ij} > 0 \)) might not be closely related due to hurdles such as search and transaction costs. This observation led to the development of the hurdle model, further elaborated

---

3 These naming conventions are derived from LeSage and Pace (2008). For a detailed breakdown: \( I_r \odot W_a \) represents Channel 1 (destination-based dependence), \( M' \odot I_r \) embodies Channel 2 (origin-based dependence), and \( M' \odot W_a \) characterizes Channel 3 (origin-to-destination dependence). Notably, \( M' \neq W_a \) distinguishes LeSage and Pace’s (2008) specification from ours.

4 An alternative SAR Tobit specification is the latent SAR Tobit model, details of which can be found in LeSage and Pace (2009, Ch. 10), Autant-Bernard and LeSage (2011), and Hoshino (2021). Unlike our focus, the latent SAR Tobit model concentrates on interactions among latent outcomes rather than actual outcomes. As Di Porto and Revelli (2013) note, the simultaneous SAR Tobit framework is better suited for depicting network/spatial interactions that present corner solution outcomes.
by Deaton and Irish (1984), and Lin and Schmidt (1984). The core concept of the hurdle model involves a two-part specification: the first part addresses the zero-value process, while the second part deals with positive values.

To elucidate the distinct mechanisms behind zero and positive flows within the spatial econometric framework, we introduce the SARF hurdle model. This model delineates the origin–destination flow, \( y_{N,j} \), through two distinct underlying processes:

First stage:

\[
y_{N,j}^p = x_{N,j}^p + \epsilon_{N,j},
\]

Second stage:

\[
y_{N,j}^o = x_{N,j}^o(\alpha_{ij}, \omega_{ij}) + \epsilon_{N,j},
\]

where \( x_{N,j} \) is a row vector of characteristics for the first stage process (e.g., \( x_{N,j}^p = [z_{N,j}, \zeta_{N,j}] \) with \( z_{N,j} = (z_{N,j,1}, \ldots, z_{N,j,L})' \)).

Absent the role of \( \epsilon_{N,j} \), the model defaults to the SARF Tobit model. Without censoring, our model would result in a linear SARF model.

Now, we examine the conditions required for the stability and coherency of the model. Given an \( N \times 1 \) vector \( \mathbf{x}_N = (x_1, \ldots, x_N)' \), we can define the operators \( F_e(x) = \{1[y_{N,j}^p > 0]x_1, \ldots, 1[y_{N,j}^o > 0]x_N\} \) and \( F_p(x) = (\max(0, x_1), \ldots, \max(0, x_N))' \) as presented. Using these defined operators, the model can be reformulated as

\[
y_N = F_e \left( F_p \left( \mathbf{A}_N \mathbf{y}_N + \mathbf{Z}_N \beta_0 + \alpha_{ij} \otimes \mathbf{l}_n + \mathbf{l}_n \otimes \eta_{ij} + \epsilon_N \right) \right),
\]
where $\eta = \text{vec}(Y_N)$, $Z_N = [Z_{N,1}, \ldots, Z_{N,L}]$ with $Z_{N,l} = \text{vec}(Z_{N,l})$ for $l = 1, \ldots, L$. $A_N = \lambda W_N + \gamma M_N + \rho R_N$, $\sigma_{\theta_0} = (a_{\theta_0}, \ldots, a_{\theta_0})'$, $n_{\theta_0} = (n_{\theta_0,0}, n_{\theta_0,1}, \ldots, n_{\theta_0,m})'$ and $\epsilon_N = (\epsilon_{N,1}, \ldots, \epsilon_{N,21}, \ldots, \epsilon_{N,m})'$. Let $y_N^{\epsilon}$ be an $N \times 1$ vector correspondingly as in (1) whose elements are inside of $F_\epsilon$ on the right-hand-side of (3). Then, the model can be written as $y_N = F_\epsilon(y_N^{\epsilon})$.

Given that the SARF hurdle model incorporates nonlinear transformations, $F_\epsilon(\cdot)$ and $F_\epsilon(\cdot)$, its stability condition could be different from that of the linear SARF model. A coherent model ensures a unique solution $y_N$ for the nonlinear system $y_N = F_\epsilon(F_\epsilon(y_N^{\epsilon}))$. The following assumption outlines a sufficient condition for the model's stability and coherency, paralleling Assumption 2 in Xu and Lee (2015):

**Assumption 2.1.** Let $\zeta = \sup_\theta \sup_\epsilon \|A_N(\epsilon)\|_{\infty}$, where $\theta_0$ is a parameter space of $\delta = (\lambda, \gamma, \rho)'$, and $A_N(\delta) = \lambda W_N + \gamma M_N + \rho R_N$. We assume $\zeta < 1$.

2.2. Log-likelihood function

Our goal is to estimate $\theta_0 = (k_{0}', \omega_0')'$ by the maximum likelihood (ML) estimation method. To derive a log-likelihood function, we rearrange the set of observed values $Y_N$ into two subsets: $y_N = \begin{pmatrix} y_{0,N_0} \backslash y_{1,N_1} \end{pmatrix}'$, where the first $N_0$ observations in $Y_{0,N_0}$ are zeros while the remaining $N_1 = N - N_0$ observations are positive. Under the normality of $\nu_{N,j}$ and $\epsilon_{N,j}$, the density of $y_N$ is

$$f(y_N) = \det(S_{N_1}) \prod_{i,j=1}^{n} \left( 1 - \phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) \cdot \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) \right) ^{1(\nu_{N,j}=0, \epsilon_{N,j}=0)} \cdot \left( \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) \cdot \phi(\epsilon_{N,j}(\kappa_0)/\sigma_0) \right) ^{1(\nu_{N,j}=0, \epsilon_{N,j}=0)} ,$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ represent respectively the CDF and PDF of the standard normal distribution, $S_{N_1}(\delta)$ is the submatrix of $S_N(\delta)$ that excludes $N - N_0$ rows corresponding to all $y_{N,j} > 0$ and $S_{N_1}(\delta_0)$.

Note that $S_{N_1}(\delta) = I_{N_1} - \lambda W_{11,N} - \gamma M_{11,N} - \rho R_{11,N}$ where $W_{11,N}, M_{11,N}$, and $R_{11,N}$ are respectively the submatrices of $W_N, M_N$, and $R_N$ corresponding to positive flows. The event $y_{N,j} = 0$ can arise in two scenarios:

- There is no participation, which has a probability of $1 - \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0)$.
- With participation, the Tobit probability of zero is realized, expressed as $1 - \Phi(\kappa_0(\kappa_0)/\sigma_0)$.

Thus, the probability of $y_{N,j} = 0$ is given by $1 - \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) + \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) \cdot (1 - \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0)) = 1 - \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) \cdot \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0)$.

For each pair $(i, j)$, let $\Delta_{N,j}(\theta_{ij}) = 1 - \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) \cdot \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0)$ and $\Delta_{N,j} = \Delta_{N,j}(\theta_{ij})$, where $\theta_{ij} = (k', \omega_j')'$ and $\theta_{0,0} = (k_0', \omega_0')'$.

It is evident from the following inequality:

$$\Delta_{N,j} \geq 1 - \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) ,$$

that $0 \leq \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) \leq 1$. This relationship implies that when the first process does not play a role, $f(y_N)$ in the SARF hurdle model is identical to that in the SARF Tobit model. Hence, the mixed density and the probability of obtaining zero in the SARF hurdle model can lead to more zero flows than in the SARF Tobit model. Further, the probability of $y_{N,j} = 0$, $\Delta_{N,j}$, is characterized by a bivariate function $\Delta(x', x^\prime) = 1 - \Phi(x') \cdot \Phi(x^\prime)$, where $x' = x_{N,j}^{N}(\kappa_0)$ and $x^\prime = x_{N,j}^{N}(\kappa_0)/\sigma_0$.

The penalized log-likelihood function from $f(y_N)$ is given by

$$\ell_N(\theta, \xi_n) = \ln \det(S_{N_1}(\delta)) + \sum_{i,j=1}^{n} 1(y_{N,j} = 0, \Delta_{N,j}(\theta_{ij}))$$

$$+ \sum_{i,j=1}^{n} 1(y_{N,j} > 0) \left( \ln \Phi(x_{N,j}^{N}(\kappa_0)/\sigma_0) + \ln \phi(\epsilon_{N,j}(\kappa_0)/\sigma_0) \right) \left( \frac{1}{2} \sum_{j=1}^{n} a_j - \sum_{j=1}^{n} b_j \right)^2 ,$$

where $\theta = (k', \omega_j')'$, $\xi_n = (\alpha_1', \eta_n')'$ with $\alpha_1 = (a_1, \ldots, a_1)'$ and $\eta_n = (n_1, \ldots, n_1)'$, and $\mu$ is an arbitrary positive constant. The penalty term $-\frac{\mu}{2} \left( \sum_{j=1}^{n} a_j - \sum_{j=1}^{n} b_j \right)^2$ is introduced to impose $\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j$ for identification purposes. Without this penalty term, the second-order derivative of $\ell_N(\theta, \xi_n)$ with respect to $\xi_n$ would not possess full rank. This approach aligns with the settings of nonlinear panel models with individual and time-fixed effects. It is noteworthy that, unlike in the linear SARF model, the fixed-effect components $\{n_j\}$ and $\{a_j\}$ are no longer linear parameters.

The MLE is obtained by maximizing the penalized log-likelihood function (4):

$$(\hat{\theta}_N, \hat{\xi}_{N,n}) = \arg\max_{\theta \in \theta, \xi_n} \ell_N(\theta, \xi_n).$$

---

7 The derivation of this expression can be found in Appendix A.
The asymptotic properties of the MLE $\hat{\theta}_N$ rely on stochastic properties of $\ln \det (S_N(\delta))$ and its derivatives with respect to $\delta$. To delve deeper into these properties, we define $G_N(y_N) = \text{diag}_{i,j=1}^n 1 (y_{N,i,j} > 0)$, $W_N = G_N(y_N)W_N G_N(y_N)$, $\tilde{M}_N = G_N(y_N)\tilde{M}_N G_N(y_N)$, and $\tilde{R}_N = G_N(y_N)\tilde{R}_N G_N(y_N)$. Then, we have the following representations:

$$\ln \det (S_N(\delta)) = -\sum_{l=1}^{\infty} \left( \text{tr} \left( (G_N(y_N)A_N(\delta)G_N(y_N))^l \right) \right).$$

$$\text{tr} \left( W_{N,i,j} S_{N,i,j}^{-1}(\delta) \right) = \text{tr} \left( W_N S_N^{-1}(\delta) \right) = \sum_{i,j=1}^{n} r_{N,i,j}(\delta).$$

$$\text{tr} \left( M_{N,i,j} S_{N,i,j}^{-1}(\delta) \right) = \text{tr} \left( M_N S_N^{-1}(\delta) \right) = \sum_{i,j=1}^{n} r_{N,i,j}(\delta),$$

$$\text{tr} \left( R_{N,i,j} S_{N,i,j}^{-1}(\delta) \right) = \text{tr} \left( R_N S_N^{-1}(\delta) \right) = \sum_{i,j=1}^{n} r_{N,i,j}(\delta),$$

where $r_{N,i,j}(\delta) = (W_N S_N^{-1}(\delta))_{i,j}$ with $S_N(\delta) = I_N - xW_N - \gamma M_N - \rho R_N$ for each $\delta$. We can similarly define $r_{N,i,j}(\delta)$ and $r_{N,i,j}(\delta)$. At $\delta_0$, let $r_{N,i,j} = r_{N,i,j}(\delta_0)$, $r_{N,i,j} = r_{N,i,j}(\delta_0)$, and $r_{N,i,j} = r_{N,i,j}(\delta_0)$. Under Assumption 2.1, note that $S_N(\delta)$ is invertible since $S_N(\delta)$ is diagonally dominant matrix.

**Remark 2.1.** For a more general specification, instead of $\Phi(\cdot)$ and $\phi(\cdot)$, we can assume the CDF and PDF of $\nu_{N,i,j}$ and/or $\epsilon_{N,i,j}/\sigma_0$ to be $G(\cdot)$ and $g(\cdot)$, respectively. One straightforward method to specify $g(\cdot)$ is by considering a parametric specification that nests a normal distribution (refer to Nadarajah (2005)). The generalized normal distribution has the density function given by $g(x) = \frac{1}{\sqrt{2\pi\gamma r_1}} e^{-\left(\frac{(x-\delta)^2}{2\gamma^2}\right)}$, where $x \in (-\infty, \infty)$, the mean is zero, and the parameters $r_1$ and $r_2$ lie in the range $(0, \infty)$. For $r_2 = 2$, $g(x)$ becomes a normal density with zero mean and variance $\sigma^2$, with $\sigma^2 = \frac{c^2}{2}$. This can be validated as the gamma function at $\frac{1}{2}$ is $\sqrt{\pi}$, i.e., $\Gamma(1/2) = \sqrt{\pi}$. The data structure intended for our model combines a point mass at zero with a bell-shaped distribution for positive flows. Testing $r_2 = 2$ for positive flows thus serves as a method to verify normal disturbances within the generalized normal distribution framework.

### 3. Asymptotic analysis

#### 3.1. Topological specification and regularity conditions

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

**Assumption 3.1.** Each cross-sectional unit $i \in \{1, \ldots, n\}$ is located in a space $D_n \subset D \subset \mathbb{R}^{d}$ ($d \geq 1$). We assume that $\lim_{n \to \infty} \#(D_n) = \infty$, where $\#(D_n)$ represents the cardinality of $D_n$. Let $d(i,j)$ denote the distance between $i$ and $j$. We further assume $\min_{i,j \neq i} d(i,j) \geq 1$.

This topological specification, introduced by Jenish and Prucha (2009, 2012), is used to establish the stochastic properties of spatial mixing and spatial near-epoch dependent (NED) processes. The set $D \subset \mathbb{R}^{d}$, with $d \geq 1$, represents an irregular lattice that contains all potential locations of the cross-sectional units $\{i\}$. We define a mapping $i \mapsto l(i) \in D$ for any $i$, and $d(i,j) = \|l(i) - l(j)\|_\infty$. The minimum distance assumption ($d(i,j) \geq 1$ for $i \neq j$) corresponds to the situation of the increasing domain asymptotic, which is natural for regional studies. The increasing domain asymptotic allows the LLN and CLT to be applicable. Based on Assumption 3.1, we introduce a distance measure for pairs $\{i,j\}$ involved in the flow outcomes $\{y_{N,i,j}\}$, allowing us to locate them in the product space $D \times D \subset \mathbb{R}^{2d}$.

We now define a distance function for pairs $(i,j)$ and $(g,h)$ as follows:

$$d_F ((i,j),(g,h)) = \max \left\{ \frac{d(i,g)}{\text{Distance between destinations}}, \frac{d(j,h)}{\text{Distance between origins}} \right\}.$$  

Due to the maximum norm property and Assumption 3.1, we have $d_F ((i,j),(g,h)) \geq 1$ when $(i,j) \neq (g,h)$. By Jenish and Prucha’s (2009) Lemma A.1, it follows that $\# \{(g,h) : d_F ((i,j),(g,h)) \leq m\} \leq Cm^{2d}$ for some constant $C > 0$. This metric ensures that $\text{Cov} (y_{N,i,j}, y_{N,g,h}) \to 0$ as $d_F ((i,j),(g,h)) \to \infty$.

Here are additional regularity assumptions for asymptotic analyses.

---

8 Proposition B.3 in the supplement provides the key properties of the terms above.

9 We shall leave nonparametric approaches in this paper because of their complexity in theory and estimation (particularly under the presence of fixed effects).
Assumption 3.2. (i) Denote \(c_{w,r} = \sup_n \|W_n\|_1\), \(c_{w,c} = \sup_n \|M_n\|_1\), \(c_{w,r} = \sup_n \|W_n\|_\infty\), and \(c_{w,c} = \sup_n \|M_n\|_\infty\). The sequences \{\(W_n\)\} and \{\(M_n\)\} satisfy \(\max\{c_{w,c}, c_{w,r}, c_{w,c}, c_{w,r}\} < \infty\), i.e., they are uniformly bounded in both row and column sum norms.

