



The effects of urban transformation on productivity spillovers in China[☆]

Ming He^a, Yang Chen^{a,*}, Charles van Marrewijk^b

^a IBSS, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu Province, 215123, China

^b Utrecht University School of Economics, P. O. Box 80125, 3508, TC, Utrecht, the Netherlands



ARTICLE INFO

Keywords:

Productivity
Urban transformation
Technology spillover
Spatial durbin model

ABSTRACT

We analyze the impact of China's massive urban-transformation process on productivity spillovers for the east-coast electrical apparatus sector by combining the sub-prefecture-level population census data in 2000 and 2010 with detailed firm-level data from 1999 to 2007. At the sub-prefecture level, we identify three types of regions (metro-core, metro-ring, and non-metro) and develop three urban transformation measures (modernization, mobility, and disparity). A tailored spatial Durbin model enables us to distinguish between intra- and inter-regional technology spillovers. The baseline model reveals strong technology spillovers between neighboring firms, particularly those in close range. All three types of urban-transformation measures have significant impacts on technology spillovers. Modernization and mobility boost spillover effects, whereas economic disparity impedes inter-regional spillovers. Based on the structural difference between the three types of regions, our findings imply that metro-ring and non-metro regions are the major beneficiaries of productivity spillovers related to urban transformation.

1. Introduction

China's economic reform, launched in the late 1970s, triggered not only remarkable economic growth, but also an unprecedented urbanization process within the country (Brakman et al., 2016). According to the *World Urbanization Prospects: The 2018 Revision* compiled by the United Nations,¹ China's urban population rose from 192 million in 1980 to 775 million in 2015, or a 304% increase. In the same period, China's urbanization rate increased from 19.4% to 59.2%. These developments are illustrated in Fig. 1. The projected data show that China's urbanization rate in 2018 was slightly higher than that of middle-income countries (52.6%), but far less than that of high-income countries (81.5%). It is estimated that China's rural population will reduce by 53% and its urbanization rate will rise to 80% by 2050. Clearly, China has contributed significantly to the world's urbanization drive, and will continue to do so in the near future.

There is a growing body of empirical studies on China's urban transformation. Researchers reveal a close link between urbanization, industrial development, and internal migration of labor (Q. Chen and Song, 2014; Liu et al., 2015; L. Zhang, LeGates and Zhao, 2016). Several topics have attracted intense scholarly interest. These include the distribu-

tional property of China's urban system (Bosker et al., 2012; Li et al., 2016; Soo, 2014), inequality between demographic groups and regions (Cai and Du, 2011; Wu and Rao, 2017; W. Zhang and Bao, 2015), and economic development (Au and Henderson, 2006b; Bond and Riezman, 2016; Kumar and Kober, 2012). However, little is known about the impact of urbanization on China's industrial productivity, especially at the micro level.

A different strand of literature aims to unveil the sources of productivity growth in China's manufacturing industry. The mechanisms identified include resource reallocation via firm turnover (Brandt et al., 2012; Ding et al., 2016; Jefferson et al., 2008), reform of state-owned enterprises (Berkowitz et al., 2017; Hsieh and Song, 2015; Walheer and He, 2020), economies of agglomeration (Howell et al., 2016; C. Hu, Xu and Yashiro, 2015; Lin et al., 2011), and the law of convergence (Deng and Jefferson, 2011; Lemoine et al., 2015). None of these studies, however, consider the effect of urbanization on firm productivity.

We combine these two strands of literature by analyzing the impact of China's urban-transformation process on the productivity spillovers between firms. Using the 2000 and 2010 population census data, we are able to measure three facets of China's urban transformation at the county level: modernization (growth of non-agricultural population),

[☆] We would like to thank the editor and an anonymous referee for their constructive comments and suggestions. All remaining errors are our own.

^{*} Corresponding author.

E-mail addresses: ming.he@xjtlu.edu.cn (M. He), yang.chen@xjtlu.edu.cn (Y. Chen), j.g.m.vanmarrewijk@uu.nl (C. van Marrewijk).

¹ The publication can be accessed at <https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf>.

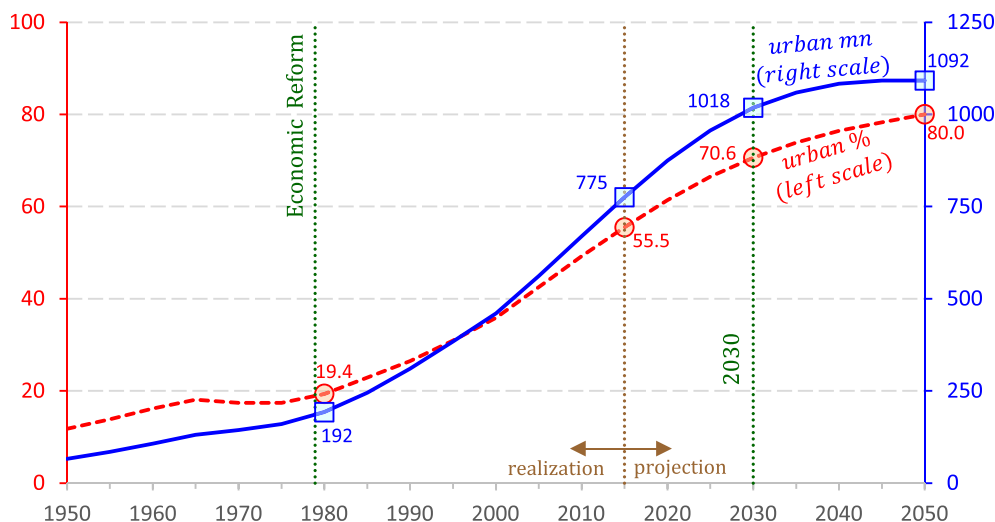


Fig. 1. Urban population in China; million and percent of total, 1950–2050³⁰.

mobility (migration), and disparity.² These data are used in conjunction with a rich firm-level dataset for 1999–2007 to examine the moderation effects of the urbanization measures on productivity spillovers. This panel dataset consists of 23,206 distinct firms and 82,376 observations for the electric apparatus sector.³

We model productivity spillovers in a spatial Durbin framework similar to those of Baltagi et al. (2015) and Z. Chen, Haynes, Zhou, and Dai (2019). The geo-referenced data allow us to distinguish between two types of spillovers by spatial distance: intra-regional spillovers (within-county) and inter-regional spillovers (across-county). The results from the baseline model align with the findings in the literature. The spillover effects attenuate rapidly in space, resulting in much stronger intra-regional effects than inter-regional effects. Research and Development (R&D)-intensive and exporting firms exhibit higher productivity. Measures of agglomeration economies, such as variety and competition, also contribute to higher productivity.

Estimates from the extended model reveal that all three types of urban transformation have a significant impact on technology spillovers: modernization and mobility boost spillovers. Conversely, economic disparity impedes inter-regional spillovers (but encourages the intra-regional spillover). Based on China's administrative classification, regional characteristics, and historical legacy, we classify regions (counties) under study into metro-core, metro-ring, and non-metro regions in descending order of centrality. Summary statistics show that metro-core regions have lower levels of modernization and mobility, but greater disparity than do metro-ring and non-metro regions. This suggests that metro-ring and non-metro regions are the major beneficiaries of productivity spillovers related to China's urban transformation.

These results have two major policy implications. First, further relaxation of the *Hukou* system, as outlined by China's *National New-Type Urbanization Plan (2014–2020)* (Wang et al., 2015), will help increase modernization and mobility in central cities and benefit technology spillovers across regions. Second, productivity spillovers across regions will not be fully materialized unless regional disparity diminishes. Thus, it is imperative to alleviate structural differences between central cities and their neighboring regions.

The rest of this article is organized as follows. Section 2 reviews

the empirical literature on urban transformation and productivity spillovers, while Section 3 presents the statistical model and discusses implementation issues. Section 4 explains the data and variable definitions and Section 5 presents the empirical results, followed by a discussion in Section 6. Finally, Section 7 concludes the article.

2. Literature review

2.1. Urban transformation

Urban transformation is viewed as one of the most important social developments in China. The sources of China's urban transformation are (1) the natural growth of urban population, (2) the rural-to-urban "migration" resulting from land reclassification, (3) the influx of mobile workers, and (4) the change of industrial employment structure (M. Chen, Liu and Tao, 2013; Q. Chen and Song, 2014). Of these, internal migration is the main driving force of China's urbanization in the early 2000s, especially in mega cities and on China's eastern coast (Q. Chen and Song, 2014; Liu et al., 2015).

Some efforts have been made to unfold the structure of urban hierarchy and pattern of urban system evolution during rapid urban transformation, although findings of the rank-size distribution are inconclusive. Soo (2014), using provincial-level data, finds that the size distribution of population in China conforms to Zipf's law. Ye and Xie (2012) demonstrate that Zipf's law captures the main trend at national level, while it masks various regional rank-size distribution patterns. Li et al. (2016) argue that Zipf's exponent is significantly less than 1, employing the 2000 and 2010 national population census data. Bosker et al. (2012) simulation analysis of spatial equilibrium of people and economic activity under different scenarios of labor mobility reveals a downward-size distortion among Chinese mega cities because of the restrictions of the *Hukou* system. From a dynamic perspective, Li et al. (2016) show that the population growth follows Gibrat's law. Conversely, both Soo (2014) and Chan (2010) assert that China's urban system evolution is highly imbalanced and affected by market forces and government intervention.

Empirical studies on the urbanization–inequality relationship deliver mixed messages. Cai and Wang (2010), Knight et al. (2011), and Cai and Du (2011) report a salient increase in the real wage of migrants and rural laborers since 2000, which results in wage convergence in various dimensions.⁴ The increasing wage trend, wage conver-

² Similar concepts are discussed in Pannell (2002), He et al. (2017) and Shang et al. (2018). These data are used in conjunction with a rich firm-level dataset for 1999–2007 to examine the moderation effects of the urbanization measures on productivity spillovers. This panel dataset consists of 23,206 distinct firms and 82,376 observations for the electric apparatus sector.³

³ The choice of the electric apparatus sector is explained in Section 4.

