RIETI Discussion Paper Series 23-E-016

Estimating Firm-level Production Functions with Spatial Dependence in Output, Input, and Productivity (Revised)

CHANG, Pao-Li
Singapore Management University

MAKIOKA, Ryo
RIETI

NG, Lin Bo
Singapore Management University

YANG, Zhenlin
Singapore Management University

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 23-E-016
First Draft: March 2023
Revised: February 2024

Estimating Firm-Level Production Functions with Spatial Dependence in Output, Input, and Productivity*

Pao-Li Chang[1], Ryo Makioka[2], Bo Lin Ng[1], Zhenlin Yang[1]
[1] Singapore Management University
[2] RIETI, Hokkaido University

Abstract

This paper proposes a three-stage efficient GMM estimation algorithm for estimating firm-level production functions given spatial dependence across firms due to supplier/customer relationships, sharing of input markets, or knowledge spillover. The procedure builds on Ackerberg, Caves and Frazer (2015) and Wooldridge (2009), but in addition, allows the productivity process to depend on the lagged output levels and lagged input usages of related firms, and potential spatially correlated productivity shocks across firms, where the set of related firms can differ across the three dimensions of spatial dependence. We establish the asymptotic properties of the proposed estimator, and conduct Monte Carlo simulations to validate these properties. The estimator is consistent under DGPs with or without spatial dependence, and with strong/weak or positive/negative spatial dependence. In contrast, the conventional estimators lead to biased estimates of the production function parameters when the underlying DGPs have spatial dependence structure, and the magnitudes of the bias increase with the strength of spatial dependence in the underlying DGPs. We apply the proposed estimation algorithm to a Japanese firm-to-firm dataset for the period 2009–2018. We find significant and positive spatial coefficients in the Japanese firm-level productivity process via all three channels proposed above.

Keywords: productivity estimation, spatial dependence, buyer-seller network, factor market pooling, knowledge spillover

JEL classification: C31; D24

*This study is conducted as a part of the Project "Comprehensive Research on Evidence Based Policy Making (EBPM)" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The draft of this paper was presented at the DP seminar of the Research Institute of Economy, Trade and Industry (RIETI). I would like to thank Daiji Kawaguchi (Univ. of Tokyo), Masayuki Morikawa (RIETI), and participants of the RIETI DP Seminar for their helpful comments. This study utilizes the micro data of the questionnaire information based on "the Basic Survey of Japanese Business Structure and Activities" which is conducted by the Ministry of Economy, Trade and Industry (METI), and TSR firmrelation data.

1 Introduction

Firm-level productivity (production function) estimation is critical to both positive and normative research, in inferring the characteristics of firm-level production activities and identifying the effect of policy/exogenous shocks (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg, Caves and Frazer, 2015; Wooldridge, 2009). Equally important, a large literature has documented/analyzed how firms interact with each other via the input-output linkages, factor markets, and knowledge spillovers (e.g., Ellison, Glaeser and Kerr, 2010; Helsley and Strange, 2014, 1990; Diamond and Simon, 1990; Helsley and Strange, 2002; Jaffe, Trajtenberg and Henderson, 1993; Audretsch and Belitski, 2020; Matray, 2021). This paper contributes to the literature by proposing methodologies for estimating firm-level productivity (production function), simultaneously taking into account potential spatial interactions across firms. In particular, a firm's productivity is allowed to depend on related firms' lagged outputs (e.g., via local input-output linkages), on related firms' lagged labor inputs (e.g., via sharing local labor pools), and on related firms' current productivity shocks (e.g., via knowledge spillovers with the boundary defined by the geographical area and/or by the network of suppliers-customers).

We develop a three-stage efficient GMM estimation algorithm, and show by theory the asymptotic properties of the proposed estimator and by Monte Carlo simulations the finite sample performance of the estimator. The procedure provides the estimates of the production function parameters (the labor and capital elasticities in value-added), the degree of autoregressive correlation in the productivity process, and the spatial parameters (characterizing the dependence of productivity on related firms' lagged outputs and lagged inputs respectively, and the strength of spatial error correlation of the productivity shocks). The Monte Carlo simulations demonstrate that the proposed estimator yields point estimates that are consistent for the true parameters both in the absence and in the presence of spatial effects. In other words, it returns statistically insignificant coefficient estimates of the spatial dependence parameters, when the underlying DGPs are free of such structures, and consistent estimates of the spatial dependence parameters when the underlying DGPs are characterized with such structures (via the lagged output, the lagged labor input, or the productivity shock channel). This finding holds for DGPs with strong or weak spatial dependence, and DGPs with positive or negative spatial dependence. The proposed estimation algorithm also generates standard error estimates of the parameters that are consistent with the Monte Carlo simulated standard deviations, and with a convergence rate (when the sample size changes) in line with the theory. In contrast, the conventional productivity estimators (which ignore potential spatial interactions across firms) lead to biased estimates of the production function parameters when the underlying DGPs exhibit spatial dependence structure (and thus, for which the conventional estimator is misspecified). The conventional estimates are upward (downward) biased when the underlying DGPs have positive (negative) spatial dependence structure, and the extents of the bias worsen when the underlying DGPs' spatial dependence strengthens.

We apply the developed methodology and estimation algorithm to the Japanese BSJBSA-TSR linked dataset for the period 2009–2018. The dataset combines the firm-level financial statement information from the Basic Survey of Japanese Business Structure and Activities (BSJBSA), and the firm-to-firm buyer-supplier relationship from Tokyo Shoko Research (TSR). The estimation sample covers 12,525 firms per year (both publicly listed and unlisted firms in Japan of medium/large sizes, across 267 commuting zones, and 14 industries), and in particular, provides information on each firm's most important domestic suppliers and customers (up to 24 connections, respectively). We find significant and positive spatial coefficients in the Japanese firm-level productivity process via all three proposed channels. In particular, a 1% increase in the average sales of a firm's customers/suppliers in the previous period in the same commuting zone helps improve the firm's current productivity by 0.005%. A larger local labor market also enhances a firm's productivity: specifically, a 1% increase in the average labor inputs in the previous period by firms located in the same commuting zone raises a firm's current productivity by 0.05%. There is also evidence of contemporary knowledge spillovers among firms located in the same commuting zone (with a positive and significant spatial error correlation coefficient of 0.38) and/or with buyer-supplier relationships. In sum, the proposed estimator suggests that spatial interactions across firms play a significant role in determining the Japanese firm-level productivity both statistically and economically.

The prior literatures on estimating firm-level production functions typically ignore potential spatial dependence across firms. The firm-level production functions are often taken to be independent and estimated, before the estimated productivities are used to analyze potential linkages across firms. For example, Javorcik (2004) and Keller and Yeaple (2009) examine how foreign direct investments affect the productivity of domestic Lithuanian and US manufacturing firms, respectively; Alfaro-Ureña et al. (2021) study how domestic Costa Rican firms' productivities change when firms start to supply to MNC. The recent work by Iyoha (2022) highlights the need to estimate firm productivities in a modified framework taking into account the presence of productivity spillovers. Her work, however, models the interdependence across firms "in reduced form" in terms of their productivities, and not directly in terms of outputs, inputs, or the productivity shocks of related firms. This leads to a rather difficult setup for estimations, and for establishing the asymptotic properties (and variance-covariance) of the estimator. Her framework, albeit with a dynamic-spatial structure for productivities, has relied heavily on static-spatial estimation techniques.

In this paper, we propose methodologies that model productivity dependence across firms structurally where the spatial effects operate via potentially lagged outputs, lagged inputs, and current productivity shocks of related firms, motivated by the mechanisms highlighted by the literatures. We draw on the approaches proposed in the productivity estimation literature (e.g., Wooldridge, 2009; Ackerberg, Caves and Frazer, 2015), and the spatial econometrics literature (e.g., Kelejian and Prucha, 1998, 1999; Kapoor, Kelejian and Prucha, 2007; Lee and Yu, 2014; Elhorst, 2014). The resulting three-stage efficient GMM estimator has standard asymptotic properties, with variance-covariance estimators that take into account the spatial interactions across firms in each of the three dimensions proposed. The sets of instruments suitable for each stage are also straightforward extensions of those suggested by each of these two individual literatures. As discussed above, our proposed estimator is shown to be consistent under DGPs with or without spatial dependence. In contrast, the conventional estimators are biased when the true DGPs are indeed characterized by spatial dependence. These findings imply that analyzing spatial interactions across firms based on the productivities estimated by the conventional estimators will lead to biased inferences. Instead, the proposed estimator in this paper offers a framework to simultaneously estimate firm production functions and spatial interactions across firms in one unified setup.

The rest of the paper is organized as follows. We set up the model in Section 2. In Section 3, we develop the estimation algorithms and establish the asymptotic properties of the proposed estimator. Section 4 introduces the Japanese firm-level and firm-to-firm datasets. Section 5 conducts Monte Carlo simulations to evaluate the performance of the proposed estimator in comparison with the conventional estimator. In Section 6, we apply the proposed methodology empirically to the Japanese dataset, and Section 7 concludes.

2 Model

Consider the following production function, where a firm's value-added depends on its primary factor inputs and productivity, while its productivity depends on its lagged productivity, and related firms' lagged outputs and input choices. In addition, the innovations to the productivity of related firms are allowed to be spatially correlated. The set of related firms that a firm's productivity depends upon can be defined by supplier-customer relationship, by ownership structure, by physical location, by industry of sales, or by combinations of them, and can differ across the three channels of spatial dependence, as the context of the study may deem appropriate. In particular, define:

$$va_{it} = \alpha_0 + \alpha_l l_{it} + \alpha_k k_{it} + \omega_{it} + \xi_{it}, \tag{1}$$

$$\omega_{it} = f(\omega_{i,t-1}) + \lambda \sum_{j \in \mathcal{N}^y_{i,t-1}} \mathtt{w}^y_{ij,t-1} y_{j,t-1} + \sum_{j \in \mathcal{N}^\Omega_{i,t-1}} \mathtt{w}^\Omega_{ij,t-1} \boldsymbol{\Omega}_{j,t-1} \boldsymbol{\beta}_{\bar{\Omega}} + \mathbf{x}_{i,t-1} \boldsymbol{\beta}_x + u_{it}, \tag{2}$$

$$u_{it} = \mu \sum_{j \in \mathcal{N}^u_{it}} \mathtt{w}^u_{ij,t} u_{jt} + v_{it}, \tag{3}$$

$$i, j = 1, 2, 3, ..., N \text{ and } t = 2, 3, ..., T.$$

where va_{it}, y_{it}, l_{it}, k_{it} and ω_{it} denote the log of: value-added, gross output, labor input, capital stock at the beginning of period, and productivity, respectively, of firm i in period t, with ξ_{it} denoting the value-added shock to firm i in period t. A firm's productivity ω_{it} is assumed to be dependent on its lagged productivity $\omega_{i,t-1}$ via an unknown function $f(\cdot)$ as in Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves and Frazer (2015), and Wooldridge (2009). One might also consider a firm's current productivity to depend on the lagged characteristics $\mathbf{x}_{i,t-1}$ of the firm (such as its lagged exporting status and R&D expenditure) à la De Loecker (2013), Doraszelski and Jaumandreu (2013), and Braguinsky, Ohyama, Okazaki and Syverson (2015).

We generalize the setup of the literature and allow spatial dependence across firms in their productivities. In particular, a firm i's current productivity could depend on the lagged output $y_{j,t-1}$ of its related firms j in the set $\mathcal{N}^y_{i,t-1}$, and the lagged inputs $\boldsymbol{\Omega}_{j,t-1} \equiv \{l_{j,t-1}, k_{j,t-1}, \mathbf{m}_{j,t-1}\}$ of a possibly different set $\mathcal{N}^\Omega_{i,t-1}$ of related firms, where $\mathbf{m}_{j,t-1}$ denotes the log of: $1 \times M$ vector of intermediate inputs of firm j in period $t-1$. Furthermore, the innovation u_{it} to the productivity of firm i in period t is allowed to be spatially correlated with those of related firms in the set $\mathcal{N}^u_{it}$ contemporarily. The weight assigned to each of the related firms in the set $\mathcal{N}^y_{i,t-1}$ is specified by $\mathtt{w}^y_{ij,t-1}$, and correspondingly those for firms in $\mathcal{N}^\Omega_{i,t-1}$ and $\mathcal{N}^u_{it}$ are specified by $\mathtt{w}^\Omega_{ij,t-1}$ and $\mathtt{w}^u_{ij,t}$, respectively.

2.1 Assumptions

We adopt standard assumptions similar to those in the productivity estimation literature, but extend them to accommodate the setup with spatial dependence across firms in productivity as introduced in Equations (1)–(3).

Assumption 1. $E(\xi_{it} | l_{it}, k_{it}, \boldsymbol{m}_{it}) = 0.$

Assumption 2. $E(\xi_{it} | l_{jt}, k_{jt}, \boldsymbol{m}_{jt}, l_{j,t-1}, k_{j,t-1}, \boldsymbol{m}_{j,t-1}, \ldots, l_{j1}, k_{j1}, \boldsymbol{m}_{j1}) = 0.$

Assumption 3. $E(u_{it} | k_{it}, l_{j,t-1}, k_{j,t-1}, \boldsymbol{m}_{j,t-1}, \boldsymbol{x}_{j,t-1}, \ldots, l_{j1}, k_{j1}, \boldsymbol{m}_{j1}, \boldsymbol{x}_{j1}) = 0.$

Assumption 4. *The residuals, ξ_{it}, are assumed to be i.i.d. across both i and t, and have finite fourth moments:*

$$\xi_{it} \overset{iid}{\sim} (0, \sigma_\xi^2),\ E|\xi_{it}^{4+\eta}| < \infty,\ \textit{for some } \eta > 0.$$

Assumption 5. *The productivity innovations, u_{it}, are spatially correlated as specified in Equation (3), with residual v_{it} assumed to be i.i.d. across both i and t, and have finite fourth moments:*

$$v_{it} \overset{iid}{\sim} (0, \sigma_v^2),\ E|v_{it}^{4+\eta}| < \infty,\ \textit{for some } \eta > 0.$$

Assumption 6. *The residuals, ξ_{it} and v_{it}, are uncorrelated.*

Assumption 7. *$(\boldsymbol{I}_N - \mu\, \boldsymbol{W}_t^u)$ are non-singular for $t = 1, 2, \ldots, T$, with $\mu \in (\frac{1}{\lambda_{\min,t}}, \frac{1}{\lambda_{\max,t}})$, where $\boldsymbol{I}_N$ is the identity matrix of size N, $\boldsymbol{W}_t^u \equiv \{w_{ij,t}^u\}$, and $\lambda_{\min,t}$ and $\lambda_{\max,t}$ are the smallest and largest eigenvalues of $\boldsymbol{W}_t^u$.*

Assumption 8. *The row and column sums of the matrices, $\boldsymbol{W}_{t-1}^y$, $\boldsymbol{W}_{t-1}^\Omega$, $\boldsymbol{W}_t^u$ and $(\boldsymbol{I}_N - \mu\, \boldsymbol{W}_t^u)$ are uniformly bounded in absolute value for $t = 2, 3, \ldots, T$, as N approaches infinity, where $\boldsymbol{W}_{t-1}^y \equiv \{w_{ij,t-1}^y\}$ and $\boldsymbol{W}_{t-1}^\Omega \equiv \{w_{ij,t-1}^\Omega\}$.*

Assumption 9. *The regressor matrices $\{\boldsymbol{\Omega}_t, y_{t-1}, \boldsymbol{\Omega}_{t-1}, \boldsymbol{x}_{t-1}\}$ have full column rank, and the elements are uniformly bounded in absolute value for $t = 2, 3, \ldots, T$.*

Assumption 1 is the standard assumption made in the literature for firm-level productivity estimations. Assumption 2 requires that the residuals ξ_{it} in the value-added equation (1) are conditionally mean independent of current and past input usages of the firm itself, *and* also those of the other firms. This is not as stringent an assumption as it might appear, because the productivity term ω_{it} in Equation (1) has absorbed potential spatial dependence across firms to the extent modelled by Equation (2). Assumption 3 basically states that the innovation u_{it} to productivity is conditionally mean independent of the state variable (capital), as well as the past input choices and characteristics of all the other firms. Together, the first three assumptions will help identify the set of moment conditions and instruments for estimating the parameters in Equations (1) and (2). Assumptions 4–6 are made to develop the variance-covariance estimator of the parameters. In particular, the finite fourth moment condition for v_{it} is required for the estimation of the spatial parameter μ in Equation (3). Assumptions 7–9 are adopted from Kelejian and Prucha (1999), Kapoor, Kelejian and Prucha (2007), and Elhorst (2014) to ensure that the spatial parameter estimates exist. Note that we will construct the connectivity matrices such that they are row-normalized (with zeros in the diagonal by construction). They will hence satisfy Assumption 8.

3 Estimation Algorithms

In this section, we propose a three-stage estimation procedure based on Generalized Method of Moments (GMM) to obtain consistent estimates of the parameters in Equations (1)–(3).

3.1 Moment Conditions

Given Assumptions 1–5, the following moment conditions hold with respect to the error terms in Equation (1) and Equation (2):

$$\begin{aligned} E(\xi_t | l_t, k_t, \mathbf{m}_t, \mathbf{\Omega}_{t-1}) &= \mathbf{0}, & (4) \\ E(\xi_t + u_t | k_t, \mathbf{\Omega}_{t-1}, \mathbf{W}^y_{t-1} y_{t-1}, \mathbf{W}^\Omega_{t-1} \mathbf{\Omega}_{t-1}, \mathbf{x}_{t-1}) &= \mathbf{0}, & (5) \end{aligned}$$

where $\xi_t \equiv (\xi_{1t}, ..., \xi_{Nt})'$ and $u_t \equiv (u_{1t}, ..., u_{Nt})'$ denote the $N \times 1$ vector of the residual terms from Equation (1) and Equation (2), respectively, across firms in period t; $l_t \equiv (l_{1t}, ..., l_{Nt})'$ denotes the $N \times 1$ vector of labor inputs across firms in period t; k_t and $\mathbf{m}_t$ are similarly defined; $\mathbf{\Omega}_{t-1} \equiv [l_{t-1}\, k_{t-1}\, \mathbf{m}_{t-1}]$; $y_{t-1} \equiv (y_{1,t-1}, ..., y_{N,t-1})'$ denotes the $N \times 1$ vector of gross outputs across firms in period $t-1$; $\mathbf{x}_{t-1}$ is similarly defined. The matrices $\mathbf{W}^y_{t-1} \equiv \{\mathtt{w}^y_{ij,t-1}\}$ and $\mathbf{W}^\Omega_{t-1} \equiv \{\mathtt{w}^\Omega_{ij,t-1}\}$ are $N \times N$ connectivity matrices in period $t-1$ that specify the dependence of firm i's productivity in period t on related firms j's lagged outputs and lagged inputs, respectively. Note that the conditional mean is defined element (firm) wise in each period t.

Furthermore, by Kelejian and Prucha (1999) and Kapoor, Kelejian and Prucha (2007), the following three moment conditions hold with respect to the error term in Equation (3):

$$E \begin{bmatrix} \frac{1}{N} v_t' v_t \\ \frac{1}{N} v_t' \mathbf{W}^{u\prime}_t \mathbf{W}^u_t v_t \\ \frac{1}{N} v_t' \mathbf{W}^u_t v_t \end{bmatrix} = \begin{bmatrix} \sigma_v^2 \\ \frac{\sigma_v^2}{N} \mathrm{tr}(\mathbf{W}^{u\prime}_t \mathbf{W}^u_t) \\ 0 \end{bmatrix}. \quad (6)$$

where $v_t \equiv (v_{1t}, ..., v_{Nt})'$ denotes the $N \times 1$ vector of the residual term from Equation (3) across firms in period t.