(ii) \(w_{n,j}\) and \(m_{n,j}\) satisfy one of the two conditions:

(ii-1) \(w_{n,j} > 0\) and \(m_{n,j} > 0\) only if \(d(i,j) \leq d\) for some \(d > 1\); otherwise \(w_{n,j} = 0\) and \(m_{n,j} = 0\).

(ii-2) \(0 \leq w_{n,j} \leq \frac{C_0}{d(i,j)}\) and \(0 \leq m_{n,j} \leq \frac{C_0}{d(i,j)}\) for some \(C_0 > 0\) and \(d > 2d\). In this case, we assume \(\lambda_0|c_{w,c} + \gamma_0|c_{w,r} + \rho_0|c_{w,c}c_{w,r} - \xi\) where \(\xi\) is the upper bound of \(\|A_N(0)\|_\infty\) as defined in Assumption 2.1; if \(c_{w,c} < c_{w,r}\), there exist at most \(K_W \geq 1\) of columns of \(W_n\) whose column sums exceed \(c_{w,c}\), where \(K_W\) is a constant that does not rely on \(n\).

Assumption 3.3. For \{\(X_{n,j}\), \(\tilde{X}_{n,j}\), \(\eta_j\)\}, \(\eta_j\) are uniformly bounded constants.

For \{\(X_{n,j}^\prime\) and \(\tilde{X}_{n,j}^\prime\)\}, we can also assume \(\sup_{j,n} \left\{ \max_k \|X_{n,j,k}\|_{L_k}, \max_j \|X_{n,j,j}\|_{L_k} \right\} < \infty\) for some \(\eta > 0\), where \(\|x\|_{L_p}\) denotes the \(L_p\)-norm of a random variable \(x\).

Assumption 3.4. \((v_{n,j}, e_{n,j})' \sim i.i.d. N(0, \begin{bmatrix} 1 & 0 \\ 0 & \sigma_0^2 \end{bmatrix})\) across pairs \((i,j)\).

Conditions in Assumption 3.2 are similar to those in Xu and Lee (2015). Assumption 3.2 (ii-2) characterizes the maximum column sum (and the sums of its powers) of \(A_N\), given that \(w_{n,j}\) and \(m_{n,j}\) decrease geometrically as a function of \(d(i,j)\). Under this condition, we deduce that \(\|A_N^\prime\|_1 \leq \tau^0 \cdot A^\prime \cdot \xi^{-1}\) for any \(l \in Z_+\), where \(A\) is a positive integer independent of \(n\) and \(\Gamma\) represents the upper bound of \(\|A_N^\prime\|\).

This leads to the conclusion that \(\sum_{j=1}^m \|A_N^\prime\|_1 \leq \tau^0 \cdot \sum_{j=1}^m \xi^{-1} < \infty\). We introduced Assumption 3.3 for the sake of analytical simplicity. The normality outlined in Assumption 3.4 ensures the uniform boundedness of the densities of \((v_{n,j}, e_{n,j})\), implying the NED properties of \((X_{n,j})\) as indicated in Proposition B.2.

Given that the estimator \(\hat{\theta}_N\) is a highly nonlinear function of \((v_{n,j}, e_{n,j})\), we use the spatial near epoch dependence (NED) concept introduced by Jenish and Prucha (2012) to verify both the LLN and CLT. The NED concept relates two random fields, \(Q = \{q_{n,j} : (i,j) \in D_n \times D_n, n \geq 1\}\) and \(E = \{v_{n,j}, e_{n,j} : (i,j) \in D_n \times D_n, n \geq 1\}\). A random field \(Q\) is \(L_p^2\)-NED on \(E\) if \(\sup_{n,j} \|q_{n,j}\|_{L_p^2} < \infty\) and \(\|q_{n,j}\|_{L_p^2} = \{q_{n,j}(s)\}_{L_p^2} \leq e_{n,j} \cdot \nu(s)\), where \(p > 1\), \(\sum_{j=1}^m \|q_{n,j}\|_{L_p^2} = \sigma (v_{n,j}, e_{n,j} : d^p((i,j), (g,h)) \leq s)\). \(v_{n,j}\) is an array of finite positive constants (NED scaling factor), and \(\nu(s)\) is a sequence such that \(\nu(s) \to 0\) as \(s \to \infty\) (NED coefficient). Note that \(Q\) is a uniform NED random field if \(\sup_{n,j} \|q_{n,j}\|_{D_n^2} < \infty\); and \(Q\) is a geometric random field if \(\nu(s) = O(r^s)\) for some \(0 < r < 1\).

3.2. Asymptotic distribution of the MLE

To examine the asymptotic properties of \(\hat{\theta}_N\), we define

\[
\tilde{x}_{n,N}(\theta) = (\tilde{a}_{n,N}(\theta)', \tilde{a}_{n,N}(\theta)', ... \tilde{a}_{n,N}(\theta)')' = \arg\max_{x \in \Theta} \frac{e_{n}(\theta, \tilde{x}_{n,N}(\theta))}{x_{n}(\theta)}
\]

where \(\Theta\) denotes the parameter space for \(\theta\). For each element in \(\tilde{x}_{n,N}(\theta)\), we let \(\tilde{a}_{n,N}(\theta) = (\tilde{a}_{N,1}(\theta), ..., \tilde{a}_{N,\Theta}(\theta))'\) and \(\tilde{\eta}_{n,N}(\theta) = (\tilde{\eta}_{N,1}(\theta), ..., \tilde{\eta}_{N,\Theta}(\theta))'\) for each \(\theta \in \Theta\). Then, we have \(\tilde{\theta}_N = \arg\max_{\Theta} e_{N}(\theta, \tilde{x}_{n,N}(\theta))\), where \(e_{N}(\theta, \tilde{x}_{n,N}(\theta))\) denotes the concentrated penalized log-likelihood function. However, due to the nonlinearity of the model, deriving a closed-form expression for \(\tilde{x}_{n,N}(\theta)\) is unattainable and leads to further analytical challenges. For notation consistency with Fernandez-Val and Weindl (2016), we represent derivatives of a function (e.g., \(e_{N}(\cdot)\)) using the derivative operators: for instance, \(\partial_\theta e_{N}(\theta, \tilde{x}_{n,N}(\theta)) = \frac{\partial e_{N}(\theta, \tilde{x}_{n,N}(\theta))}{\partial \theta}\) and \(\partial_{\tilde{x}_{n,N}} e_{N}(\theta, \tilde{x}_{n,N}(\theta)) = \frac{\partial e_{N}(\theta, \tilde{x}_{n,N}(\theta))}{\partial \tilde{x}_{n,N}}\) for the \(i\)th element of \(\theta\). At the true parameter values, we exclude the function's arguments, e.g., \(\partial_\theta e_{N}(\theta, \tilde{x}_{n,N}(\theta)) = \frac{\partial e_{N}(\theta)}{\partial \theta}\).

We aim to demonstrate two primary results:

(i) \(\tilde{\theta}_N \overset{p}{\to} \theta_0\), and

(ii) \(\sqrt{N} (\tilde{\theta}_N - \theta_0) - \Sigma_N^{-1} \cdot A_N \overset{d}{\to} N(0, \Sigma^{-1})\) as \(n \to \infty\).

Here, \(\Sigma_N\) is defined as:

\[
\Sigma_N = \mathbb{E} \left( -\frac{1}{N} \partial_{\tilde{x}_{n,N}} e_{N} \right) - \frac{1}{\sqrt{N}} \left\{ \mathbb{E} \left( \frac{1}{\sqrt{N}} \partial_{\tilde{x}_{n,N}} e_{N} \right)^{-1} \mathbb{E} \left( \frac{1}{\sqrt{N}} \partial_{\tilde{x}_{n,N}} e_{N} \right)^{-1} \right\}.
\]

10 To control \(\|A_N^\prime\|_1\) for each \(l \in Z_+\), the column sum restriction is imposed on \(W_n\) as specified in Assumption 3.2 (ii-2). This is because an upper bound of \(\|A_N^\prime\|_1\) can be expressed by \(\|W_n^\prime\|_1\) and \(\|M_n^\prime\|_\infty\), where \(p + q + r = t\). We enforce the column sum restriction on \(W_n\) to establish an upper bound for \(\|W_n^\prime\|_1\) (as seen in Lemma B.1).

11 If \(\tilde{X}_{n,j}\) and \(\tilde{X}_{n,j}\) are stochastic, we can define \(E = \{X_{n,j}, \tilde{X}_{n,j}, v_{n,j}, e_{n,j} : (i,j) \in D_n \times D_n, n \geq 1\}\).
The term $A_N$ represents the asymptotic bias resulting from the inclusion of fixed effects, while $\Sigma = \lim_{n \to \infty} \Sigma_N$ is the limiting variance of $\frac{1}{\sqrt{N}} \partial_0 \epsilon_N + E \left( \frac{1}{\sqrt{N}} \partial_0 \epsilon_N \right) \Xi - \frac{1}{\sqrt{N}} \partial_0 \epsilon_N$. \(^{12}\) We will later provide the derivation for $A_N$.

This section provides the basic ideas of proving the two objects. Proposition C.2(i) yields the Taylor expansion of the first-order condition around $\theta_0$ as follows:

$$
\theta = \frac{1}{\sqrt{N}} \partial_0 \epsilon_N (\hat{\theta}_N, \hat{\xi}_{N,0}) = \frac{1}{\sqrt{N}} \partial_0 \epsilon_N (\theta_0, \hat{\xi}_{N,0}(\theta_0)) - \Sigma_N \cdot \sqrt{N} \left( \hat{\theta}_N - \theta_0 \right) + o_p(1). 
$$

(5)

Our main focus is to examine $\frac{1}{\sqrt{N}} \partial_0 \epsilon_N (\theta_0, \hat{\xi}_{N,0}(\theta_0))$ in order to investigate the asymptotic distribution of $\sqrt{N} \left( \hat{\theta}_N - \theta_0 \right)$. Using the second-order Taylor expansion of $\frac{1}{\sqrt{N}} \partial_0 \epsilon_N (\theta_0, \hat{\xi}_{N,0}(\theta_0))$ around $\xi_{0,0}$, Proposition C.2(ii) yields

$$
\frac{1}{\sqrt{N}} \partial_0 \epsilon_N (\theta_0, \hat{\xi}_{N,0}(\theta_0)) = \frac{1}{\sqrt{N}} \partial_0 \epsilon_N + \frac{1}{\sqrt{N}} \sum_{j=1}^n \partial_{\theta_j} \epsilon_N \left( \hat{\theta}_j(N)(\theta_0) - a_0 \right) + \frac{1}{\sqrt{N}} \sum_{i=1}^n \partial_{\theta_0} \epsilon_N \left( \hat{\xi}_i(N)(\theta_0) - \eta_0 \right)
$$

$$
+ \frac{1}{\sqrt{N}} \sum_{j=1}^n \partial_{\theta_j} \epsilon_N \left( \hat{a}_{j,N}(\theta_0) - a_0 \right) \left( \hat{a}_{j,K}(\theta_0) - a_K \right)
$$

$$
+ \frac{1}{\sqrt{N}} \sum_{j=1}^n \partial_{\theta_j} \epsilon_N \left( \hat{\theta}_j(N)(\theta_0) - \eta_j \right) \left( \hat{\theta}_j(N)(\theta_0) - \eta_j \right)
$$

$$
+ \frac{1}{\sqrt{N}} \sum_{j=1}^n \partial_{\theta_0} \epsilon_N \left( \hat{\xi}_i(N)(\theta_0) - \eta_0 \right) \left( \hat{\xi}_i(N)(\theta_0) - \eta_0 \right) + o_p(1). 
$$

(6)

In (6), note that $\frac{1}{\sqrt{N}} \partial_0 \epsilon_N$ characterizes the main part of the asymptotic variance $\Sigma$ and has an expected value of zero. As indicated by the 2nd through 7th terms in the right-hand-side expansion above, $\frac{1}{\sqrt{N}} \partial_0 \epsilon_N (\theta_0, \hat{\xi}_{N,0}(\theta_0))$ is not centered at zero even as $n \to \infty$. These components generate asymptotic bias terms. The first source of this bias comes from using $\hat{\xi}_{N,0}(\theta_0)$ instead of $\xi_{0,0}$, whose components have a convergence rate of $\sqrt{8}$, which is slower than $\sqrt{N} = n$, i.e., the convergence rate of $\hat{\theta}_N$. \(^{13}\) The second source of bias originates from the correlation between $\hat{\theta}_j(N)(\theta_0)$ and $\hat{\theta}_j(N)(\theta_0)$ and the second and third-order derivatives of the log-likelihood in (6), which are related to the inference of $\hat{\theta}_N$. Lastly, the variances of $\hat{a}_{j,N}(\theta_0)$ and $\hat{\theta}_j(N)(\theta_0)$ affect the asymptotic bias term. \(^{14}\)

To approximate the asymptotic bias terms, we define some components relevant to the log-likelihood function:

$$
\epsilon_{N,jj}(\theta_j) = 1 \left( y_{N,j} = 0 \right) \ln \Delta_{N,jj}(\theta_j) - 1 \left( y_{N,j} > 0 \right) \left( \ln \Phi (x_{N,j} \phi + \ln \phi) \right) - \frac{\mu_j}{N} \left( \sum_{i=1}^n a_j - \sum_{i=1}^n \eta_i \right), \quad
$$

(5)

$$
H_N = E \left( \frac{1}{N} \partial_0 \epsilon_N \right) \Xi,
$$

$$
\epsilon_{N,jj} = \partial_{\theta_j} \epsilon_{N,jj}, \quad \eta_{N,jj} = \partial_{\eta_j} \epsilon_{N,jj}, \quad\quad \rho_{N,jj} = \partial_{\rho_j} \epsilon_{N,jj}, \quad \gamma_{N,jj} = \partial_{\gamma_j} \epsilon_{N,jj},
$$

$$
\theta_{N,jj} = \partial_{\theta_{jj}} \epsilon_{N,jj}, \quad \phi_{N,jj} = \partial_{\phi_j} \epsilon_{N,jj}, \quad \epsilon_{N,j} = \partial_{\epsilon_j} \epsilon_{N,j}, \quad \eta_{N,j} = \partial_{\eta_j} \epsilon_{N,j},
$$

$$
\rho_{N,j} = \partial_{\rho_j} \epsilon_{N,j}, \quad \gamma_{N,j} = \partial_{\gamma_j} \epsilon_{N,j}, \quad\quad a_{N,j} = \partial_{a_j} \epsilon_{N,j}, \quad b_{N,j} = \partial_{b_j} \epsilon_{N,j}
$$

$$
\big( 5 \big)
$$

$$
\bullet \quad q_{N,j} = \partial_{q_j} \epsilon_{N,j}, \quad q_{N,j} = \partial_{q_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j},
$$

$$
\big( 6 \big)
$$

$$
\bullet \quad q_{N,j} = \partial_{q_j} \epsilon_{N,j}, \quad q_{N,j} = \partial_{q_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j},
$$

$$
\big( 7 \big)
$$

$$
\bullet \quad q_{N,j} = \partial_{q_j} \epsilon_{N,j}, \quad q_{N,j} = \partial_{q_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j},
$$

$$
\big( 7 \big)
$$

$$
\bullet \quad q_{N,j} = \partial_{q_j} \epsilon_{N,j}, \quad q_{N,j} = \partial_{q_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j}, \quad h_{N,j} = \partial_{h_j} \epsilon_{N,j},
$$

$$
\big( 7 \big)
$$

\(^{12}\) Note that $\Sigma_N = O(1)$ since $\| \frac{1}{\sqrt{N}} \partial_0 \epsilon_N \| = O_p(1)$, and $\| \frac{1}{\sqrt{N}} \partial_0 \epsilon_N \| = O_p(1)$, and $\| \frac{1}{\sqrt{N}} \partial_0 \epsilon_N \| = O_p(1)$, where $\| \cdot \|$ denotes the Euclidean norm. Details can be found in Lemma C.1 of the supplement file.