⁴ Wage convergence is observed between migrants and urban residents, skilled and unskilled jobs, and rural non-agricultural and agricultural jobs.

gence and migrant worker scarcity are interpreted as evidence supporting the notion that China's urban transformation has reached or passed 'the Lewis point' by 2010. However, [Golley and Meng \(2011\)](#) show different findings and report that migrant stock could be doubled by removing institutional and policy-induced migration barriers. Further to these depictions of labor market development and wage structure change, some researchers estimate the association between urbanization and rural–urban income inequality. For example, [M. Lu and Chen \(2006\)](#) find that urbanization reduces income inequality (using data from 1987 to 2001). [Zhou \(2009\)](#) study suggests that China's overall inequality reached its peak value between 2006 and 2009. Alternatively, [Wu and Rao \(2017\)](#) estimate that income inequality is maximized when the urbanization rate is 0.53. In addition to the study of rural–urban inequality, [W. Zhang and Bao \(2015\)](#) study the inter- and intra-province inequality.

There is less controversy on the role of urbanization in economic growth. Theoretical developments, such as [Duranton \(2008\)](#), emphasize the role of urbanization in bolstering productivity and economic growth. [Au and Henderson \(2006a\)](#), and [Au and Henderson \(2006b\)](#) find an inverted-U relationship between labor productivity and urban size among Chinese cities. In addition, their finding suggests that most Chinese cities will benefit from further urbanization. [Bond and Riezman \(2016\)](#) demonstrate that the primary driving force of China's economic growth is migration-induced urbanization. At the country-level, [Henderson \(2003b\)](#) result suggests there is an optimal level of urbanization for productivity growth. Using global data, [Annez and Buckley \(2009\)](#) and [Kumar and Kober \(2012\)](#) provide clear evidence that urbanization explains cross-country productivity differences.

These studies, however, are conducted at the macro (national, regional, or provincial) level. Few studies link the performance of micro-level production units with China's urbanization features in much smaller geographical scopes. This article focuses on three facets of China's urban transformation: change in urbanization rate, composition of local migrant population, and economic disparity. These measures are constructed for county-level jurisdictions, which enables examination of their impact on the productivity of local firms. To our knowledge, few attempts have been made to analyze China's urbanization process at such a disaggregate level.

2.2. Productivity spillovers

Krugman's statement that "productivity isn't everything, but in the long run it is almost everything" ([Krugman, 1997](#), p.11) is widely supported in the literature. [Z. Song, Storesletten, and Zilibotti \(2011\)](#) and [Jefferson et al. \(2006\)](#) find that productivity is the main driver of China's high output growth. [Brandt et al. \(2012\)](#) argue that productivity growth dominates input accumulation as the major source of output growth for China's manufacturing industry.

At the macro level, [Hsieh and Klenow \(2009\)](#) and [Z. Song et al. \(2011\)](#) study resource allocation and aggregate total factor productivity (TFP). [Brandt et al. \(2012\)](#) document the impact of China's entry into the World Trade Organization (WTO) on her productivity. [Bournakis, Christopoulos, and Mallick \(2018\)](#) examine how a country's productivity benefits from both domestic and international R&D spillovers through imports. In the field of industrial organization, [Gao and Van Biesebroeck \(2014\)](#) estimate the efficiency gains from deregulation in China's electricity sector. [Brandt et al. \(2012\)](#) find that firm turnover has been the dominant force in productivity growth in Chinese manufacturing compared with other mechanisms.

At the firm level, sources of productivity growth include both firm- and regional-specific factors. Firm-specific determinants include ownership ([Berkowitz et al., 2017](#); [Ding et al., 2016](#); [Jefferson et al., 2008](#)), trade ([Bao et al., 2015](#); [Syverson, 2011](#); [Yang and He, 2014](#)) and R&D ([Aiello et al., 2014](#); [Boeing et al., 2016](#); [Wei and Liu, 2006](#)). Micro-level productivity differences are also attributed to market conditions such

as financing or fiscal burden ([Bournakis and Mallick, 2018](#); [Demurger, 2001](#); [Vijverberg et al., 2011](#)), the presence of foreign direct investment ([Anwar and Sun, 2014](#); [Eapen et al., 2019](#); [Haskel et al., 2007](#)), and economies of agglomeration ([Glaeser et al., 1992](#); [Henderson, 2003a](#); [Y. Hu, Fisher-Vanden and Su, in press](#)). The externalities from agglomeration are usually treated as a "spillover effect" and captured by agglomeration measures, such as competition, specialization, and variety in regression models ([Combes, 2000](#); [C. Hu et al., 2015](#); [Sheng and Song, 2013](#)).

Our work is closely related to the productivity studies at the firm level. Focusing on the role of urban transformation in shaping productivity spillovers, we account for firm behavioral measures (such as R&D activity and foreign market engagement) and regional characteristics, including specialization, variety, competition, and public spending. Methodologically, our study is related to [Baltagi et al. \(2015\)](#) and [Z. Chen et al. \(2019\)](#), who use spatial econometric methods to model productivity spillover. We add to their approach by identifying two types of spillover effects: that within the same jurisdiction (intra-regional) and that between neighboring jurisdictions (inter-regional). With geo-referenced firm-level data, we model both types of spillovers in a spatial Durbin model. This treatment sheds new light on the spatial interactions of firm productivity.

3. Statistical methodology

At the core of this study is the estimation of spillover effects between individual firms. To this end, we employ a spatial econometric model and gauge spillover effects by the spatial autocorrelation in firm productivity. This section addresses three technical obstacles to our analysis: the endogeneity in spatial lags of the dependent variable; time-varying spatial weights and relatively large sample size; and modeling the moderating effects of auxiliary variables.

3.1. The spatial Durbin model

The basic framework for our analysis is the spatial Durbin model. We assume that in period $t = 1, \dots, T$, the number of observations (firms) is N_t . The spatial model for period t is⁵:

$$\mathbf{y}_t = \lambda \mathbf{W}_t \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \alpha \mathbf{1}_t + \boldsymbol{\mu}_t + \mathbf{u}_t. \quad (1)$$

Here, \mathbf{y}_t is an $N_t \times 1$ vector of the dependent (firm-level productivity), \mathbf{W}_t is an $N_t \times N_t$ maximum row-normalized spatial weight matrix whose main diagonal are all zeros, \mathbf{X}_t is an $N_t \times K$ matrix of time-varying exogenous regressors, $\mathbf{1}_t$ is an $N_t \times 1$ vector of ones, and $\boldsymbol{\mu}_t$ is an $N_t \times 1$ vector of individual effects. The error terms \mathbf{u}_t are assumed to be i.i.d. with zero mean and finite variance. For expositional convenience, we stack Equation (1) into its panel representation

$$\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \alpha \mathbf{1} + \boldsymbol{\mu} + \mathbf{u}, \quad (2)$$

where $N = \sum_{t=1}^T N_t$ is the total number of observations and $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_T)'$. Moreover, $\mathbf{W} = \text{diag}(\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_T)$ is an $N \times N$ block-diagonal matrix with \mathbf{W}_t on the main diagonal.

The spatial lag term $\mathbf{W} \mathbf{y}$ on the right hand side of (2) is endogenous because

$$E((\mathbf{W} \mathbf{y})' \mathbf{u}') = E(\mathbf{W}(\mathbf{I}_N - \lambda \mathbf{W})^{-1} \mathbf{u} \mathbf{u}') = \mathbf{W}(\mathbf{I}_N - \lambda \mathbf{W})^{-1} E(\mathbf{u} \mathbf{u}') \neq \mathbf{0}.$$

Thus, estimating (2) by ordinary least square (OLS) will produce biased estimates. A popular treatment for the endogeneity problem in (2) is the maximum likelihood (ML) estimator ([Anselin, 1988](#); [Elhorst, 2003](#)). The quasi-maximum likelihood (QML) estimators were proposed by [Lee and Yu \(2010\)](#) to tackle spatial models with fixed effects. Despite the

⁵ Our empirical model features two spatial lags, $\mathbf{W}_{1t} \mathbf{y}_t$ and $\mathbf{W}_{2t} \mathbf{y}_t$, so there are two spatial AR coefficients: λ_1 and λ_2 . The theoretical arguments developed here carry over to the extended model.

analytical advantage of the ML estimator—it can be adapted to most types of spatial models (LeSage and Pace, 2009)—it requires spatial transforming of all variables and computing the determinants of relevant matrices when constructing the likelihood function. Using Model (2) as an example, the transformation takes the form $S(\lambda) = \mathbf{I}_N - \lambda \mathbf{W}$, which involves the unknown parameter λ and the dimension of the matrix equals the sample size.⁶ Thus, the size of the sample becomes a major threat to the implementation of the ML method. Our dataset (see Section 4) comprises over 23,000 firms over a nine-year period, with a total of over 82,000 observations. Since the panel is unbalanced, a customized algorithm is necessary.⁷ There is no easy work around in computing the determinant of $S(\lambda)$. In practice, evaluating $|S(\lambda)|$ demands special numeric methods and is computationally burdensome (LeSage and Pace, 2009).

Compared with the ML approach, Kelejian and Prucha (1998) 2SLS estimator is far easier to implement on our data. Their solution is to instrument $\mathbf{W}\mathbf{y}$ by a polynomial of \mathbf{W} pre-multiplied to \mathbf{X} . To justify, let us denote the matrix of within-transformation by \mathbf{Q}_0 and multiply it on both sides of (2) to obtain:

$$\mathbf{Q}_0 \mathbf{y} = \lambda \mathbf{Q}_0 \mathbf{W} \mathbf{y} + \mathbf{Q}_0 \mathbf{X} \boldsymbol{\beta} + \mathbf{Q}_0 \mathbf{u}. \quad (3)$$

The expectation of the RHS endogenous variable in (3) can be expressed as

$$\begin{aligned} E(\mathbf{Q}_0 \mathbf{W} \mathbf{y}) &= \mathbf{Q}_0 \mathbf{W} (\mathbf{I}_N - \lambda \mathbf{W})^{-1} (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\alpha} \mathbf{1}) \\ &= \mathbf{Q}_0 \sum_{k=0}^{\infty} \lambda^k \mathbf{W}^{k+1} (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\alpha} \mathbf{1}). \end{aligned} \quad (4)$$

Equation (4) suggests

$$\mathbf{G}_0 = (\mathbf{Q}_0 \mathbf{X}, \mathbf{Q}_0 \mathbf{W} \mathbf{X}, \mathbf{Q}_0 \mathbf{W}^2 \mathbf{X}, \dots) \quad (5)$$

as valid instruments for $\mathbf{Q}_0 \mathbf{W} \mathbf{y}$ in (3). The resulting fixed-effects 2SLS estimator for $(\lambda, \boldsymbol{\beta}', \boldsymbol{\alpha})$ is shown to be consistent (Kelejian and Prucha, 1998). Let us denote it by $\hat{\theta}_w$.