3.2 Estimation Strategy

3.2.1 Stage 1

Following the productivity estimation literature (e.g., Levinsohn and Petrin, 2003; Wooldridge, 2009), the productivity ω_{it} is assumed to be observable to the firm (or managers of the firm), but not to the econometrician. However, since a firm would in theory choose the optimal level of intermediate input $\mathbf{m}_{it}$ to maximize profits, given its initial capital stock k_{it}, labor force l_{it}, and realized productivity level ω_{it}, the econometrician could invert the relationship to infer a firm's productivity given its initial capital stock and observed input choices:

$$\omega_{it} = g(l_{it}, k_{it}, \mathbf{m}_{it}), \tag{7}$$

where $g(\cdot, \cdot, \cdot)$ is some unknown general function in the observed input levels. Equation (7), together with Equation (1), imply the following reduced-form value-added function:

$$\begin{aligned} va_t &= \alpha_0 \iota_N + \alpha_l l_t + \alpha_k k_t + \omega_t + \xi_t \\ &= \alpha_0 \iota_N + \alpha_l l_t + \alpha_k k_t + \mathbf{g}(l_t, k_t, \mathbf{m}_t) + \xi_t \\ &\equiv \mathbf{h}(l_t, k_t, \mathbf{m}_t) + \xi_t, \end{aligned} \tag{8}$$

where $va_t \equiv (va_{1t}, ..., va_{Nt})'$ denotes the $N \times 1$ vector of value-added across firms in period t; and ι_N is a $N \times 1$ vector of one's. The shorthand $\mathbf{g}(l_t, k_t, \mathbf{m}_t)$ is a $N \times 1$ column vector with $g(l_{it}, k_{it}, \mathbf{m}_{it})$ as its i-th entry; similarly, $\mathbf{h}(l_t, k_t, \mathbf{m}_t)$ is a $N \times 1$ column vector with $h(l_{it}, k_{it}, \mathbf{m}_{it})$ as its i-th entry, where $h(l_{it}, k_{it}, \mathbf{m}_{it}) \equiv \alpha_0 + \alpha_l l_{it} + \alpha_k k_{it} + g(l_{it}, k_{it}, \mathbf{m}_{it})$. As in Ackerberg, Caves and Frazer (2015) and Wooldridge (2009), one could approximate $h(\cdot, \cdot, \cdot)$ in Equation (8) by a n-degree polynomial that contain at least l_{it}, k_{it} and $\mathbf{m}_{it}$. For example, in the case where $\mathbf{m}_{it}$ contains only one type of intermediate input and is hence a scalar, $h(l_{it}, k_{it}, m_{it})$ can be approximated by $\sum_{p,q,r} (l_{it}^p k_{it}^q m_{it}^r)\ \delta_{p,q,r}$, with nonnegative integers p, q and r such that $p + q + r \leq n$. That is:

$$h(l_{it}, k_{it}, \mathbf{m}_{it}) = \alpha_0 + c(l_{it}, k_{it}, \mathbf{m}_{it})\boldsymbol{\delta}, \tag{9}$$

where $c(l_{it}, k_{it}, \mathbf{m}_{it})$ is a $1 \times Q$ vector of functions in $(l_{it}, k_{it}, \mathbf{m}_{it})$ and $\boldsymbol{\delta}$ a $Q \times 1$ vector of parameters. For example, for a 2nd-order polynomial h function ($n = 2$), $c(l_{it}, k_{it}, \mathbf{m}_{it}) = [l_{it}, k_{it}, m_{it}, l_{it}^2, l_{it}k_{it}, l_{it}m_{it}, k_{it}^2, k_{it}m_{it}, m_{it}^2]$.

Given the moment condition (4), Equation (8) given Equation (9) can be estimated using the following set of instrument variables (IVs) for period t:

$$\mathcal{Z}_{t,I} = (\iota_N, l_t, k_t, \mathbf{m}_t, \mathbf{\Omega}_{t-1}). \tag{10}$$

Note that since $g(\cdot, \cdot, \cdot)$ is allowed to be a general function (including linearity in the arguments as a special case), the slope coefficients (α_l, α_k) on the inputs are not identified from Equation (8), as highlighted by Ackerberg, Caves and Frazer (2015). However, it enables an estimate $\hat{h}(l_{it}, k_{it}, \mathbf{m}_{it})$ of $h(l_{it}, k_{it}, \mathbf{m}_{it})$. In turn, the slope coefficients of the production function can be identified in a later stage, along with the other parameters, as laid out in the next section.

The set of IVs listed in (10)—and in the moment condition (4)—includes the input variables only up to one lag, and hence corresponds to weaker conditions than stated in Assumption 2. One could potentially enlarge the set and include longer lags of the input variables in the conditioning set, given Assumption 2

3.2.2 Stage 2

Next, given the productivity process's dynamic and spatial dependence structure specified in Equation (2), and Equation (7), we can also write the value-added function in the following alternative reduced form:

$$
\begin{aligned}
va_t &= \alpha_0 \iota_N + \alpha_l l_t + \alpha_k k_t + \omega_t + \xi_t \\
&= \alpha_0 \iota_N + \alpha_l l_t + \alpha_k k_t + \mathbf{f}[\mathbf{g}(l_{t-1}, k_{t-1}, \mathbf{m}_{t-1})] \\
&\quad + \lambda \mathbf{W}^y_{t-1} y_{t-1} + \mathbf{W}^\Omega_{t-1} \boldsymbol{\Omega}_{t-1} \boldsymbol{\beta}_{\bar{\Omega}} + \mathbf{x}_{t-1} \boldsymbol{\beta}_x + u_t + \xi_t, \qquad (11)
\end{aligned}
$$

where the shorthand $\mathbf{f}[\mathbf{g}(l_{t-1}, k_{t-1}, \mathbf{m}_{t-1})]$ is a $N \times 1$ column vector with $f[g(l_{i,t-1}, k_{i,t-1}, \mathbf{m}_{i,t-1})]$ as its i-th entry. Recall that the matrices $\mathbf{W}^y_{t-1} \equiv \{\mathtt{w}^y_{ij,t-1}\}$ and $\mathbf{W}^\Omega_{t-1} \equiv \{\mathtt{w}^\Omega_{ij,t-1}\}$ are $N \times N$ connectivity matrices in period $t-1$ that specify the dependence of firm i's productivity in period t on related firms j's lagged outputs and lagged inputs, respectively. In deriving Equation (11), we have used Equation (2) to replace ω_t and Equation (7) to replace $\omega_{i,t-1}$ in the $f(\cdot)$ function such that $f(\omega_{i,t-1}) = f[g(l_{i,t-1}, k_{i,t-1}, \mathbf{m}_{i,t-1})]$. As suggested by Wooldridge (2009), one could use a G-th degree polynomial to approximate $f(\cdot)$ such that:

$$
f(\nu) = \rho_1 \nu + \rho_2 \nu^2 + \ldots + \rho_G \nu^G. \qquad (12)
$$

Given the moment condition in (5), Equation (11) can be estimated using the following set of IVs for period t:

$$
\mathcal{Z}_{t,II} = (\iota_N, k_t, \boldsymbol{\Omega}_{t-1}, \mathbf{W}^y_{t-1} y_{t-1}, \mathbf{W}^\Omega_{t-1} \boldsymbol{\Omega}_{t-1}, \mathbf{x}_{t-1}), \qquad (13)
$$

with up to one lag of the variables (or a longer past history of the variables). Additional spatio-temporal lags of explanatory variables, such as $(\mathbf{W}^y_{t-1})^2 y_{t-1}$, $(\mathbf{W}^y_{t-1})^3 y_{t-1}$, $(\mathbf{W}^\Omega_{t-1})^2 \boldsymbol{\Omega}_{t-1}$ and $(\mathbf{W}^\Omega_{t-1})^3 \boldsymbol{\Omega}_{t-1}$, may also be added to the set of IVs to help identify the spatial coefficients.

While Ackerberg, Caves and Frazer (2015) propose to estimate Equations (8) and (11)—without the spatial structure—sequentially, by plugging in estimates from Equation (8) into Equation (11), we adopt the approach proposed by Wooldridge (2009) and estimate them jointly, as it leads to more efficient estimators. In particular, denote the parameters of the system by $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\delta}', \alpha_l, \alpha_k, \lambda, \boldsymbol{\beta}'_{\bar{\Omega}}, \boldsymbol{\beta}'_x, \rho_1, ..., \rho_G)'$. The residuals from Equations (8) and (11) given the parameters are, respectively:

$$
\begin{aligned}
r_{t,I}(\boldsymbol{\theta}) &= va_t - \alpha_0 \iota_N - \mathbf{c}_t \boldsymbol{\delta}, \qquad (14) \\
r_{t,II}(\boldsymbol{\theta}) &= va_t - \alpha_0 \iota_N - \alpha_l l_t - \alpha_k k_t - \mathbf{f}[\mathbf{c}_{t-1} \boldsymbol{\delta} - \alpha_l l_{t-1} - \alpha_k k_{t-1}] \\
&\quad - \lambda \mathbf{W}^y_{t-1} y_{t-1} - \mathbf{W}^\Omega_{t-1} \boldsymbol{\Omega}_{t-1} \boldsymbol{\beta}_{\bar{\Omega}} - \mathbf{x}_{t-1} \boldsymbol{\beta}_x, \qquad (15)
\end{aligned}
$$

where recall that $g(l_{i,t-1}, k_{i,t-1}, \mathbf{m}_{i,t-1}) = h(l_{i,t-1}, k_{i,t-1}, \mathbf{m}_{i,t-1}) - \alpha_0 - \alpha_l l_{i,t-1} - \alpha_k k_{i,t-1} = c(l_{i,t-1}, k_{i,t-1}, \mathbf{m}_{i,t-1}) \boldsymbol{\delta} - \alpha_l l_{i,t-1} - \alpha_k k_{i,t-1}$, given Equations (8) and (9). The shorthand $\mathbf{c}_t$ is a $N \times Q$

matrix with $c(l_{it}, k_{it}, \mathbf{m}_{it})$ as its i-th row entry. The moment conditions in (4) and (5) imply that:

$$E[\mathcal{Z}'_{it}\, r_{it}(\boldsymbol{\theta})] \equiv E\left[\begin{pmatrix} \mathcal{Z}'_{it,I} & \mathbf{0} \\ \mathbf{0} & \mathcal{Z}'_{it,II} \end{pmatrix} \begin{pmatrix} r_{it,I}(\boldsymbol{\theta}) \\ r_{it,II}(\boldsymbol{\theta}) \end{pmatrix} \right] = \mathbf{0}, \tag{16}$$

where $\mathcal{Z}_{it,I}$, $\mathcal{Z}_{it,II}$, $r_{it,I}(\boldsymbol{\theta})$, and $r_{it,II}(\boldsymbol{\theta})$ are the i-th row entry of $\mathcal{Z}_{t,I}$, $\mathcal{Z}_{t,II}$, $r_{t,I}(\boldsymbol{\theta})$, and $r_{t,II}(\boldsymbol{\theta})$, respectively.

3.2.3 Stage 3

We estimate the spatial error structure in Equation (3) based on the GMM approach of Kelejian and Prucha (1999) and Kapoor, Kelejian and Prucha (2007). Specifically, given the parameter estimates $\hat{\boldsymbol{\theta}}$ from the previous stages, we impute estimates of the productivity innovation term, $\hat{u}_t$, by taking the difference between (15) and (14), since the residuals from the second stage is $\widehat{\xi_t + u_t}$ and the residuals from the first stage is $\widehat{\xi}_t$:

$$\begin{aligned} \hat{u}_t \;\; \equiv \;\; & \mathbf{c}_t\hat{\boldsymbol{\delta}} - \hat{\alpha}_l l_t - \hat{\alpha}_k k_t \\ & -\mathbf{f}[\mathbf{c}_{t-1}\hat{\boldsymbol{\delta}} - \hat{\alpha}_l l_{t-1} - \hat{\alpha}_k k_{t-1}] \\ & -\hat{\lambda}\mathbf{W}^y_{t-1} y_{t-1} - \mathbf{W}^{\Omega}_{t-1}\boldsymbol{\Omega}_{t-1}\hat{\boldsymbol{\beta}}_{\bar{\Omega}} - \mathbf{x}_{t-1}\hat{\boldsymbol{\beta}}_x, \end{aligned} \tag{17}$$

and use the moment conditions in (6) to estimate μ and σ^2_v jointly by GMM. Note that if we define $\overline{u}_t \equiv \mathbf{W}^u_t u_t$, $\overline{v}_t \equiv \mathbf{W}^u_t v_t$, and $\overline{\overline{u}}_t = (\mathbf{W}^u_t)^2 u_t$, it follows that $v_t = u_t - \mu\overline{u}_t$ and $\overline{v}_t = \overline{u}_t - \mu\overline{\overline{u}}_t$. By replacing v_t in the moment condition (6) with $u_t - \mu\overline{u}_t$ rewrites the three moment conditions in terms of u_t and $\mu\overline{u}_t$. This simplifies the computation greatly by avoiding taking the inverse of the $N \times N$ matrix $(\mathbf{I}_N - \mu\mathbf{W}^t_u)$ to compute $v_t = (\mathbf{I}_N - \mu\mathbf{W}^u_t)^{-1} u_t$ (Kelejian and Prucha, 1999), which is particular meaningful as our data have a large N. By replacing v_t in (6) with u_t and $\mu\overline{u}_t$, we can follow similar steps as in Kelejian and Prucha (1999) and Kapoor, Kelejian and Prucha (2007) to derive Equation (18) below:

$$\boldsymbol{\gamma}_t = \boldsymbol{\Gamma}_t \begin{pmatrix} \mu \\ \mu^2 \\ \sigma^2_v \end{pmatrix}, \tag{18}$$

where

$$\boldsymbol{\gamma}_t \;\; = \;\; \frac{1}{N} \begin{pmatrix} E(u'_t u_t) \\ E(\overline{u}'_t \overline{u}_t) \\ E(u'_t \overline{u}_t) \end{pmatrix}, \tag{19}$$

$$\boldsymbol{\Gamma}_t \;\; = \;\; \frac{1}{N} \begin{pmatrix} 2E(u'_t\overline{u}_t) & -E(\overline{u}'_t\overline{u}_t) & N \\ 2E(\overline{\overline{u}}'_t\overline{u}_t) & -E(\overline{\overline{u}}'_t\overline{\overline{u}}_t) & \mathrm{tr}(\mathbf{W}^{u\prime}_t\mathbf{W}^u_t) \\ E(u'_t\overline{\overline{u}}_t + \overline{u}'_t\overline{u}_t) & -E(\overline{u}'_t\overline{\overline{u}}_t) & 0 \end{pmatrix}. \tag{20}$$

Use the estimates of the productivity innovation term from Equation (17), $\hat{u}_t$, to construct the sample counterparts of the $\boldsymbol{\gamma}_t$ vector and the $\boldsymbol{\Gamma}_t$ matrix:[1]

$$\varsigma_t \equiv \frac{1}{N}\begin{pmatrix} \hat{u}_t{}'\hat{u}_t \\ \hat{u}_t{}'\mathbf{W}_t^{u\prime}\mathbf{W}_t^u\hat{u}_t \\ \hat{u}_t{}'\mathbf{W}_t^u\hat{u}_t \end{pmatrix}, \tag{21}$$

$$\boldsymbol{F}_t \equiv \frac{1}{N}\begin{pmatrix} 2\hat{u}_t{}'\mathbf{W}_t^u\hat{u}_t & -\hat{u}_t{}'\mathbf{W}_t^{u\prime}\mathbf{W}_t^u\hat{u}_t & N \\ 2\hat{u}_t{}'\mathbf{W}_t^{u\prime}\mathbf{W}_t^{u\prime}\mathbf{W}_t^u\hat{u}_t & -\hat{u}_t{}'\mathbf{W}_t^{u\prime}\mathbf{W}_t^{u\prime}\mathbf{W}_t^u\mathbf{W}_t^u\hat{u}_t & \mathrm{tr}(\mathbf{W}_t^{u\prime}\mathbf{W}_t^u) \\ \hat{u}_t'\mathbf{W}_t^u\mathbf{W}_t^u\hat{u}_t + \hat{u}_t'\mathbf{W}_t^{u\prime}\mathbf{W}_t^u\hat{u}_t & -\hat{u}_t'\mathbf{W}_t^{u\prime}\mathbf{W}_t^u\mathbf{W}_t^u\hat{u}_t & 0 \end{pmatrix}, \tag{22}$$

and form the sample counterpart of the condition in Equation (18):

$$\varsigma_t = \boldsymbol{F}_t \begin{pmatrix} \mu \\ \mu^2 \\ \sigma_v^2 \end{pmatrix} + \epsilon_t, \tag{23}$$

where ϵ_t is a 3×1 vector of residuals. We can then estimate μ and σ_v^2 by the transformed moment condition $E(\epsilon_t) = \mathbf{0}$. Specifically,

$$\epsilon_t(\mu, \sigma_v^2) = \frac{1}{N}\begin{pmatrix} \hat{u}_t{}'(\mathbf{I}_N - 2\mu\mathbf{W}_t^u + \mu^2\mathbf{W}_t^{u\prime}\mathbf{W}_t^u)\hat{u}_t - \sigma_v^2 N \\ \hat{u}_t{}'\mathbf{W}_t^{u\prime}(\mathbf{I}_N - 2\mu\mathbf{W}_t^u + \mu^2\mathbf{W}_t^{u\prime}\mathbf{W}_t^u)\mathbf{W}_t^u\hat{u}_t - \sigma_v^2\mathrm{tr}(\mathbf{W}_t^{u\prime}\mathbf{W}_t^u) \\ \hat{u}_t{}'\left(\mathbf{I}_N - \mu(\mathbf{W}_t^u + \mathbf{W}_t^{u\prime}) + \mu^2\mathbf{W}_t^{u\prime}\mathbf{W}_t^u\right)\mathbf{W}_t^u\hat{u}_t \end{pmatrix}. \tag{24}$$

The algorithm above provides a set of estimates consistent for $\boldsymbol{\theta}$, μ and σ_v^2. We can improve the efficiency of the estimators by deriving the weighting matrix for the GMM estimator, and repeat the procedure until the parameter estimates converge. Section 3.3 characterizes the algorithm to obtain the efficient GMM estimator.

3.3 Efficient GMM Estimator

This section itemizes the steps to implement the proposed estimation strategy and obtain the efficient GMM estimator of $\boldsymbol{\theta}$ and $\psi \equiv \{\mu, \sigma_v^2\}$.

1. Minimize the objective function: $\left[\frac{1}{N(T-1)}\sum_{i=1}^N\sum_{t=2}^T \mathcal{Z}_{it}' r_{it}(\boldsymbol{\theta})\right]' \boldsymbol{\mathcal{W}}_{\boldsymbol{\theta}} \left[\frac{1}{N(T-1)}\sum_{i=1}^N\sum_{t=2}^T \mathcal{Z}_{it}' r_{it}(\boldsymbol{\theta})\right]$ with respect to $\boldsymbol{\theta}$ by setting $\boldsymbol{\mathcal{W}}_{\boldsymbol{\theta}} = \mathbf{I}_{\mathcal{M}}$ to obtain the one-step estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, where $\boldsymbol{\mathcal{W}}_{\boldsymbol{\theta}}$ of dimension $\mathcal{M} \times \mathcal{M}$ refers to the weighting matrix for the moment conditions used in the estimation of $\boldsymbol{\theta}$, and $\mathcal{M}$ is the combined number of moment conditions (instruments) from Stage 1 and Stage 2.

[1] The derivations are provided in Section A.1 of the Theoretical Appendix.

2. Given the one-step estimate of $\boldsymbol{\theta}$, obtain the residuals $\{\hat{u}_t\}_{t=2}^T$ by Equation (17). This in turn can be used to obtain an estimator of μ and σ_v^2 based on Equations (23) and (24):

$$\arg\min_{\mu,\sigma_v^2} \frac{1}{T}\left(\varsigma_t - \boldsymbol{F}_t \begin{pmatrix} \mu \\ \mu^2 \\ \sigma^2 \end{pmatrix}\right)' \boldsymbol{\mathcal{W}}_\psi \frac{1}{T}\left(\varsigma_t - \boldsymbol{F}_t \begin{pmatrix} \mu \\ \mu^2 \\ \sigma^2 \end{pmatrix}\right), \tag{25}$$

by setting $\boldsymbol{\mathcal{W}}_\psi = \mathbf{I}_3$ to obtain the one-step estimator of μ and σ_v^2, where $\boldsymbol{\mathcal{W}}_\psi$ of dimension 3×3 refers to the weighting matrix for the moment conditions used in the estimation of ψ, and there are three moment conditions in this case.

3. Derive a variance-covariance estimator $\widehat{\mathbf{V}}_{\boldsymbol{\theta}}$ of the moment conditions used in the estimation of $\boldsymbol{\theta}$, $\mathbf{V}_{\boldsymbol{\theta}} = \text{Var}\left(\frac{1}{\sqrt{N(T-1)}} \sum_{i=1}^N \sum_{t=2}^T Z_{it}' r_{it}\right)$, noting that:[2]

$$\begin{aligned}
\mathbf{V}_{\boldsymbol{\theta}} &= \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T \sum_{j=1}^N \sum_{s=2}^T E[(Z_{it}' r_{it})(Z_{js}' r_{js})'] \\
&= \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T \begin{pmatrix} \mathcal{Z}_{it,I}' \mathcal{Z}_{it,I} E(\xi_{it}^2) & \mathcal{Z}_{it,I}' \mathcal{Z}_{it,II} E(\xi_{it}^2) \\ \mathcal{Z}_{it,II}' \mathcal{Z}_{it,I} E(\xi_{it}^2) & \mathcal{Z}_{it,II}' \mathcal{Z}_{it,II} E(\xi_{it}^2) \end{pmatrix} \\
&\quad + \frac{1}{N(T-1)} \sum_{t=2}^T \boldsymbol{\mathcal{Z}}_t' \begin{pmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \sigma_v^2[(\mathbf{I}_N - \mu \mathbf{W}_t^u)^{-1}(\mathbf{I}_N - \mu \mathbf{W}_t^u)^{-1\prime}]) \end{pmatrix} \boldsymbol{\mathcal{Z}}_t,
\end{aligned} \tag{26}$$

by replacing the residuals ξ_t with the sample counterpart $\widehat{\xi}_t$, and the parameters (μ and σ_v^2) with their estimates obtained from Steps 1–2 above.