\(^{13}\) We have $\sqrt{N} (\hat{a}_{j,N}(\theta_0) - a_0) = O_p(1)$ for $j = 1, \ldots, n$ and $\sqrt{N} (\hat{\theta}_j(N)(\theta_0) - \eta_0) = O_p(1)$ for $j = 1, \ldots, n$ according to Lemma C.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.

\(^{14}\) Those three points are raised by Hahn and Newey (2004).

\(^{15}\) That is, $\epsilon_{N,0}(\theta, \xi) = \sum_{j=1}^n \epsilon_{N,jj}(\theta_j) + \ln \det(N_{0,j} \psi_j \delta_j)$. Each component, $\epsilon_{N,jj}(\theta_j)$, is a function of $\theta_j = (\theta', \phi', \phi, \gamma, \eta, \eta)'$. Notably, $\epsilon_{N,jj}(\theta_j)$ incorporates only the main parameters $\theta$ and two fixed-effect components, $a_j$ and $\eta_j$, rather than the entire set of fixed-effect parameters $\xi_j$. 

8
This additive separation is a consequence of the additive separability of \( \alpha \) and invertibility of \( \nu \) studied in Section 2.2 of the supplementary material. The fourth and seventh terms are respectively approximated by \( \nu_{ij}^{\alpha} \) whose inverse is a component of the Taylor approximation of \( \nu_{ij}^{\alpha} \) provided that \( W_n, M_n \) adhere to Assumption 3.5 (ii-2).

**Theorem 3.1.** Suppose Assumptions 2.1, 3.1-3.5 hold. In addition, we assume

(i) \( E \left( -\frac{1}{\sqrt{N}} \partial_{\theta} \psi_N \right) \) is positive definite under large \( n \). Furthermore, \( \Sigma = \lim_{n \to \infty} \Sigma_N \) is positive definite.

(ii) The parameter space \( \Theta \) of \( \theta \) is compact, and \( \theta_0 \in \text{int}(\Theta) \). Assume each element of \( \xi_n \) is a bounded constant. Let \( \theta_n^* = (k, \delta', \beta', \tau, \sigma') \) be a vector of the transformed parameters. Additionally, assume that possible \( \theta \in \Theta \) and \( \xi_n \) satisfy \( \partial \theta_n^* \psi_n^*(\theta_n^*) < 0 \), where \( \psi_n^*(\theta_n^*) \) denotes the log-likelihood evaluated at \( \theta_n^* \).

Then, we have (i) \( \hat{\theta}_N \xrightarrow{d} \theta_0 \), and (ii) \( \sqrt{N} (\hat{\theta}_N - \theta_0) \xrightarrow{d} N (\Sigma^{-1} A_{\infty}, \Sigma^{-1}) \) as \( n \to \infty \), where \( A_{\infty} = \lim_{n \to \infty} A_N \) with \( A_N = A_{1,N} + A_{2,N} + A_3,N + A_{4,N} + A_{5,N} + A_{6,N} \).

Proving Theorem 3.1 relies on the asymptotic expansion of \( \frac{1}{\sqrt{N}} \partial_{\theta} \psi_N (\theta_N, \xi_n) \) by (5) and (6). Condition (i) guarantees the invertibility of \( E \left( -\frac{1}{\sqrt{N}} \partial_{\theta} \psi_N \right) \), whose inverse is a component of the Taylor approximation of \( \frac{1}{\sqrt{N}} \partial_{\theta} \psi_N (\theta_0, \xi_n) \). Condition (i) also ensures that the asymptotic variance of \( \hat{\theta}_N \) is well-defined. Condition (ii) restricts the parameter space and is based on the reparameterization by Olsen (1978), which is essential for achieving the concavity of the log-likelihood function. However, additional restrictions on the parameter space are needed for our model due to its complexity compared to traditional Tobit models. The first feature arises from the specification of the zero probability using the bivariate function \( \Delta(x', x') \) instead of the Tobit probability of zero. This results in the second-order derivatives of \( \ln \Delta(x', x') \) having indefinite signs, while the derivative of the Inverse Mills Ratio for Tobit models is always negative.16 Intuitively, in situations without any zero flow, parameters responsible for generating zero flows might not be identifiable. Hence, a parameter set generating both large values of \( \Pr(y_{ij}^{\alpha} \geq 0) = \Phi(y_{ij}^{\alpha} k) \) and \( \Pr(y_{ij}^{\alpha} \geq 0) \neq 0 \) should be avoided. Another feature is the inclusion of \( \ln \det(S_{N,\xi_n}(\delta)) \) in \( \epsilon_n(\theta, \xi_n) \), an additional component not present in conventional Tobit models.17 These challenges necessitate the inclusion of Condition (ii) to ensure the consistency of \( \hat{\xi}_n \) and the structure of (6). Consequently, these conditions together imply that \( \sqrt{N} (\hat{\theta}_N - \theta_0) - \Sigma^{-1} A_{\infty} \xrightarrow{d} N (0, \Sigma^{-1}) \) as \( n \to \infty \).

By Theorem 3.1, we can define a bias-corrected MLE as follows:

\[
\hat{\theta}_N^* = \hat{\theta}_N - \frac{1}{\sqrt{N}} \hat{\xi}_N^{-1} \cdot A_N.
\]

16 The second-order derivatives of \( \ln \Delta(x', x') \) in the log-likelihood function contain \( \Delta(x', x') \) as their denominator terms. When \( x' \) and \( x'' \) are both large, \( \Delta(x', x') \) becomes close to zero. Then, the second-order derivatives of \( \ln \Delta(x', x') \) can be large positive values. Detailed properties of \( \Delta(x', x') \) are numerically studied in Section 2.2 of the supplementary material.

17 Under the reparameterization, the Jacobian term relies on \( \sigma^2 \) as well as \( \delta \). If all eigenvalues of \( W_n \) and \( M_n \) are real-valued, we do not need to consider the second feature (see Lemma 2 of Liu et al. (2022)).
where $\hat{\Sigma}^{-1}_N$ denotes the asymptotic variance matrix evaluated at $(\hat{\theta}_N, \hat{\xi}_{n,N})$ and $\hat{A}_N$ is a consistent estimator of $A_{\infty}$ obtained using $(\hat{\theta}_N, \hat{\xi}_{n,N})$.

Using (7), we examine the structure of each component in $A_N$ to consider $\hat{A}_N$:

$$A_N^A = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{i,k,l = 1} A_{N,i,j,k}^A B_{N,j,l}^A,$$

and

$$A_N^B = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{i,j,k,l = 1} A_{N,i,j}^B B_{N,j,l}^B,$$

where $A_N^A$ denotes the structure of $A_{1,N}, A_{3,N},$ and $A_{5,N}$, while $A_N^B$ represents the form of $A_{2,N}, A_{4,N},$ and $A_{6,N}$. Here,

- $d_{N,i,j}^A, d_{N,i,j}^B$ are non-stochastic bounded weights, e.g., $d_{N,i,j}^A = a_{N,i,j}$ for $A_{1,N}$ and $d_{N,i,j}^B = c_{N,i,j}$ for $A_{2,N}$.
- $A_{N,i,j}^A A_{N,j}^A B_{N,j}^A$ and $B_{N,j}^B$ are random components, e.g., $A_{N,i,j}^A = h_{N,i,j}^A B_{N,j}^A = q_{N,i,j}^A$ for $A_{1,N}$ and $A_{N,i,j}^B = h_{N,i,j}^B B_{N,j}^B = q_{N,i,j}^B$ for $A_{2,N}$.

To obtain $\hat{A}_N$, we employ a truncation technique, which is akin to methods used in the time series literature for obtaining truncated sums of sample covariances since $A_N^A$ and $A_N^B$ consist of the sum of many cross moments.\(^{18}\) To avoid too many imprecisely estimated cross moments, consistent estimators for $A_{\infty}^A = \lim_{n \to \infty} A_N^A$ and $A_{\infty}^B = \lim_{n \to \infty} A_N^B$ with truncation can be derived as:

$$\hat{A}_N^A = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1 \in \text{nbd}(i,s_n)} A_{N,i,j,k}^A \hat{B}_{N,k}^A,$$

and

$$\hat{A}_N^B = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1 \in \text{nbd}(i,s_n)} A_{N,i,j}^B \hat{B}_{N,j}^B,$$

where nbd$(i, s_n)$ indicates a set of $i$'s $s_n$-th order neighboring units induced by an $n$-dimensional square matrix $D_n$, based on additional geographic information. For example, if $D_n = 1$ for $i$ and $j$ are connected, then $k \in \text{nbd}(i, s_n)$ when $D_n_k \neq 0$ for $l \in \{0, 1, \ldots, s_n\}$. The evaluations of $A_{N,i,j}^A, A_{N,i,j}^B, B_{N,j}^A, B_{N,j}^B, d_{N,i,j}^A$, and $d_{N,i,j}^B$ at $(\hat{\theta}_N, \hat{\xi}_{n,N})$ are respectively represented by $\hat{A}_{N,i,j}^A, \hat{A}_{N,i,j}^B, \hat{B}_{N,j}^A, \hat{B}_{N,j}^B, \hat{d}_{N,i,j}^A$, and $\hat{d}_{N,i,j}^B$. Then, # $\{(k : k \in \text{nbd}(i, s_n))\}$ should be a small value in practice.

The asymptotic properties of the bias-corrected estimator $\hat{\theta}_N$ are presented in Theorem 3.2.

**Theorem 3.2.** Assume that the conditions of Theorem 3.1 hold. If $s_n \to \infty$ as $n \to \infty$ such that $\sup \# \text{nbd}(i, s_n)) \to 0$ and $\text{inf}_i \# \{(k : k \in \text{nbd}(i, s_n))\} \to \infty$,\(^{19}\) we have $\hat{A}_N \to A_\infty$ and

$$\sqrt{N} (\hat{\theta}_N - \theta_0) \xrightarrow{d} N (0, \Sigma^{-1})$$

as $n \to \infty$.

When compared to the bias correction for the linear SARF model, we must use the estimates $\hat{\xi}_{n,N}$ when evaluating $\hat{\Sigma}_N$ and $\hat{A}_N$, given that a closed-form expression for $\hat{\xi}_{n,N}$ does not exist. Notably, our model indicates that # of origin units $= \frac{p}{n} = 1$. Consequently, there are no constraints on the sample size $n$, in contrast to the requirements in Lee and Yu (2010) concerning panel settings with large cross-sectional units and time periods.\(^{20}\)

4. Monte Carlo simulations

4.1. Finite sample performance of the MLE

We conducted Monte Carlo simulations to assess the finite sample performance of the MLE. To align with the empirical application detailed in Section 5, we used the same variable components: $\{x_{n,1}\}$ (representing logged population levels) and $\{z_{N,1}\}$ (indicating logged geographic distances). Hence, we consider $n = 49$. Our simulations also incorporate the specification $(W_n, M_n) = (W_n, M_n^O)$, which mirrors the empirical application. Specifically, $W_n^O$ represents the shares of historical migration inflows from state $i$ to state $j$ relative to the total migration inflows to state $i$. Conversely, $M_n^O$ details the shares of historical migration inflows from state $j$ to state $i$ in relation to the total outflows from state $j$. A more comprehensive explanation of $W_n^O$ and $M_n^O$ will be presented in the subsequent subsection.

In the first stage of the process, we set $\gamma_{n,j} = (1, z_{n,j,1}, x_{n,1,1}, x_{n,1,j})$. For the second stage process, we consider $\mu_{n,j} = z_{N,j,1}$. The true parameters for the first stage are given by $\gamma_0 = (\alpha_0, \beta_0, \gamma_0, \beta_0, \mu_0) = (-14.076, 0.707, 0.707)'$, and for the second stage, the

\(^{18}\) 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 for the time dimension. Implementations of this truncation concept in the spatial econometric context can be found in Kelejian and Prucha (2007), Kim and Sun (2011), and Conley et al. (2023).

\(^{19}\) The validity of Theorem 3.2 is also supported by $\hat{\Sigma}_N \to \Sigma$ as $n \to \infty$. This conclusion stems from the continuous mapping theorem with $||\theta_n - \theta_0||_2 \to 0$ and $\hat{\xi}_{n,N} - \hat{\xi}_{n,N} \to 0$ as $n \to \infty$. For details, see the proof of Theorem 3.2 in the supplement file (Section 2.2.1).

\(^{20}\) To deduce the asymptotic properties of the bias-corrected MLE for spatial dynamic panel data models, Lee and Yu (2010) established that both $\frac{n}{p} \to 0$ and $\frac{p}{n} \to 0$ are necessary. These conditions were introduced to confirm the asymptotic equivalence between the infeasible (ideal) bias-corrected MLE and the feasible bias-corrected MLE.
Table 1
Simulation results for the SARF hurdle model with fixed effects.

| $x_0 = (-14, -0.76, 0.7, 0.7)'$ and $\omega_0 = (0.2, 0.2, 0.1, -0.12, 0.6)'$
| --- |
| $\alpha'$ | $\rho'$ | $\rho''$ | $\lambda$ | $\gamma$ | $\rho$ | $\beta$ | $\sigma^2$
| --- | --- | --- | --- | --- | --- | --- | --- |
| True values | $-14$ | $-0.76$ | $0.7$ | $0.2$ | $0.2$ | $0.1$ | $-0.12$ | $0.6$
| MLE | 0.4885 | $-0.0113$ | $-0.0133$ | $-0.0140$ | $-0.0097$ | $-0.0107$ | 0.0290 | 0.0061 | $-0.0278$
| STD | 1.1714 | 0.0764 | 0.0462 | 0.0477 | 0.0169 | 0.0167 | 0.0313 | 0.0166 | 0.0181
| CP | 0.9690 | 0.9490 | 0.9480 | 0.9420 | 0.8890 | 0.8920 | 0.8060 | 0.9170 | 0.6410
| Bias corrected MLE with $s_n = 1$ | 0.4885 | $-0.0113$ | $-0.0133$ | $-0.0140$ | $-0.0055$ | $-0.0105$ | 0.0265 | 0.0068 | $-0.0149$
| STD | 1.1714 | 0.0764 | 0.0462 | 0.0477 | 0.0171 | 0.0168 | 0.0311 | 0.0164 | 0.0177
| CP | 0.9690 | 0.9490 | 0.9480 | 0.9420 | 0.9200 | 0.8920 | 0.8290 | 0.9210 | 0.8500
| Bias corrected MLE with $s_n = 2$ | 0.4885 | $-0.0113$ | $-0.0133$ | $-0.0140$ | $-0.0007$ | $-0.0099$ | 0.0229 | 0.0074 | $-0.0156$
| STD | 1.1714 | 0.0764 | 0.0462 | 0.0477 | 0.0172 | 0.0168 | 0.0314 | 0.0165 | 0.0178
| CP | 0.9690 | 0.9490 | 0.9480 | 0.9420 | 0.9350 | 0.9040 | 0.8440 | 0.9130 | 0.8440
| Bias corrected MLE with $s_n = 3$ | 0.4885 | $-0.0113$ | $-0.0133$ | $-0.0140$ | $-0.0007$ | $-0.0099$ | 0.0229 | 0.0074 | $-0.0156$
| STD | 1.1714 | 0.0764 | 0.0462 | 0.0477 | 0.0172 | 0.0168 | 0.0314 | 0.0165 | 0.0178
| CP | 0.9690 | 0.9490 | 0.9480 | 0.9420 | 0.9060 | 0.8480 | 0.9090 | 0.8410 |

$a_0 = (0.2, 0.2, 0.1, -0.12, 0.6)'$. We draw $(v_{N,i}, e_{N,i}) \sim \text{i.i.d. } N((0, 0), (1, 0) \sigma^2_0)$ for $i = 1, \ldots, n$. The fixed-effect components are modeled by:

$$a \sim \text{i.i.d. } N(0, 0.1^2)$$

where $c^2 = -2.3$, $e_{N,i} \sim \text{i.i.d. } N(0, 0.1^2)$ for $j = 1, \ldots, n$, and $\eta_0 = a_0$ for all $i = 1, \ldots, n$. In both Cases 1 and 2, about 91% of the flows are positive. The proportion of nonzero observations can be controlled by adjusting $a_0$, $\lambda$, and $\eta_0$. A more negative value for $\lambda$ or $\rho''$ reduces $\gamma_{N,ij}^{\ast}$, increasing the proportion of zero flows in the first stage. Conversely, increasing $\rho''$ or $\lambda$ decreases $\gamma_{N,ij}^{\ast}$, leading to a higher proportion of zero flows in the second stage. Intuitively, when either $\rho''$ or $\lambda$ increases, larger geographic distances ($z_{N,ij}$) result in a greater number of zero flows.