Under the random-effects assumption $E(\boldsymbol{\mu}|\mathbf{X}) = \mathbf{0}$, $\hat{\theta}_w$ remains consistent but efficiency can be improved if we perform the familiar Cochrane-Orcutt transformation on (2) before running 2SLS. Specifically, the generalized least square (GLS) transformation of a variable ξ is given by

$$\tilde{\xi}_{it} = \xi_{it} - \frac{\hat{\sigma}_u}{(T_i \hat{\sigma}_\mu^2 + \hat{\sigma}_u^2)^{1/2}} \bar{\xi}_{it}, \quad (6)$$

where $\hat{\sigma}_\mu$ and $\hat{\sigma}_u$ are consistent estimates of σ_μ and σ_u , $\bar{\xi}_{it}$ is the group mean of individual i , and T_i is number of observations for individual i . The random-effects 2SLS estimator $\hat{\theta}_R$ is simply the 2SLS estimator of:

$$\tilde{\mathbf{y}} = \lambda \tilde{\mathbf{W}} \mathbf{y} + \tilde{\mathbf{X}} \boldsymbol{\beta} + \boldsymbol{\alpha} \mathbf{1} + \mathbf{v}, \quad (7)$$

using:

$$\tilde{\mathbf{G}} = (\tilde{\mathbf{X}}, \tilde{\mathbf{W}} \mathbf{X}, \tilde{\mathbf{W}}^2 \mathbf{X}, \dots) \quad (8)$$

as instruments (Baltagi, 2008). The Hausman test can be used to select the appropriate specification.

The main objective of our study is to investigate whether and to what extent the strength of spillover effects, measured by λ , is influenced by measures of urban transformation. This can be achieved by introducing interactions between the urban transformation variable and $\mathbf{W}\mathbf{y}$. Suppose we have an exogenous variable \mathbf{i} that varies across indi-

viduals but is constant across time. If we allow it to interact with $\mathbf{W}\mathbf{y}$, (2) and (3) become

$$\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \gamma \mathbf{i} \cdot \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\alpha} \mathbf{1} + \boldsymbol{\mu} + \mathbf{u}, \quad (2')$$

and

$$\mathbf{Q}_0 \mathbf{y} = \lambda \mathbf{Q}_0 \mathbf{W} \mathbf{y} + \gamma \mathbf{i} \cdot \mathbf{Q}_0 \mathbf{W} \mathbf{y} + \mathbf{Q}_0 \mathbf{X} \boldsymbol{\beta} + \mathbf{Q}_0 \mathbf{u}, \quad (3')$$

where \cdot denotes pairwise multiplication. The interaction term in (3') requires another set of instruments. In light of (5), the natural candidates are:

$$\mathbf{H}_0 = \mathbf{i} \cdot \mathbf{G}_0 = (\mathbf{i} \cdot \mathbf{Q}_0 \mathbf{X}, \mathbf{i} \cdot \mathbf{Q}_0 \mathbf{W} \mathbf{X}, \mathbf{i} \cdot \mathbf{Q}_0 \mathbf{W}^2 \mathbf{X}, \dots). \quad (9)$$

A separate set of instruments similar to (9) must be used in conjunction with every interaction term introduced to (2').

3.2. Implementation

The basic geographic unit in our study is an urban district or a rural county. The data allow us to match each firm to a district/county but provide no further information about the exact location of the firm. In this regard, each firm has two types of neighbors: those in the same jurisdiction (intra-regional), and those in neighboring jurisdictions (inter-regional).⁸ Therefore, we introduce two spatial weight matrices in the spatial Durbin model (1): for any variable ξ ,

$$\mathbf{W}_{wt} \xi_{ikt} = \sum_{\substack{j \in I_t(k) \\ j \neq i}} l_{jkt} \xi_{jkt} / \sum_{\substack{j \in I_t(k) \\ j \neq i}} l_{jkt}, \quad (10)$$

$$\mathbf{W}_{nt} \xi_{ikt} = \sum_{\substack{j \in I_t(k') \\ k' \in N(k)}} l_{jk't} \xi_{jk't} / \sum_{\substack{j \in I_t(k') \\ k' \in N(k)}} l_{jk't}. \quad (11)$$

Here, the subscripts i and j denote individuals (firms), while k and k' denote jurisdictions (district or county). Moreover, $I_t(k)$ is the set of all firms in jurisdiction k in time t , and $N(k)$ is the set of all contiguous neighbors of jurisdiction k . Finally, l_{ikt} is the employment of firm i in jurisdiction k at time t . The construction is based on the premise that larger firms (measured by employment) exert a stronger influence on their neighbors than do smaller ones. Both \mathbf{W}_{wt} and \mathbf{W}_{nt} are maximum row normalized, with zeros in their main diagonals. We have an unbalanced panel, so $I_t(k)$ varies in time. Therefore, the spatial weight matrices are not fixed.

With two spatial lags, (1) becomes

$$\mathbf{y}_t = \lambda_1 \mathbf{W}_{wt} \mathbf{y}_t + \lambda_2 \mathbf{W}_{nt} \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\alpha} \mathbf{1}_t + \boldsymbol{\mu}_t + \mathbf{u}_t. \quad (12)$$

It remains to narrow down the number of instruments (e.g., \mathbf{G}_0) into a finite set. Practically, researchers use spatial lags to order 2 (Kelejian and Prucha, 1998; Kelejian et al., 2004, among others), which is the strategy we adopt here. We have two spatial weights, so their interactions must be included. Therefore, (5) becomes:

$$\mathbf{G}_0 = \mathbf{Q}_0 (\mathbf{X}, \mathbf{W}_w \mathbf{X}, \mathbf{W}_n \mathbf{X}, \mathbf{W}_w^2 \mathbf{X}, \mathbf{W}_n^2 \mathbf{X}, \mathbf{W}_w \mathbf{W}_n \mathbf{X}, \mathbf{W}_n \mathbf{W}_w \mathbf{X}), \quad (5')$$

where \mathbf{W}_w and \mathbf{W}_n are block-diagonal matrices constructed from \mathbf{W}_{wt} and \mathbf{W}_{nt} .

The number of instruments in (5') could be large, even if we have a limited number of exogenous regressors. For instance, if \mathbf{X} consists of six variables, we will have a total of 42 instruments. Luckily, \mathbf{W}_{wt} and \mathbf{W}_{nt} defined in (10) and (11) offer an opportunity to simplify. Let us divide \mathbf{X} into variables that are firm-specific (varying across firms and in time) and jurisdiction-specific (varying in time but are invariant in each jurisdiction), denoted by \mathbf{X}_1 and \mathbf{X}_2 , respectively. Suppose that the sample has no islands (firms without intra-regional or inter-regional

⁶ With random effects, the likelihood function also computes the inverse of large-scale matrices containing unknown parameters.

⁷ To our knowledge, spatial econometric routines in mainstream software packages all require a balanced panel and a time-invariant spatial weight matrix.

⁸ Two jurisdictions are considered neighbors if they share a contiguous border.

neighbors). We can show that

$$W_{wt}X_{2t} = X_{2t}, \quad \text{and} \quad W_{wt}W_{nt}X_t = W_{nt}X_t. \quad (13)$$

Thus, G_0 in (5') is made up of two groups:

$$G_{01} = Q_0(X_1, W_wX_1, W_nX_1, W_w^2X_1, W_n^2X_1, W_nW_wX_1),$$

and

$$G_{02} = Q_0(X_2, W_nX_2, W_n^2X_2). \quad (5'')$$

If there are two variables in X_1 and four in X_2 , the total number of instruments is reduced to 24.⁹

4. Data description

4.1. Firm-level data

The firm-level data used in this study are extracted from the *China Industry Survey* (CIS) dataset. The survey, conducted by the National Bureau of Statistics (NBS), covers all state-owned enterprises and medium- or large-sized non-state firms in China's manufacturing industry over 1998–2007.¹⁰ The resulting dataset comprises over 2 million observations for more than 0.5 million distinct firms. In addition, it provides rich information about each firm, including location, industrial classification, and inputs and outputs. As a result of its comprehensiveness, the CIS dataset is highly visible in the recent literature on China's industry (Brandt et al., 2012; M. Chen and Guariglia, 2013; Hsieh and Song, 2015; Yu, 2015).¹¹

In compliance with the *National Standard of Economic Classification* (GB/T 4754-2007), the CIS dataset assigns a four-digit industry classification code to each firm, with which we define 30 two-digit sectors. Similarly, each firm receives a six-digit (12 digit since 2004) administrative code according to the *Codebook of Administrative Divisions* (GB/T 2260-2007). This enables us to pinpoint the county (or urban district) where the firm operates. This study focuses on the electrical apparatus sector (code range 3900–3990) in 10 coastal provinces in Eastern China.¹² The choice of sector is based on several considerations. First, the electrical apparatus sector is an important component of China's manufacturing industry.¹³ Second, this sector is relatively technology-intensive.¹⁴ Hence we expect stronger and more sig-

nificant productivity spillovers in this sector. Third, the sector offers a large sample that facilitates our econometric analysis.¹⁵ We focus on the coastal provinces because of their critical role in China's economic development and urbanization process. In terms of land area, the 10 provinces are approximately 1.03 million square kilometers, which is about the size of Texas and California combined. In terms of economic size and population, they host 41% of China's population and generate 62% of the nation's gross domestic product (GDP, 2010 values). In addition, most of the electrical apparatus sector (86% of all observations) are found in the 10 provinces. Finally, the literature documents that cities on the export-oriented eastern coast experienced the most vibrant urbanization in the new millennium (Chan, 2010; Li et al., 2016).

4.2. Variables in the baseline model

Information on inputs and outputs enables us to estimate firm-level productivity. In this article, TFP is estimated using Levinsohn and Petrin (2003) two-step procedure (LP). Specifically, we adopt their value-added specification. Details of the estimation procedure can be found in their original work or other empirical studies (C. Hu et al., 2015; Sheng and Song, 2013, among others).

The spatial analysis employs the administrative codes and a shapefile for Chinese county-level jurisdictions. As we describe in Section 3.2, two types of neighbors are considered: intra-regional (firms sharing the same administrative code) and inter-regional neighbors (firms in neighboring jurisdictions). To facilitate the estimation strategy discussed in Section 3.2, we retain observations in any jurisdiction that satisfies the following conditions: (1) It hosts at least two firms, (2) it has at least one contiguous neighbor that hosts at least two firms, and (3) it is located in one of the 10 coastal provinces (see Footnote 12). This procedure leaves us with a sample of 23,206 firms in 700 districts/counties: a total of 82,376 observations.