4. Repeat Step 1, but update the weighting matrix by $\boldsymbol{\mathcal{W}}_{\boldsymbol{\theta}} = \widehat{\mathbf{V}}_{\boldsymbol{\theta}}^{-1}$.

5. To obtain an estimate of the variance-covariance matrix for the moment conditions used in the estimation of $\boldsymbol{\psi}$, we extend the framework of Kapoor, Kelejian and Prucha (2007) to allow for non-normal errors and time-varying connectivity matrices.[3] The variance-covariance matrix for the moment conditions in Equation (6) is given by:

$$\mathbf{V}_\psi = \begin{pmatrix} V_{\psi,11} & V_{\psi,12} & 0 \\ V_{\psi,21} & V_{\psi,22} & V_{\psi,23} \\ 0 & V_{\psi,32} & V_{\psi,33} \end{pmatrix}, \tag{27}$$

[2]The derivations of $\mathbf{V}_{\boldsymbol{\theta}}$ are provided in Section A.2 of the Theoretical Appendix.
[3]The derivations of $\mathbf{V}_\psi$ are provided in Section A.3 of the Theoretical Appendix.

where

$$
\begin{aligned}
V_{\psi,11} &= \sigma_v^4[\kappa_v + 2]; \\
V_{\psi,12} &= \frac{\sigma_v^4}{N(T-1)}[\kappa_v + 2]\mathrm{tr}(\mathbf{W}^{u\prime}\mathbf{W}^u); \\
V_{\psi,21} &= V_{\psi,12}; \\
V_{\psi,22} &= \frac{\sigma_v^4}{N(T-1)}[\kappa_v \ \mathrm{diagv}(\mathbf{W}^{u\prime}\mathbf{W}^u)'\mathrm{diagv}(\mathbf{W}^{u\prime}\mathbf{W}^u); \\
&+ \ \mathrm{tr}(\mathbf{W}^{u\prime}\mathbf{W}^u(\mathbf{W}^{u\prime}\mathbf{W}^u + \mathbf{W}^u\mathbf{W}^{u\prime}))]; \\
V_{\psi,23} &= \frac{\sigma_v^4}{N(T-1)}\mathrm{tr}((\mathbf{W}^{u\prime}\mathbf{W}^u)(\mathbf{W}^u + \mathbf{W}^{u\prime})); \\
V_{\psi,32} &= 2\frac{\sigma_v^4}{N(T-1)}\mathrm{tr}((\mathbf{W}^u\mathbf{W}^{u\prime}\mathbf{W}^u); \\
V_{\psi,33} &= \frac{\sigma_v^4}{N(T-1)}\mathrm{tr}(\mathbf{W}^u(\mathbf{W}^u + \mathbf{W}^{u\prime}));
\end{aligned}
$$

κ_v is the finite excess kurtosis of v_{it}, and $\mathbf{W}^u$ is a $N(T-1) \times N(T-1)$ block-diagonal matrix with $\mathbf{W}_2^u, \mathbf{W}_3^u, \ldots, \mathbf{W}_T^u$ on the diagonal; the operator 'diagv' takes the diagonal elements of a matrix and converts them to a column vector.

The excess kurtosis κ_v can be estimated using the following formula given the estimates of μ and σ_v^2:

$$\hat{\kappa}_v = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{v}_{it} - \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T \hat{v}_{it})^4}{N(T-1)\hat{\sigma}_v^4} - 3, \tag{28}$$

where $\hat{v}_{it}$ is the i-th element of $\hat{v}_t = (\mathbf{I}_N - \hat{\mu}\mathbf{W}_t^u)\hat{u}_t$ and $\hat{\sigma}_v^4 = (\hat{\sigma}_v^2)^2$.

6. Repeat Step 2, this time setting $\boldsymbol{\mathcal{W}}_\psi = \widehat{\mathbf{V}}_\psi^{-1}$.

7. Repeat Steps 3–6 until convergence in the estimates: $\widehat{\mathbf{V}}_{\boldsymbol{\theta}}, \widehat{\mathbf{V}}_\psi, \widehat{\boldsymbol{\theta}}, \hat{\mu}$ and $\hat{\sigma}_v^2$.

8. Obtain a variance-covariance matrix estimator of the parameters $\boldsymbol{\theta}$ based on the asymptotic property established for the efficient GMM estimator (Lee and Yu, 2014):

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \boldsymbol{\mathcal{N}}(0, \mathrm{plim}_{N\to\infty}\frac{1}{T-1}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}), \tag{29}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = (\boldsymbol{\mathcal{Z}}_\Delta' \mathbf{V}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\mathcal{Z}}_\Delta)^{-1}$ and $\boldsymbol{\mathcal{Z}}_\Delta = E\left[\frac{d}{d\boldsymbol{\theta}'} Z_{it}' r_{it}(\boldsymbol{\theta})\right]$. The sample counterpart is correspondingly:

$$\widehat{\boldsymbol{\mathcal{Z}}}_\Delta = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T \left[\frac{d}{d\boldsymbol{\theta}'} Z_{it}' r_{it}(\boldsymbol{\theta})\right].$$

By the Slutsky theorem, we have $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} \equiv (\widehat{\boldsymbol{\mathcal{Z}}}_\Delta' \widehat{\mathbf{V}}_{\boldsymbol{\theta}}^{-1} \widehat{\boldsymbol{\mathcal{Z}}}_\Delta)^{-1} \xrightarrow{p} (\boldsymbol{\mathcal{Z}}_\Delta' \mathbf{V}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\mathcal{Z}}_\Delta)^{-1} \equiv \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$.

9. Similarly, the variance-covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\psi}}$ (*in the original scale*) for the parameters $\boldsymbol{\psi}$ can be estimated by:

$$\widehat{\mathbf{\Sigma}}_{\boldsymbol{\psi}} = \frac{1}{N(T-1)}(\widehat{\mathcal{G}}_{\Delta}' \widehat{\mathbf{V}}_{\psi}^{-1} \widehat{\mathcal{G}}_{\Delta})^{-1}, \tag{30}$$

where

$$\tag{31}$$

$$\widehat{\mathcal{G}}_{\Delta} = \frac{1}{(T-1)} \sum_{t=2}^{T} \frac{d\epsilon_t(\widehat{\boldsymbol{\psi}})}{d\boldsymbol{\psi}'} \tag{32}$$

$$= \frac{1}{N(T-1)} \sum_{t=2}^{T} \begin{pmatrix} 2\hat{u}_t'(\widehat{\mu}\mathbf{W}_t^{u\prime} - \mathbf{I}_N)\mathbf{W}_t^u \hat{u}_t & -N \\ 2\hat{u}_t'\mathbf{W}_t^{u\prime}(\widehat{\mu}\mathbf{W}_t^{u\prime} - \mathbf{I}_N)\mathbf{W}_t^u\mathbf{W}_t^u \hat{u}_t & -\mathrm{tr}(\mathbf{W}_t^{u\prime}\mathbf{W}_t^u) \\ \hat{u}_t'(2\widehat{\mu}\mathbf{W}_t^{u\prime}\mathbf{W}_t^u - (\mathbf{W}_t^u + \mathbf{W}_t^{u\prime}))\mathbf{W}_t^u \hat{u}_t & 0 \end{pmatrix}.$$

4 Data

Our dataset is constructed by combining two Japanese datasets. The first dataset is the Basic Survey of Japanese Business Structure and Activities (BSJBSA), provided by the Ministry of Economy, Trade and Industry (METI), Japan. The data include a firm-level annual survey of detailed business information, such as sales, employment, capital stock, industry classification (Japan Standard Industry Classification, JSIC) and intermediate purchases. The data cover both manufacturing and non-manufacturing firms that have: (1) more than 50 employees, and (2) capital stocks of more than 30 million yens (approximately 350 thousand USD in 2012).

The second dataset contains information on firm-to-firm relationship provided by Tokyo Shoko Research (TSR), a major credit reporting company in Japan. It provides a firm's most important domestic suppliers and customers (up to 24 connections in each direction) and covers both publicly listed and unlisted firms in Japan with all sizes and industries. Because these two datasets do not use the same firm identification codes, we match them on the basis of firm name, address, phone number, and postal code. Using the BSJBSA as the denominator (since it provides the required firm-level variables for productivity estimations), the percentage of firms in BSJBSA that are matched with its counterpart in TSR is very high, typically at 93%–94%, across years during the sample period 2009–2018. Table 1 provides the detailed firm counts.

We supplement the BSJBSA-TSR linked dataset with the JIP database 2021 provided by the Research Institute of Economy, Trade and Industry (RIETI), and with the information on commuting zones (CZs) constructed by Adachi, Fukai, Kawaguchi and Saito (2021). We impute the industry-level price deflators based on the JIP database. In particular, it contains the nominal and real values of outputs, intermediate inputs, investment, and value added for the 100 industries classified by the JIP database. We construct the deflators by the ratios of the nominal and real values for each of these variables, and merge them with the BSJBSA data (based on concordance between the BSJBSA JSIC industries and the JIP industries, provided in the JIP database). There are in total 433 JSIC 3-digit industries. A JIP industry is matched on average with 4.8 JSIC 3-digit industries. These deflators are then used to convert the BSJBSA corresponding variables into real terms. We also impute the average work hours per person in a year in an industry based on the JIP database and merge the variable with the BSJBSA data (using again the concordance between the JIP and JSIC industries).

The information on CZs is used as one criterion below in defining connectivity matrices across firms. Adachi, Fukai, Kawaguchi and Saito (2021) construct the CZ information for Japan using the hierarchical agglomerative clustering (HAC) method of Tolbert and Killian (1987) and Tolbert and Sizer (1996) for constructing the US CZs. By Adachi et al. (2021), there are 267 CZs in 2010 and 265 in 2015 in Japan. We use the 2010 CZ definition, and merge the CZ information with the BSJBSA observations based on the prefecture and city names of a firm's address, and examine/adjust manually if necessary.

The variables at the firm-year level used for the production function estimation is constructed in the following manners. The number of workers is constructed as the sum of regular workers and part-time workers (excluding temporary workers) in headquarter, head office, branch office, and assignee company (available from the BSJBSA). The labor hour is constructed as the number of workers (from the BSJBSA), times the average work hours per person in each industry (from the JIP database). The real physical capital stock is constructed by the perpetual inventory method with 2007 as the base year, using the initial real physical capital value in 2007 and the real investment in physical capital in each year from the BSJBSA, together with the depreciation rates at the industry level from the JIP database. The real intermediate inputs are constructed by the sum of the cost of goods sold, and general and administrative expenses, minus wages, rental costs, depreciation, and taxes reported in the BSJBSA, deflated by the input deflator (constructed using the JIP database as documented above). The real revenue is measured by sales, deflated by the output deflator imputed from the JIP database. The real value added is constructed by the difference between the real sales and the real intermediate inputs, following Wakasugi et al. (2008).

4.1 Summary Statistics

Table 2(a) provides the summary statistics of the key variables for the BSJBSA-TSR linked sample in year 2015 (based on the denominator of BSJBSA firms, not all of which have corresponding entries in TSR). The effective number of observations differs from Table 1 due to potentially missing observations on the variable of interest.

A few remarks are in order. First, the average firms tend to be large (e.g., having 459 workers, and 5.1 billion JPY physical capital, roughly equivalent to 40 million USD). This is due to the fact that the BSJBSA only covers medium and large firms. Second, the average firms report 6.8 customers and 6.7 suppliers, suggesting that the TSR's limit of reporting the top suppliers and customers up to 24 connections in each direction is not practically binding for most of the firms.

Figures 1(a)–3(a) illustrate the number of firms, their average size in terms of employment, and their average number of customers and suppliers for each 1-digit JSIC rev12 industry. Figure 1(a) indicates that most of the firms in the sample are in the manufacturing, wholesale & retail trade, and information & communications industries. Among them, those in the service industries tend to be large in terms of employment (e.g., accommodation and food & beverage services, electricity, gas, heat supply & water, and finance & insurance) (Figure 2(a)). In terms of connectedness with other firms, those in the construction, manufacturing, and mining & quarrying industries tend to have a larger number of customers and suppliers, while those in the service industries tend to have a smaller number of business customers but just as many suppliers as in other industries (Figure 3(a)).

Figures 4(a)–6(a) show a large heterogeneity across prefectures in terms of the number of firms, their average size, and their average number of customers and suppliers. Most of the firms in the sample are located in economically large prefectures, such as Tokyo, Osaka, Aichi, Kanagawa, and Hyogo (Figure 4(a)). Firms in these large prefectures tend also to be large in terms of employment (Figure 5(a)) and connected with a larger number of both customers and suppliers (Figure 6(a)).

The potential spatial dependence across firms through the supply chain and the input markets can be more local than the prefecture level. Bernard, Moxnes and Saito (2019) show that the median distance of any customer-supplier pair in the TSR data is 30 km and thus smaller than the typical size of prefectures. Thus, the following figures further provide the characterization at the commuting zone level. Figure 7(a) shows the number of CZs within each prefecture in 2015. Prefectures with large areas (e.g., Hokkaido and Nagano) tend to have many CZs, while those with small areas and economic sizes (e.g., Kagawa and Fukui) tend to have few CZs. Figures 8(a)–10(a) show the counterparts of Figures 4(a)–6(a) at the commuting-zone level. The commuting zone with the largest number of firms (8993) is CZ89 that covers the busiest areas around Tokyo (parts of Tokyo, Kanagawa, Chiba, and Saitama). Economically large CZs also tend to have larger firms and more connected firms. For example, the same CZ89 ranks top 5th in terms of average firm's employment size (643.99), and top 24th in terms of average firm's customer connections (7.44). Note also that the average firm size is much more dispersed at the right tail when we zoom in at the commuting-zone level (Figure 9(a)) compared to that at the prefecture level (Figure 5(a)). Similarly, the distributions of supplier/customer connections are much more dispersed at the commuting-zone level (Figure 10(a)) than at the prefecture level (Figure 6(a)), suggesting a large degree of heterogeneity across CZs within prefectures.

4.2 Definition of Connectivity Matrices

To model the spatial-temporal lag dependence in outputs, we define the output connectivity matrix $\mathbf{W}_{t-1}^{y}$ based on the set of a firm's customers and/or suppliers located in the same commuting zone. The ij-th element of $\mathbf{W}_{t-1}^{y}$ takes on the value one if both firms i and j are located in the same commuting zone, and in addition, firm j is a customer or supplier of firm i, in period $t-1$. This is in line with the research conducted by Ellison, Glaeser and Kerr (2010). They find that the geographical proximity of firms and the input-output linkages across firms play an important role in productivity spillovers.

Next, for productivity dependence across firms through the input markets, we restrict our focus to the labor market channel, and define input connectivity matrix $\mathbf{W}_{t-1}^{\Omega}$ such that the ij-th element of $\mathbf{W}_{t-1}^{\Omega}$ takes on the value one if firms i and j are located in the same commuting zone in period $t-1$. Firms located in the same commuting zone are more likely to tap into the same labor pool, considering the potential labor mobility frictions across zones. Multiple theories have been proposed about the benefits associated with a large labor pool. When firms in the same location employ more workers, the potential pool of labor in the location increases. This facilitates better worker-firm matches (e.g., Helsley and Strange, 1990); allows risk sharing and worker turnover across firms (e.g., Diamond and Simon, 1990; Krugman, 1991); and induces stronger incentives for workers to invest in human capital knowing that they do not face ex post appropriation (Rotemberg and Saloner, 2000). As a result, conditional on the amount of labor input hired by a firm, the quality of labor input (and hence firm productivity) is likely higher when the total labor employed in the same location in the past period is larger. Relatedly, Greenstone, Hornbeck and Moretti (2010) find that estimated spillover effects resulting from the opening of Million Dollar Plants are larger for other plants that share labor pools and similar technologies with the new plant.

To model the spatial diffusion of the productivity shock u_t, we consider three variants of the connectivity matrix $\mathbf{W}_t^u$, depending on whether two firms are located in the same commuting zone, whether they have supplier/customer relationships, or both. In particular, the ij-th element of $\mathbf{W}_t^u$ takes on the value one: (i) if both firms i and j are located in the same commuting zone in period t, (ii) if both firms i and j are located in the same commuting zone, and firm j is a customer or supplier of firm i, in period t, and (iii) if firm j is a customer or supplier of firm i in period t, respectively. Supporting evidence of the criteria used above includes the work by Audretsch and Belitski (2020) and Matray (2021). The latter shows that local innovation spillovers decline rapidly with distance. In the second variant, the spillover is further restricted specifically to firms in supplier-customer relationships. An unexpected productivity shock experienced by a firm may trickle down to its buyers via the provision of higher quality inputs, allowing its buyers to scale up their productivities. Alternatively, the technology innovation or discovery may occur simultaneously to the network of firms belong to the same supply or value chain. The third variant instead focuses on the supply chain as the conduit of innovation spillovers, but disregards the potential distance between the buyers/sellers. Note that the second variant is a relatively sparse matrix compared to the other two variants.

The connectivity matrices defined above are then row-normalized, such that each row has a row sum equal to one (and zero if all elements in a row are zeros).

5 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to assess the consistency and efficiency of the estimator we proposed in Section 3 (which allows for spatial dependence in productivity), and compare it with the conventional estimators (that assume no such spatial dependence). We consider five data generating processes (DGPs). The first DGP (DGP1) is favorable to the conventional estimator and assumes that the productivity ω_t follows an AR(1) process. The remaining DGPs consider spatial dependence of various mechanisms and strengths across firms. The second DGP (DGP2) assumes the productivity ω_t to depend on own lagged productivity and the lagged outputs/inputs of connected firms as specified in Equation (2). The third DGP (DGP3) further allows the productivity shock u_t to be spatially correlated as specified in Equation (3). The fourth DGP (DGP4) is the same as DGP3 but assumes stronger spatial dependence in the lagged output/inputs of connected firms. The fifth DGP (DGP5) is the same as DGP3 but assumes instead negative spatial dependence in the lagged output/inputs of connected firms.

We generate the simulation data based on the empirical sample statistics of the Japanese BSJBSA-TSR linked dataset. Appendix B provides detailed documentations of the simulation setups, which we summarize below. We follow Ackerberg, Caves and Frazer (2015) and adopt a Leontief production function such that:

$$VA_{it} = \min\{e^{\alpha_0} L_{it}^{\alpha_l} K_{it}^{\alpha_k} e^{\omega_{it}}, e^{\alpha_m} M_{it}\} e^{\xi_{it}},$$

which implies Equation (1). In turn, gross output is linear in value-added. In particular, we set $e^{\alpha_m} = 1$ in simulating the gross output. The firm-level productivity is simulated based on Equations (2)–(3), with variations in the parameter values across the DGPs studied.

The firm-level input variables (labor and capital inputs) and the firm-to-firm connectivity matrices are simulated based on the firm-level statistics and the supplier-customer network statistics of the BSJBSA-TSR linked dataset. For example, based on the BSJBSA-TSR linked dataset, we tabulate the distribution of firms that supply to one, two, three, …, and up to 24 other firms; and respectively, the distribution of firms that purchase from one, two, three, …, and up to 24 other firms. We use these distribution statistics across years to simulate time-varying supplier-customer networks, which takes into account network addition, attrition, and persistency observed in the data.

Given the model structure, we assume that the error terms (ξ_{it}, v_{it}) are normally distributed with mean zeros and standard deviations of $\sigma_\xi = 0.3$ and $\sigma_v = 0.7$. We simulate a balanced panel of 500, 750 or 1000 firms for 10 or 19 time periods. For each DGP, 1000 simulated samples are drawn and estimated. We report the mean (Mean) and the standard deviation (SD) of the parameter point estimates across the 1000 Monte Carlo simulations, together with the estimated standard errors (SE) derived from the variance-covariance matrices of the estimators. The exact parameter values used in the DGPs are listed in the first row of Tables 3–7. The parameter values that are common across DGPs are: $\alpha_0 = 0$, $\alpha_l = 0.6$, $\alpha_k = 0.4$; $\rho_1 = 0.5$ and $\rho_2 = \ldots = \rho_G = 0$. To simplify the Monte Carlo exercises, we drop $\mathbf{x}_{i,t-1}$ (a firm's lagged exporting status and/or R&D expenditure) from consideration in the simulation. In DGP2, the strength of spillovers in terms of lagged outputs and lagged labor inputs of related firms is set at: $\lambda = \beta_l = 0.01$. DGP4 considers stronger spillovers such that $\lambda = \beta_l = 0.1$, while DGP5 considers negative spillovers such that $\lambda = \beta_l = -0.1$. In DGP3–DGP5, with spatial error dependence, we set $\mu = 0.25$.