To generate the data, follow these steps:

**Step 1:** Define $y_{N}^{(k)} = \text{vec}(y_{N}^{(k)})$ as an $N \times 1$ outcome vector from the $k$th iteration. Initialize $y_{N}^{(0)} = 0$. Then, for each $k = 0, 1, \ldots$,

$$y_{N}^{(k+1)} = y_{N}^{(k)} \text{ if } y_{N}^{(k)} > 0 \text{ and } y_{N}^{(k+1)} > 0$$

$$y_{N}^{(k+1)} = 0 \text{ otherwise}$$,

where $y_{N}^{(k)} = x_{N,ij}^{(k)} + v_{N,ij}$ and $y_{N}^{(k+1)} = \chi_{N,ij}^{(k)} + \varepsilon_{N,ij}$ with

$$\chi_{N,ij}^{(k)} = \lambda_0 \sum_{g=1}^{n} w_{a,gg} y_{N,gg}^{(k)} + \rho_0 \sum_{g=1}^{n} w_{a,gg} m_{a,gg} + \rho_0 \sum_{g=1}^{n} w_{a,gg} y_{N,gg}^{(k)} m_{a,gg} + z_{N,ij}^{(k)} + \beta_0 + \alpha_0 + \eta_0$$.

**Step 2:** Conclude the iteration when $\gamma$ meets the condition $\max_{i,j=1,\ldots,n} |y_{N,i}^{(k+1)} - y_{N,j}^{(k+1)}| < 10^{-6}$. This contraction is supported by Assumption 2.1. Finally, assign $y_{N} = y_{N}^{(k)}$.

Three criteria are used to evaluate the finite sample performance of $\hat{\beta}_N$ and $\tilde{\beta}_N$: (i) empirical bias, (ii) empirical standard deviation, and (iii) the coverage probability of a nominal 95% confidence interval (CP). Each experiment consists of 1,000 sample repetitions. For the trimming spatial order $s_n$, we define

$$(D_{N})_{ij} = 1 \text{ if } i = j \text{ and } D_{N}=1 \text{ for } i \neq j \text{ and } (D_{N})_{ij} = 1$$.

This definition restricts the number of connected units to minimize the bias correction terms’ reliance on numerous cross-moments for well-connected units. Then, $k \in \text{nd}(i, s_n)$ if $[D_{N}]_{ik} \neq 0 \text{ for some } l = 0, 1, \ldots, s_n$ and $\# \{k : k \in \text{nd}(i, s_n)\} \leq s_n + 1$ for all $i = 1, \ldots, n$. We report the bias-corrected MLE with $s_n = 1, 2, 3$. Table 1 presents the simulation results. Downward biases are observed in the MLEs of $\rho''$, $\rho''$, $\rho''$, $\rho''$, $\rho''$, $\rho''$, and $\rho''$. We verify that the biases in the MLEs for the first-stage parameters show the same patterns.

21 We report the simulation results for other parameter values in Section 3.1 of the supplement, where similar patterns to those in Table 1 are observed.

22 For example, Missouri and Tennessee, the U.S. states with the most adjacent states (eight each), contrast sharply with Maine, which has only one adjacent state. If we apply the states’ adjacency matrix for $D_{N}$, $\# \{k : k \in \text{nd} (\text{Missouri, } s_n = 1)\} = 9$, while $\# \{k : k \in \text{nd} (\text{Maine, } s_n = 1)\} = 2$. 
as those from the model without fixed effects (see Section 3.2 in the supplement). These biases diminish notably as $n$ grows. Hence, the presence of fixed effects in the SARP hurdle model does not affect the finite sample performance of the MLE for the first-stage parameters. After correcting the asymptotic biases, the magnitudes of the biases in the MLEs for $\lambda_0$, $Y_0$, and $\sigma^2_\epsilon$ decrease, and their CIs show improvement, indicating improved accuracy. As for the trimming parameter, we observe the improvements of the MLEs of all the second-stage parameters (improvement of the CP for the estimate of $\beta_0$) when $s_n = 1$. Increasing $s_n$ enhances the bias correction performance for the estimates of $\lambda_0$, $Y_0$, and $\sigma_0^2$, but the performance deteriorates for the estimates of $\beta_0$ and $\sigma^2_\epsilon$. As Fernandez-Val and Weidner (2016) point out, we suggest (i) conducting a sensitivity analysis by presenting estimates for various $s_n$ values and (ii) considering a small value for $\#\{k: k \in \text{nbhd}(i, s_n)\}$ to avoid many cross-moments for the bias correction terms.

We provide a summarized overview of supplementary simulation results, detailed in Section 3 of the supplemental file. First, we explore two distribution specifications for the disturbances: logistic and gamma. To ensure a fair comparison, the generated random variables for these distributions were standardized to a zero mean and a variance of one. Specifically for the gamma distribution, we adopted a specification of $\Gamma(4, \frac{1}{2})$, characterized by a pronounced right skew. The MLEs tend to perform well under both distribution specifications. However, the MLEs for the first-stage parameters under the gamma distribution were less accurate in terms of biases and CIs. Second, we assess the finite sample properties of the MLEs in the model without fixed effects. Substantial biases in the MLEs of $\lambda_0$, $Y_0$, and $\sigma_0$ were observed when the model was estimated without fixed effects, even though the true DGPs incorporate these fixed effects.

### 4.2. Selecting spatial weighting matrices

This subsection addresses the issue of model comparison when selecting $(W_n, M_n)$ in our model, which distinguishes between origin- and destination-based spatial dependencies. While a researcher might have some prior knowledge about the specifications of these matrices, the challenge of model selection can still emerge. To address this, we propose the use of Akaike weight, based on the Akaike Information Criterion (AIC) as suggested by Akaike (1973). It is worth noting that the selection of spatial weighting matrices has been studied in contexts such as comparing non-nested model specifications, with references such as Kelejian (2008), Kelejian and Piras (2011), Liu et al. (2014), and Zhang and Yu (2018). A notable advantage of using Akaike weights over hypothesis tests is the ability for practitioners to compare multiple models at once. In empirical applications, the Akaike weight is especially valuable because it eliminates the unknown true AIC from both the numerator and denominator, facilitating more effective model comparisons.

In Section 5, we consider our baseline specification, denoted as Specification 1, in which $(W_n, M_n) = (W_n^0, M_n^0)$. In this specification, $W_n^0$ signifies the shares of five-year historical migration to destination location $i$, while $M_n^0$ captures the shares of five-year historical migrations originating from location $j$. Consequently, $w_{nk}^0$ (associated with $y_{nk}$) represents the force acting on destination location $i$, and $m_{nk}^0$ (associated with $y_{nk}^0$) indicates the force emanating from origin location $j$. Additionally, in Section 5, we explore two alternative specifications. Specification 2 uses $(W_n, M_n) = (W_n^0, W_n^f)$, where $W_n^f$ is the states’ adjacency matrix. Meanwhile, Specification 3 adopts $(W_n, M_n) = (W_n^{R,R}, W_n^{w,R})$, which is based on the specification by LeSage and Pace (2008), with $W_n^{R,R}$ denoting the row-normalized version of $W_n^f$.

We prefer Specification 1 as the baseline over Specifications 2 and 3 for two primary reasons. Firstly, Specification 1 distinguishes between origin- and destination-based spatial dependencies. Secondly, it captures directional forces effectively. In this specification, the term $\lambda_0 \sum_{k=1}^{n} w_{nk}^0 y_{nk}$ represents the outflow effect from location $i$ influenced by the forces $w_{nk}^0$. The term $Y_0 \sum_{k=1}^{n} m_{nk}^0 w_{nk}^0$ represents the inflow effect to location $i$ influenced by the forces $m_{nk}^0$. Additionally, $\rho_0 \sum_{k=1}^{n} w_{nk}^0 y_{nk}^0 m_{nk}^0$ illustrates the effect from flows among third-party units through forces $w_{nk}^0$ and $m_{nk}$.

For both Specifications 1 and 3, given the row-normalization of matrices $W_n^0$, $M_n^0$, and $W_n^{R,R}$, terms such as $\sum_{k=1}^{n} w_{nk}^0 y_{nk}^0$ and its analogues can be seen as weighted averages of neighboring flows. In contrast, Specification 2, due to the absence of row-normalization in $W_n^f$, has these terms representing locally aggregated flows. Consequently, the parameter space for $\delta$ in Specification 2 is constrained in comparison to Specifications 1 and 3. In our data generation, with $n = 49$, we consider parameter vectors $\kappa_0 = (-14, -0.76, 0.7, 0.7)^T$ and $\alpha_0 = (0.2, 0.2, 0.1, -0.12, 0.6)^T$ for Specifications 1 and 3. Meanwhile, for Specification 2, we use $\kappa_0 = (-14, -0.76, 0.7, 0.7)^T$ and $\alpha_0 = (0.02, 0.02, 0.01, -0.12, 0.6)^T$. These parameter vectors produce roughly 12% zero flows.

Table 2 displays the average Akaike weights, accompanied by the empirical probabilities associated with each model selection. For every scenario, the true model is always selected from the 1,000 generated data. This underscores the Akaike weight’s effectiveness in pinpointing the true model, particularly when the data-generating process (DGP) closely aligns with the application’s data environment.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Akaike weights.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True DGP</strong></td>
<td>Specification 1</td>
</tr>
<tr>
<td>Estimated model</td>
<td>1.0000</td>
</tr>
<tr>
<td>Specification 1</td>
<td>(100.00%)</td>
</tr>
<tr>
<td>Specification 2</td>
<td>0.0000</td>
</tr>
<tr>
<td>(0.00%)</td>
<td>(100.00%)</td>
</tr>
<tr>
<td>Specification 3</td>
<td>0.0000</td>
</tr>
<tr>
<td>(0.00%)</td>
<td>(0.00%)</td>
</tr>
</tbody>
</table>

Note: Each number denotes the average Akaike weight. In the parenthesis, we report the empirical probability of choosing a model.
5. Application: Migration flows across U.S. states

This section examines the migration flows among the 48 contiguous U.S. states (excluding Alaska and Hawaii) and Washington DC. We consider the following variables:

- Logged population levels of the states, denoted as \( \{x_{n,1}\} \),
- Percentage growth rates of personal incomes from 2010, represented by \( \{x_{n,2}\} \),
- Logged degrees of nodes in the states’ adjacency network, given by \( \{x_{n,3}\} \),
- Logged geographic distances, with an addition of 1, denoted as \( \{z_{N,ij,1}\} \),
- Differential of the insured unemployment rate, represented by \( \{z_{N,ij,2}\} \), and
- Housing burden ratio differential, given by \( \{z_{N,ij,3}\} \).

All data for this study are sourced from the U.S. Census.

The first panel of Fig. 1 depicts the distribution of U.S. migration flows from state \( j \) in 2009 to state \( i \) in 2010, represented by \( mflow_{N,ij} \). This distribution is highly right-skewed. Consequently, we use the logged migration flows (incremented by 1) as the dependent variable, denoted as \( y_{N,ij} = \ln(mflow_{N,ij} + 1) \) for each pair \((i, j)\). Given that the minimum nonzero migration flow is 9, we have \( y_{N,ij} = 0 \) only when \( mflow_{N,ij} = 0 \) in this study. The second panel of Fig. 1 demonstrates that the distribution of logged positive flows closely aligns with the normal distribution (mean = 7.0731 and standard deviation = 1.7902). Given the generalized normal distribution assumption on \( \epsilon_{N,ij} \) as noted in Remark 2.1, the estimated \( \tau_2 \) is 2.0099 and its standard error is 0.0856. We find no reason to reject the normal distribution assumption (as discussed in Section 4.2 of the supplement). Descriptive statistics are provided in Table 3, where we note that 91.30% of the observations are positive. This observation leads us to consider the SARF hurdle model. For the first stage, we define \( x'_{N,ij} = (1, z_{N,ij,1}, z_{N,ij,2}, z_{N,ij,3}, x_{n,1,i}, x_{n,2,i}, x_{n,3,i}, x_{n,1,j}, x_{n,2,j}, x_{n,3,j}) \), and for the second stage.

---

**Table 3**

<table>
<thead>
<tr>
<th>Description</th>
<th>Mean</th>
<th>Std.dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration flows (x10⁴)</td>
<td>1.8686</td>
<td>17.3857</td>
<td>0.0000</td>
<td>541.3287</td>
</tr>
<tr>
<td>(% of nonzero observations)</td>
<td>(91.30%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_{N,ij} = \ln(mflow_{N,ij} + 1) )</td>
<td>6.4574</td>
<td>2.6274</td>
<td>0.0000</td>
<td>15.5044</td>
</tr>
<tr>
<td>Intrastate migration flows (x10⁴)</td>
<td>78.2708</td>
<td>94.9532</td>
<td>5.9032</td>
<td>541.3287</td>
</tr>
<tr>
<td>(% of nonzero observations)</td>
<td>(100%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interstate migration flows (x10⁴)</td>
<td>0.2769</td>
<td>0.5350</td>
<td>0.0000</td>
<td>6.8959</td>
</tr>
<tr>
<td>(% of nonzero observations)</td>
<td>(91.11%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic dist. (km)</td>
<td>1537.7266</td>
<td>962.2910</td>
<td>26.7745</td>
<td>4283.9987</td>
</tr>
<tr>
<td>Population (x10⁶)</td>
<td>6.2248</td>
<td>6.8407</td>
<td>0.5443</td>
<td>36.9617</td>
</tr>
<tr>
<td>Personal income growth (%)</td>
<td>1.4598</td>
<td>1.4948</td>
<td>−1.2748</td>
<td>7.2790</td>
</tr>
<tr>
<td>Insured unemployment rate (%)</td>
<td>4.0485</td>
<td>1.1770</td>
<td>1.5156</td>
<td>6.5631</td>
</tr>
<tr>
<td>Housing burden ratio (%)</td>
<td>33.6293</td>
<td>5.0449</td>
<td>26.4516</td>
<td>48.0656</td>
</tr>
<tr>
<td>Degree of ( W^*_a )</td>
<td>4.4490</td>
<td>1.6338</td>
<td>1.0000</td>
<td>8.0000</td>
</tr>
</tbody>
</table>

---

23 Given that the states’ adjacency matrix is symmetric, its outdegree matches its indegree, and thus, is referred to as its degree.
flows. SARF hurdle and Tobit models. Key observations for model comparisons include: Akaike weight, provides the most accurate prediction of the distribution of $y$ mean and standard deviation of positive flows, as generated by Specifications 1, 2, and 3. Notably, Specification 1, favored by the $(\sigma, \zeta, \gamma)$ stage, we use $z_{N,j,i} = \{z_{N,j,i,1}, z_{N,j,i,2}, z_{N,j,i,3}\}$. Based on the simulation results, we employ the same $D_n$ in (8) for the trimming spatial order $s_n$ and report the results when $s_n = 2$.