Following the discussions in Section 3.2, we divide the exogenous variables into firm-specific (X_1) and region-specific variables (X_2). X_1 consists of measures of R&D activity (*R&D*) and export activity (*export*). *R&D* is defined as the share of new products in gross output and *export* is defined as the share of exports in total sales revenue.¹⁶ X_2 includes three measures of agglomeration economies (specialization, variety, and competition) and local public expenditure. Our measures of *specialization* and *variety* follow those of Glaeser et al. (1992), whereas *competition* is defined as the Herfindahl–Hirschman index of local employment in the sector. More precisely, for any jurisdiction k and year t , we define:

$$specialization_{kt} = \frac{\text{employment of the sector in region } k \text{ and year } t}{\frac{\text{total industrial employment in region } k \text{ and year } t}{\frac{\text{national employment of the sector in year } t}{\text{national industrial employment in year } t}}}$$

⁹ Despite the simplification, the number of instruments could be large when we add interactions such as $i \cdot W_w y$ or $i \cdot W_n y$ to (12).

¹⁰ This means that the firm must have a sales revenue greater than RMB 5 million.

¹¹ Brandt et al. (2014) provide an excellent review of this dataset.

¹² The 10 provinces are Beijing, Fujian, Guangdong, Hebei, Jiangsu, Liaoning, Shandong, Shanghai, Tianjin, and Zhejiang.

¹³ It comprises eight vital sub-sectors: electric motor manufacturing (code range 3911–3919), electrical transmitting and controlling device manufacturing (code range 3921–3929), electrical and optical cable and electrical equipment manufacturing (code range 3931–3939), battery manufacturing (code 3940), home electrical appliances manufacturing (code range 3951–3959), home non-electrical appliances manufacturing (code range 3961–3969), lighting device manufacturing (code range 3971–3979), and other electrical machinery and equipment manufacturing (code range 3991–3999).

¹⁴ According to the classification of the Organization for Economic Co-operation and Development (OECD), the technology intensity of the electrical apparatus sector is medium high (<http://www.oecd.org/sti/ind/48350231.pdf>).

¹⁵ Over the study period 1999–2007, we have 26,516 distinct firms and 93,452 observations.

¹⁶ The literature (C. Hu et al., 2015; Sheng and Song, 2013) identifies other determinants of productivity, including firm size and ownership structure, but they are controlled for by the fixed effects.

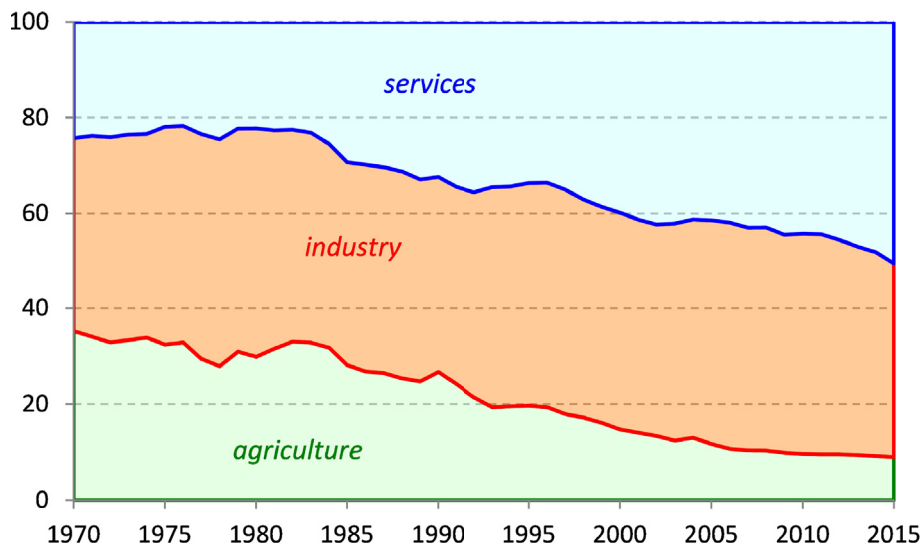


Fig. 2. China income composition; main sectors, value added (% of GDP)³¹.

$$variety_{kt} = \frac{\text{employment of the other five largest sectors in region } k \text{ and year } t}{\text{total industrial employment excluding the sector in region } k \text{ and year } t}$$

$$competition_{kt} = \sum_{i \in I_t(k)} \left(\frac{l_{ikt}}{\sum_{i \in I_t(k)} l_{ikt}} \right)^2,$$

where “the sector” refers to the electrical apparatus sector, and $I_t(k)$ and l_{ikt} are defined in the same way as in (10) and (11). Finally, local public expenditure (*public*) is measured by total fiscal expenditure by the local government (log-transformed) using data from the *Fiscal Statistics of Cities and Counties*.

4.3. Urban-transformation variables and classification of regions

We construct the urban-transformation variables using demographic data from the *China Population Census*. The census is conducted by NBS every 10 years. In this study, we use sub-prefecture data from the 2000 and 2010 census to reflect the demographic transition during the period of study. With the fixed-effects assumption, it is impossible to include any demographic variable as a conventional regressor in (1). Our strategy is to introduce them as interaction terms in (2). The coefficient of the interaction term γ then informs us whether the urban-transformation variable has any impact on the strength of spillover effects and by how much.¹⁷

This study considers three aspects of China’s urban-transformation process: *modernization*, *mobility*, and *disparity*. With the rapid urban transition and economic growth, there is rapid modernization of the demographic structure, driven initially by industrialization, and subsequently by a move toward services as shown in Fig. 2.

Data show that the economic transition is accompanied by a radical demographic change: nationwide total non-agricultural population increased from 437.5 million in 1999 to 606.3 million in 2007 (China Population and Employment Statistical Yearbook, 2018). We use the variable *modernization* to capture this demographic transition. In the literature, H. Lu, Xie, and Yao (2019) and Y. Zhang, Shao, and Dong (2018) measure modernization using the change of non-agricultural

labor supply. Thus, we define *modernization* as the change of non-agricultural population as a fraction of total population from 2000 to 2010 in the district/county. Across the nation, the variable is small in most jurisdictions, with a mean of 1.97% and a standard deviation of 7.35%.

The census also provides data on three types of immigrants: those from other districts/counties within the same prefecture, those from other prefectures within the same province, and those from other provinces. The literature emphasizes the role of migration of skilled labor in creating local economic growth and spillovers (Faggian et al., 2017; Fu and Gabriel, 2012). Census data clearly show that inter-provincial migrants along China’s eastern coast are mostly from rural areas of less-developed inland provinces (Cao et al., 2018; Shen, 2013). Further, intra-provincial migrants are, on average, more likely to originate from cities and towns, and are more educated than inter-provincial migrants (Liang, 2001; Su et al., 2018). In addition, Eriksson (2011) shows that only the inflow of skilled labor from within 50 km has a positive effect on the performance of plants. Based on these findings, we construct the second urban-transformation variable—*mobility*—using the fraction of short-distance (within prefecture) migrants in the total migrant population. Finally, regional disparity captures the dissimilarity in economic structure between neighboring regions. It is measured by the difference in employment structure between a jurisdiction and its neighbors:

$$disparity_i = \sqrt{\sum_{k=1}^3 (emp_{ki} - emp_{nki})^2},$$

where emp_{ki} ($k = 1, 2, 3$) are the employment shares of the primary, secondary, and tertiary industries in jurisdiction i , while emp_{nki} is the average of emp_{kj} in neighboring jurisdictions, weighted by their total employment. A large value indicates a wide gap in economic structure.

After discussing the three aspects of urban transformation, we classify regions into three types to capture the concepts of the central city (metro-core), suburban area (metro-ring), and conventional rural area (non-metro). Chinese jurisdictions at the sub-prefecture level are officially named urban districts or counties (including county-level cities). This dichotomy is a legacy from the Mao era, when urban infrastructure, such as paved roads, power and sewage, was assigned to urban dis-

¹⁷ Since (12) considers two spatial lags, two interaction terms may be used. The number of interaction terms increases when we consider multiple demographic factors simultaneously. However, using too many interaction terms makes the 2SLS procedure computationally challenging.

tricts only. Almost all industrial production was conducted in urban districts, whereas counties were populated exclusively by peasants. These urban districts were characterized by small land areas and high population density. Many still exist, but may have been renamed or experienced boundary changes. These urban districts constitute the old central cities (metro-core). Starting from the 1990s many counties close to the central cities were promoted to urban districts to cope with rapid economic growth and urban sprawl. Land in these jurisdictions was reclaimed for industrial and residential use, although a substantial amount of rural land remained. These newly established urban districts are usually much larger in area than the old central city, while population density is usually lower despite strong population growth. Finally, there was also economic development and population growth in more remote counties. Some were also promoted to urban districts with more administrative power, although they are less developed because of geographical disadvantages.

Our tripartite classification is not only based on the administrative hierarchy; it is also based on size, population density and historical legacy. As mentioned previously, our concept of metro-core is close to the central city concept proposed by Baum-Snow (2007) and Baum-Snow et al. (2017). It comprises the traditional city center that features a high level of urbanization and high population density. Since the literature does not provide any guidance, our classification is performed manually. In most situations, we define a jurisdiction as part of a metro-core if (1) it is an urban district by administrative hierarchy, (2) it is no greater than 300-square kilometers in area, and (3) the population density is no less than 2000 people per square kilometer. However, these rules are not strict. Sometimes the decision is also based on the administrative divisions created in the 1980s. During the process we consult the socioeconomic information provided by Wikipedia. In this way, we identify 355 metro-core districts across the nation. The metro-rings are then defined as the contiguous neighbors of the metro-core, while those not neighboring any metro-core are defined as non-metro regions. Statistically, the metro-cores occupy only a small fraction of the territory (0.54% nationwide and 2% in the east). Fig. 3 shows the metro-cores in our sample regions.

The distinctions between the three types of regions are illustrated for Shanghai, the largest city of the nation and the economic locomotive of the Yangtze River Delta. Fig. 4 shows the administrative division of Shanghai and Fig. 5 plots the key variables of interest. First, it is clear that metro-core districts are indeed much smaller than others in land area. Second, while the urbanization rate in the metro-core is higher than in the metro-ring and non-metro jurisdictions (see Fig. 5b), changes in the urbanization rate (see Fig. 5c) are more pronounced in the latter. This is probably because of the agglomeration of manufacturing firms (see Fig. 5a) that attracts long-distance migrants into these regions (see Fig. 5e). Third, the metro-core attracts more short-distance migrants rather than it does long-distance migrants (see Fig. 5d). Finally, regional disparity in economic structure is more pronounced in the metro-core than elsewhere (see Fig. 5f). With these observations, we expect that the impact of urbanization variables on productivity spillover varies across these regions.