Given the simulated sample, we use the Wooldridge (2009) GMM estimator (henceforth WGMM) to represent the conventional estimators (that assume no spatial dependence in productivity across firms). The Wooldridge (2009) procedure estimates the production function parameters in one step, in contrast with the two-step procedures proposed by LP and ACF. For our proposed estimator (SGMM), we use the instruments indicated in Equations (10) and (13) in estimations. In particular, the current and first lag of labor, capital and material inputs are used as the instruments for the first-stage equation (8), with a degree-1 h function in labor, capital and material inputs (à la Ackerberg, Caves and Frazer, 2015). For the second-stage equation (11), the current capital along with the first lag of labor, capital and material inputs, and the lagged-one-period outputs and labor inputs of related firms ($\mathbf{W}^y_{t-1}y_{t-1}$, $(\mathbf{W}^y_{t-1})^2 y_{t-1}$, $\mathbf{W}^l_{t-1}l_{t-1}$, $(\mathbf{W}^l_{t-1})^2 l_{t-1}$) are used as instruments. The connectivity matrices are as defined in Section 4.2. In particular, the connectivity matrix $\mathbf{W}^u_t$ specifying the spatial correlation of productivity shocks is defined based on the customer-supplier relationships across firms. The same set of instruments are used for the WGMM estimations, but excluding the related firms' lagged outputs and lagged labor inputs ($\mathbf{W}^y_{t-1}y_{t-1}$, $(\mathbf{W}^y_{t-1})^2 y_{t-1}$, $\mathbf{W}^l_{t-1}l_{t-1}$, $(\mathbf{W}^l_{t-1})^2 l_{t-1}$).

5.1 Simulation Results

Table 3 reports the results for DGP1. The conventional estimator (WGMM) performs well as it should, when the DGP has no spatial dependence across firms in productivity. Importantly, our proposed estimator (SGMM) performs just as well. The point estimates of both estimators are close to the true parameter values, and the 95% confidence intervals (CIs) include the true parameter values for the input coefficients of the production function (α_l, α_k). While our estimator has wider confidence intervals than the conventional estimator for the input coefficients, it returns mean estimates of the spatial coefficients (λ, β_l, μ) nearly identical to zeros, consistent with the true parameter values of the underlying DGP. Both estimators obtain estimates for the autoregressive parameter (ρ_1) that are close to the true parameter value, even when the duration of the panel is relatively short. Both the conventional estimator and our proposed estimator yield standard error estimates (SE) that are close to their Monte Carlo standard deviations (SD). The SEs also reduce as the sample size increases at a rate consistent with the asymptotic properties laid out in Section 3.3.

Table 4 reports the findings for the second set of simulations based on DGP2. When spatial dependence in productivity across firms via lagged outputs and lagged labor inputs are indeed present, the conventional estimator leads to biased estimates of the input coefficients. In particular, its mean estimates for α_l across variations in N and T are higher than the true parameter value. The bias does not shrink with a larger sample size, suggesting the inconsistency of the conventional estimator when spatial dependence is present in the underlying DGP. In contrast, our proposed SGMM estimator yields estimates that are close to the true parameter values for both input elasticities (α_l, α_k) and the spatial coefficients (λ, β_l), with 95% CIs that well cover the true parameter values. Finally, our proposed SGMM estimator reports statistically insignificant estimates of μ, consistent with the underlying DGP where no spatial correlation in the error terms (i.e., the productivity shocks u_t) is present.

In DGP3, the data generating process for the productivity term further allows for spatial error correlation across related firms. Table 5 shows that the conventional estimator of the labor coefficient of the production function remains to be upward biased, while our proposed estimator yields consistent estimates that are close to the true parameter values for all the coefficients of interest. In particular, we note that the SGMM estimator returns estimates of the spatial error coefficient (μ) that are close to its true parameter value when it is indeed non-zero.

Table 6 reports the simulation results for DGP4. With larger spatial coefficients ($\lambda = \beta_l = 0.1$, instead of 0.01), the conventional estimator of all coefficients (α_l, α_k, ρ_1) are upward biased, and the extents of bias are substantial (by around 23 percentage points for α_l, 4-8 percentage points for α_k, and 5-13 percentage points for ρ_1). Furthermore, the standard errors (SE) obtained by the conventional estimator deviate significantly from the Monte Carlo standard deviations (SD). In contrast, our proposed SGMM estimator continues to yield consistent estimates for the true parameters, with estimates of the standard errors (SE) very close to the Monte Carlo standard deviations (SD).

Table 7 indicates that if the underlying DGP has larger, negative, spatial coefficients ($\lambda = \beta_l = -0.1$, instead of 0.01), the conventional estimator of the input coefficients (α_l, α_k) are instead downward biased, and the extents of bias continues to be substantial (by around 21 percentage points for α_l, 3-4 percentage points for α_k). The standard errors (SE) obtained by the conventional estimator also deviate significantly from the Monte Carlo standard deviations (SD) for these two input coefficients. In contrast, our proposed SGMM estimator yields consistent estimates for the true parameters, with estimates of the standard errors (SE) very close to the Monte Carlo standard deviations (SD). In particular, it is able to capture the negative signs of the two spatial coefficients ($\lambda = \beta_l = -0.1$) and their magnitudes.

In sum, across all the DGPs, we find that the proposed SGMM estimator yields point estimates that are consistent for the true parameters both in the absence and in the presence of spatial effects. By the SGMM estimator, the standard error estimates (SE) of the parameters are also very close to the Monte Carlo standard deviations (SD). As the sample size $N(T-1)$ doubles (either due to doubling of N or $T-1$), both the standard error estimates (SE) and the Monte Carlo standard deviations (SD) shrink at a rate close to $1/\sqrt{2}$, consistent with a convergence rate of $1/\sqrt{N(T-1)}$.

6 Empirical Analysis

6.1 Estimation Sample

We apply the methodology and estimation algorithms proposed in Section 3 to the Japanese dataset introduced in Section 4. Given the BSJBSA-TSR linked data, we further restrict the sample to a balanced panel of firms with observations on the set of variables required for productivity estimations. In particular, the sample is based on firms that were surveyed for 10 consecutive years from 2009 to 2018.[4] Second, the sample of firms used for analysis also need to have non-missing values for log of labor hours (used to measure l_{it}), log of real capital stock (k_{it}), log of real intermediate inputs (m_{it}), log of real revenues (y_{it}), and log of real value added (va_{it}), during the entire sample period 2009–2018. Recall that the real capital stock is calculated based on the perpetual inventory method with the real capital stock in 2007 as the initial value. Observations on real capital stock for a firm could be missing, for example, if the firm was not observed in 2007.

The resulting sample is a balanced panel of 12,525 firms for the period 2009–2018. Given the balanced panel of firms, the set of a firm's customers/suppliers identified via the TSR entries is effectively restricted to those whose firm-level data also exist in BSJBSA. In particular, firm j is regarded effectively as a customer/supplier of firm i in the estimation if firm i reports firm j as a customer or supplier, *and* if firm j exists in the BSJBSA dataset (with consecutive observations on output and inputs, as required for the estimation of Equation (2)).

Table 2(b) provides the summary statistics for the estimation sample. Relative to the raw sample reported in Table 2(a), the firms in the estimation sample tends to be larger in terms of both inputs and output, and have more customers and suppliers. This is expected, as larger firms are more likely to be surveyed consecutively throughout the years and have positive inputs/output. Although larger firms tend to have more customers/suppliers, the orders of magnitude in the number of connections on average do not differ substantially between the raw and estimation samples. Despite the much smaller set of firms covered, the estimation sample accounts for 60.45% of aggregate real value added and 56.90% of real gross output of the raw sample in 2015 (and a majority of the other economic activities in terms of employment, labor hours, real capital stock, and real spending on intermediate inputs).

Figures 1(b)–10(b) repeat the characterization as in Figures 1(a)–10(a), and show that the estimation sample has similar patterns as documented for the raw sample. Notably, four industries are not present in the estimation sample: construction, mining & quarrying, agriculture & forestry, and fisheries (so firms in these industries tend not to be large enough to be consecutively surveyed by BSJBSA). Focusing on the remaining industries, the rank across industries is almost identical in terms of the number of firms: manufacturing, wholesale & retail trade, and information & communications remain to be the top three industries with the largest numbers of firms (Figure 1).

[4] This excludes, for example, firms whose number of employees fell under 50 at some point during the sample period.

The set of prefectures with the largest numbers of firms is also similar to that previously documented (Figure 4). Basically, firms in the estimation sample tend to be larger in terms of employment size (Figures 2, 5, and 9), and are slightly more connected in terms of customers/suppliers (Figures 3, 6, and 10), in comparison with the raw sample.

6.2 Estimation Results

We estimate the model proposed in Equations (1)–(3) based on the estimation methodology laid out in Section 3 and the connectivity matrices defined in Section 4.2. In short, we define the output connectivity matrix $\mathbf{W}^{y}_{t-1}$ based on the set of a firm's customers/suppliers located in the same commuting zone. The ij-th element of $\mathbf{W}^{y}_{t-1}$ takes on the value one if both firms i and j are located in the same commuting zone, and in addition, firm j is a customer or supplier of firm i, in period $t-1$. Second, we define input connectivity matrix $\mathbf{W}^{\Omega}_{t-1}$ such that the ij-th element of $\mathbf{W}^{\Omega}_{t-1}$ takes on the value one if both firms i and j are located in the same commuting zone in period $t-1$. The input variable being analyzed corresponds to the lagged labor inputs of the connected firms defined by $\mathbf{W}^{\Omega}_{t-1}$. Third, we consider three variants of the spatial error connectivity matrix $\mathbf{W}^{u}_{t}$ and define it such that the ij-th element of $\mathbf{W}^{u}_{t}$ takes on the value one: (i) if both firms i and j are located in the same commuting zone in period t, (ii) if both firms i and j are located in the same commuting zone, and firm j is a customer or supplier of firm i, in period t, and (iii) if firm j is a customer or supplier of firm i in period t, respectively. The connectivity matrices are then row-normalized, such that each row has a row sum equal to one (and zero if all elements in a row are zeros).

Table 8 reports the estimation results. Column 1 (based on the first definition of $\mathbf{W}^{u}_{t}$) indicates that all three spatial coefficients are significant and positive. A 1% increase in the sales of customers/suppliers in the same commuting zone in the previous period helps improve a firm's current productivity by 0.005%. A larger local labor market also enhances a firm's productivity: specifically, a 1% increase in the employment of firms located in the same commuting zone in the previous period raises a firm's current productivity by 0.05%. Finally, there is evidence of contemporary knowledge spillovers across firms located in the same commuting zone: the productivity innovations u_{it} are spatially correlated with a positive and significant slope coefficient of 0.38. The other production function parameters are also precisely estimated and fall within the typical range. Column 1 reports a labor value-added share of 0.79, a capital value-added share of 0.06, and a partial AR(1) coefficient of 0.964 for the productivity process.

The estimates for key parameters of interest remain similar if we adopt alternative definitions of spatial error connectivity matrices $\mathbf{W}^{u}_{t}$, as reported in Column 2 and Column 3. The key difference is the strength of contemporary spatial correlation in the productivity shocks u_{it}. They tend to be quantitatively smaller by an order of magnitude if $\mathbf{W}^{u}_{t}$ is defined in a more restricted manner, relative to the findings in Column 1 based on common commuting zone alone.

To compare the spatial GMM findings with those if one ignores the spatial dependence in productivity across firms, Column 4 reports the estimates based on conventional estimators (restricting the spatial coefficients to be zeros but otherwise adopting the same GMM approach). We find that the labor value-added share tends to be downward biased (0.55), the capital value-added share upward biased (0.15), and the AR(1) coefficient upward biased (0.979), in comparison with the spatial GMM estimates reported above.

7 Conclusion

In this paper, we develop a framework to simultaneously estimate firm production functions and spatial interactions across firms in one unified setup. We propose a three-stage efficient GMM estimation algorithm, and show by theory the asymptotic properties of the proposed estimator and by Monte Carlo simulations the finite sample performance of the estimator. The Monte Carlo simulations demonstrate that the proposed estimator is consistent under DGPs with or without spatial dependence across firms. In contrast, the conventional estimators are biased when the true DGPs are indeed characterized by spatial dependence. By applying the developed methodology and estimation algorithm to the Japanese BSJBSA-TSR linked dataset for the period 2009–2018, we find that spatial interactions across firms play a significant role in determining the Japanese firm-level productivity both statistically and economically.

The paper can be extended in several directions in future research. First, the connectivity matrices in our setup are allowed to differ across different mechanisms of spatial interactions. One can potentially hypothesize alternative candidates for the connectivity matrices and conduct specification tests that select the specification that best fits the model. This will also provide insights into the nature of spatial interactions across firms and tests for competing hypotheses. Second, the current framework allows for time-varying connectivity matrices. This is useful, as we can use the framework to analyze how shocks (such as high-speed rails and earthquakes) affect the connectivity matrices across firms, and in turn, the firm-level performance measures (such as productivity, and production technology). Third, the current framework could also be used to analyze the centrality of firms in the sense that a firm is more central if by increasing connectivity (links) to this firm, the average productivity of all firms is improved by more than if by increasing connectivity (links) to another firm. This is useful for policy design that aims to target subsidies at the critical links of a firm network structure for the greater benefits of the economy.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83*, 2411–2451.

Adachi, Daisuke, Taiyo Fukai, Daiji Kawaguchi, and Yukiko Umeno Saito, "Commuting Zones in Japan," RIETI Discussion Paper 20-E-021 2021.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P. Vasquez, "The Effects of Multinationals on Workers: Evidence from Costa Rica Microdata," Griswold Center for Economic Policy Studies Working Paper No. 285 2021.

Audretsch, David B. and Maksim Belitski, "The Role of R&D and Knowledge Spillovers in Innovation and Productivity," *European Economic Review*, 2020, *123* (103391).

Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito, "Production Networks, Geography and Firm Performance," *Journal of Political Economy*, 2019, *127*, 639–688.

Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson, "Acquisitions, Productivity, and Profitability: Evidence from the Japanese Cotton Spinning Industry," *American Economic Review*, 2015, *105* (7), 2086–2119.

De Loecker, Jan, "Detecting Learning by Exporting," *American Economic Journal: Microeconomics*, 2013, *5* (3), 1–21.

Diamond, Charles A. and Curtis J. Simon, "Industrial Specialization and the Returns to Labor," *Journal of Labor Economics*, 1990, *8* (2), 175–201.

Doraszelski, Ulrich and Jordi Jaumandreu, "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies*, 2013, *80*, 1338–1383.

Elhorst, Paul, *From Cross-Sectional Data to Spatial Panels*, Heidelberg: Springer, 2014.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 2010, *100* (3), 1195–1213.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Helsley, Robert W. and William C. Strange, "Matching and Agglomeration Economies in a System of Cities," *Regional Science and Urban Economics*, 1990, *20* (2), 189–212.

_ **and** _ , "Innovation and input sharing," *Journal of Urban Economics*, 2002, *51* (1), 25—-45.

_ **and** _ , "Coagglomeration, clusters, and the scale and composition of cities," *Journal of Political Economy*, 2014, *122* (5), 1064–1093.

Iyoha, Ebehi, "Estimating Productivity in the Presence of Spillovers: Firm-level Evidence from the US Production Network," manuscript 2022.

Jaffe, A. B., M. Trajtenberg, and R. Henderson, "Geographic localization of knowledge spillovers as evidenced by patent citations," *Quarterly Journal of Economics*, 1993, *108* (3), 577–598.

Javorcik, Beata Smarzynska, "Does Foreign Direct Investment increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages," *American Economic Review*, 2004, *94*, 605–627.

Kapoor, Mudit, Harry Kelejian, and Ingmar Prucha, "Panel Data Models with Spatially Correlated Error Components," *Journal of Econometrics*, 2007, *140*, 97–130.

Kelejian, Harry and Ingmar R Prucha, "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *The Journal of Real Estate Finance and Economics*, 1998, *17*, 99–121.

Kelejian, Harry H and Ingmar R Prucha, "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 1999, *40* (2), 509–533.

Keller, Wolfgang and Stephen R Yeaple, "Multinational Enterprises, International Trade, and Productivity Growth: Firm-level Evidence from the United States," *Review of Economic and Statistics*, 2009, *91*, 821–831.

Krugman, Paul, *Geography and Trade*, Cambridge, MA: MIT Press, 1991.

Lee, Lung-Fei and Jihai Yu, "Efficient GMM Estimation of Spatial Dynamic Panel Data Models with Fixed Effects," *Journal of Econometrics*, 2014, *180*, 174–197.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, *70*, 317–341.

Matray, Adrien, "The Local Innovation Spillovers of Listed Firms," *Journal of Financial Economics*, 2021, *141*, 395–412.

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64*, 1263–1297.

Rotemberg, Julio J. and Garth Saloner, "Competition and Human Capital Accumulation: A Theory of Interregional Specialization and Trade," *Regional Science and Urban Economics*, 2000, *30* (4), 373–404.

Tolbert, Charles M. and Molly Sizer, "U.S. Commuting Zones and Labor Market Areas: A 1990 Update," Staff Reports 278812, USDA 1996.

_ **and Molly Sizer Killian**, "Labor Market Areas for the United States," ERS staff report, USDA 1987.

Wakasugi, Ryuhei, Yasuyuki Todo, Hitoshi Sato, Shuichiro Nishioka, Toshiyuki Matsuura, Banri Ito, and Ayumu Tanaka, "The Internationalization of Japanese Firms: New Findings Based on Firm-Level Data," RIETI Discussion Paper Series 08-E-036 2008.

Wooldridge, Jeffrey M., "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 2009, *104*, 112–114.