Table 4 and Fig. 2 present the estimation results. Based on the Akaike weights, Specification 1, characterized by $(W_n, M_n) = (W_n^1, M_n^1)$, stands out as the preferred choice. The latter section of Table 4 displays the proportions of zero flows, alongside the $\lambda$ and $\gamma$'s population, $\psi$'s degree, $\delta$'s unemployment rate and housing burden ratio. Theoretical standard errors are in parenthesis. Estimates that are significant at the 10%, 5%, and 1% levels are respectively marked by ‘*’, ‘**’, and ‘***’. Last, here is information in the actual data: proportion of zero flows $= 0.0870$, mean of positive flows $= 7.0731$, and standard deviation of positive flows $= 1.7902$.

Table 4 and Fig. 2 present the estimation results. Based on the Akaike weights, Specification 1, characterized by $(W_n, M_n) = (W_n^1, M_n^1)$, stands out as the preferred choice. The latter section of Table 4 displays the proportions of zero flows, alongside the $\lambda$ and $\gamma$'s population, $\psi$'s degree, $\delta$'s unemployment rate and housing burden ratio. Theoretical standard errors are in parenthesis. Estimates that are significant at the 10%, 5%, and 1% levels are respectively marked by ‘*’, ‘**’, and ‘***’. Last, here is information in the actual data: proportion of zero flows $= 0.0870$, mean of positive flows $= 7.0731$, and standard deviation of positive flows $= 1.7902$.

Table 4 and Fig. 2 present the estimation results. Based on the Akaike weights, Specification 1, characterized by $(W_n, M_n) = (W_n^1, M_n^1)$, stands out as the preferred choice. The latter section of Table 4 displays the proportions of zero flows, alongside the $\lambda$ and $\gamma$'s population, $\psi$'s degree, $\delta$'s unemployment rate and housing burden ratio. Theoretical standard errors are in parenthesis. Estimates that are significant at the 10%, 5%, and 1% levels are respectively marked by ‘*’, ‘**’, and ‘***’. Last, here is information in the actual data: proportion of zero flows $= 0.0870$, mean of positive flows $= 7.0731$, and standard deviation of positive flows $= 1.7902$.

Fig. 2 contrasts the proportions of zero flows and distributions of positive flows in the actual data with those generated by the SARF hurdle and Tobit models. Key observations for model comparisons include:

- The SARF Tobit model yields fewer zero flows compared to both the actual data and the flows generated by the SARF hurdle model.
- In the SARF hurdle model, 7.71% of flows are predicted as zero, with all attributed to the model’s first stage. This underscores the unique nature of the two-stage processes underpinning U.S. migrations.

---

**Note:** Recall that $x_{z,ij} = $ state $i$’s population, $x_{z,ij} = $ state $i$’s personal income growth, $z x_{N,ij} = $ logged state $i$’s degree, $z x_{N,ij} = \ln(dx_{N,ij}+1)$, $z x_{N,ij} = [\text{unemp}_{ij} - \text{unemp}_{ij}]$, and $z x_{N,ij} = [\text{hbr}_{ij} - \text{hbr}_{ij}]$, where $\text{unemp}_{ij}$ and $\text{hbr}_{ij}$ denote respectively state $i$’s unemployment rate and housing burden ratio. Theoretical standard errors are in parenthesis. Estimates that are significant at the 10%, 5%, and 1% levels are respectively marked by ‘*’, ‘**’, and ‘***’. Last, here is information in the actual data: proportion of zero flows $= 0.0870$, mean of positive flows $= 7.0731$, and standard deviation of positive flows $= 1.7902$.

---

**Fig. 2.** Distributions of the actual and generated data.
The second and third panels in Fig. 2 display the distributions of positive flows in the actual data versus those generated by the models. In terms of data fit, the SARF hurdle model outperforms the SARF Tobit model, particularly in variance estimation. The exaggerated variance in the SARF Tobit model is likely due to its inherent assumption that a single latent variable process governs both zero and positive flows, leading to an overestimated variance parameter, $\sigma_0^2$. Part I of Table A.4 in the supplement shows that a significant upward bias in the estimate of $\sigma_0^2$ is detected when we use the SARF Tobit model, but the true DGP follows the SARF hurdle model.

From the estimated values of $\hat{\lambda}_0$ and $\gamma_0$, we note significant positive spatial effects on $y_{N,jj}$ for a pair $(i,j)$ arising from outflows from state $j$ (represented as $y_{N,jj}$ for $g \neq i$) and inflows to state $i$ (depicted as $y_{N,ij}$ for $h \neq j$). The estimate for $\rho_0$ lacks statistical significance. Under the SARF Tobit specification, the estimate for $\rho_0$ is significantly positive, but this result can be attributed to misspecification bias. Both geographic and economic distances influence both stages of the migration process. Specifically, the economic distance, measured by housing burden ratios, and the geographic distance have a significant impact. Additionally, the insured unemployment rate differential specifically impacts the second stage process. Given these insights, policymakers might focus on assisting long-distance migrants or addressing economic disparities to encourage migration.

Lastly, we calculated the simulated marginal effects, represented by $\frac{\partial y_{N,jj}}{\partial z}$, utilizing the DGP formulated by the estimated parameters. In the linear SARF model, this quantity is synonymous with $\{S'[S(N-1)^{i-1}+1(i-1)^{i-1}]\} \cdot \beta$, where $S = I_N - A_N$. Specifying the marginal effect of $z_{N,jj}$ on $y_{N,jj}$ within the SARF hurdle specification presents challenges due to several factors: (i) the dual coefficients, $\beta_j^y$ and $\beta_j$, associated with $z_{N,jj}$, (ii) the model’s inherent nonlinearity, and (iii) the varying effects across $i$ and $j$ stemming from $W_\delta$ and $M_\delta$. Table 5 outlines the primary statistics of these computed marginal effects. The mean marginal effects for $z_{N,jj}$, $z_{N,jj}^2$, and $z_{N,jj}^3$ are $-0.7090$, $-0.0295$, and $-0.0424$ respectively. Their absolute values are marginally smaller than their respective median values, which are $-0.7675$, $-0.0320$, and $-0.0459$.

### 6. Conclusion

We introduce a SAR model tailored for an origin–destination flow variable, which we term the SARF model. Specifically, in data scenarios where a flow variable might contain zero values, we propose SARF hurdle models. These models outline a two-stage process: (i) the participation of a flow unit in the first stage and (ii) the latent amount of flow in the second stage.

To robustly control unit-specific unobservables, our model incorporates a two-way fixed effect specification. This model’s main parameters and fixed-effect components are estimated by the ML estimation method. We delve into the asymptotic properties of the MLE, drawing upon the concept of NED. Through Monte Carlo simulations, we present the finite sample properties of the MLEs. Finally, when applied to migration flows between U.S. states, our models identify significant spatial influences from neighboring migration flows. We find that the SARF hurdle models fit the data more adeptly compared to the SARF Tobit models.

### Appendix. Mathematical proofs

To streamline our analysis, we adopt the following notations: For each pair $(i, j)$, where $i, j \in \{1, \ldots, n\}$, there exists a unique index $f \in \{1, \ldots, N\}$ such that $f = (j-1)n + i$. We denote $e_{i,j}$ as an $n$-dimensional unit vector in which only the $j$th component is one, and all other entries are zero. We set $S_0$ to be $(i, j) : y_{N,ij} = 0$ and $S_1$ as $(i, j) : y_{N,ij} > 0$. Furthermore, we define the following: $c_{W,F} = \sup_n \|W_\delta\|_1$, $c_{M,F} = \sup_n \|M_\delta\|_1$, $c_{W,F} = \sup_n \|W_\delta\|_{\infty}$, and $c_{M,F} = \sup_n \|M_\delta\|_{\infty}$.

### Appendix A. The density of $y_N$ for the SARF hurdle models

The density of $y_N$ is given by $f(y_N) = f(y_{N_0}, y_{1,N_1}) \cdot f(y_{1,N_1})$, where $N_0$ represents the number of zero flows and $N_1$ is the number of positive flows; and thus $y_{N_0}$ represents the subvector of flows being 0, and $y_{1,N_1}$ is the subvector of positive flows. Let

---

25 Part II of Table A.4 in the supplement verifies that substantial upward biases in the estimate of $\rho_0$ are observed when using the SARF Tobit model even though the true DGP follows the SARF hurdle model.
us consider the first part, \( f(y_{0,0}) = 0 \). Due to the independence of \( u_{N,j} \) and \( \epsilon_{N,j} \), we have:

\[
f(y_{0,0}) = \prod_{i,j \in S_0} \left( \frac{1}{y_{N,i,j} = 0} \right) + \prod_{i,j \not\in S_0} \left( \frac{1}{y_{N,i,j} > 0} \right)
\]

as derived in the main text. The fourth equality holds because for \( (i,j) \in S_0 \), \( u_{N,j} \) and \( \epsilon_{N,j} \) are independent events. The last equality above holds since \( 1 - \Phi(x_{N,j,i}^*) \cdot \Phi(x_{N,j,i}^*/\sigma) = 1 - \Phi(x_{N,j,i}^*/\sigma) \cdot \Phi(x_{N,j,i}^*/\sigma) \).

The resulting density of \( y_N \) is

\[
f(y_N) = \text{det}(S_N) \cdot \prod_{i,j=1}^n \left( 1 - \Phi(x_{N,j,i}^*/\sigma) \cdot \Phi(x_{N,j,i}^*/\sigma) \right)^{1_{\{y_{N,i,j} = 0\}}} \cdot \phi(\epsilon_{N,j}(\omega_{j,0})/\sigma).
\]

### Appendix B. NED properties

This section verifies the NED properties of \( \{y_{N,j}\} \). Recall that \( y_N \) can be represented by

\[
y_N = F(A_Ny_M + t_N),
\]

where \( F(\cdot) = F_f(F_j(\cdot)) \) and \( t_N = Z_Nb_0 + \alpha_0 \otimes l_s + \eta_0 + \epsilon \). The \((i,j)\)-component of \( t_N \) is denoted by \( t_{N,j} \), i.e., \( t_{N,j} = (t_{N,1}, \ldots, t_{N,i-1}, t_{N,i,j}, t_{N,i+1}, \ldots, t_{N,n}) \). To show the NED property of \( \{y_{N,j}\} \), let \( \|y_{N,j} - E(y_{N,j})[F_{N,j}^s(s)]\|_\infty \leq c_{N,j}(s) \), we need to characterize the NED coefficient \( \nu(s) \), where \( s \) denotes a threshold distance between pairs. The key part of characterizing \( \nu(s) \) is to control the partial sum

\[
sup_{n,j,i} \sum_{g,h} \sum_{(i,j) \not\in S_0} |A_N| \sum_{l=1}^\infty \left[ |N,i,j| \right]^{(l-1)n+i+(j-1)n+g+j} \to 0 \text{ as } s \to \infty,
\]

where the notation \( |A_N| \) for a matrix \( A_N = [a_{N,f,f'}] \) means \( |A_N| = [a_{N,f,f'}^*] \), since \( (I_N - |A_N|)^{-1} = \sum_{n=0}^\infty |A_N|^n \) characterizes an upper bound of the total spatial spillovers by Assumption 2.1.26 Lemma B.1 shows the result above, and it will be employed to prove Proposition B.1.

#### Lemma B.1

Let \( a_{N,(i,j),(g,h)} \) be the \((f, f')\)-element of \( |A_N| \), where \( f = (j-1)n+i \) and \( f' = (h-1)n+g \). Assumption 2.1 holds.

(i) Under Assumption 3.2 (ii-1), \( a_{N,(i,j),(g,h)} \) is bounded above by \( d_{f,f'} \), and \( a_{N,(i,j),(g,h)} = 0 \) otherwise. Then,

\[
sup_{n,j,i} \sum_{g,h} \sum_{(i,j) \not\in S_0} |A_N| \sum_{l=1}^\infty \left[ |N,i,j| \right]^{(l-1)n+i+(j-1)n+g+j} \leq \sum_{l=1}^\infty \frac{\epsilon^l}{a-2d} \to 0 \text{ as } s \to \infty,
\]

where \( [s/d] \) is the biggest integer that is less than or equal to \( s/d \).

(ii) Under Assumption 3.2 (ii-2), \( a_{N,(i,j),(g,h)} \leq C_0 \cdot d_{f,f'} \) for some \( C_0 > 0 \). In consequence,

\[
sup_{n,j,i} \sum_{g,h} \sum_{(i,j) \not\in S_0} |A_N| \sum_{l=1}^\infty \left[ |N,i,j| \right]^{(l-1)n+i+(j-1)n+g+j} \leq C_1 \cdot C_2 \cdot \frac{3^l}{a-2d} \cdot s^{-a+2d} \to 0 \text{ as } s \to \infty,
\]

for some constants \( C_1, C_2 > 0 \) satisfying \( C_1 = C_0 \cdot K_A \cdot \Gamma \cdot \sum_{l=1}^\infty l^2(l+1)^2 \), \( C_2 \) is an upper bound of \( m^{-2d+1} \#(g,h) \), \( m \leq d_{f,f'} \), \( (i,j) \not\in S_0 \) obtained by Jenish and Prucha (2009)’s Lemma A.1, where \( \Gamma \) denotes an upper bound of \( \|A_N\|_1 \) and \( K_A \) is a positive constant satisfying \( \frac{\varphi_n}{m^{d_a}} \leq K_A \cdot \Gamma \). Note that \( \frac{3^l}{a-2d} \cdot s^{-a+2d} \) is an upper bound of \( \sum_{m=1}^{m^{d_a}} m^{d_a-1} \).

#### Proof of Lemma B.1

Note that \( \left( e_{n,i} \otimes e_{n,i}^* \right) (I_n \otimes W_n) (e_{n,h} \otimes e_{n,g}) = e_{n,i}^* e_{n,i}^* w_{n,i,j} \), \( \left( e_{n,j} \otimes e_{n,j}^* \right) (M_n^* \otimes I_n) (e_{n,h} \otimes e_{n,g}) = m_{n,h} e_{n,j} e_{n,j} \), and \( \left( e_{n,i} \otimes e_{n,i}^* \right) (M_n^* \otimes W_n) (e_{n,h} \otimes e_{n,g}) = w_{n,i,j} m_{n,h,j} \). Then,

\[
a_{N,(i,j),(g,h)} = |\lambda_i (I + h) w_{n,i,j} + \gamma_0 (i = g) m_{n,h,j} + \rho_0 w_{n,j} m_{n,h,j} |,
\]

26 If \( l = 0 \), \( |A_N|^0 = I_n \). Thus, \( |A_N|^0 \in \{1\}^{1-2(n+1-1)+1} \) if \( (i,j) \not\in (g,h) \) by Assumption 3.1.
so there exist four mutually exclusive cases of $a_{N,(i,j),(g,h)}$. Then, the column sum of $|A_N|$ for each pair $(g,h)$ is

$$\sum_{i,j=1}^n a_{N,(i,j),(g,h)} = \sum_{i,j=1}^n |\lambda_0| \sum_{i,j=1}^n \gamma_0 |m_{i,j}| + \rho_0 |w_{i,j}| m_{i,j}$$

$$\leq |\lambda_0| \sum_{i=1}^n |m_{i,j}| + \left| \gamma_0 \right| \sum_{j=1}^n m_{i,j} + \left| \rho_0 \right| \sum_{i=1}^n |w_{i,j}| m_{i,j}$$

$$\leq |\lambda_0| |e_{w,c} + \left| \gamma_0 \right| |e_{m,r} + \left| \rho_0 \right| |e_{w,c} |e_{m,r}|.$$ 

Set $\Gamma = |\lambda_0| |e_{w,c} + \left| \gamma_0 \right| |e_{m,r} + \left| \rho_0 \right| |e_{w,c} |e_{m,r}|$. Then, $|A_N| \leq \Gamma < \infty$ by Assumption 3.2 (i).