The case of Shanghai may not be representative, so we show the distributional properties of the three urban transformation variables across the nation. In Fig. 6a, we plot the kernel density of *modern* for the three types of jurisdictions. All three curves appear similar in shape. However, the metro-core curve is to the left of the other two, with a larger spread. This is justified by the comparison of means and standard deviations (see Table 1). The graph shows that, on average, the percentage of non-agricultural population decreases in the metro-core but increases in the metro-ring and the non-metro regions, partially because of the strict population control resulting from the *Hukou* system in major cities (Chan, 2010; Y. Song, 2014). Meanwhile, there are large variations within each group. Except for the difference in mean, which is insignificant, there seems to be no systematic difference between the three types of jurisdictions. We also construct the variable *modern_n* as the average

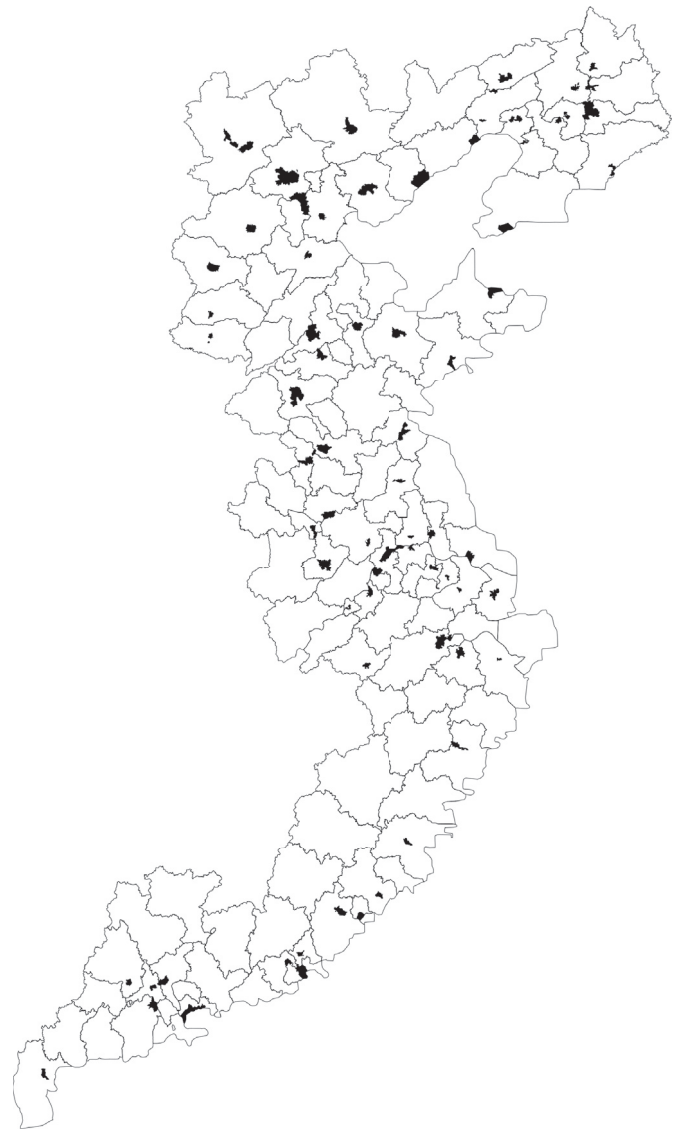


Fig. 3. The metro-cores (dark shade) in the 10 eastern provinces³².

of *modern* in neighboring jurisdictions, weighted by their population. These two variables, *modern* and *modern_n*, will interact with W_{wtfp} and $W_{n}tfp$ to gauge the impact of *modernization* on intra- and inter-regional spillovers, respectively. The mean and median of *modern_n* (see Table 1) are similar across the three types, and the density functions (not shown) are close to each other.

A pattern emerges in the density functions of *mobility* (see Fig. 6b). The metro-core sample has the lowest mean (42.4%), followed by the metro-ring sample (50.1%) and the non-metro sample (60.3%). Judging by the mean or median, the proportion of short-distance migrants is lowest in the metro-core but highest in the non-metro regions. We also construct a variable *mobility_n* as the weighted (by total migrant population) average of *mobility* in neighboring jurisdictions. The density plots of the new variable appear similar to those of *mobility*, but the differences are smaller.

The density functions of *disparity* reveal a clear distinction between metro-core and the other two types of regions in employment structure (see Fig. 6c). The distribution of the metro-core sample has a fat right tail. Judging by the mean, this variable is far larger in the metro-core sample (0.76), than in the metro-ring sample (0.33) or the non-metro sample (0.27). A large fraction of the metro-core sample (68%) has values of 0.5 or higher. This observation echoes the finding in Fig. 5f

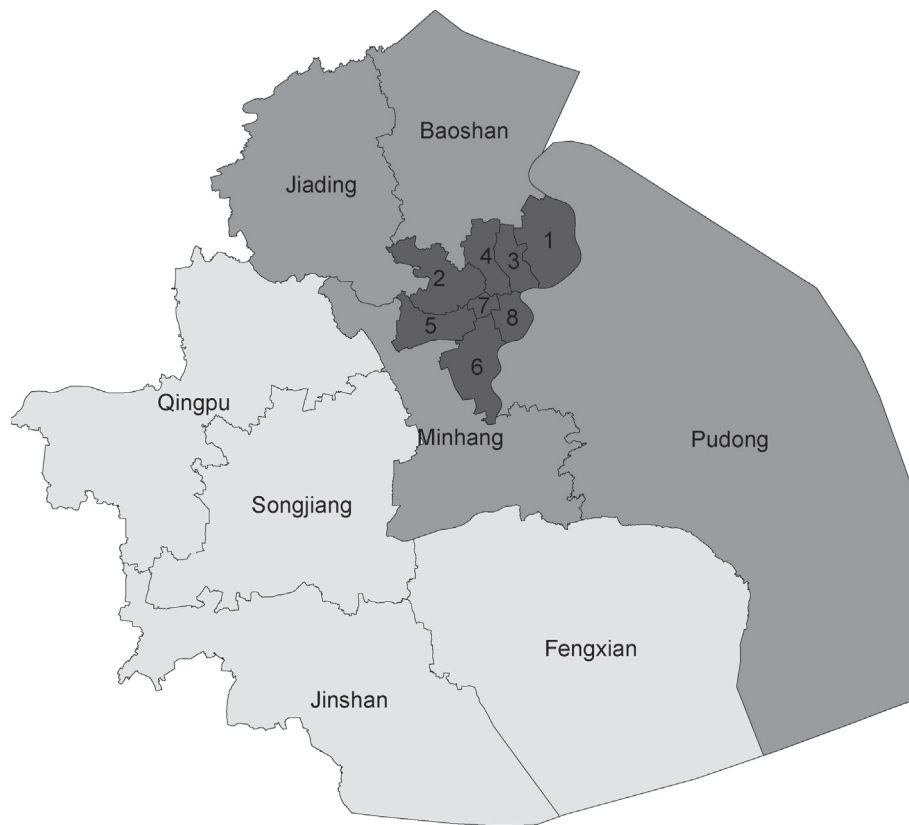


Fig. 4. The metro-core (dark), metro-ring (gray) and non-metro jurisdictions (light gray) of Shanghai³³.

that the economic structure of the metro-core is very different from that of the metro-ring. Economic disparity continues to fall as one moves from the metro-ring further into non-metro regions.

To summarize, the metro-core, metro-ring, and non-metro regions have systematic differences in *mobility* and *disparity*. If we find empirical evidence that technological spillovers benefit from any of these variables, there must be systematic differences in the magnitude of spillover effects among the three types of regions.

5. Empirical result

This section presents the main empirical findings. We first assess the strength of the spillover effects and present evidence that both effects are moderated by regional characteristics, such as land area, administrative type, and population density. We then extend the framework to study the impact of urban transformation on productivity spillovers, and conduct a robustness check.

5.1. The baseline results

We regress the baseline model (12) using the full sample with $X_1 = (R\&D, export)$ and $X_2 = (specialization, variety, competition, public)$. The instrumenting strategy follows (5'). The results are reported in Table 2. The first three columns present the FE, FE-2SLS and RE-2SLS estimates. Columns 4–6 report the estimation results controlling for land area, administrative type, and population density, respectively.

The spatial autoregressive (AR) coefficients $\hat{\lambda}_1$ and $\hat{\lambda}_2$ estimated by the conventional fixed-effects regression in column 1 are different from those of the FE-2SLS estimator in column 2, indicating the presence of endogeneity bias, as discussed in Section 3.1. The FE-2SLS estimates in column 1 are, in general, different from those of the RE-2SLS estimator in column 3. A test for $H_0 : \mu = 0$ gives an F -value of 7.33, which is highly significant, suggesting important omitted vari-

ables without the fixed effects. We also find strong correlations between the estimated individual effects $\hat{\mu}$ and the regressors $Z = (Wy, X)$ in (2): $\text{corr}(\hat{\mu}, Z\hat{\delta}_w) = -0.34$. In addition, the standard error of $\hat{\mu}$ (1.12) is large in size relative to that of \hat{Q}_0u (0.659). The evidence suggests that the random-effects assumption $E(\mu|X) = 0$ is likely to be invalid. The Hausman test statistic is 4773.56, which is also highly significant. Thus, we conclude that the fixed-effects assumption is appropriate, and the platform on which further analyses are based.¹⁸

The estimated spatial AR coefficients in column 2 are $\hat{\lambda}_1 = 0.506$ and $\hat{\lambda}_2 = 0.011$, respectively. According to these estimates, the productivity of a firm is expected to increase by 5% if the productivity of all intra-regional neighbors uniformly increases by 10%, and 0.1% if the productivity of all inter-regional neighbors uniformly increases by 10%. The intra-regional spillover effect is large in size and highly significant, but the inter-regional effect is small and insignificant. Thus, the result supports the general observation that spatial interactions attenuate rapidly with distance (Melo et al., 2013; Rosenthal and Strange, 2008, among others).