A Theoretical Appendix

A.1 Deriving the Moment Condition in Stage 3

Replacing $v_t = u_t - \mu \mathbf{W}_t^u u_t$ in Equation (6) yields:

$$\frac{1}{N}E\begin{bmatrix} (u_t - \mu\mathbf{W}_t^u u_t)'(u_t - \mu\mathbf{W}_t^u u_t) \\ (u_t - \mu\mathbf{W}_t^u u_t)\mathbf{W}_t^{u\prime}\mathbf{W}_t^u(u_t - \mu\mathbf{W}_t^u u_t) \\ (u_t - \mu\mathbf{W}_t^u u_t)'\mathbf{W}_t^u(u_t - \mu\mathbf{W}_t^u u_t) \end{bmatrix} = \frac{1}{N}E\begin{bmatrix} (u_t - \mu\overline{u}_t)'(u_t - \mu\overline{u}_t) \\ (u_t - \mu\overline{u}_t)'\mathbf{W}_t^{u\prime}\mathbf{W}_t^u(u_t - \mu\overline{u}_t) \\ (u_t - \mu\overline{u}_t)'\mathbf{W}_t^u(u_t - \mu\overline{u}_t)' \end{bmatrix}$$

$$= \frac{1}{N}E\begin{bmatrix} (u_t'u_t - 2\mu\overline{u}_t{}'u_t + \mu^2\overline{u}_t'\overline{u}_t) \\ (\overline{u}_t'\overline{u}_t - 2\mu\overline{\overline{u}}_t'\overline{u}_t + \mu^2\overline{\overline{u}}_t'\overline{\overline{u}}_t) \\ \overline{u}_t'u_t - \mu(\overline{u}_t'\overline{u}_t + \overline{\overline{u}}_t'u_t) + \mu^2\overline{\overline{u}}_t'\overline{u}_t \end{bmatrix}.$$

Furthermore, Equation (6) implies:

$$\frac{1}{N}E\begin{bmatrix} (u_t'u_t - 2\mu\overline{u}_t{}'u_t + \mu^2\overline{u}_t'\overline{u}_t) \\ (\overline{u}_t'\overline{u}_t - 2\mu\overline{\overline{u}}_t'\overline{u}_t + \mu^2\overline{\overline{u}}_t'\overline{\overline{u}}_t) \\ \overline{u}_t'u_t - \mu(\overline{u}_t'\overline{u}_t + \overline{\overline{u}}_t'u_t) + \mu^2\overline{\overline{u}}_t'\overline{u}_t \end{bmatrix} = \begin{bmatrix} \sigma_v^2 \\ \frac{\sigma_v^2}{N}\mathrm{tr}(\mathbf{W}_t^{u\prime}\mathbf{W}_t^u) \\ 0 \end{bmatrix}. \tag{33}$$

Rearranging terms, we have:

$$\frac{1}{N}E\begin{bmatrix} u_t'u_t \\ \overline{u}_t'\overline{u}_t \\ \overline{u}_t'u_t \end{bmatrix} = \frac{1}{N}\begin{bmatrix} 2\mu\overline{u}_t{}'u_t - \mu^2\overline{u}_t'\overline{u}_t + \sigma_v^2 \\ 2\mu\overline{\overline{u}}_t'\overline{u}_t - \mu^2\overline{\overline{u}}_t'\overline{\overline{u}}_t + \frac{\sigma_v^2}{N}\mathrm{tr}(\mathbf{W}_t^{u\prime}\mathbf{W}_t^u) \\ \mu(\overline{u}_t'\overline{u}_t - \overline{\overline{u}}_t'u_t) - \mu^2\overline{\overline{u}}_t'\overline{u}_t \end{bmatrix}$$

$$= \frac{1}{N}\begin{bmatrix} 2\overline{u}_t{}'u_t & -\overline{u}_t'\overline{u}_t & 1 \\ 2\overline{\overline{u}}_t'\overline{u}_t & -\overline{\overline{u}}_t'\overline{\overline{u}}_t & \frac{1}{N}\mathrm{tr}(\mathbf{W}_t^{u\prime}\mathbf{W}_t^u) \\ (\overline{u}_t'\overline{u}_t - \overline{\overline{u}}_t'u_t) & -\overline{\overline{u}}_t'\overline{u}_t & 0 \end{bmatrix}\begin{bmatrix} \mu \\ \mu^2 \\ \sigma_v^2 \end{bmatrix}, \tag{34}$$

which yields the following relationship:

$$\gamma_t = \mathbf{\Gamma}_t \begin{bmatrix} \mu \\ \mu^2 \\ \sigma_v^2 \end{bmatrix}.$$

A.2 Deriving the Variance-Covariance Matrix for the Moment Conditions in Stages 1 and 2

Using the definition of the variance-covariance matrix of the moment conditions in the first and second stages, we have:

$$
\begin{aligned}
\mathbf{V}_{\boldsymbol{\theta}} &= \text{Var}\left(\frac{1}{\sqrt{N(T-1)}}\sum_{i=1}^{N}\sum_{t=2}^{T} Z'_{it} r_{it}\right) \\
&= \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{s=2}^{T} E[(Z'_{it} r_{it})(Z'_{js} r_{js})'] \\
&= \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{s=2}^{T} Z'_{it} E[r_{it} r'_{js}] Z_{js} \\
&= \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{s=2}^{T} Z'_{it} E\left[\begin{pmatrix} \xi_{it} \\ \xi_{it} + u_{it} \end{pmatrix} (\xi_{js} \quad \xi_{js} + u_{js})'\right] Z_{js} \\
&= \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{s=2}^{T} Z'_{it} E\begin{pmatrix} \xi_{it}\xi_{js} & \xi_{it}(\xi_{js}+u_{js}) \\ (\xi_{it}+u_{it})\xi_{js} & (\xi_{it}+u_{it})(\xi_{js}+u_{js}) \end{pmatrix} Z_{js}.
\end{aligned} \tag{35}
$$

To derive $E\begin{pmatrix} \xi_{it}\xi_{js} & \xi_{it}(\xi_{js}+u_{js}) \\ (\xi_{it}+u_{it})\xi_{js} & (\xi_{it}+u_{it})(\xi_{js}+u_{js}) \end{pmatrix}$, consider the following 4 cases:

1. $i = j, t = s$

$$
\begin{aligned}
E\begin{pmatrix} \xi_{it}\xi_{js} & \xi_{it}(\xi_{js}+u_{js}) \\ (\xi_{it}+u_{it})\xi_{js} & (\xi_{it}+u_{it})(\xi_{js}+u_{js}) \end{pmatrix} &= E\begin{pmatrix} \xi_{it}\xi_{it} & \xi_{it}(\xi_{it}+u_{it}) \\ (\xi_{it}+u_{it})\xi_{it} & (\xi_{it}+u_{it})(\xi_{it}+u_{it}) \end{pmatrix} \\
&= \begin{pmatrix} \sigma_\xi^2 & \sigma_\xi^2 \\ \sigma_\xi^2 & \sigma_\xi^2 + E(u_{it}^2) \end{pmatrix}.
\end{aligned}
$$

2. $i = j, t \neq s$

$$
\begin{aligned}
E\begin{pmatrix} \xi_{it}\xi_{js} & \xi_{it}(\xi_{js}+u_{js}) \\ (\xi_{it}+u_{it})\xi_{js} & (\xi_{it}+u_{it})(\xi_{js}+u_{js}) \end{pmatrix} &= E\begin{pmatrix} \xi_{it}\xi_{is} & \xi_{it}(\xi_{is}+u_{is}) \\ (\xi_{it}+u_{it})\xi_{is} & (\xi_{it}+u_{it})(\xi_{is}+u_{is}) \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}
$$

3. $i \neq j, t = s$

$$E\begin{pmatrix} \xi_{it}\xi_{js} & \xi_{it}(\xi_{js}+u_{js}) \\ (\xi_{it}+u_{it})\xi_{js} & (\xi_{it}+u_{it})(\xi_{js}+u_{js}) \end{pmatrix} = E\begin{pmatrix} \xi_{it}\xi_{jt} & \xi_{it}(\xi_{jt}+u_{jt}) \\ (\xi_{it}+u_{it})\xi_{jt} & (\xi_{it}+u_{it})(\xi_{jt}+u_{jt}) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & E(u_{it}u_{jt}) \end{pmatrix}.$$

4. $i \neq j, t \neq s$

$$E\begin{pmatrix} \xi_{it}\xi_{js} & \xi_{it}(\xi_{js}+u_{js}) \\ (\xi_{it}+u_{it})\xi_{js} & (\xi_{it}+u_{it})(\xi_{js}+u_{js}) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

To estimate $E(u_{it}^2)$, note that u_{it} can be obtained by taking the product of the i-th row of $(I_N - \mu\mathbf{W}_t^u)^{-1}$ and v_t, since $u_t = (\mathbf{I}_N - \mu\mathbf{W}_t^u)^{-1}v_t$. Then, we have:

$$\begin{aligned} E(u_{it}^2) &= E([(I_N - \mu\mathbf{W}_t^u)^{-1}]_i v_t v_t'([(I_N - \mu\mathbf{W}_t^u)^{-1}]_i)']) \\ &= [(I_N - \mu\mathbf{W}_t^u)^{-1}]_i E[v_t v_t']([(I_N - \mu\mathbf{W}_t^u)^{-1}]_i)' \\ &= \sigma_v^2[(I_N - \mu\mathbf{W}_t^u)^{-1}]_i([(I_N - \mu\mathbf{W}_t^u)^{-1}]_i)', \end{aligned}$$

where $[X]_i$ refers to the i-th row of matrix X and $E(v_t v_t') = \sigma_v^2\mathbf{I}_N$. Similarly, $E(u_{it}u_{jt})$ is given by:

$$\begin{aligned} E(u_{it}u_{jt}) &= E([(I_N - \mu\mathbf{W}_t^u)^{-1}]_i v_t v_t'([(I_N - \mu\mathbf{W}_t^u)^{-1}]_j)']) \\ &= [(I_N - \mu\mathbf{W}_t^u)^{-1}]_i E[v_t v_t']([(I_N - \mu\mathbf{W}_t^u)^{-1}]_j)' \\ &= \sigma_v^2[(I_N - \mu\mathbf{W}_t^u)^{-1}]_i([(I_N - \mu\mathbf{W}_t^u)^{-1}]_j)'. \end{aligned}$$

Substituting all the terms back into $\mathbf{V}_{\boldsymbol{\theta}}$, we obtain:

$$
\begin{aligned}
\mathbf{V}_{\boldsymbol{\theta}} =& \frac{1}{N(T-1)} \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{s=2}^{T} \mathcal{Z}'_{it} E \begin{pmatrix} \xi_{it}\xi_{js} & \xi_{it}(\xi_{js}+u_{js}) \\ (\xi_{it}+u_{it})\xi_{js} & (\xi_{it}+u_{it})(\xi_{js}+u_{js}) \end{pmatrix} \mathcal{Z}_{js} \\
=& \frac{1}{N(T-1)} \sum_{i=1}^{N}\sum_{t=2}^{T} \begin{pmatrix} \mathcal{Z}'_{it,I} & 0 \\ 0 & \mathcal{Z}'_{it,II} \end{pmatrix} E \begin{pmatrix} \xi_{it}^2 & \xi_{it}^2 \\ \xi_{it}^2 & \xi_{it}^2 + u_{it}^2 \end{pmatrix} \begin{pmatrix} \mathcal{Z}_{it,I} & 0 \\ 0 & \mathcal{Z}_{it,II} \end{pmatrix} \\
&+ \frac{1}{N(T-1)} \sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=2}^{T} \begin{pmatrix} \mathcal{Z}'_{it,I} & 0 \\ 0 & \mathcal{Z}'_{it,II} \end{pmatrix} E \begin{pmatrix} 0 & 0 \\ 0 & u_{it}u_{jt} \end{pmatrix} \begin{pmatrix} \mathcal{Z}_{it,I} & 0 \\ 0 & \mathcal{Z}_{it,II} \end{pmatrix} \\
=& \frac{1}{N(T-1)} \sum_{i=1}^{N}\sum_{t=2}^{T} \begin{pmatrix} \mathcal{Z}'_{it,I}\mathcal{Z}_{it,I}E(\xi_{it}^2) & \mathcal{Z}'_{it,I}\mathcal{Z}_{it,II}E(\xi_{it}^2) \\ \mathcal{Z}'_{it,II}\mathcal{Z}_{it,I}E(\xi_{it}^2) & \mathcal{Z}'_{it,II}\mathcal{Z}_{it,II}E(\xi_{it}^2) \end{pmatrix} \\
&+ \frac{1}{N(T-1)} \sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \begin{pmatrix} \mathcal{Z}'_{it,I} & 0 \\ 0 & \mathcal{Z}'_{it,II} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \sigma_v^2[(I_N-\mu\mathbf{W}_t^u)^{-1}]_i([(I_N-\mu\mathbf{W}_t^u)^{-1}]_j)' \end{pmatrix} \begin{pmatrix} \mathcal{Z}_{it,I} & 0 \\ 0 & \mathcal{Z}_{it,II} \end{pmatrix} \\
=& \frac{1}{N(T-1)} \sum_{i=1}^{N}\sum_{t=2}^{T} \begin{pmatrix} \mathcal{Z}'_{it,I}\mathcal{Z}_{it,I}E(\xi_{it}^2) & \mathcal{Z}'_{it,I}\mathcal{Z}_{it,II}E(\xi_{it}^2) \\ \mathcal{Z}'_{it,II}\mathcal{Z}_{it,I}E(\xi_{it}^2) & \mathcal{Z}'_{it,II}\mathcal{Z}_{it,II}E(\xi_{it}^2) \end{pmatrix} \\
&+ \frac{1}{N(T-1)} \sum_{t=2}^{T} \begin{pmatrix} \mathbf{0}_{\mathcal{M}_1\times\mathcal{M}_1} & \mathbf{0}_{\mathcal{M}_1\times\mathcal{M}_2} \\ \mathbf{0}_{\mathcal{M}_2\times\mathcal{M}_1} & \sigma_v^2\mathcal{Z}'_{t,II}[(\mathbf{I}_N-\mu\mathbf{W}_t^u)^{-1}(\mathbf{I}_N-\mu\mathbf{W}_t^u)^{-1\prime}]\mathcal{Z}_{t,II} \end{pmatrix},
\end{aligned}
$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ are the number of moment conditions (instruments) used in the first and second stages respectively (such that $\mathcal{M}_1+\mathcal{M}_2=\mathcal{M}$), and $\mathcal{Z}_{t,II} = \begin{pmatrix} Z_{1t,II} \\ Z_{2t,II} \\ \vdots \\ Z_{Nt,II} \end{pmatrix}$ is a $N\times\mathcal{M}_2$ matrix.

We can estimate $\mathbf{V}_{\boldsymbol{\theta}}$ by its sample counterpart:

$$
\begin{aligned}
\widehat{\mathbf{V}}_{\boldsymbol{\theta}} =& \frac{1}{N(T-1)} \sum_{i=1}^{N}\sum_{t=2}^{T} \begin{pmatrix} \mathcal{Z}'_{it,I}\mathcal{Z}_{it,I}\widehat{\xi}_{it}^2 & \mathcal{Z}'_{it,I}\mathcal{Z}_{it,II}\widehat{\xi}_{it}^2 \\ \mathcal{Z}'_{it,II}\mathcal{Z}_{it,I}\widehat{\xi}_{it}^2 & \mathcal{Z}'_{it.II}\mathcal{Z}_{it,II}\widehat{\xi}_{it}^2 \end{pmatrix} \\
&+ \frac{1}{N(T-1)} \sum_{t=2}^{T} \boldsymbol{\mathcal{Z}}'_t \begin{pmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \widehat{\sigma_v^2}[(\mathbf{I}_N-\widehat{\mu}\mathbf{W}_t^u)^{-1}(\mathbf{I}_N-\widehat{\mu}\mathbf{W}_t^u)^{-1\prime}]) \end{pmatrix} \boldsymbol{\mathcal{Z}}_t,
\end{aligned}
$$

where $\widehat{\xi}_{it}$, $\widehat{\mu}$ and $\widehat{\sigma_v^2}$ are obtained from Steps 1–2 in Section 3.3.

A.3 Deriving the Variance-Covariance Matrix for the Moment Conditions in Stage 3

In this section, we derive the variance-covariance matrix for the moment conditions used in Stage 3. Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ non-stochastic matrices. For a $n \times 1$ random vector $\mathbf{e}$ with mean 0, variance σ_e^2 and finite excess kurtosis κ_e:

$$\text{cov}(\mathbf{e}'\mathbf{A}\mathbf{e}, \mathbf{e}'\mathbf{B}\mathbf{e}) = \sigma_e^4 \kappa_e a'b + \sigma_e^4 \text{tr}(\mathbf{A}(\mathbf{B}' + \mathbf{B})), \tag{36}$$

where $a = \text{diagv}(\mathbf{A})$ and $b = \text{diagv}(\mathbf{B})$. The operator 'diagv' takes the diagonal elements of a matrix and converts them to a column vector.

Define $\mathbf{v} = [v_2', v_3', \ldots, v_T']$ to be a $N(T-1) \times 1$ vector, and $\overline{\mathbf{v}} = \mathbf{W}^u\mathbf{v}$. Let κ_v be the finite excess kurtosis of $\mathbf{v}$. The following computes each cell of the variance-covariance matrix of the moment conditions in Equation (6):

$$\begin{aligned}
V_{\psi,11} &= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{v}, \frac{1}{N(T-1)}\mathbf{v}'\mathbf{v}\right) \\
&= N(T-1)\frac{\sigma_v^4}{(N(T-1))^2}[\kappa_v \iota_{N(T-1)}' \iota_{N(T-1)} + \text{tr}(\mathbf{I}_{N(T-1)}(\mathbf{I}_{N(T-1)}' + \mathbf{I}_{N(T-1)}))] \\
&= \sigma_v^4[\kappa_v + 2], \\
V_{\psi,12} &= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{v}, \frac{1}{N(T-1)}\overline{\mathbf{v}}'\overline{\mathbf{v}}\right) \\
&= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{v}, \frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^{u\prime}\mathbf{W}^u\mathbf{v}\right) \\
&= \frac{\sigma_v^4}{N(T-1)}[\kappa_v \text{tr}(\mathbf{W}^{u\prime}\mathbf{W}^u) + \text{tr}(\mathbf{I}_{N(T-1)}(\mathbf{W}^{u\prime}\mathbf{W}^u + \mathbf{W}^u\mathbf{W}^{u\prime}))] \\
&= \frac{\sigma_v^4}{N(T-1)}[\kappa_v + 2]\text{tr}(\mathbf{W}^{u\prime}\mathbf{W}^u), \\
V_{\psi,13} &= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{v}, \frac{1}{N(T-1)}\mathbf{v}'\overline{\mathbf{v}}\right) \\
&= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{v}, \frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^{u\prime}\mathbf{v}\right) \\
&= \frac{\sigma_v^4}{N(T-1)}[\kappa_v \text{tr}(\mathbf{W}^u) + \text{tr}(\mathbf{W}^u)] \\
&= 0,
\end{aligned}$$

$$
\begin{aligned}
V_{\psi,22} &= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\overline{\mathbf{v}}'\overline{\mathbf{v}}, \frac{1}{N(T-1)}\overline{\mathbf{v}}'\overline{\mathbf{v}}\right) \\
&= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^{u\prime}\mathbf{W}^u\mathbf{v}, \frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^{u\prime}\mathbf{W}^u\mathbf{v}\right) \\
&= \frac{\sigma_v^4}{N(T-1)}[\kappa_v \text{ diagv}(\mathbf{W}^{u\prime}\mathbf{W}^u)' \text{ diagv}(\mathbf{W}^{u\prime}\mathbf{W}^u)) + \text{tr}((\mathbf{W}^{u\prime}\mathbf{W}^u)(\mathbf{W}^{u\prime}\mathbf{W}^u + \mathbf{W}^u\mathbf{W}^{u\prime}))], \\
V_{\psi,23} &= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\overline{\mathbf{v}}'\overline{\mathbf{v}}, \frac{1}{N(T-1)}\mathbf{v}'\overline{\mathbf{v}}\right) \\
&= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^{u\prime}\mathbf{W}^u\mathbf{v}, \frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^u\mathbf{v}\right) \\
&= \frac{\sigma_v^4}{N(T-1)}[\kappa_v \text{ diagv}(\mathbf{W}^{u\prime}\mathbf{W}^u)' \text{ diagv}(\mathbf{W}^u) + \text{tr}((\mathbf{W}^{u\prime}\mathbf{W}^u)(\mathbf{W}^u + \mathbf{W}^{u\prime}))] \\
&= \frac{\sigma_v^4}{N(T-1)}\text{tr}((\mathbf{W}^{u\prime}\mathbf{W}^u)(\mathbf{W}^u + \mathbf{W}^{u\prime})), \\
V_{\psi,32} &= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\overline{\mathbf{v}}, \frac{1}{N(T-1)}\overline{\mathbf{v}}'\overline{\mathbf{v}}\right) \\
&= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^u\mathbf{v}, \frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^{u\prime}\mathbf{W}^u\mathbf{v}\right) \\
&= \frac{\sigma_v^4}{N(T-1)}[\kappa_v \text{ diagv}(\mathbf{W}^u)' \text{ diagv}(\mathbf{W}^{u\prime}\mathbf{W}^u) + \text{tr}(\mathbf{W}^u(\mathbf{W}^{u\prime}\mathbf{W}^u + \mathbf{W}^{u\prime}\mathbf{W}^u))] \\
&= 2\frac{\sigma_v^4}{N(T-1)}\text{tr}(\mathbf{W}^u\mathbf{W}^{u\prime}\mathbf{W}^u), \\
V_{\psi,33} &= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\overline{\mathbf{v}}, \frac{1}{N(T-1)}\mathbf{v}'\overline{\mathbf{v}}\right) \\
&= N(T-1)\text{cov}\left(\frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^u\mathbf{v}, \frac{1}{N(T-1)}\mathbf{v}'\mathbf{W}^u\mathbf{v}\right) \\
&= \frac{\sigma_v^4}{N(T-1)}[\kappa_v \text{ diagv}(\mathbf{W}^u)' \text{ diagv}(\mathbf{W}^u) + \text{tr}(\mathbf{W}^u(\mathbf{W}^u + \mathbf{W}^{u\prime}))] \\
&= \frac{\sigma_v^4}{N(T-1)}\text{tr}(\mathbf{W}^u(\mathbf{W}^u + \mathbf{W}^{u\prime})).
\end{aligned}
$$

B Simulation Appendix

B.1 Simulation of Connectivity Matrices

The BSJBSA-TSR linked dataset provides the distribution of the number of customers (and respectively suppliers) that a firm has, up to 24 customers (and suppliers). We assign a time-invariant random number for each firm, r_i, which is uniformly distributed in $[0, 1]$, for $i \in \{1, 2, \ldots, N\}$. For the initial period, we use a weakly monotonic mapping function, $q_t^{num}(\cdot)$, to map the firm random number $r_i \in [0, 1]$ to the number of customers, given the empirical distribution. In other words, $q_t^{num}(r_i) = num_{it}$, where $q_t^{num}(\cdot)$ is the inverse of the empirical distribution function of the number of customers in period t. Given the number of customers assigned to each firm in the initial period, we randomly draw its customers from the pool of firms. Subsequently, given the mapping from the firm random number to the number of customers that firm i has at time t, $q_t^{num}(r_i) = num_{it}$, we randomly drop firms from the set of customers that a firm initially has in the previous period if $num_{it} < num_{i,t-1} * persistency_{t-1}$, where $persistency_{t-1}$ is the fraction of firm-to-firm relationships in period $t-1$ that survive in period t as observed in the data. Alternatively, we add firms (randomly drawn from the pool of unrelated firms) to the set of customers that a firm has in the previous period after attrition (the identity of the connections dropped being randomly drawn from the pool of existing customers of a firm) if $num_{it} > num_{i,t-1} * persistency_{t-1}$. The number of suppliers that a firm has across time is simulated in similar manner.