Case 1, Assumption 3.2 (ii-1): Suppose Assumption 3.2 (ii-1) holds. First, if $i = g$ and $j = h$, we have $a_{N,(i,j),(g,h)} = 0$. Second, when $i \neq g$ and $j = h$, $a_{N,(i,j),(g,h)} = |\lambda_0| |w_{i,j}| > 0$ only if $d_F((i,j),(g,h)) \leq \overline{d}$ since $d_F((i,j),(g,h)) = \max\{d(i,g),d(j,h)\} = d(i,g) \leq \overline{d}$. Third, when $i = g$ and $j \neq h$, $a_{N,(i,j),(g,h)} = |\gamma_0| |m_{i,j}| > 0$ only if $d_F((i,j),(g,h)) \leq \overline{d}$ since $d_F((i,j),(g,h)) = \max\{d(i,g),d(j,h)\} = d(j,h) \leq \overline{d}$. Fourth, when $i \neq g$ and $j \neq h$, $a_{N,(i,j),(g,h)} = |\rho_0| |w_{i,j}m_{i,j}| > 0$ only if both $d_F((i,j),(g,h)) \leq \overline{d}$ and $d(j,h) \leq \overline{d}$, because both $|w_{i,j}| > 0$ and $|m_{i,j}| > 0$ only if both $d_F((i,j),(g,h)) \leq \overline{d}$ and $d(j,h) \leq \overline{d}$, then $d_F((i,j),(g,h)) = \max\{d(i,g),d(j,h)\} \leq \overline{d}$. Hence, we can show that $a_{N,(i,j),(g,h)}>0$ only if $d_F((i,j),(g,h)) \leq \overline{d}$ and $a_{N,(i,j),(g,h)} = 0$ otherwise.

For some large $s > 0$, we observe

$$\sum_{g,h:d_F((i,j),(g,h))>s} \sum_{i=1}^n \left| A_N \right| = \sum_{i=1}^n \sum_{j=1}^n \sum_{g,h=1}^n a_{N,(i,j),(g,h)} \sum_{i=1}^n \left| A_N \right|$$

$$\leq \sum_{i=1}^n \left| \sum_{j=1}^n \sum_{g,h=1}^n a_{N,(i,j),(g,h)} \right|$$

The first equality comes from Claim C.2.3 of Qu and Lee (2015). By Assumption 2.1, the following inequality justifying the first inequality above holds:

$$\sum_{g,h,d_F((i,j),(g,h))>s} \sum_{i=1}^n \left| A_N \right| = \sum_{i=1}^n \sum_{j=1}^n \sum_{g,h=1}^n a_{N,(i,j),(g,h)} \sum_{i=1}^n \left| A_N \right|$$

$$\leq \sum_{i=1}^n \left| \sum_{j=1}^n \sum_{g,h=1}^n a_{N,(i,j),(g,h)} \right|$$

Case 2, Assumption 3.2 (ii-2): First, $a_{N,(i,j),(g,h)} = 0$ if $i = g$ and $j = h$. Second, $a_{N,(i,j),(g,h)} = |\lambda_0| |w_{i,j}|$ if $i \neq g$ and $j = h$. Then, $a_{N,(i,j),(g,h)} = |\lambda_0| |w_{i,j}|$ and $a_{N,(i,j),(g,h)} = |\gamma_0| |m_{i,j}|$ if $i = g$ and $j \neq h$. Last, if $i \neq g$ and $j \neq h$, $a_{N,(i,j),(g,h)} = |\rho_0| |w_{i,j}m_{i,j}|$ and $d_F((i,j),(g,h)) = \max\{d(i,g),d(j,h)\}$ if $d_F((i,j),(g,h)) \geq 1$ and $d(j,h) \geq 1$ with $a > 1$. Hence, we have $a_{N,(i,j),(g,h)} \leq C_0 \cdot d_F((i,j),(g,h))^{-a}$ for some $C_0 > 0$.

The next step is showing $\|A_N\|_1 \leq |K_A| \cdot \Gamma \cdot \zeta^{l-1}$ for $l \in \mathbb{Z}_n$ where $K_A$ is a positive integer that does not depend on $n$. First, if $c_{w,c} > c_{w,r}$, $\|A_n\|_1 \leq \|A_N\|_1 \leq \|A_N\|_\infty \leq \zeta^l$. Consider the case $c_{w,c} > c_{w,r}$. Then, we have $\|W^p\|_1 \leq p \cdot c_{w,c} \cdot K_W \cdot c_{w,r}^{-1}$ for $p \in \mathbb{Z}_n$ by Claim C.1.2 of Qu and Lee (2015). For $l = 2, 3, \ldots$, by the trinomial expansion, we have

$$A_N = \sum_{p=q+r=l} \frac{1}{p!q!r!} \lambda_0^p \gamma_0^q \rho_0^r W_{N}^{p} M_{N}^{q} R_{N}^{r} = \sum_{p=q+r=l} \frac{1}{p!q!r!} \lambda_0^p \gamma_0^q \rho_0^r (M_{w,c}^{q+r} \otimes W_{w,c}^{p+r}).$$
Hence, by the triangle inequality,
\[\|A_N^l\| \leq \sum_{p+q+r+l} \left( \frac{l}{p!q!r!} \right)^d |\lambda_0|^p |\lambda_0|^q |\rho_0|^r \|W^\perp_\alpha\|_1 \|M_\alpha\|_\infty \]
\[\leq \sum_{p+q+r+l} \left( \frac{l}{p!q!r!} \right)^d |\lambda_0|^p |\lambda_0|^q (p+r) K_W c_{w,x} e_{w,x}^{-1} \|q^r\|_2 \]
\[\leq l \cdot K_W \cdot \varepsilon \left( \sum_{p+q+r+l} \left( \frac{l}{p!q!r!} \right)^d |\lambda_0|^p |\lambda_0|^q (p+r) K_W c_{w,x} e_{w,x}^{-1} \|q^r\|_2 \right) \leq l \cdot K_W \cdot \varepsilon \varepsilon^{-1},\]
where \(K_A\) is a positive constant satisfying \(K_W \cdot \varepsilon \varepsilon^{-1} \leq K_A \cdot \varepsilon\) and \(\varepsilon > 1\) such that \(c_{w,x} = \varepsilon c_{w,x}\). The second and third inequalities hold since \(\|W^\perp_\alpha\| \leq (p+r) \cdot K_W \cdot c_{w,x}^{-1} \|q^r\|_2\) and \(\left( \frac{l}{p!q!r!} \right)^d \leq \frac{l!}{p!q!r!} \) for \(p, q, r \in \mathbb{Z}_+\) such that \(p + q + r = l \in \mathbb{Z}_+\).

Consider the case of \(d_f((i,j),(g,h)) > s\) with \(f = (j-1)n+i, f' = (h-1)n+g\), and sufficiently large \(s\). For any \(l \in \mathbb{Z}_+\), we construct two matrices \(A_{1N} = [a_{N,i,j}(g,h)]\) and \(A_{2N} = [a_{N,i,j}(g,h)]\) as follows: \(a_{1N,i,j}(g,h) = a_{N,i,j}(g,h) \cdot 1\left( a_{N,i,j}(g,h) \leq C_0 \left( \frac{d_f((i,j),(g,h))}{l} \right)^{-\alpha} \right)\) and \(a_{2N,i,j}(g,h) = a_{N,i,j}(g,h) \cdot 1\left( a_{N,i,j}(g,h) \leq C_0 \left( \frac{d_f((i,j),(g,h))}{l} \right)^{-\alpha} \right)\), then \(|A_{1N}| = A_{1N} + A_{2N}\) and \(a_{1N,i,j}(g,h) = a_{2N,i,j}(g,h) = 0\). At least one of the items \(a_{1N,i,j}(g,h)\), \(a_{2N,i,j}(g,h)\), \(\cdots\), and \(a_{N,i,j}(g,h)\) would be less than or equal to \(C_0 \left( \frac{d_f((i,j),(g,h))}{l} \right)^{-\alpha}\), because there exist at least two neighboring points in the chain \((i,j) \rightarrow (i+1,j) \rightarrow \cdots \rightarrow (i-1,j-1) \rightarrow (g,h)\) such that their distance is at least \(d_f((i,j),(g,h))\).

Hence,
\[\|A_{1N}^l\|_{f^\alpha} = \sum_{i,j=1}^n \cdots \sum_{i,j-1,j+1}^n a_{N,i,j}(g,h) \cdot 1\left( a_{N,i,j}(g,h) \leq C_0 \left( \frac{d_f((i,j),(g,h))}{l} \right)^{-\alpha} \right) = 0,\]
and we have
\[\|A_{1N}^l\|_{f^\alpha} = \|A_{1N}^l\|_1 - \|A_{2N}^l\|_1 \leq \|A_{1N}^l\|_1 \cdot \sum_{m=0}^{l-1} \|A_{2N}^m\|_\infty \cdot \|A_{N}^l-m-1\|_1 \]
\[\leq \tilde{C}_0 \cdot d_f((i,j),(g,h))^{-\alpha} \cdot \varepsilon^{-1} \cdot K_A \cdot \varepsilon^{-1} \cdot \sum_{m=0}^{l-1} (l-m-1) \]
\[\leq \tilde{C}_0 \cdot K_A \cdot \varepsilon^{-1} \cdot d_f((i,j),(g,h))^{-\alpha} \cdot l^{2+a} \cdot \varepsilon^{-1},\]
where \(\|A_{1N}\|_1 = \max f^\alpha \cdot \|A_{1N}\|_{f^\alpha}\). The first inequality follows by Xu and Lee’s (2015) Lemma A.3 and \(\|A_{1N}\|_1 \leq l \cdot K_A \cdot \varepsilon^{-1}\) for \(l \in \mathbb{Z}_+\) by Assumption 3.2 (ii-2), all elements in \(A_{1N}\) are less than or equal to \(\tilde{C}_0 \left( \frac{d_f((i,j),(g,h))}{l} \right)^{-\alpha}\), and \(\sum_{m=0}^{l-1} (l-m-1) = \sum_{m=0}^{l-1} m = \frac{l(l-1)}{2} \leq l^2\). Then,
\[\sum_{i=1}^{\infty} \|A_{1N}^l\|_{(j-1)n+i,(i,j-1),g,h} \leq \sum_{i=1}^{\infty} \sum_{i,j=1}^n \cdots \sum_{i,j-1,j+1}^n a_{N,i,j}(g,h) \cdot 1\left( a_{N,i,j}(g,h) \leq C_0 \left( \frac{d_f((i,j),(g,h))}{l} \right)^{-\alpha} \right) \]
\[\leq \tilde{C}_0 \cdot K_A \cdot \varepsilon^{-1} \cdot d_f((i,j),(g,h))^{-\alpha} \cdot \varepsilon^{-1} \cdot \sum_{i=1}^{\infty} l^{2+a} \varepsilon^{-1} \leq C_1 \cdot d_f((i,j),(g,h))^{-\alpha},\]
where \(C_1 = \tilde{C}_0 \cdot K_A \cdot \varepsilon^{-1} \cdot \sum_{i=1}^{l} l^{2+a} \varepsilon^{-1} < \infty\).

This result implies
\[\sum_{g,h: d_f((i,j),(g,h)) > s} \sum_{i=1}^{\infty} \|A_{1N}^l\|_{(j-1)n+i,(i,j-1),g,h} \leq \sum_{m=0}^{\infty} \sum_{m=0}^{\infty} C_1 \cdot d_f((i,j),(g,h))^{-\alpha} \]
\[\leq \sum_{m=0}^{\infty} C_2 \cdot m^{-2a-1} \cdot C_1 \cdot m^{-a} \]
\[\leq \sum_{m=0}^{\infty} C_1 \cdot C_2 \cdot m^{-2} \cdot \left( \frac{m+1}{3} \right)^{-a} \]
\[\leq C_1 \cdot C_2 \cdot 3^a \int_1^\infty \frac{x^{-a+2d-1}}{x} \, dx \]
\[\leq C_1 \cdot C_2 \cdot 3^a (2d-1)^{-a} \leq (m+2)^{2d-1},\]
The second inequality above holds since \(#((g,h) : m \leq d_f((i,j),(g,h)) < m+1) \leq C_2 \cdot m^{2d-1}\) for some constant \(C_2 > 0\) by Jenish and Prucha’s (2009) Lemma A.1. Since \(m+1 \leq 3m \) for \(m \in \mathbb{Z}_+\) and \(a > 0\), the third inequality above holds since \(m^{2d-1} \leq (m+2)^{2d-1}\) and
\[ m^{-\alpha} \leq \left( \frac{m + 1}{3} \right)^{-\alpha}. \]

The fourth inequality comes from the relation \( m^{-a+2d-1} \leq f_{m-1} x^{-a+2d-1} dx \) since \( m^{-a+2d-1} \) is decreasing in \( m \). The last inequality holds because \( s \leq |t| + 1 \) with \( a + 2d < 0 \) by Assumption 3.2 (ii-2). Then, we have the desired result.

First, consider the sequences \( \{y_{N,j}\}, \{\sum_{n=1}^{m} w_{n,j} y_{N,n}\}, \{\sum_{n=1}^{m} w_{n,j} y_{N,n} m_{n}\}, \{y_{N,j}^a\}, \{y_{N,j}^b\}, \) and \( \{\epsilon_{N,j}(o_{ij})\} \), where \( \epsilon_{N,j}(o_{ij}) = y_{N,j} - y_{N,j}^a(o_{ij}) \). For the propositions below, the notation \( A_N \leq B_N \), where \( A_N = [a_{N,f}] \) and \( B_N = [b_{N,f}] \), means \( |a_{N,f}| \leq |b_{N,f}| \) for all \( f, f' = 1, \ldots, N \).

**Proposition B.1.** Assumptions 2.1, 3.1, and 3.3 hold.

(i) If \( \sup_{n,i,j} E[|y_{N,j}|]^p < \infty \) and \( \sup_{n,i,j} E[|\epsilon_{N,j}(o_{ij})|^p] < \infty \) for some \( p \geq 1 \), we have uniform \( L_p \)-boundedness of \( \{y_{N,j}\} \), \( \{\sum_{n=1}^{m} w_{n,j} y_{N,n}\} \), \( \{\sum_{n=1}^{m} w_{n,j} y_{N,n} m_{n}\} \), \( \{y_{N,j}^a\} \), \( \{y_{N,j}^b\} \), and \( \{\epsilon_{N,j}(o_{ij})\} \).

(ii) Under Assumption 3.2 (ii-1), \( \{y_{N,j}\}, \{\sum_{n=1}^{m} w_{n,j} y_{N,n}\}, \{\sum_{n=1}^{m} w_{n,j} y_{N,n} m_{n}\}, \{y_{N,j}^a\}, \{y_{N,j}^b\}, \) and \( \{\epsilon_{N,j}(o_{ij})\} \) are uniformly and geometrically \( L_2 \)-NED on \( \mathcal{E} \). For example, \( \|y_{N,j} - \mathbb{E}(y_{N,j}|F_N(s))\|_2 \leq C \cdot \varepsilon^d/2 \) when \( C \) is a constant, and \( d \) and \( a \) are constants such that \( a > 2d \) in Assumption 3.2 (ii-2).