The coefficients of other regressors are mostly significant, indicating a strong influence on productivity of firm behavior (X_1) and regional characteristics (X_2). Their interpretations, however, demand special care. These estimates reflect the direct effect on a firm but ignore the indirect effect, which includes the circular feedback between the productivity of the firm (y_t) and that of neighbors ($W_{wt}y_t$ and $W_{nt}y_t$) through spatial interaction. Individual firms have different sets of neighbors, so the total effect differs across individual firms and its computation requires spatial transformation. For any exogenous regressor x_k in (12), its total effect on all firms is given by

¹⁸ The choice between spatial fixed-effects and random-effects models follows the same logic as in the non-spatial case. Please see Elhorst (2010) for a full discussion over this topic.

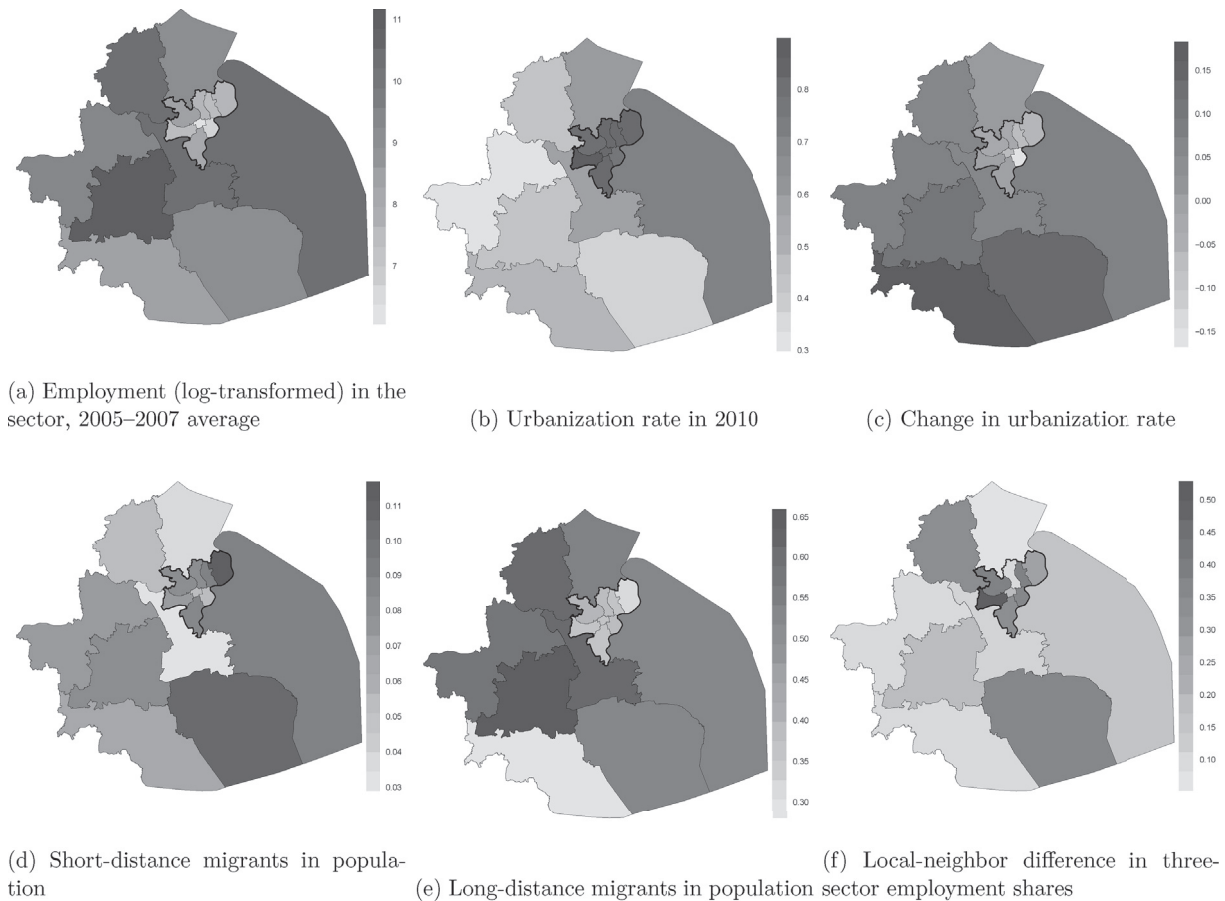


Fig. 5. Urban transformation in Shanghai, 2000–2010. Darker shades represent larger values. The thick solid black line encircles the metro-core.

$(I_{N_t} - \lambda_1 W_{wt} - \lambda_2 W_{nt})^{-1} \beta_k$.¹⁹ Despite its theoretical attractiveness, the total effect is computationally infeasible, and practically less appealing because it is firm-specific. For these reasons, the subsequent interpretations on β are all based on the direct effect.

The FE-2SLS estimates for R&D and export suggest a 1.7% or 0.8% increase in productivity if the firm increases its R&D or export share by 10%. Both parameters are highly significant. These results align with other findings on Chinese manufacturing firms (C. Hu et al., 2015; Sheng and Song, 2013). The coefficients of the three agglomeration measures are all significant. According to these estimates, firm-level productivity is harmed by *specialization*, but benefits from *variety* and *competition*. The findings are qualitatively similar to those of Glaeser et al. (1992). We also find strong evidence that local public expenditures contribute to firm's productivity, consistent with the findings of Demurger (2001) and Vijverberg et al. (2011). Finally, the spatial lags of all regional variables bear the same signs as their local counterparts, but the coefficients are smaller in size. They are all significant except for $W_{n, specialization}$. These results suggest that firms also benefit from public expenditures and agglomeration economies in neighboring regions, though their effects are weaker.

If spatial dependence attenuates in distance, the parameters of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ may not be the same in regions with different size. Thus, it is necessary to control for the impact of land area.²⁰ A firm is closer to its cross-border neighbors if it is located in a small jurisdiction. A small area also implies a small number of intra-regional neighbors and a rela-

tively large number of inter-regional neighbors. Therefore, we expect stronger inter-regional spillovers (larger and more significant λ_2) in small jurisdictions. That is, λ_2 becomes larger but λ_1 becomes smaller.

We also control for the effects of administrative hierarchy. Subprefectures in China are divided into urban districts or rural counties (including county-level cities). Urban districts comprise the metro-core and in many cases, most of the metro-ring. They host the prefecture government, most business activities, and non-agricultural population in the prefecture. As a result of their socioeconomic importance, they receive favorable considerations in regional planning.²¹ Statistically, urban districts are often smaller in land area and host disproportionately more firms than counties.²² Further, urban districts are far more likely to have urban districts as neighbors.²³ For these reasons, firms in urban districts are more likely to interact with firms located in neighboring jurisdictions. Thus, we expect a smaller λ_1 and a larger λ_2 in urban districts than we do in counties.

²¹ Pilot plans of industrialization in a given jurisdiction are usually accompanied by an escalation of the administration hierarchy, from rural county to urban district. For example, when the central government decided to forge Chongqing into the development engine of Western China in 1997, three counties and one county-level city were promoted to urban districts.

²² Of the 700 jurisdictions in our sample, 341 are urban districts. 312 of them are identified as small, but only 187 of the 359 counties are small. The mean land area of urban districts is 549 square kilometers; that of counties is 1672 square kilometers. Over the sample period, urban districts hosted 14,813 firms in our sample, compared to 8393 firms in counties.

²³ In the 10 provinces on the eastern coast (not limited to the sample area), an urban district on average has 4.79 contiguous neighbors, of which 2.85 are urban districts. In comparison, a county on average has 6.19 contiguous neighbors, but only 1.34 are urban districts.

¹⁹ LeSage and Pace (2009) provide an excellent exposition of the technical details.

²⁰ Taking Shanghai as an example, the smallest district (Jing'an) is only 7.26 square kilometers, but the largest (Pudong) is 1839 square kilometers.

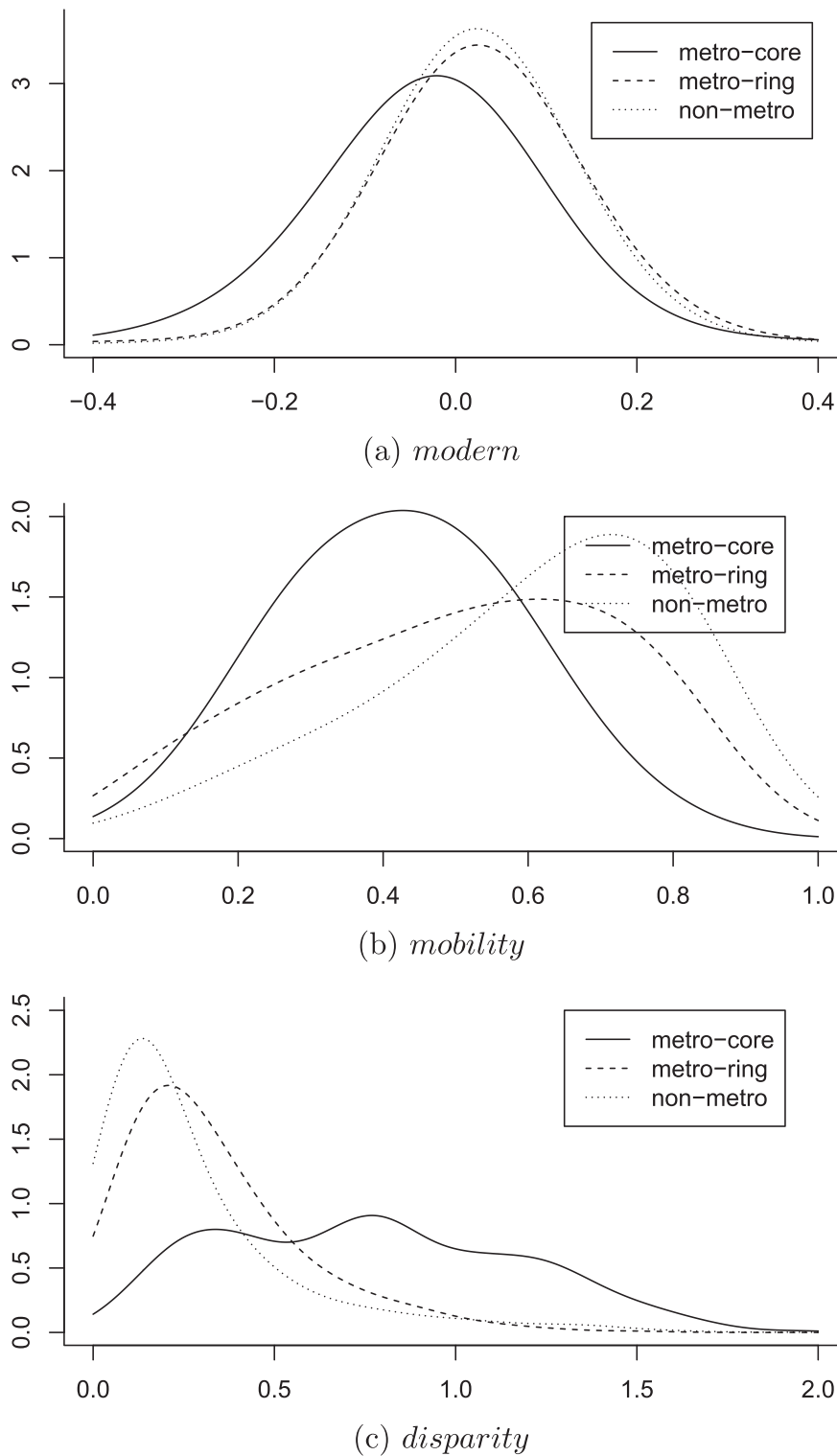


Fig. 6. Kernel density functions of urban transformation variables in metro-core districts (352), districts/counties in the metro-ring (416), and non-metro jurisdictions (2102). The bandwidth is set to 0.1 throughout.