To generate the connectivity matrix based on common commuting zone, we target the average density of connections per firm. In particular, start with the scenario of $N = 500$. Given data on the distribution of firms across commuting zones, we use the inverse of the empirical distribution function of commuting zones, $q_t^{cz}(\cdot)$, to map each firm $r_i \in [0, 1]$ to commuting zone in each period, such that $q_t^{cz}(r_i) = cz_{it}$. We then generate the connectivity matrix based on common commuting zone. The ij-th element of the matrix is set equal to 1, if firms i and j are located in the same commuting zone in period t and 0 otherwise. We compute the average density of connections per firm, that is, the proportion of simulated connections per firm on average relative to the maximum number of possible connections across the 500 firms in period t. Denote the simulated proportion of connections per firm on average as d_t. As we vary the size N to the case of $N = 750$ or $N = 1000$, we target the same d_t. Specifically, we generate random numbers $r_{ij,t} \in [0, 1]$ from a uniform distribution for each firm-pair and time period. If $r_{ij,t} > 1 - d_t$, then the ij-th element of the connectivity matrix is set equal to 1, and 0 otherwise.

B.2 Simulation of Input and Output Variables

Based on the BSJBSA-TSR linked dataset, we obtain the mean and standard deviation of labor input (and respectively, capital) across firms in each year from 2009 to 2018. We then simulate the usage of labor input (and respectively, capital) for each firm, by drawing randomly from Normal distributions that have the same mean and standard deviation as empirically observed specific to each input variable and year.

For the Monte Carlo simulations, we adopt a Leontief production function as in Ackerberg, Caves and Frazer (2015) such that:

$$VA_{it} = \min\{e^{\alpha_0} L_{it}^{\alpha_l} K_{it}^{\alpha_k} e^{\omega_{it}}, e^{\alpha_m} M_{it}\} e^{\xi_{it}}, \tag{37}$$

which gives rise to the following relationship between material inputs and productivity after taking logs:

$$\alpha_m + m_{it} = \alpha_0 + \alpha_l l_{it} + \alpha_k k_{it} + \omega_{it}. \tag{38}$$

Setting $e^{\alpha_m} = 1$ as in ACF, we have: $m_{it} = \alpha_0 + \alpha_l l_{it} + \alpha_k k_{it} + \omega_{it}$. The logged output, $y_{it} = \ln Y_{it}$, is then derived using the sum of value-added and material inputs: $y_{it} = \ln(VA_{it} + M_{it})$.

B.3 Simulation Procedure

Given simulated data on $\{l_{it}\}_{i=1,t=1}^{i=N,t=T}$, $\{k_{it}\}_{i=1,t=1}^{i=N,t=T}$, $\{\mathbf{W}_t^y\}_{t=1}^{T-1}$, $\{\mathbf{W}_t^l\}_{t=1}^{T-1}$ and $\{\mathbf{W}_t^u\}_{t=1}^{T}$ and the parameter values for $\{\alpha_0, \alpha_l, \alpha_k, \lambda, \beta_l, \rho_1, \mu, \sigma_\xi, \sigma_v\}$, the data used for the simulations are generated as follows:

1. Set $\omega_{i,t-1} = 0$, for $t = 1$.

2. Generate $va_{i,t-1}$ based on the simulated $l_{i,t-1}$ and $k_{i,t-1}$, the parameter values for $\{\alpha_0, \alpha_l, \alpha_k\}$, the productivity $\omega_{i,t-1}$, and the random draw of $\xi_{i,t-1}$ from a Normal distribution with mean 0 and variance σ_ξ^2.

3. Set $m_{i,t-1}$ according to Equation (38) and derive $y_{i,t-1} = \ln(VA_{i,t-1} + M_{i,t-1})$.

4. Generate ω_{it} based on Equations (2)–(3), given $y_{i,t-1}$, simulated data on $\{\mathbf{W}_t^y\}_{t=1}^{T-1}$, $\{\mathbf{W}_t^l\}_{t=1}^{T-1}$ and $\{\mathbf{W}_t^u\}_{t=1}^{T}$, the parameter values for $\{\lambda, \beta_l, \rho_1, \mu\}$, and the random draw of v_{it} from a Normal distribution with mean 0 and variance σ_v^2.

5. Iterate Steps 2–4 for $t = 1, 2, \ldots, T$ to generate simulated data on $\{va_{it}\}_{i=1,t=1}^{i=N,t=T}$, $\{m_{it}\}_{i=1,t=1}^{i=N,t=T}$, $\{y_{it}\}_{i=1,t=1}^{i=N,t=T}$, and $\{\omega_{it}\}_{i=1,t=1}^{i=N,t=T}$.

Table 1: BSJBSA and TSR Matching Percentage

Sample	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
# firms in BSJBSA	29096	29570	30647	30584	30217	30180	30231	30151	29530	29780
# TSR firms matched	26947	27559	28486	28557	28237	28263	28196	28448	27715	27978
Percentage	92.61	93.20	92.95	93.37	93.45	93.65	93.27	94.35	93.85	93.95

Notes: This table reports the percentage of firms in BSJBSA that are matched with its counterpart in TSR. The BSJBSA set of firms is used as the denominator, since it provides the required firm-level variables for productivity estimations.

Table 2: Summary Statistics (for 2015 cross section)

(a) BSJBSA-TSR Linked Sample

	Observations	Mean	Std	Min	Max
Labor headcounts	29044	458.93	1826.51	0	130725
Labor hours	29044	794921.01	3111792	0	206818495
Real capital	28861	5149.71	53513.25	0	4275886
Real spending on intermediate inputs	29044	14004.56	106970.4	0	7086646
Real revenue	29044	21954.38	137121.6	0.93	1.06+07
Real valued added	29044	3128.06	16631.27	-208309.8	1411724
Number of customers	27788	6.77	5,54	0	24
Number of suppliers	27788	6.73	4.75	0	24
Number of customers existing in BSJBSA	27788	4.09	3.95	0	23
Number of suppliers existing in BSJBSA	27788	3.73	3.17	0	22

(b) Estimation Sample

	Observations	Mean	Std	Min	Max
Labor headcounts	12525	540.40	2040.51	3	81740
Labor hours	12525	948590.6	3602148	5621.48	162990234
Real capital	12525	5789.81	36734.81	0.13	1849815
Real spending on intermediate inputs	12525	23700.94	162126.7	12.36	9208361
Real revenue	12525	27879.81	177634.3	123.84	1.06+07
Real valued added	12525	4178.87	20506.37	0.48	1411722
Number of customers	12414	7.24	5.74	0	24
Number of suppliers	12414	7.35	4.92	0	24
Number of customers existing in BSJBSA	12414	4.50	4.15	0	23
Number of suppliers existing in BSJBSA	12414	4.24	3.36	0	22

Notes: Refer to Section 4 for the measurement of the variables. The revenue, value added, capital, and intermediate inputs are in million Japanese yens. The number of customers / suppliers is identified by the BSJBSA firm's corresponding entries in TSR, which provides the list of a firm's top 24 customers / suppliers. The number of customers / suppliers existing in BSJBSA refers to the subset of a firm's top 24 customers / suppliers listed in TSR that also have firm-level information in BSJBSA.

Table 3: DGP1 – No Spatial Dependence in Productivity

Estimator	N	T-1	Stat.	α_l 0.6	α_k 0.4	λ 0	β_l 0	ρ_1 0.5	μ 0	σ_v^2 0.49
WGMM	500	9	Mean	0.6004	0.3997	-	-	0.4993	-	-
			SD	(0.0070)	(0.0102)	-	-	(0.0139)	-	-
			SE	(0.0071)	(0.0105)	-	-	(0.0140)	-	-
	500	18	Mean	0.6001	0.4000	-	-	0.4995	-	-
			SD	(0.0050)	(0.0073)	-	-	(0.0095)	-	-
			SE	(0.0050)	(0.0075)	-	-	(0.0095)	-	-
	750	9	Mean	0.6004	0.3992	-	-	0.4996	-	-
			SD	(0.0059)	(0.0085)	-	-	(0.0119)	-	-
			SE	(0.0058)	(0.0087)	-	-	(0.0114)	-	-
	750	18	Mean	0.6001	0.3998	-	-	0.4998	-	-
			SD	(0.0041)	(0.0059)	-	-	(0.0077)	-	-
			SE	(0.0041)	(0.0061)	-	-	(0.0077)	-	-
	1000	9	Mean	0.6001	0.4001	-	-	0.4996	-	-
			SD	(0.0050)	(0.0073)	-	-	(0.0097)	-	-
			SE	(0.0049)	(0.0073)	-	-	(0.0099)	-	-
	1000	18	Mean	0.5999	0.4001	-	-	0.4998	-	-
			SD	(0.0040)	(0.0051)	-	-	(0.0069)	-	-
			SE	(0.0040)	(0.0053)	-	-	(0.0067)	-	-
SGMM	500	9	Mean	0.6001	0.3997	0.0000	0.0001	0.4991	-0.0010	0.4885
			SD	(0.0211)	(0.0105)	(0.0022)	(0.0109)	(0.0139)	(0.0360)	(0.0118)
			SE	(0.0213)	(0.0110)	(0.0022)	(0.0110)	(0.0140)	(0.0353)	(0.0103)
	500	18	Mean	0.6013	0.4001	0.0001	-0.0001	0.4994	-0.0008	0.4895
			SD	(0.0157)	(0.0075)	(0.0016)	(0.0079)	(0.0095)	(0.0243)	(0.0083)
			SE	(0.0152)	(0.0077)	(0.0016)	(0.0078)	(0.0095)	(0.0254)	(0.0072)
	750	9	Mean	0.6003	0.3993	0.0001	0.0000	0.4995	-0.0011	0.4899
			SD	(0.0169)	(0.0090)	(0.0020)	(0.0088)	(0.0119)	(0.0283)	(0.0097)
			SE	(0.0172)	(0.0090)	(0.0019)	(0.0087)	(0.0115)	(0.0284)	(0.0084)
	750	18	Mean	0.6000	0.3997	0.0001	0.0000	0.4998	-0.0005	0.4898
			SD	(0.0118)	(0.0061)	(0.0014)	(0.0060)	(0.0077)	(0.0213)	(0.0067)
			SE	(0.0122)	(0.0063)	(0.0013)	(0.0062)	(0.0089)	(0.0207)	(0.0059)
	1000	9	Mean	0.6002	0.4001	-0.0001	0.0000	0.4995	0.0005	0.4894
			SD	(0.0143)	(0.0076)	(0.0017)	(0.0073)	(0.0097)	(0.0240)	(0.0084)
			SE	(0.0142)	(0.0077)	(0.0018)	(0.0073)	(0.0099)	(0.0242)	(0.0073)
	1000	18	Mean	0.5997	0.4000	0.0000	0.0001	0.4997	-0.0002	0.4899
			SD	(0.0097)	(0.0053)	(0.0012)	(0.0050)	(0.0069)	(0.0173)	(0.0059)
			SE	(0.0099)	(0.0055)	(0.0012)	(0.0051)	(0.0067)	(0.0177)	(0.0051)

Notes: For each DGP, 1000 simulated samples are drawn and estimated. We report the mean (Mean) and the standard deviation (SD) of the parameter point estimates across the 1000 Monte Carlo simulations, together with the estimated standard errors (SE) derived from the variance-covariance matrices of the estimators. The exact parameter values used in the DGPs are listed in the first row of the table.

Table 4: DGP2 – Spatial Dependence in Productivity via Lagged Outputs and Lagged Labor Inputs of Related Firms

Estimator	N	T-1	Stat.	α_l 0.6	α_k 0.4	λ 0.01	β_l 0.01	ρ_1 0.5	μ 0	σ_v^2 0.49
WGMM			Mean	0.6233	0.4045	-	-	0.5014	-	-
	500	9	SD	(0.0070)	(0.0103)	-	-	(0.0137)	-	-
			SE	(0.0071)	(0.0106)	-	-	(0.0138)	-	-
			Mean	0.6232	0.4047	-	-	0.5022	-	-
	500	18	SD	(0.0050)	(0.0073)	-	-	(0.0095)	-	-
			SE	(0.0050)	(0.0075)	-	-	(0.0094)	-	-
			Mean	0.6230	0.4026	-	-	0.5011	-	-
	750	9	SD	(0.0059)	(0.0085)	-	-	(0.0118)	-	-
			SE	(0.0059)	(0.0088)	-	-	(0.0113)	-	-
			Mean	0.6231	0.4027	-	-	0.5016	-	-
	750	18	SD	(0.0041)	(0.0059)	-	-	(0.0078)	-	-
			SE	(0.0041)	(0.0061)	-	-	(0.0077)	-	-
			Mean	0.6219	0.4033	-	-	0.5003	-	-
	1000	9	SD	(0.0050)	(0.0074)	-	-	(0.0098)	-	-
			SE	(0.0049)	(0.0073)	-	-	(0.0098)	-	-
			Mean	0.6221	0.4029	-	-	0.5011	-	-
	1000	18	SD	(0.0035)	(0.0051)	-	-	(0.0069)	-	-
			SE	(0.0035)	(0.0053)	-	-	(0.0067)	-	-
SGMM			Mean	0.6003	0.3996	0.0100	0.0101	0.4990	-0.0012	0.4886
	500	9	SD	(0.0209)	(0.0106)	(0.0021)	(0.0108)	(0.0137)	(0.0359)	(0.0117)
			SE	(0.0213)	(0.0110)	(0.0022)	(0.0110)	(0.0138)	(0.0353)	(0.0103)
			Mean	0.6015	0.4001	0.0100	0.0092	0.4994	-0.0008	0.4896
	500	18	SD	(0.0156)	(0.0075)	(0.0015)	(0.0079)	(0.0093)	(0.0242)	(0.0083)
			SE	(0.0152)	(0.0077)	(0.0015)	(0.0078)	(0.0094)	(0.0254)	(0.0072)
			Mean	0.6004	0.3992	0.0101	0.0100	0.4995	-0.0009	0.4899
	750	9	SD	(0.0169)	(0.0089)	(0.0019)	(0.0088)	(0.0117)	(0.0282)	(0.0098)
			SE	(0.0172)	(0.0090)	(0.0019)	(0.0087)	(0.0113)	(0.0284)	(0.0084)
			Mean	0.6001	0.3998	0.0101	0.0100	0.4998	-0.0005	0.4898
	750	18	SD	(0.0118)	(0.0061)	(0.0014)	(0.0060)	(0.0077)	(0.0213)	(0.0067)
			SE	(0.0122)	(0.0063)	(0.0013)	(0.0062)	(0.0077)	(0.0207)	(0.0059)
			Mean	0.6002	0.4001	0.0099	0.0100	0.4995	0.0004	0.4893
	1000	9	SD	(0.0143)	(0.0077)	(0.0017)	(0.0073)	(0.0097)	(0.0240)	(0.0084)
			SE	(0.0142)	(0.0077)	(0.0018)	(0.0073)	(0.0098)	(0.0242)	(0.0073)
			Mean	0.5997	0.4000	0.0100	0.0101	0.4998	-0.0002	0.4898
	1000	18	SD	(0.0097)	(0.0054)	(0.0012)	(0.0050)	(0.0068)	(0.0173)	(0.0059)
			SE	(0.0100)	(0.0055)	(0.0012)	(0.0051)	(0.0067)	(0.0177)	(0.0051)

Notes: For each DGP, 1000 simulated samples are drawn and estimated. We report the mean (Mean) and the standard deviation (SD) of the parameter point estimates across the 1000 Monte Carlo simulations, together with the estimated standard errors (SE) derived from the variance-covariance matrices of the estimators. The exact parameter values used in the DGPs are listed in the first row of the table.

Table 5: DGP3 – Spatial Dependence in Productivity via Lagged Outputs and Lagged Labor Inputs of Related Firms, and via Productivity Shocks

Estimator	N	T-1	Stat.	α_l 0.6	α_k 0.4	λ 0.01	β_l 0.01	ρ_1 0.5	μ 0.25	σ_v^2 0.49
WGMM	500	9	Mean	0.6233	0.4045	-	-	0.5012	-	-
			SD	(0.0071)	(0.0104)	-	-	(0.0139)	-	-
			SE	(0.0071)	(0.0106)	-	-	(0.0138)	-	-
	500	18	Mean	0.6232	0.4047	-	-	0.5022	-	-
			SD	(0.0050)	(0.0074)	-	-	(0.0095)	-	-
			SE	(0.0050)	(0.0076)	-	-	(0.0094)	-	-
	750	9	Mean	0.6230	0.4026	-	-	0.5011	-	-
			SD	(0.0060)	(0.0086)	-	-	(0.0120)	-	-
			SE	(0.0059)	(0.0088)	-	-	(0.0113)	-	-
	750	18	Mean	0.6185	0.4013	-	-	0.5054	-	-
			SD	(0.0054)	(0.0035)	-	-	(0.0099)	-	-
			SE	(0.0057)	(0.0037)	-	-	(0.0094)	-	-
	1000	19	Mean	0.6219	0.4033	-	-	0.5003	-	-
			SD	(0.0051)	(0.0074)	-	-	(0.0099)	-	-
			SE	(0.0050)	(0.0074)	-	-	(0.0098)	-	-
	1000	18	Mean	0.6221	0.4030	-	-	0.5011	-	-
			SD	(0.0036)	(0.0052)	-	-	(0.0070)	-	-
			SE	(0.0036)	(0.0053)	-	-	(0.0067)	-	-
SGMM	500	9	Mean	0.6004	0.3996	0.0100	0.0100	0.4988	0.2483	0.4885
			SD	(0.0213)	(0.0107)	(0.0021)	(0.0110)	(0.0139)	(0.0359)	(0.0117)
			SE	(0.0215)	(0.0110)	(0.0022)	(0.0111)	(0.0140)	(0.0353)	(0.0103)
	500	18	Mean	0.6015	0.4001	0.0101	0.0093	0.4993	0.2489	0.4896
			SD	(0.0158)	(0.0075)	(0.0015)	(0.0079)	(0.0094)	(0.0240)	(0.0082)
			SE	(0.0153)	(0.0077)	(0.0015)	(0.0078)	(0.0095)	(0.0254)	(0.0072)
	750	9	Mean	0.6003	0.3992	0.0101	0.0100	0.4994	0.2486	0.4899
			SD	(0.0170)	(0.0090)	(0.0019)	(0.0088)	(0.0119)	(0.0282)	(0.0098)
			SE	(0.0172)	(0.0090)	(0.0019)	(0.0088)	(0.0115)	(0.0284)	(0.0084)
	750	18	Mean	0.6000	0.3997	0.0101	0.0100	0.4997	0.2493	0.4898
			SD	(0.0118)	(0.0062)	(0.0014)	(0.0060)	(0.0077)	(0.0213)	(0.0066)
			SE	(0.0122)	(0.0063)	(0.0013)	(0.0062)	(0.0078)	(0.0207)	(0.0059)
	1000	9	Mean	0.6002	0.4001	0.0099	0.0101	0.4994	0.2501	0.4894
			SD	(0.0143)	(0.0077)	(0.0017)	(0.0074)	(0.0099)	(0.0240)	(0.0083)
			SE	(0.0143)	(0.0077)	(0.0018)	(0.0074)	(0.0099)	(0.0242)	(0.0074)
	1000	18	Mean	0.5997	0.4000	0.0100	0.0102	0.4998	0.2498	0.4899
			SD	(0.0098)	(0.0054)	(0.0012)	(0.0050)	(0.0070)	(0.0173)	(0.0058)
			SE	(0.0100)	(0.0055)	(0.0012)	(0.0051)	(0.0067)	(0.0177)	(0.0051)

Notes: For each DGP, 1000 simulated samples are drawn and estimated. We report the mean (Mean) and the standard deviation (SD) of the parameter point estimates across the 1000 Monte Carlo simulations, together with the estimated standard errors (SE) derived from the variance-covariance matrices of the estimators. The exact parameter values used in the DGPs are listed in the first row of the table.