**Proof of Proposition B.1.** (i). For an arbitrary pair \((i,j)\), choose \( x^{(1)}_{N,j} \) and \( x^{(2)}_{N,j} \) in \( \mathbb{R} \). As a first property, \( F(\cdot) \) is non-decreasing since

\[
F(x^{(1)}_{N,j}) = \mathbb{1}\{y_{N,j} > 0\} \cdot \max(0, x^{(1)}_{N,j}) \leq \mathbb{1}\{y_{N,j}^a > 0\} \cdot \max(0, x^{(2)}_{N,j}) = F(x^{(2)}_{N,j})
\]

if \( x^{(1)}_{N,j} \leq x^{(2)}_{N,j} \). Second, \( F(\cdot) \) is convex since

\[
F(c \cdot x^{(1)}_{N,j} + (1-c) \cdot x^{(2)}_{N,j}) = \mathbb{1}\{y_{N,j} > 0\} \cdot \max\left(0, c \cdot x^{(1)}_{N,j} + (1-c) \cdot x^{(2)}_{N,j}\right) \leq c \cdot \mathbb{1}\{y_{N,j} > 0\} \cdot \max(0, x^{(1)}_{N,j}) + (1-c) \cdot \mathbb{1}\{y_{N,j} > 0\} \cdot \max(0, x^{(2)}_{N,j}) = c \cdot F(x^{(1)}_{N,j}) + (1-c) \cdot F(x^{(2)}_{N,j})
\]

for \( c \in [0,1] \). Third, \( y_N \to F(A_N y_N + t_N) \) is a contraction mapping. To show this, choose \( y^{(1)}_N \) and \( y^{(2)}_N \) in \( \mathbb{R}^N \). Then, this mapping is a contraction since

\[
\|F(A_N y^{(1)}_N + t_N) - F(A_N y^{(2)}_N + t_N)\|_\infty \leq \|A_N\|_\infty \cdot \|y^{(1)}_N - y^{(2)}_N\|_\infty \leq \zeta \cdot \|y^{(1)}_N - y^{(2)}_N\|_\infty
\]

by Assumption 2.1. For the relation above, we use \( \max(0, x_s) - \max(0, x_s) \leq |x_1 - x_2| \) (i.e., Lipschitz).

Consequently, \( y_N \) can be represented by a unique explicit function of \( t_N \). Denote the fixed point of \( y_N = F(A_N y_N + t_N) \) as \( y_N(t_N) \) with \( j \)th element \( \epsilon_N(t_N) \). By the mean value theorem for a convex function (see Wegge 1974), we have

\[
y_N(t_N) - y_N(0) = \nabla F \cdot (A_N y_N(t_N) + t_N - A_N y_N(0)) + \epsilon_N(t_N),
\]

where \( \nabla F = \text{diag}(\nabla F_1, \ldots, \nabla F_N) \) and \( \nabla F \) is a subgradient of \( F(\cdot) \) at a point lying between \( \epsilon_{N,f}(A_N y_N(t_N) + t_N) \) and \( \epsilon_{N,f}(A_N y_N(0) + 0) \).\footnote{The mean value theorem is applied to each element of \( F(A_N y_N) \).

28. The mean value theorem is applied to each element of \( F(A_N y_N) \).

29. Note that the difference between \( \nabla F \) and \( \nabla F \) is their evaluation points, i.e., \( \nabla F \) is the subgradient evaluated between \( A_N y_N(t_N) + t_N \) and \( A_N y_N(0) + 0 \), while \( \nabla F \) is evaluated using the points lying between \( A_N y_N(t_N) + t_N \) and \( A_N y_N(0) + t_N \).} Note that \( \mathbb{E}(y_N(t_N)|F_N(s)) \) is an approximation of \( y_N(t_N) \), which is a function

\[
I_N - \nabla F A_N = \sum_{k=0}^{\infty} \left( \nabla F A_N^k \right) \nabla F \leq \sum_{k=0}^{\infty} \|A_N\|^k \equiv \mathbb{M} \leq \mathbb{M} = \|m_{N,j-1+1+i-1+1+i+1+j+1+k+1} \|
\]

Then, we have \( \|\epsilon_{N,f} \otimes \epsilon_{N,f}\|_N \leq \sum_{n=1}^{m} \|m_{N,j-1+1+i-1+1+j+1+k+1} \| \|y_{N,j}\|_N \|_N \). By the Minkowski's inequality, we have \( \|\epsilon_{N,f} \otimes \epsilon_{N,f}\|_N \leq \sum_{n=1}^{m} \|m_{N,j-1+1+i-1+1+j+1+k+1} \| \|y_{N,j}\|_N \|_N \).
of \(\{(u_{N,gh}, \epsilon_{N,gh}) : d_F((i,j),(g,h)) \leq s\}\). Then, we have
\[
\|y_{N,j} - \mathbb{E}\{y_{N,j}|\mathcal{F}_{N,j}(s)\}\|_{L_2} \leq \sum_{g,h : d_F((i,j),(g,h)) > s} \|m_{N,(i,j),(g,h)}\|_{L_2} \cdot \|\epsilon_{N,gh}\|_{L_2}
\leq \sup_{n,j} \|y_{N,j}\|_{L_2} \cdot \sup_{n,j,g,h : d_F((i,j),(g,h)) > s} \|m_{N,(i,j),(g,h)}\|_{L_2}.
\]
Note that \(\sup_{n,j} \|y_{N,j}\|_{L_2} < \infty\) by Assumptions 3.3 and 3.4. To show the NED properties of \(\{y_{N,j}\}\), we need to show \(\sup_{n,i,j} \sum_{g,h : d_F((i,j),(g,h)) > s} \|m_{N,(i,j),(g,h)}\|_{L_2} \to 0\) as \(s \to \infty\). Using Lemma B.1 with the similar argument of Xu and Lee's (2015) Proposition 1, we finish the proof. The details can be found in the supplement file.

Next, we consider the NED properties of \(\{1\{y_{N,j} > 0\}\}\) which is a component of \(\xi_{N,j}(\theta_j)\). The normality assumption (Assumption 3.4) helps to restrict an upper bound of PDFs of \(y_{N,j}^i\) and \(y_{N,j}^s\). Here are relevant lemmas and propositions. Since the ideas of the proofs are the same as Xu and Lee's (2015) Lemma 2 and Proposition 2, we do not provide their proofs in the main draft. Modified proofs for our framework can be found in the supplement file.

Lemma B.2. When \(A_n\) is an \(n\)-dimensional symmetric matrix, \(x_n'A_nx_n \geq \min_{i=1,\ldots,n} \phi_i(A_n)x_n'x_n\) where \(\phi_i(A_n)\) denotes the \(i\)th eigenvalue of \(A_n\) and \(x_n\) is a nonzero \(n\)-dimensional vector.

Lemma B.3. Assumption 3.4 additionally holds. The essential supremums of densities of \(y_{N,j}^s\) and \(y_{N,j}^a\) are uniformly bounded in \(i, j, n\).

Proposition B.2. Assumptions 2.1, 3.1, 3.3, and 3.4 hold.
(i) Under Assumption 3.2 (ii-1), \(\{y_{N,j} \geq 0\}\) is uniformly and geometrically \(L_2\)-NED on \(\Theta\). That is, \(\|1\{y_{N,j} > 0\} - \mathbb{E}\{1\{y_{N,j} > 0\}|\mathcal{F}_{N,j}(s)\}\|_{L_2} \leq C \cdot \zeta^{|s|/3}\) where \(C\) is a constant, and \(d\) is a constant defined in Assumption 3.2 (ii-1).
(ii) Under Assumption 3.2 (ii-2), \(\{y_{N,j} > 0\}\) is uniformly \(L_2\)-NED on \(\Theta\). That is, \(\|1\{y_{N,j} > 0\} - \mathbb{E}\{1\{y_{N,j} > 0\}|\mathcal{F}_{N,j}(s)\}\|_{L_2} \leq C \cdot s^{(2d-a)/3}\) where \(C\) is a constant, and both \(d\) and \(a\) are constants such that \(a > 2d\) in Assumption 3.2 (ii-2).

Appendix C. Asymptotic properties of the MLE
Fernandez-Val and Weidner (2016) investigate the asymptotic distribution of estimated parameters in nonlinear panel models with both individual and time-fixed effects in the context where both \(n\) and \(T\) grow large; this is the large-\(T\) variant of the incidental parameter problem. In deriving the asymptotic distribution of \(\hat{\theta}_N\), we draw upon concepts from their work, given the similarities in the framework between our models.

- They consider the scenario where \(0 < \lim_{n,T \to \infty} \frac{n}{T} < \infty\) (as per Fernandez-Val and Weidner (2016)'s Assumption 4.1 (i)). This aligns with our context, given that the number of units for origins and destinations is \(n\), which is equivalent to having \(T = n\).
- The fixed-effect specification in our model falls within the additive separable two-way fixed-effect category as described by Assumption 4.1 (iii) in Fernandez-Val and Weidner (2016).
- Our statistical objective function, \(\ell_N(\theta, \xi_n)\), is infinitely differentiable, ensuring it complies with the required smoothness conditions (Fernandez-Val and Weidner, 2016's Assumption 4.1 (iv)). It guarantees the higher-order stochastic expansions of \(\sqrt{n} \partial \ell_N(\theta, \xi_n, \delta_N)\).

However, there are some notable differences when compared to the framework in Fernandez-Val and Weidner (2016). Primarily, our model accounts for weak cross-sectional dependence among \(\{y_{N,j}\}\) across origins and destinations, as delineated by the NED concept. Additionally, our model necessitates an extra condition to ensure the strict concavity of \(\ell_N(\theta, \xi_n)\), aligning with Assumption 4.1 (v) in Fernandez-Val and Weidner (2016). In the main draft, we briefly describe arguments for consistency and asymptotic normality. Section 2 in the supplement file contains the arguments and proofs in detail.

The proposition presented below details the structure of \(\mathbb{E}\left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} \ell_N \right)\). The diagonal elements of \(\mathbb{E}\left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} \ell_N \right)^{-1}\) are of \(O(1)\), while its off-diagonal ones are of \(O\left( \frac{1}{\sqrt{N}} \right)\). If \(\mathbb{E}\left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} \ell_N \right)^{-1}\) is invertible for large \(n\), it can be approximated by a diagonal matrix.

Proposition C.1. Denote \(\mathbb{E}\left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} \ell_N \right)^{-1} \equiv \mathcal{H}_N = \begin{bmatrix} \mathcal{H}_N^a & \mathcal{H}_N^b \\ \mathcal{H}_N^b & \mathcal{H}_N^c \end{bmatrix}\), a block matrix. Under the regularity conditions for Theorem 3.1,
(i) \(a_{N,j} = [\mathcal{H}_N^a]_{ij}\) for \(j = 1, \ldots, n\) and \(\epsilon_{N,ij} = [\mathcal{H}_N^c]_{ij}\) for \(i = 1, \ldots, n\) are of \(O(1)\). Furthermore, \(\|\mathcal{H}_N - \tilde{\mathcal{H}}_N\|_{\max} = O\left( \frac{1}{\sqrt{N}} \right)\), where \(\tilde{\mathcal{H}}_N = \text{diag} \begin{bmatrix} a_{N,j} \varepsilon_{N,ij} \end{bmatrix}_{j=1}^{n} \varepsilon_{N,ij} \), is an approximation of \(\mathcal{H}_N\). It implies \(b_{N,ij} = [\mathcal{H}_N^b]_{ij} = O\left( \frac{1}{\sqrt{N}} \right)\) for \(i, j = 1, \ldots, n\).

\[\text{Note that the event } \{y_{N,j} > 0\} \text{ is the same as } \{y_{N,j}^s > 0\} \cap \{y_{N,j}^a > 0\}\]
Note that the dimension of $\hat{\xi}_{n,N}(\theta)$ (as well as $\xi_{30}$) is 2n, which grows with increasing $n$. To evaluate a $2n \times 1$ vector and a $2n \times 2n$ matrix (e.g., $-\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N$), we employ the $q$-norm $\| \cdot \|_q$ for $2 \leq q \leq \infty$. To describe parameters proximate to the true ones, we define closed balls with a radius $r \geq 0$: (i) for $\theta_0$, let $B(\theta_0,r) = \{ \theta : \| \theta - \theta_0 \| \leq r \}$ and (ii) for $\xi_{30}$, let $B_3(\xi_{30},r) = \{ \xi_n : \| \xi_n - \xi_{30} \| \leq r \}$.

**Step 1:** The initial step in deriving the asymptotic distribution of $\hat{\theta}_N$ involves obtaining Taylor approximations of $\hat{\xi}_{n,N}(\theta) - \xi_{30}$ and $\frac{1}{\sqrt{N}} \partial_{\theta} e_N^N(\theta, \xi_{n,N}(\theta))$ for a given $\theta$. To deduce the Taylor expansions of these elements, certain regularity conditions must be verified (refer to Lemma C.1 in the supplementary material). These conditions are the counterpart of Fernandez-Val and Weidner’s (2016) Assumption B.1.

The subsequent proposition characterizes the bounds of the Taylor approximation’s remainder terms when utilizing an $r_0$-consistent estimator $\hat{\theta}_N$ for $-\frac{1}{\sqrt{N}} \partial_{\theta} e_N^N(\theta, \xi_{n,N}(\theta))$ and $\hat{\xi}_{n,N}(\theta) - \xi_{30}$.

**Proposition C.2.** Let $q = 4 + \eta$ for some $\eta > 0$ and $0 \leq \eta < \frac{1}{2} - \frac{1}{q}$. Assume the results of Lemma C.1 in the supplement hold and suppose $\| B(\theta_0, r_0) \|_q = 0$, $\| \theta - \theta_0 \|_q \leq \frac{1}{q}$. For $\theta \in B(\theta_0, r_0)$ and $\xi_n \in B_3(\xi_{30}, r_0)$, we have the two results below:

(i) For a given $\theta \in B(\theta_0, r_0)$, the Taylor expansion of $-\frac{1}{\sqrt{N}} \partial_{\theta} e_N^N(\theta, \xi_{n,N}(\theta))$ can be represented by

\[
\frac{1}{\sqrt{N}} \partial_{\theta} e_N^N(\theta, \xi_{n,N}(\theta)) = -\Sigma_N \cdot \sqrt{N} (\theta - \theta_0) + U^{(0)}_N + U^{(1)}_N + R_N(\theta),
\]

where $\Sigma_N = E \left( -\frac{1}{\sqrt{N}} \partial_{\theta} e_N^N \right) - \frac{1}{\sqrt{N}} \left( E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right) \right)^{-1} \left( E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right) \right),$

\[
U^{(0)}_N = \frac{1}{\sqrt{N}} \partial_{\theta} e_N^N + E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right) E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N,
\]

\[
U^{(1)}_N = U^{(1,a)}_N + U^{(1,b)}_N \text{ with } U^{(1,a)}_N = U^{(1,a,1)}_N + U^{(1,a,2)}_N,
\]

\[
U^{(1,a,1)}_N = \frac{1}{\sqrt{N}} \partial_{\theta} e_N^N \underbrace{E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)}_{\in \Sigma_N} \cdot \left( E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right),
\]

\[
U^{(1,a,2)}_N = \left[ \underbrace{\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N} - E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right) \right] \cdot \left( E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)
\]

\[
\cdot \underbrace{E \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N}_{\in \Sigma_N}.
\]

(ii) For a given $\theta \in B(\theta_0, r_0)$, the Taylor expansion of $\xi_{n,N}(\theta)$ around $\xi_{30}$ is

\[
\xi_{n,N}(\theta) - \xi_{30} = \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right) \frac{1}{\sqrt{N}} \partial_{\theta} e_N^N \cdot (\theta - \theta_0)
\]

\[
+ \frac{1}{2} \left( \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \cdot \sum_{j=1}^n u_{N,j}^{(a)} \cdot \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \cdot \frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)
\]

\[
+ \frac{1}{2} \left( \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \cdot \sum_{i=1}^n u_{N,i}^{(b)} \cdot \left( -\frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right)^{-1} \cdot \frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N \right) + R_N(\theta),
\]

where $u_{N,j}^{(a)}$ is the $j$th element of $\overline{N}_N \frac{1}{\sqrt{N}} \partial_{\theta} e_N^N + \overline{N}_N \frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N$, $u_{N,i}^{(b)}$ denotes the $i$th element of $\overline{N}_N \frac{1}{\sqrt{N}} \partial_{\theta} e_N^N + \overline{N}_N \frac{1}{\sqrt{N}} \partial_{\xi_N} e_N^N$, $R_N(\theta)$ denotes the remainder term.