Summary statistics show large variations in population density in the sample, ranging from 45 to 59,370 people per square kilometer. In this article, we use 2000 people per square kilometer as the cutoff value for population density based on our knowledge of metro-core areas which are known to be densely populated. This gives us 152 densely and 548 sparsely populated jurisdictions. As expected, most metro-core

districts are densely populated according to this definition. The correlation coefficient is high (0.79). It follows that densely populated jurisdictions are more likely to have neighbors with high population density (metro-core districts). Thus, we expect a closer link with their neighbors. Therefore, we postulate that λ_1 is smaller and λ_2 is larger in densely populated jurisdictions.

Table 1
Summary statistics of demographic variables by group.

		mean	median	sd	min	max
<i>modern</i>	metro-core	−0.026	−0.020	0.102	−0.392	0.494
	metro-ring	0.029	0.020	0.084	−0.413	0.696
	non-metro	0.025	0.019	0.062	−0.484	0.584
<i>modern_n</i>	metro-core	0.031	0.021	0.061	−0.092	0.464
	metro-ring	0.039	0.033	0.038	−0.141	0.189
	non-metro	0.032	0.025	0.038	−0.315	0.317
<i>mobility</i>	metro-core	0.424	0.426	0.156	0.032	0.832
	metro-ring	0.501	0.515	0.222	0.024	0.889
	non-metro	0.603	0.645	0.202	0.011	0.950
<i>mobility_n</i>	metro-core	0.424	0.430	0.161	0.035	0.808
	metro-ring	0.481	0.523	0.169	0.034	0.781
	non-metro	0.571	0.609	0.177	0.039	0.897
<i>disparity</i>	metro-core	0.761	0.751	0.407	0.032	1.946
	metro-ring	0.333	0.274	0.244	0.012	1.489
	non-metro	0.274	0.178	0.280	0.003	2.024

FE-2SLS estimates with controls of land area, administrative hierarchy and population density are reported in columns 4–6 of Table 2. In column 4, we control for difference in land area using the dummy variable *large*.²⁴ When we add interactions to the baseline model, the FE-2SLS regression uses the instrumenting strategy described in Section 3.1. λ_1 and λ_2 in the new model become the spatial AR coefficients for small jurisdictions; the coefficients of the interaction terms (γ_1 and γ_2) are the differences in the spatial AR coefficients between the two groups. The estimates verify our conjecture. We observe a sharp decline in $\hat{\lambda}_1$ and an increase in $\hat{\lambda}_2$ in column 4. The intra-regional spillover effects are stronger ($\gamma_1 > 0$) in large jurisdictions, but the inter-regional effects are weaker ($\gamma_2 < 0$). In columns 5 and 6, we interact $W_w tfp$ and $W_n tfp$ with *county* and *sparse*. Coefficients λ_1 and λ_2 denote spillover effects of urban districts or densely populated jurisdictions, while γ_1 and γ_2 measure the differences in the spillovers between different types of jurisdictions. The estimates for λ_1 and γ_1 support our arguments in both cases. However, neither λ_2 nor γ_2 are significant. These observations suggest that firms located in larger regions, counties or sparsely populated jurisdictions benefit more from intra-regional productivity spillovers.

Finally, we observe that the coefficients of exogenous regressors are qualitatively the same as those in column 2, $W_n specialization$ being the only exception. It is not significant in column 2 but significant in columns 4–6. Nevertheless, they share the same sign, which is consistent with that of *specialization*. These parameters are also comparable in size across the specifications. Judging by these estimates, the baseline model is robust to the inclusion of additional controls.

5.2. Urbanization and technological spillovers

In this section we extend the baseline model to investigate whether and to what extent urban-transformation variables moderate productivity spillovers among firms. As discussed in Section 4, we introduce three variables—*modern*, *mobility*, and *disparity*—to measure three aspects of urban transformation.

We first study *modern* and *modern_n*. The coefficient of $W_w tfp \times modern$ measures how the modernization process, measured by the change of non-agricultural population in a jurisdiction, influences the spillover among firms within the jurisdiction. Likewise, the coefficient of $W_n tfp \times modern_n$ measures how the inter-regional spillover effects are affected by modernization in neighboring jurisdictions. To assess the robustness of the results, we introduce these

interaction terms one at a time, then together. We further control for area size of the jurisdiction and repeat the procedure. In all specifications, we use the full set of exogenous regressors from the baseline model. The estimates are reported in Table 3.

In column 1, the coefficient of the interaction term is positive and highly significant, suggesting that progressive modernization in the jurisdiction strengthens spillovers among local firms. Similarly, in column 2, we find that modernization in neighboring jurisdictions significantly contributes to spillovers from inter-regional neighbors. There is little change to these results when both interaction terms are introduced together. However, when we control for land area in columns 4–6 in Table 3, the coefficient of $W_w tfp \times modern$ becomes not significant, while that of $W_n tfp \times modern_n$ remains significant. Both estimates also become smaller in size, probably because of the correlation between *large* and *modern* (0.17). In all specifications, the remaining estimates change little from those of the baseline model, either qualitatively or quantitatively.²⁵

According to the estimates in column 6, an increase in *modern* by one standard deviation (7.35%) results in an increase in λ_1 by 1.36 percentage points, equivalently a 4% increase. An increase in *modern_n* by one standard deviation (4.14%) causes λ_2 to increase by 3.06 percentage points, or 21%.²⁶ The second effect is larger in size and statistically significant. Thus, *modernization* is more effective on the inter-regional spillover effect than on the intra-regional one.

We perform a similar analysis on *mobility* and *mobility_n*, which measure the percent of short-distance migrants in the jurisdiction and neighboring jurisdictions, respectively. The former is interacted with $W_w tfp$ and the latter with $W_n tfp$. Table 4 reports the estimates.

When the two interaction terms are introduced to the baseline model separately (see columns 1 and 2), they are each positive and highly significant. The two coefficients become smaller but remain significant when introduced together. When we control for land area in columns 4 and 5, the results are quantitatively similar to those in columns 1 and 2. However, in the complete model (see column 6), the coefficient of $W_w tfp \times mobility$ becomes smaller in size and less significant. Strong correlation between *mobility* and *mobility_n* (0.86) may explain the instability of estimates when they are considered together. Nevertheless, $W_n tfp \times mobility_n$ remains highly significant. As for the other covariates, the coefficients of $W_w tfp \times large$ and $W_n tfp \times large$ are quite stable and close to their values in the baseline model. The same is true for all exogenous regressors (not reported). Finally, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ exhibit noticeable fluctuations, and are often different from the values reported in Table 2. This is probably an artifact since *mobility* and *mobility_n* have not been re-centered.²⁷

According to the complete model (see column 6), an increase in *mobility* by one standard deviation (0.21) increases λ_1 by 2.58 percentage points, which amounts to an 8% increase on the estimate in the baseline model. Similarly, an increase in *mobility_n* by one standard deviation (0.18) increases λ_2 by 6.84 percentage points, or 48%. The effect is far greater on λ_2 than on λ_1 . Thus, mobility benefits inter-regional spillovers.

Finally, we introduce interactions with *disparity*, which measures the difference in three-industry employment composition between a jurisdiction and its neighbors. This time, we interact the same variable with $W_w tfp$ and $W_n tfp$.

²⁵ The results on exogenous regressors are suppressed to save space.

²⁶ The standard deviations of demographic variables are taken from Table 1. The percentage changes are based on the small-jurisdiction estimates in column 4 of Table 2 (i.e., $\hat{\lambda}_1 = 0.340$ and $\hat{\lambda}_2 = 0.143$).

²⁷ The sample means of *mobility* (0.29) and *mobility_n* (0.28) are greater than zero, so using them without recentering will shift $\hat{\lambda}_1$ and $\hat{\lambda}_2$ in the opposite direction, as we observe in Table 4. The models in Table 3 are barely affected because the sample means of *modern* (0.02) and *modern_n* (0.03) are both close to zero. We did not recenter because the variables of interest are the interaction terms. The caveat is that $\hat{\lambda}_1$ and $\hat{\lambda}_2$ must be interpreted with caution.

²⁴ We use a threshold value of 1552 square kilometers, which is the median of all 2870 jurisdictions in the nation. Of the 700 jurisdictions in the eastern sample, 201 are identified as large.