Table 6: DGP4 – Stronger Spatial Dependence in Productivity via Lagged Outputs and Lagged Labor Inputs of Related Firms, and via Productivity Shocks

Estimator	N	T	Var	α_l 0.6	α_k 0.4	λ 0.1	β_l 0.1	ρ_1 0.5	μ 0.25	σ_v^2 0.49
WGMM	500	9	Mean	0.8358	0.4801	-	-	0.5938	-	-
			SD	(0.0077)	(0.0124)	-	-	(0.0079)	-	-
			SE	(0.0101)	(0.0172)	-	-	(0.0093)	-	-
	500	18	Mean	0.8277	0.4949	-	-	0.6280	-	-
			SD	(0.0055)	(0.0089)	-	-	(0.0062)	-	-
			SE	(0.0076)	(0.0130)	-	-	(0.0069)	-	-
	750	9	Mean	0.8337	0.4492	-	-	0.5688	-	-
			SD	(0.0063)	(0.0096)	-	-	(0.0072)	-	-
			SE	(0.0080)	(0.0132)	-	-	(0.0080)	-	-
	750	18	Mean	0.8329	0.4513	-	-	0.5945	-	-
			SD	(0.0045)	(0.0070)	-	-	(0.0054)	-	-
			SE	(0.0058)	(0.0098)	-	-	(0.0059)	-	-
	1000	9	Mean	0.8247	0.4408	-	-	0.5457	-	-
			SD	(0.0053)	(0.0083)	-	-	(0.0066)	-	-
			SE	(0.0064)	(0.0105)	-	-	(0.0071)	-	-
	1000	18	Mean	0.8232	0.4478	-	-	0.5772	-	-
			SD	(0.0038)	(0.0061)	-	-	(0.0051)	-	-
			SE	(0.0048)	(0.0080)	-	-	(0.0053)	-	-
SGMM	500	9	Mean	0.6002	0.3996	0.1000	0.1002	0.4993	0.2480	0.4885
			SD	(0.0212)	(0.0108)	(0.0018)	(0.0110)	(0.0074)	(0.0360)	(0.0108)
			SE	(0.0214)	(0.0110)	(0.0018)	(0.0111)	(0.0078)	(0.0352)	(0.0103)
	500	18	Mean	0.6015	0.4001	0.1000	0.0993	0.4998	0.2488	0.4896
			SD	(0.0157)	(0.0075)	(0.0013)	(0.0080)	(0.0058)	(0.0242)	(0.0077)
			SE	(0.0153)	(0.0077)	(0.0013)	(0.0079)	(0.0057)	(0.0254)	(0.0072)
	750	9	Mean	0.6004	0.3992	0.1001	0.1000	0.5001	0.2484	0.4899
			SD	(0.0170)	(0.0090)	(0.0016)	(0.0088)	(0.0067)	(0.0282)	(0.0091)
			SE	(0.0172)	(0.0090)	(0.0016)	(0.0088)	(0.0067)	(0.0283)	(0.0084)
	750	18	Mean	0.6000	0.3998	0.1001	0.1001	0.4998	0.2492	0.4898
			SD	(0.0118)	(0.0061)	(0.0011)	(0.0061)	(0.0048)	(0.0213)	(0.0062)
			SE	(0.0122)	(0.0063)	(0.0011)	(0.0063)	(0.0048)	(0.0207)	(0.0059)
	1000	9	Mean	0.6002	0.4001	0.1000	0.1001	0.4996	0.2500	0.4896
			SD	(0.0144)	(0.0077)	(0.0014)	(0.0074)	(0.0060)	(0.0240)	(0.0079)
			SE	(0.0143)	(0.0077)	(0.0015)	(0.0074)	(0.0060)	(0.0242)	(0.0073)
	1000	18	Mean	0.5997	0.4000	0.1000	0.1001	0.5001	0.2496	0.4899
			SD	(0.0097)	(0.0054)	(0.0010)	(0.0051)	(0.0043)	(0.0173)	(0.0056)
			SE	(0.0100)	(0.0055)	(0.0010)	(0.0052)	(0.0043)	(0.0177)	(0.0051)

Notes: For each DGP, 1000 simulated samples are drawn and estimated. We report the mean (Mean) and the standard deviation (SD) of the parameter point estimates across the 1000 Monte Carlo simulations, together with the estimated standard errors (SE) derived from the variance-covariance matrices of the estimators. The exact parameter values used in the DGPs are listed in the first row of the table.

Table 7: DGP5 – Negative Spatial Dependence in Productivity via Lagged Outputs and Lagged Labor Inputs of Related Firms, and via Productivity Shocks

Estimator	N	T	Var	α_l 0.6	α_k 0.4	λ -0.1	β_l -0.1	ρ_1 0.5	μ 0.25	σ_v^2 0.49
WGMM	500	9	Mean	0.3802	0.3566	-	-	0.5133	-	-
			SD	(0.0072)	(0.0107)	-	-	(0.0096)	-	-
			SE	(0.0083)	(0.0131)	-	-	(0.0097)	-	-
	500	18	Mean	0.3819	0.3553	-	-	0.5324	-	-
			SD	(0.0051)	(0.0076)	-	-	(0.0074)	-	-
			SE	(0.0059)	(0.0095)	-	-	(0.0077)	-	-
	750	9	Mean	0.3820	0.3710	-	-	0.5032	-	-
			SD	(0.0061)	(0.0091)	-	-	(0.0084)	-	-
			SE	(0.0067)	(0.0105)	-	-	(0.0081)	-	-
	750	18	Mean	0.3816	0.3713	-	-	0.5204	-	-
			SD	(0.0043)	(0.0064)	-	-	(0.0061)	-	-
			SE	(0.0048)	(0.0075)	-	-	(0.0062)	-	-
	1000	9	Mean	0.3877	0.3738	-	-	0.4956	-	-
			SD	(0.0052)	(0.0079)	-	-	(0.0070)	-	-
			SE	(0.0055)	(0.0086)	-	-	(0.0071)	-	-
	1000	18	Mean	0.3872	0.3727	-	-	0.5124	-	-
			SD	(0.0037)	(0.0057)	-	-	(0.0057)	-	-
			SE	(0.0040)	(0.0063)	-	-	(0.0054)	-	-
SGMM	500	9	Mean	0.6003	0.3996	-0.1000	-0.0999	0.5001	0.2479	0.4886
			SD	(0.0212)	(0.0108)	(0.0027)	(0.0110)	(0.0087)	(0.0360)	(0.0110)
			SE	(0.0214)	(0.0110)	(0.0028)	(0.0111)	(0.0089)	(0.0352)	(0.0103)
	500	18	Mean	0.6014	0.4001	-0.0999	-0.1008	0.4996	0.2488	0.4894
			SD	(0.0157)	(0.0075)	(0.0021)	(0.0080)	(0.0069)	(0.0242)	(0.0079)
			SE	(0.0153)	(0.0077)	(0.0021)	(0.0079)	(0.0068)	(0.0254)	(0.0072)
	750	9	Mean	0.6004	0.3992	-0.0999	-0.1001	0.4994	0.2485	0.4898
			SD	(0.0170)	(0.0089)	(0.0025)	(0.0088)	(0.0077)	(0.0282)	(0.0092)
			SE	(0.0172)	(0.0090)	(0.0024)	(0.0088)	(0.0075)	(0.0283)	(0.0084)
	750	18	Mean	0.6000	0.3998	-0.0999	-0.0999	0.5000	0.2492	0.4898
			SD	(0.0118)	(0.0062)	(0.0018)	(0.0060)	(0.0055)	(0.0213)	(0.0064)
			SE	(0.0122)	(0.0063)	(0.0018)	(0.0063)	(0.0056)	(0.0207)	(0.0059)
	1000	9	Mean	0.6002	0.4001	-0.1001	-0.1000	0.5001	0.2500	0.4895
			SD	(0.0143)	(0.0077)	(0.0022)	(0.0074)	(0.0064)	(0.0240)	(0.0081)
			SE	(0.0142)	(0.0077)	(0.0022)	(0.0074)	(0.0066)	(0.0241)	(0.0073)
	1000	18	Mean	0.5997	0.4000	-0.1000	-0.0999	0.4997	0.2497	0.4899
			SD	(0.0097)	(0.0054)	(0.0016)	(0.0051)	(0.0051)	(0.0173)	(0.0056)
			SE	(0.0100)	(0.0055)	(0.0016)	(0.0052)	(0.0050)	(0.0177)	(0.0051)

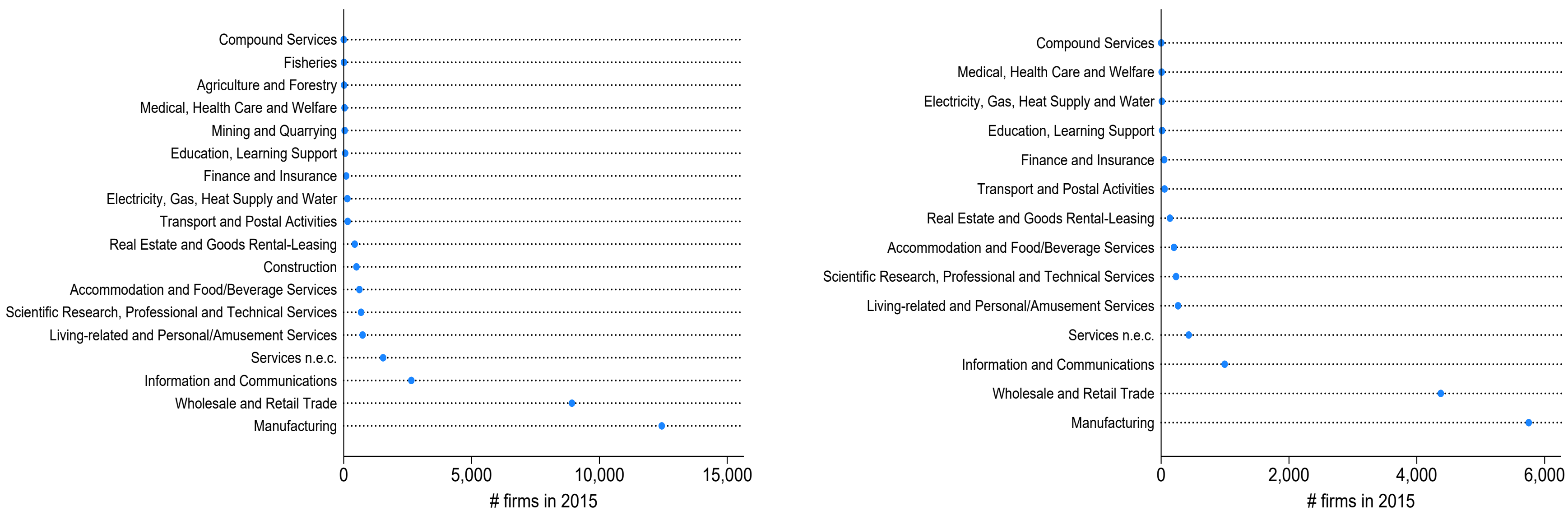
Notes: For each DGP, 1000 simulated samples are drawn and estimated. We report the mean (Mean) and the standard deviation (SD) of the parameter point estimates across the 1000 Monte Carlo simulations, together with the estimated standard errors (SE) derived from the variance-covariance matrices of the estimators. The exact parameter values used in the DGPs are listed in the first row of the table.

Table 8: Production Function Estimations (Japanese Firms 2009–2018)

	SGMM			WGMM
	(1)	(2)	(3)	(4)
$\mathbf{W}_{t-1}^{y}$	buyer-seller*CZ	buyer-seller*CZ	buyer-seller*CZ	
$\mathbf{W}_{t-1}^{\Omega}$	CZ	CZ	CZ	
$\mathbf{W}_{t}^{u}$	CZ	buyer-seller*CZ	buyer-seller	
α_0	-23.4378	-24.5051	-23.9971	-0.5286
	(1.2315)	(1.2957)	(1.2233)	(1.9240)
α_l	0.7922	0.7945	0.7651	0.5472
	(0.0320)	(0.0337)	(0.0322)	(0.1408)
α_k	0.0632	0.0616	0.0718	0.1480
	(0.0132)	(0.0135)	(0.0131)	(0.0384)
ρ_1	0.9644	0.9663	0.9650	0.9789
	(0.0026)	(0.0026)	(0.0025)	(0.0112)
λ	**0.0047**	**0.0044**	**0.0042**	
	(0.0005)	(0.0005)	(0.0005)	
β	**0.0537**	**0.0536**	**0.0553**	
	(0.0036)	(0.0036)	(0.0036)	
μ	**0.3833**	**0.0184**	**0.0353**	
	(0.0196)	(0.0036)	(0.0035)	
σ_v^2	0.0343	0.0357	0.0346	
	(0.0026)	(0.0027)	(0.0027)	
no. of observations	125,250	125,250	125,250	125,250
no. of firms	12,525	12,525	12,525	12,525

Notes: This table reports the estimations of Equations (1)–(3) based on the estimation methodology laid out in Section 3 and the connectivity matrices defined in Section 4.2. The function $h(l_{it}, k_{it}, m_{it})$ in Equation (9) is approximated by a second-order polynomial function: $l_{it}^{p} k_{it}^{q} m_{it}^{r}$ for $p + q + r \leq 2$, with non-negative integers p, q and r. The slope coefficient estimates $\boldsymbol{\delta}$ are omitted from the table above. The function $f(\nu)$ in Equation (12) is assumed to be of first order as in the conventional estimator. The list of instruments used for SGMM is: $\mathcal{Z}_{t,I} = (\iota_N, l_t, k_t, m_t, l_{t-1}, k_{t-1}, m_{t-1}, l_{t-2}, k_{t-2}, m_{t-2})$ and $\mathcal{Z}_{t,II} = (\iota_N, k_t, l_{t-1}, k_{t-1}, m_{t-1}, l_{t-2}, k_{t-2}, m_{t-2}, \mathbf{W}_{t-1}^{y} y_{t-1}, \mathbf{W}_{t-1}^{l} l_{t-1}, (\mathbf{W}_{t-1}^{y})^2 y_{t-1}, (\mathbf{W}_{t-1}^{l})^2 l_{t-1})$. The list of instruments used for WGMM is the same as those for SGMM, but excluding the related firms' outputs and labor inputs ($\mathbf{W}_{t-1}^{y} y_{t-1}$, $(\mathbf{W}_{t-1}^{y})^2 y_{t-1}$, $\mathbf{W}_{t-1}^{l} l_{t-1}$, $(\mathbf{W}_{t-1}^{l})^2 l_{t-1}$). We iterate the efficient GMM estimation procedure until the set of parameter estimates converges.

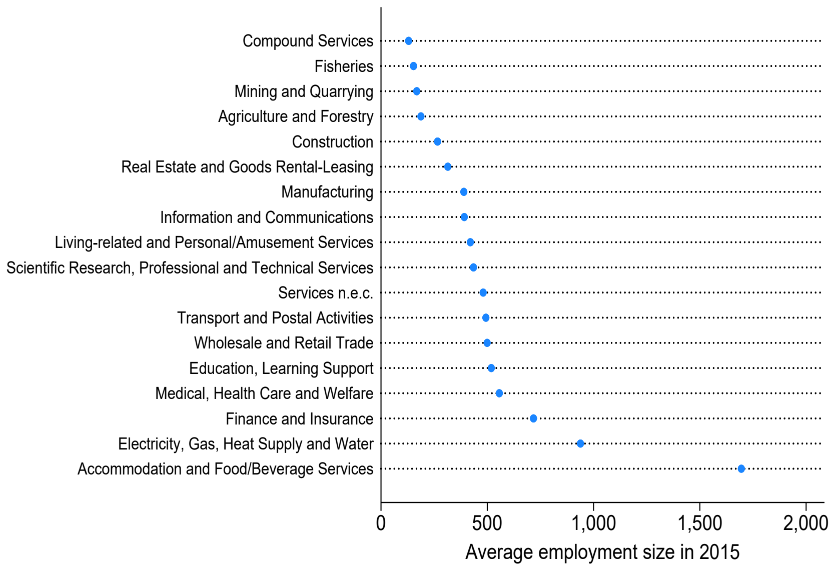
Figure 1: Number of firms in each industry in 2015