Note that $\| R_N(\theta) \|_q = o_p \left( \frac{1}{n^\eta} \right) + o_p \left( \frac{1}{n^{\frac{1}{2} \cdot (\| \theta - \theta_0 \|)} \right)$ for $\theta \in B(\theta_0, r_0)$. 

---

31 Since $\overline{N}_N$ is symmetric, note that $\overline{N}_N = \overline{N}_N'$. 

21
Step 2: Another condition for Proposition C.2 specifies that \( \| \hat{\xi}_{N,N}(\theta) - \xi_{ml} \|_q = O_p \left(n^{-\frac{1}{2}+\frac{1}{q}} \right) \), provided \( \| \theta - \theta_0 \| = O_p \left(n^{-\frac{1}{2}} \right) \). Lemmas C.2 and C.3 in the supplement demonstrate that \( \| \hat{\delta} - \delta_0 \| = O_p \left(n^{-\frac{1}{2}} \right) \) and \( \| \hat{\xi}_{N}(\hat{\delta}) - \xi_{ml} \| = O_p \left(n^{-\frac{1}{2}+\frac{1}{q}} \right) \), under the condition \( \Sigma = \lim_{n \to \infty} \Sigma_N \) with \( \Sigma > 0 \). These results are based on the strict concavity of the log-likelihood function. For this, we restrict the parameter space for \( \theta \) and \( \xi \), using Olsen’s (1978) reparameterization, represented by \( \Theta : (\theta, \xi) \to (\hat{\theta}_n^\ast, \hat{\xi}_n^\ast) \). Here, \( \theta_n = (\kappa', \delta', \beta', \xi_n', \sigma^2) \) and \( \theta_n^* = (\kappa', \delta', \beta', \xi_n', \sigma^2) / \sigma \) with \( \delta = \delta / \sigma \), \( \beta = \beta / \sigma \), and \( \xi_n' = \xi_n / \sigma \). The transformation function \( \Theta \) is bijective since

\[
\partial_{\theta_n} \Theta(\theta_n, \xi_n) = \begin{bmatrix}
\frac{f_{\text{dim}(\xi)}}{\sigma} & 0 \\
0 & \frac{-1}{\sigma^2} I_{L+2n} & 0 \\
-(2\sigma^3)^{-1} (\lambda, \gamma, \beta, \xi_n', \sigma^2) & -\frac{1}{\sigma^2}I_n
\end{bmatrix},
\]

is non-singular for finite \( \sigma \) values. The function \( \Theta \) is infinitely differentiable with respect to its arguments, and all its second-order partial derivatives are finite. By applying the results to the re-parameterized MLEs and utilizing the functional invariance property of the MLE, \( \hat{\theta}_{n,N} \) is unique under the parameter restriction stated in Theorem 3.1 (ii) (see discussions after Proposition C.2 in the supplement, pages 29–34).

Finally, \( U_{n}^{(1)} \) serves as the key component in the asymptotic distribution of \( \hat{\theta}_{n} \), whereas \( U_{n}^{(1)} \) characterizes its asymptotic bias. Subsequently, we will establish that \( U_{n}^{(0)} \overset{d}{\to} N(0, \Sigma) \), where \( \Sigma = \lim_{n \to \infty} \Sigma_N \), and \( U_{n}^{(1)} \to \hat{N} = \sigma_p(1) \) for a certain \( \hat{N} \).

Step 3: Proposition C.3 delineates the primary components of the asymptotic expansion of \( \hat{\theta}_n^\ast \): (i) \( U_{n}^{(0)} \) and (ii) \( U_{n}^{(1)} \).

Proposition C.3. (i) \( U_{n}^{(0)} \overset{d}{\to} N(0, \Sigma) \) as \( n \to \infty \); and (ii) \( U_{n}^{(1,a,1)} - (A_{1,N} + A_{2,N}) \overset{p}{\to} 0 \), \( U_{n}^{(1,a,2)} - (A_{3,N} + A_{4,N}) \overset{p}{\to} 0 \), and \( U_{n}^{(1,b)} - (A_{5,N} + A_{6,N}) \overset{p}{\to} 0 \) as \( n \to \infty \).

Step 4: Employing Propositions C.2 (ii) and C.3, we deduce

\[
\sqrt{n} (\hat{\theta}_n - \theta_0) = \Sigma \left( \frac{U_{n}^{(0)}}{U_{n}^{(1)} + O(n)} \right) + o_p(1) \overset{d}{\to} N \left( \Sigma A_{\infty}, \Sigma^{\prime} \right) \quad \text{as} \quad n \to \infty,
\]

where \( \Sigma_A = \lim_{n \to \infty} \Sigma_A \) with \( A_{\infty} = A_{1,N} + A_{2,N} + A_{3,N} + A_{4,N} + A_{5,N} + A_{6,N} \). Further insights on this can be found in the supplement’s Lemma C.4. Noting that \( \| \hat{\theta}_n - \theta_0 \| = O_p \left(n^{-\frac{1}{2}+\frac{1}{q}} \right) \) is achievable, it follows that \( \| \hat{\xi}_{N}(\hat{\theta}_n) \|_q = O_p \left(n^{-\frac{1}{2}+\frac{1}{q}} \right) \) and \( \| R_N(\hat{\theta}_n) \| = o_p(1) \).

Derivatives: Detailed forms of the pivotal derivative components are presented. These are essential for the computation of \( \Lambda_N \). Regarding the first-order derivatives, we have \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = \sum_{n=1}^{N} q_{n}^{\ast} \ast_i \langle \theta_n \rangle \). When evaluated at the true parameter values, we let \( q_{n}^{\ast} \ast_i \langle \theta_n \rangle = q_{n}^{\ast} \ast_i \langle \theta_n \rangle \) for all \( i, j = 1, \ldots, n \). Other quantities evaluated at true values are defined in a similar fashion.

Consider \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = \left( \partial_{\theta_n} \xi_{N}(\theta, \xi_n), \ldots, \partial_{\theta_n} \xi_{N}(\theta, \xi_n) \right) \). For notational convenience, we define \( \xi_n' = \xi_n / \sigma \) for each \( \omega \), and \( \xi_{n,j} = \xi_{n,j} / \sigma \). Observe that

\[
\begin{bmatrix}
\partial_{\theta_n} \xi_{N}(\theta, \xi_n) \\
\partial_{\theta_n} \xi_{N}(\theta, \xi_n)
\end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix}
q_{n,i}^{\ast} \ast_{i} \langle \theta_n \rangle \\
q_{n,i}^{\ast} \ast_{i} \langle \theta_n \rangle
\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}
\partial_{\theta_n} \xi_{N}(\theta, \xi_n) \\
\partial_{\theta_n} \xi_{N}(\theta, \xi_n)
\end{bmatrix} = \sum_{j=1}^{n} \begin{bmatrix}
q_{n,j}^{\ast} \ast_{j} \langle \theta_n \rangle \\
q_{n,j}^{\ast} \ast_{j} \langle \theta_n \rangle
\end{bmatrix}
\]

where \( q_{n,j}^{\ast} \ast_{j} \langle \theta_n \rangle = q_{n,j}^{\ast} \ast_{j} \langle \theta_n \rangle - \sum_{i=1}^{n} \langle \theta_n \rangle \ast_{i} \langle \theta_n \rangle = q_{n,j}^{\ast} \ast_{j} \langle \theta_n \rangle + \langle \theta_n \rangle \ast_{j} \langle \theta_n \rangle - \sum_{i=1}^{n} \langle \theta_n \rangle \ast_{i} \langle \theta_n \rangle \ast_{j} \langle \theta_n \rangle, \)

\[
q_{n,j}^{\ast} \ast_{j} \langle \theta_n \rangle = -1 \langle y_{N,j} = 0 \rangle \ast_{j} \langle \theta_n \rangle \ast_{j} \langle \theta_n \rangle + 1 \langle y_{N,j} > 0 \rangle \ast_{j} \langle \theta_n \rangle \ast_{j} \langle \theta_n \rangle \quad \text{for each} \quad (i, j).
\]

Consider the relevant components of the second-order derivatives. Define

\[
h_{n,j}(\theta) = -1 \langle y_{N,j} = 0 \rangle \ast_{j} \langle \theta_n \rangle \ast_{j} \langle \theta_n \rangle \quad \text{and} \quad h_{n,j}(\theta) = h_{n,j}(\theta) - \xi_n' \ast_{j} \langle \theta_n \rangle - \sum_{i=1}^{n} h_{n,i}(\theta) - \xi_n' \ast_{j} \langle \theta_n \rangle
\]

Using this, we have \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = \sum_{i=1}^{n} h_{n,i}(\theta) \) for \( j = 1, \ldots, n \); \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = -\mu \) for \( j \neq k \); \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = h_{n,j}(\theta) + \mu \) for \( i, j = 1, \ldots, n \); \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = \sum_{i=1}^{n} h_{n,i}(\theta) \) for \( i, j = 1, \ldots, n \); \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = -\mu \) for \( i \neq k \); \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = h_{n,j}(\theta) \) for \( i, j = 1, \ldots, n \); \( \partial_{\theta_n} \xi_{N}(\theta, \xi_n) = \sum_{i=1}^{n} h_{n,i}(\theta) \).

For the first part of \( h_{n,j}(\theta) \) and \( h_{n,j}(\theta) \),

\[
h_{n,j}(\theta) = -1 \langle y_{N,j} = 0 \rangle \ast_{j} \langle \theta_n \rangle \ast_{j} \langle \theta_n \rangle \quad \text{and} \quad h_{n,j}(\theta) = h_{n,j}(\theta) - \xi_n' \ast_{j} \langle \theta_n \rangle - \sum_{i=1}^{n} h_{n,i}(\theta) - \xi_n' \ast_{j} \langle \theta_n \rangle
\]
which is the part for $\kappa$. Note that $h_{N,i,j}^{\omega}(\theta_j) = \partial_{\omega_j} \ell_{N,i,j}(\theta_j) = \partial_{\omega_j}^{\epsilon} \ell_{N,i,j}(\theta_j) = h_{N,i,j}^{\epsilon}(\theta_j)$. Second, the part of $h_{N,i,j}^{\omega}(\theta_j)$ and $h_{N,i,j}^{\epsilon}(\theta_j)$ for $\omega$ is given as follows:

$$h_{N,i,j}^{\omega}(\theta_j) = -\mathbf{I}(Y_{N,i,j} > 0)\sigma^2 \frac{\Phi(\chi_{N,i,j}^{\omega}(o_{ij}))}{\Delta_{N,i,j}(\theta_j)} \left( e_{N,i,j}^{\omega}(o_{ij}) + \Phi(\chi_{N,i,j}^{\omega}(o_{ij})) \right) \partial_{\omega_j} X_{N,i,j}(o_{ij})$$

$$- \mathbf{I}(Y_{N,i,j} > 0)\sigma^2 \cdot \partial_{\omega_j} X_{N,i,j}(o_{ij}) \text{ if } o_{ij} \in \{ \lambda, \gamma, \rho, \beta_1, \ldots, \beta_L \};$$

$$h_{N,i,j}^{\omega}(\theta_j) = \mathbf{I}(Y_{N,i,j} > 0) \frac{1}{2\sigma^2} \left( \Phi(\chi_{N,i,j}^{\omega}(o_{ij})) \right) \left( e_{N,i,j}^{\omega}(o_{ij}) + \Phi(\chi_{N,i,j}^{\omega}(o_{ij})) \right) \left( e_{N,i,j}^{\omega}(o_{ij}) \right)^2 \Delta_{N,i,j}(\theta_j)$$

Note that $\partial_{\theta_k} \ell_{N}(\theta, \xi_n) = \sum_{i=1}^{n} h_{N,i,j}^{\omega}(\theta_j)$ for all $i = 1, \ldots, n$, i.e., $h_{N,i,j}^{\omega}(\theta_j) = h_{N,i,j}^{\omega}(\theta_j)$. For the part of $\omega$, if $o_{ij} \in \{ \lambda, \gamma, \rho, \beta_1, \ldots, \beta_L \}$,

$$\ell_{N,i,j}^{\omega}(\theta_j) = \mathbf{I}(Y_{N,i,j} > 0) \sigma^3 \left( \Phi(\chi_{N,i,j}^{\omega}(o_{ij})) \right) \left( e_{N,i,j}^{\omega}(o_{ij}) \right)^2 \Delta_{N,i,j}(\theta_j)$$

Then, $\partial_{\omega_j} \ell_{N}(\theta, \xi_n) = \sum_{i=1}^{n} \ell_{N,i,j}^{\omega}(\theta_j)$. For $i = 1, \ldots, n$, note that $\partial_{\omega_k} \ell_{N}(\theta, \xi_n) = \sum_{i=1}^{n} \ell_{N,i,j}^{\omega}(\theta_j)$ and $\partial_{\omega_k} \ell_{N}(\theta, \xi_n) = \sum_{i=1}^{n} \ell_{N,i,j}^{\omega}(\theta_j)$. For $\partial_{\omega_j} \ell_{N}(\theta, \xi_n)$ and $\partial_{\omega_j} \ell_{N}(\theta, \xi_n)$, the relevant term is

$$\ell_{N,i,j}(\theta_j) = \mathbf{I}(Y_{N,i,j} > 0) \sigma^3 \frac{\Phi(\chi_{N,i,j}^{\omega}(o_{ij}))}{\Delta_{N,i,j}(\theta_j)} \left( e_{N,i,j}^{\omega}(o_{ij}) + \Phi(\chi_{N,i,j}^{\omega}(o_{ij})) \right) \partial_{\omega_j} X_{N,i,j}(o_{ij})$$

$$- \mathbf{I}(Y_{N,i,j} > 0) \sigma^3 \cdot \partial_{\omega_j} X_{N,i,j}(o_{ij}) \text{ if } o_{ij} \in \{ \lambda, \gamma, \rho, \beta_1, \ldots, \beta_L \};$$

$$\ell_{N,i,j}(\theta_j) = \mathbf{I}(Y_{N,i,j} > 0) \frac{1}{2\sigma^4} \left( \Phi(\chi_{N,i,j}^{\omega}(o_{ij})) \right) \left( e_{N,i,j}^{\omega}(o_{ij}) \right)^3 \Delta_{N,i,j}(\theta_j)$$

Then, $\partial_{\omega_j} \ell_{N}(\theta, \xi_n) = \sum_{i=1}^{n} \ell_{N,i,j}^{\omega}(\theta_j)$. For $i = 1, \ldots, n$, note that $\partial_{\omega_k} \ell_{N}(\theta, \xi_n) = \sum_{i=1}^{n} \ell_{N,i,j}^{\omega}(\theta_j)$ and $\partial_{\omega_k} \ell_{N}(\theta, \xi_n) = \sum_{i=1}^{n} \ell_{N,i,j}^{\omega}(\theta_j)$.
Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2024.105790.

References


Lee, L., Yu, J., 2010. A spatial dynamic panel data model with both time and individual fixed effects. Economometric Theory 26, 564–597.