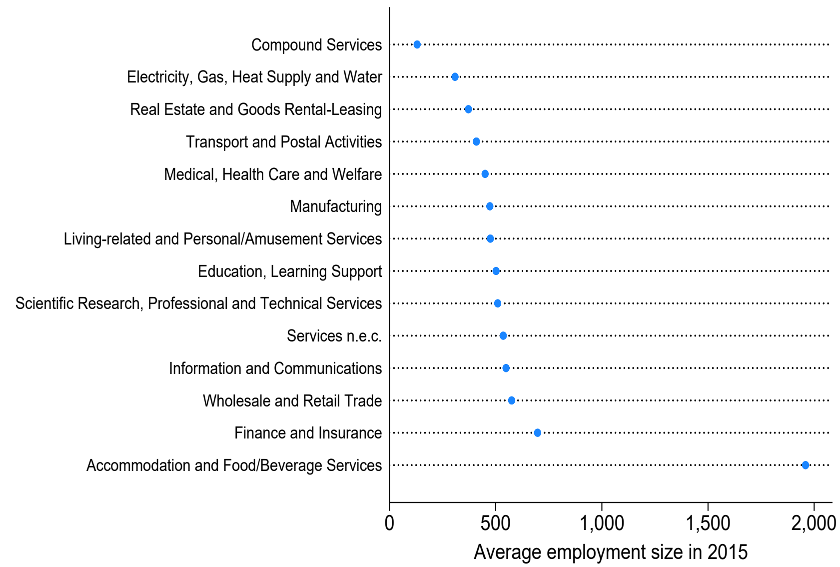
(a) BSJBSA-TSR Linked Sample

(b) Estimation Sample

Figure 2: Average firm's employment in each industry in 2015

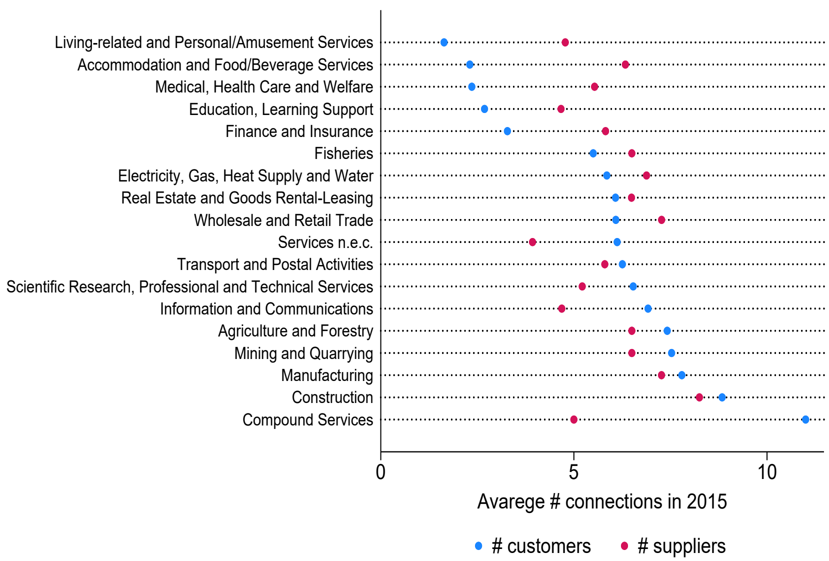


(a) BSJBSA-TSR Linked Sample

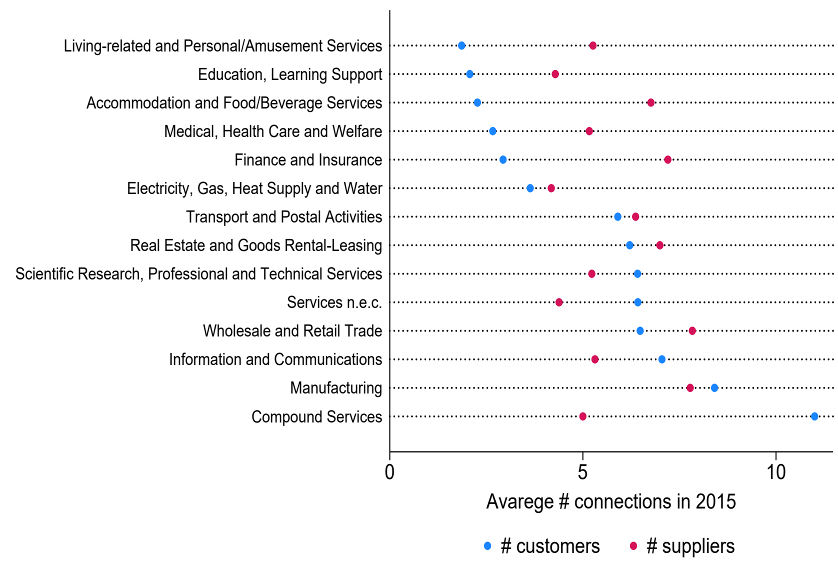


(b) Estimation Sample

Figure 3: Average number of connections in each industry in 2015

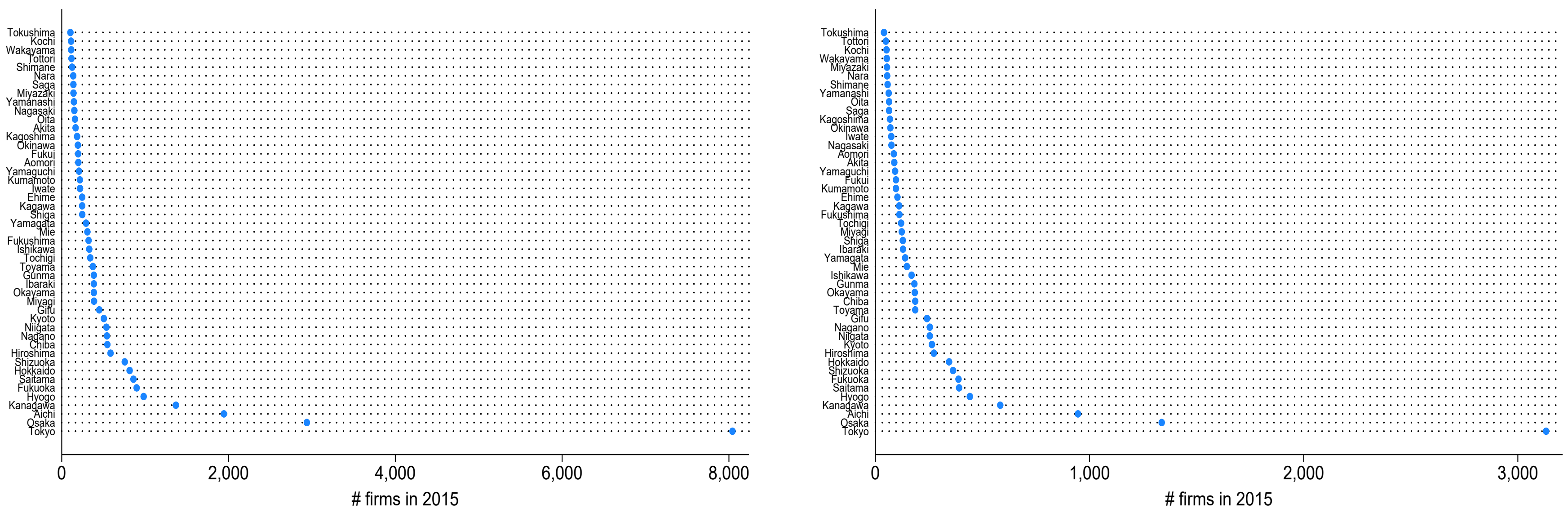


(a) BSJBSA-TSR Linked Sample



(b) Estimation Sample

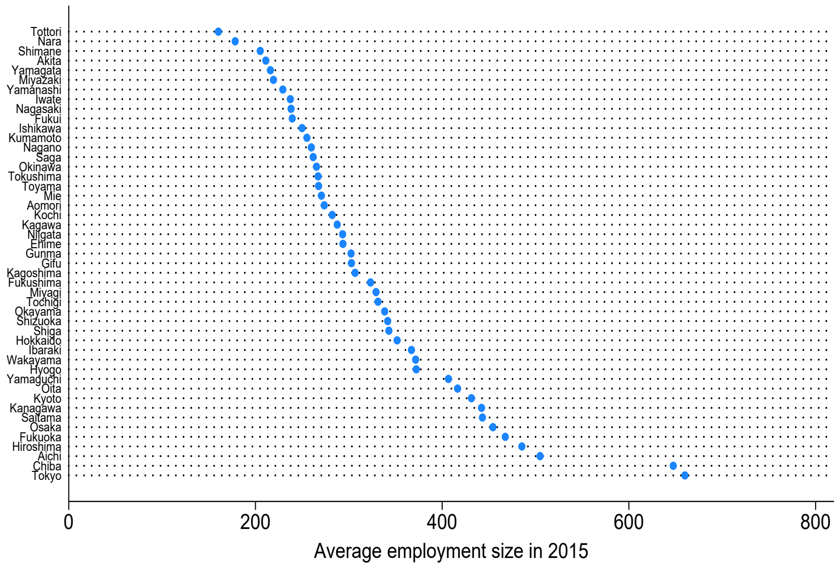
Figure 4: Number of firms in each prefecture in 2015



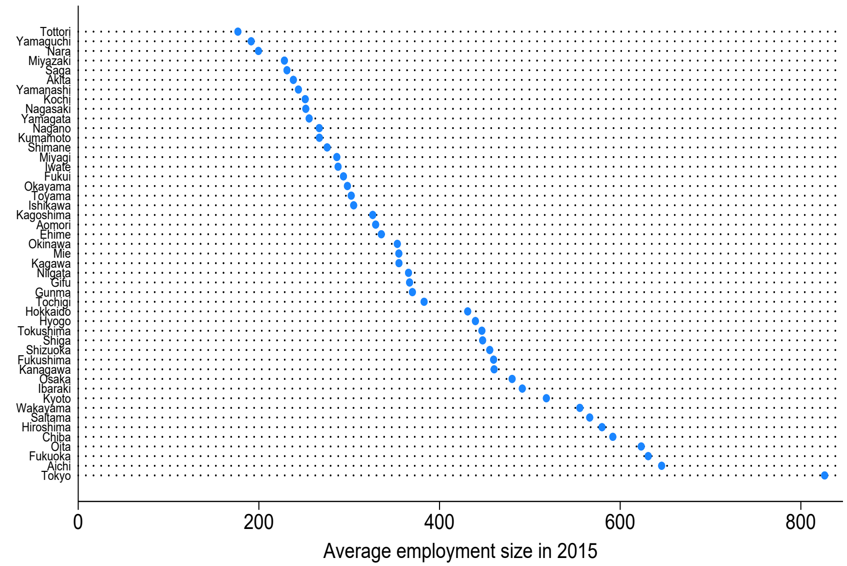
(a) BSJBSA-TSR Linked Sample

(b) Estimation Sample

Figure 5: Average firm's employment in each prefecture in 2015

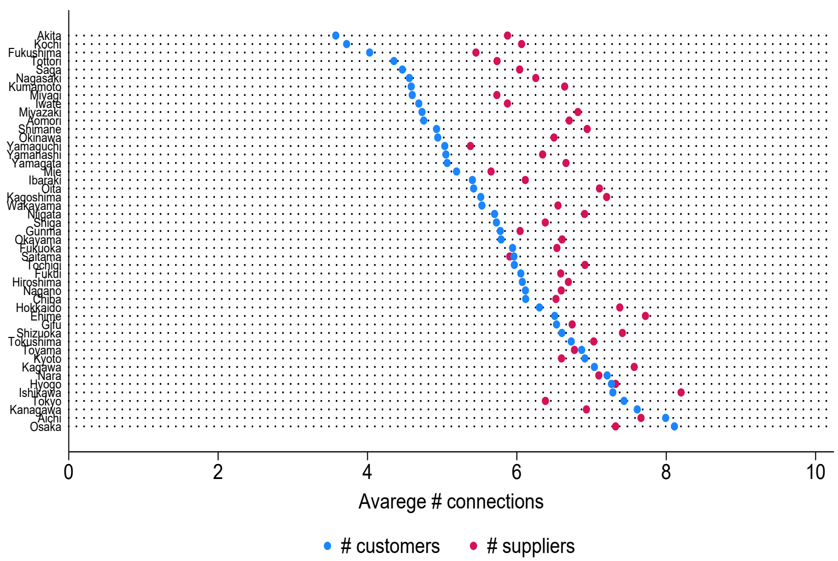


(a) BSJBSA-TSR Linked Sample

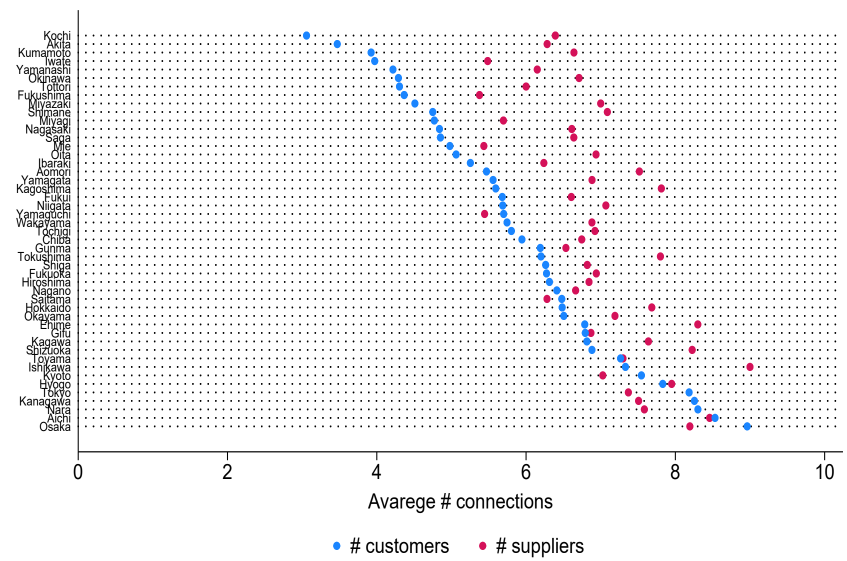


(b) Estimation Sample

Figure 6: Average number of connections in each prefecture in 2015

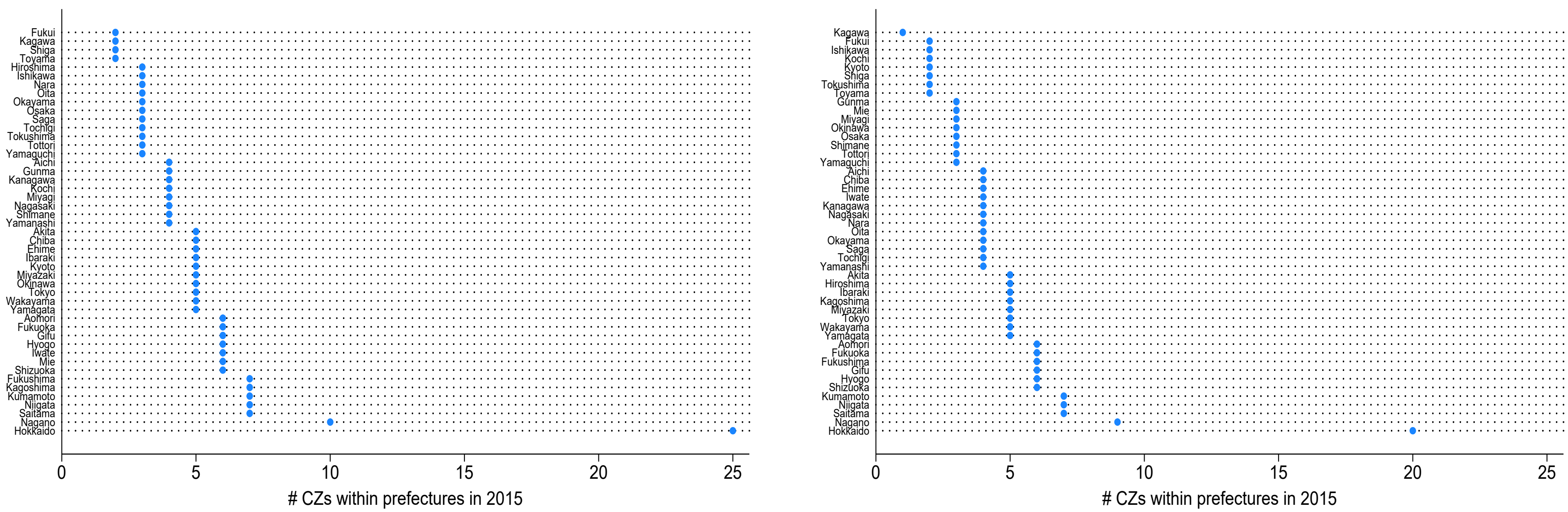


(a) BSJBSA-TSR Linked Sample



(b) Estimation Sample

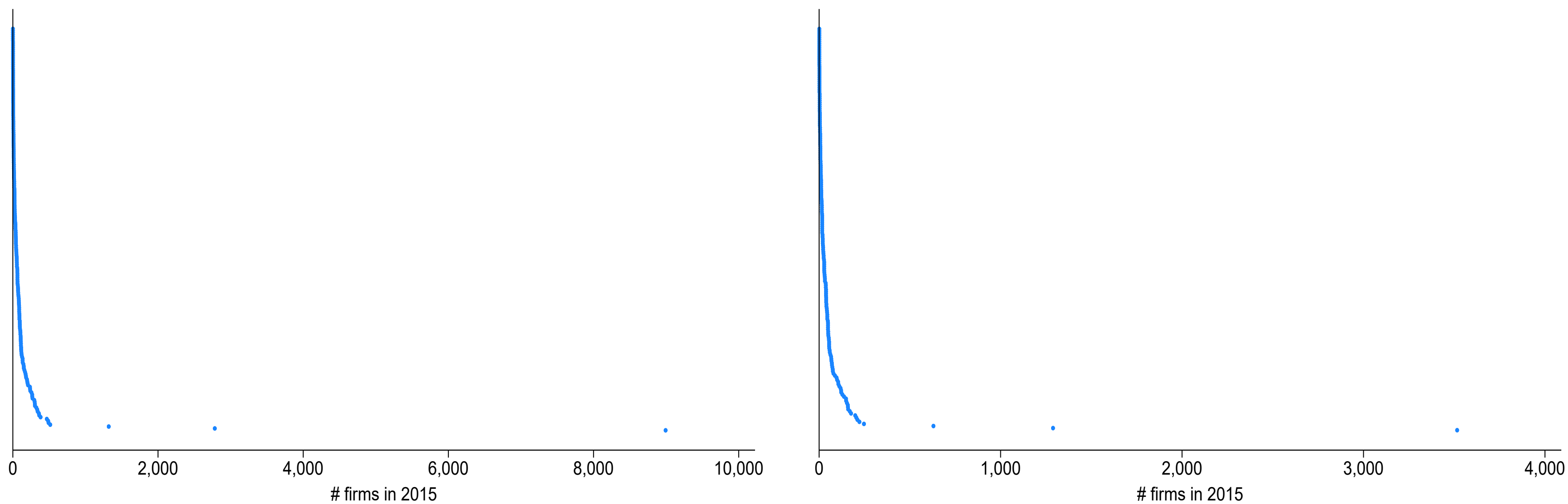
Figure 7: Number of commuting zones in each prefecture in 2015



(a) BSJBSA-TSR Linked Sample

(b) Estimation Sample

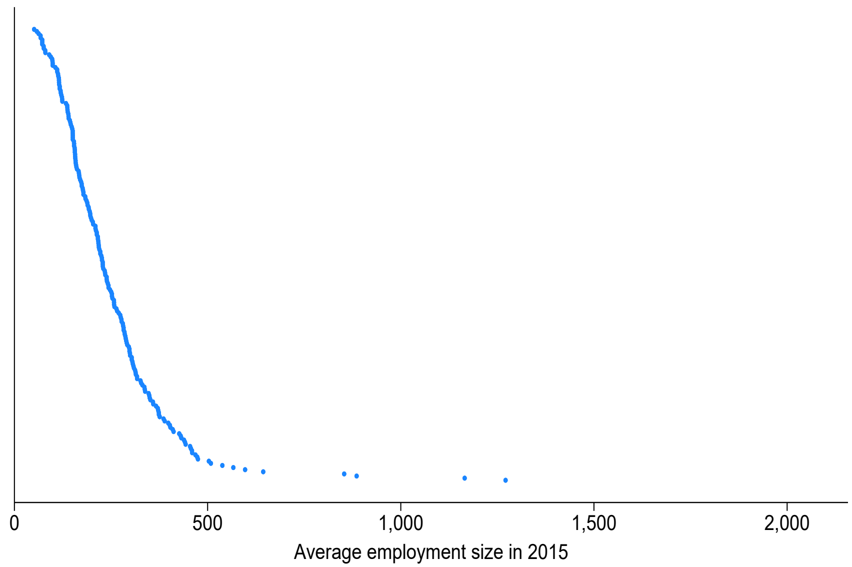
Figure 8: Number of firms in each commuting zone in 2015



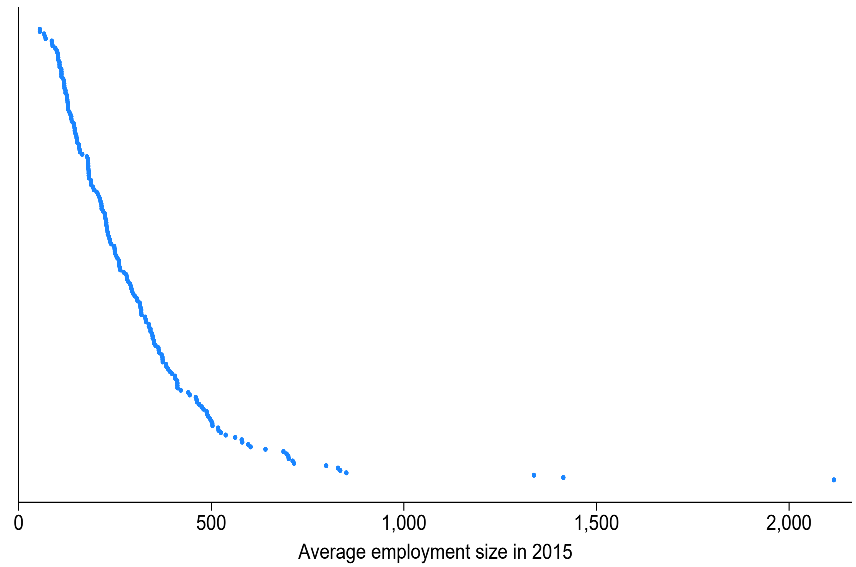
(a) BSJBSA-TSR Linked Sample

(b) Estimation Sample

Figure 9: Average firm's employment in each commuting zone in 2015

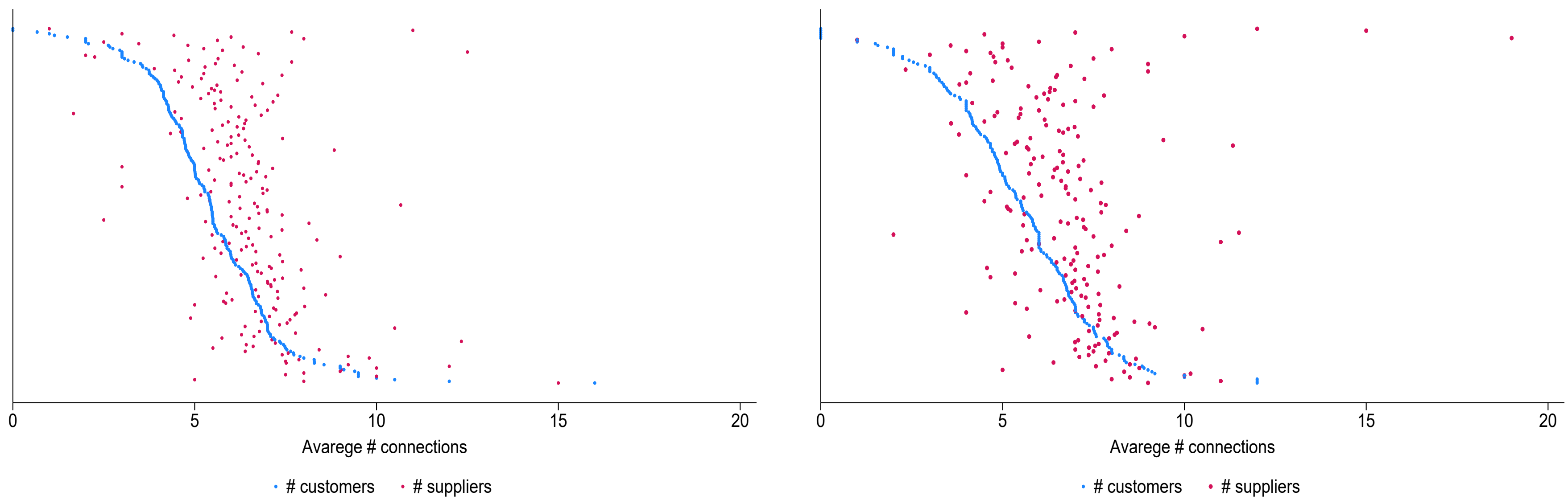


(a) BSJBSA-TSR Linked Sample



(b) Estimation Sample

Figure 10: Average number of connections in each commuting zone in 2015



(a) BSJBSA-TSR Linked Sample

(b) Estimation Sample