Table 2
Baseline model and controls for jurisdiction types.

regressor	Dependent variable: <i>tfp</i>					
	1	2	3	4	5	6
	FE	FE-2SLS	RE-2SLS	FE-2SLS with controls		
				<i>large</i>	<i>county</i>	<i>sparse</i>
$W_w tfp$	0.221** (0.009)	0.506** (0.054)	0.284** (0.030)	0.340** (0.050)	0.396** (0.051)	0.420** (0.045)
$W_n tfp$	0.147** (0.013)	0.011 (0.061)	0.387** (0.034)	0.143** (0.058)	0.017 (0.057)	0.001 (0.048)
$W_w tfp \times control$	–	–	–	0.364** (0.046)	0.254** (0.044)	0.164** (0.043)
$W_n tfp \times control$	–	–	–	–0.124** (0.053)	–0.009 (0.050)	0.036 (0.047)
<i>R&D</i>	0.165** (0.024)	0.173** (0.024)	0.345** (0.021)	0.166** (0.024)	0.163** (0.024)	0.168** (0.024)
<i>export</i>	0.075** (0.019)	0.080** (0.019)	0.115** (0.014)	0.083** (0.019)	0.077** (0.019)	0.080** (0.019)
<i>specialization</i>	–0.017** (0.007)	–0.046** (0.010)	–0.021** (0.006)	–0.042** (0.009)	–0.040** (0.009)	–0.048** (0.009)
<i>variety</i>	–0.918** (0.103)	–0.664** (0.114)	0.145** (0.068)	–0.505** (0.112)	–0.660** (0.112)	–0.642** (0.111)
<i>competition</i>	–0.565** (0.059)	–0.916** (0.091)	–0.333** (0.059)	–0.772** (0.084)	–0.832** (0.088)	–0.850** (0.083)
<i>public</i>	0.194** (0.010)	0.128** (0.023)	0.000 (0.008)	0.109** (0.021)	0.135** (0.021)	0.114** (0.020)
$W_n specialization$	–0.032** (0.009)	–0.021 (0.010)	–0.070** (0.006)	–0.046** (0.009)	–0.035** (0.010)	–0.034** (0.010)
$W_n variety$	–0.213 (0.146)	–0.397** (0.169)	0.593** (0.096)	–0.298* (0.162)	–0.462** (0.165)	–0.476** (0.160)
$W_n competition$	–0.677** (0.073)	–0.338** (0.108)	–0.254** (0.066)	–0.453** (0.099)	–0.316** (0.102)	–0.297** (0.097)
$W_n public$	0.024** (0.004)	0.015** (0.005)	0.009** (0.004)	0.010** (0.005)	0.022** (0.004)	0.016** (0.004)
<i>Intercept</i>	2.284** (0.164)	2.035** (0.209)	1.191** (0.147)	1.899** (0.199)	2.020** (0.200)	1.961** (0.197)
R^2	0.132	0.115	0.115	0.128	0.120	0.115
N	72,805	72,805	72,805	72,805	72,805	72,805

Significance codes: ‘***’ 0.05, ‘**’ 0.10.

Table 3
Modernization on intra- and inter-regional spillover effects.

regressor	Dependent variable: <i>tfp</i>					
	1	2	3	4	5	6
$W_w tfp$	0.446** (0.008)	0.486** (0.040)	0.468** (0.034)	0.329** (0.038)	0.323** (0.039)	0.333** (0.034)
$W_n tfp$	0.008 (0.045)	0.022 (0.047)	0.008 (0.040)	0.102** (0.045)	0.141** (0.045)	0.101** (0.039)
$W_w tfp \times large$	–	–	–	0.359** (0.043)	0.367** (0.043)	0.357** (0.042)
$W_n tfp \times large$	–	–	–	–0.117** (0.051)	–0.129** (0.050)	–0.118** (0.050)
$W_w tfp \times modern$	0.350** (0.144)	–	0.313** (0.142)	0.214 (0.145)	–	0.185 (0.143)
$W_n tfp \times modern_n$	–	1.206** (0.269)	1.047** (0.265)	–	0.920** (0.269)	0.740** (0.264)
other regressors	all	all	all	all	all	all
fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.122	0.118	0.121	0.130	0.130	0.129
N	72,805	72,805	72,805	72,805	72,805	72,805

Significance codes: ‘***’ 0.05, ‘**’ 0.10.

When *disparity* is interacted with $W_w tfp$ alone (see columns 1 and 4), its coefficient is not significant, no matter whether we control for land area or not. The interaction between *disparity* and $W_n tfp$ always has a negative sign and is highly significant. This is true whether it is introduced alone, or with $W_w tfp \times disparity$, or when we control for land area. When we consider the two interaction terms together (see columns 3 and 6), both coefficients become larger. Further, the

coefficient of $W_w tfp \times disparity$ becomes highly significant. The results indicate that large regional disparity inhibits inter-regional spillovers. Consequently, firms rely more on their close neighbors, resulting in a smaller λ_2 and larger λ_1 . The mechanism is similar to that of land area, discussed in Section 5.1. Note that $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are again unstable because

Table 4
Short-distance immigration on intra- and inter-regional spillover effects.

regressor	Dependent variable: tfp					
	1	2	3	4	5	6
$W_w tfp$	0.364** (0.045)	0.451** (0.040)	0.400** (0.043)	0.151** (0.045)	0.244** (0.042)	0.191** (0.043)
$W_n tfp$	−0.021 (0.041)	−0.073 (0.046)	−0.087* (0.045)	0.116** (0.047)	0.050 (0.050)	0.023 (0.048)
$W_w tfp \times large$	—	—	—	0.374** (0.043)	0.393** (0.043)	0.391** (0.042)
$W_n tfp \times large$	—	—	—	−0.111** (0.049)	−0.135** (0.049)	−0.126** (0.048)
$W_w tfp \times mobility$	0.342** (0.053)	—	0.207** (0.069)	0.322** (0.051)	—	0.123* (0.065)
$W_n tfp \times mobility_n$	—	0.418** (0.063)	0.254** (0.082)	—	0.455** (0.062)	0.380** (0.078)
other regressors	all	all	all	all	all	all
fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.115	0.121	0.118	0.131	0.133	0.134
N	72,805	72,805	72,805	72,805	72,805	72,805

Significance codes: ‘***’ 0.05, ‘**’ 0.10.

Table 5
Regional disparity on intra- and inter-regional spillover effects.

regressor	Dependent variable: tfp					
	1	2	3	4	5	6
$W_w tfp$	0.548** (0.047)	0.166** (0.014)	0.489** (0.053)	0.367** (0.044)	0.075** (0.016)	0.262** (0.051)
$W_n tfp$	0.009 (0.022)	0.233** (0.054)	0.098 (0.066)	0.072** (0.023)	0.303** (0.052)	0.272** (0.062)
$W_w tfp \times large$	—	—	—	0.372** (0.043)	0.447** (0.043)	0.405** (0.045)
$W_n tfp \times large$	—	—	—	−0.133** (0.049)	−0.174** (0.050)	−0.172** (0.051)
$W_w tfp \times disparity$	−0.110 (0.078)	—	0.261** (0.125)	0.048 (0.078)	—	0.498** (0.122)
$W_n tfp \times disparity$	—	−0.334** (0.096)	−0.577** (0.155)	—	−0.227** (0.094)	−0.711** (0.149)
other regressors	all	all	all	all	all	all
fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.113	0.132	0.110	0.125	0.136	0.124
N	72,797	72,797	72,797	72,797	72,797	72,797

Significance codes: ‘***’ 0.05, ‘**’ 0.10.

of the centering issue discussed previously.²⁸ Finally, the estimates of the other regressors (controls for land area and exogenous regressors) are all qualitatively similar to those of the baseline model.

According to column 6, an increase in *disparity* by one standard deviation (0.33) increases λ_1 by 16.4 percentage points, which is a 48% increase on the estimate in the baseline model. The change in λ_2 is −23.5 percentage points, equivalently a 164% decrease from the estimate in the baseline model. Judging by these numbers, the effect of *disparity* is strong on both types of spillover effects.

5.3. Robustness check

In this section, we test whether the results on urban-transformation variables are robust to different model specifications. We first include all interaction terms with the three urbanization variables, and then control for differences in land area, administrative type, and population density. The results are reported in Table 6.

Comparing the estimates in column 3 of Tables 3–5 to those of column 1 in Table 6, we observe that all previous results regarding the

interaction terms remain unchanged when introduced together. The effects of modernization on inter-regional spillovers (e.g., the coefficient on $W_n tfp \times modern_n$) are stronger than on intra-regional effects. When land area is also controlled for in column 2, $W_w tfp \times modern$ and $W_w tfp \times mobility$ lose their significance, similar to prior observations (see column 6 of Tables 3 and 4). When we control for *county* or *sparse*, the results are similar to those in column 2, except for the coefficient of $W_w tfp \times modern$, which becomes weakly significant. Finally, the estimated coefficients of all exogenous variables are robust throughout, so they are omitted in Table 6.

To summarize, the main results on urban-transformation variables are robust to the specifications considered here. We find evidence that *modernization* and *mobility* in neighboring jurisdictions significantly improve spillovers from inter-regional neighbors. Regional economic disparity suppresses inter-regional spillovers and make each jurisdiction more self-contained (stronger intra-regional spillovers). We also find weak evidence that *modernization* or *mobility* in the local jurisdiction also contributes to spillovers among local firms, but these effects are usually small in size and sensitive to specification change. Judging by statistical significance and relative size, the effects of urban transformation are mostly experienced in the inter-regional scope.

²⁸ The sample mean of *disparity* is 0.30.

Table 6
Other specifications.

regressor		Dependent variable: <i>tfp</i>		
		further controls		
		<i>large</i>	<i>county</i>	<i>sparse</i>
$W_w tfp$	0.296** (0.040)	0.119** (0.041)	0.214** (0.040)	0.205** (0.044)
$W_n tfp$	0.051 (0.047)	0.160** (0.050)	0.062 (0.047)	0.073 (0.049)
$W_w tfp \times control$	–	0.397** (0.039)	0.344** (0.043)	0.117** (0.037)
$W_n tfp \times control$	–	–0.151** (0.045)	–0.148** (0.046)	–0.028 (0.041)
$W_w tfp \times modern$	0.265* (0.144)	0.167 (0.143)	0.237* (0.143)	0.274* (0.142)
$W_n tfp \times modern_n$	1.424** (0.272)	1.062** (0.273)	1.215** (0.273)	1.167** (0.271)
$W_w tfp \times mobility$	0.149** (0.067)	0.077 (0.064)	–0.111 (0.077)	0.101 (0.068)
$W_n tfp \times mobility_n$	0.338** (0.080)	0.409** (0.076)	0.384** (0.082)	0.327** (0.085)
$W_w tfp \times disparity$	0.373** (0.117)	0.561** (0.116)	0.720** (0.124)	0.481** (0.119)
$W_n tfp \times disparity$	–0.854** (0.143)	–0.897** (0.140)	–0.887** (0.145)	–0.808** (0.144)
other regressors	all	all	all	all
fixed effects	yes	yes	yes	yes
R^2	0.123	0.131	0.124	0.125
N	72,797	72,797	72,797	72,797

Significance codes: *** 0.05, ** 0.10.

6. Further discussion and policy implications

In Section 5.2, when we consider all regions together, we find that regional differences in technological spillovers are mostly explained by regional disparity in economic structure (*disparity*), followed by short-distance labor mobility (*mobility*), and by the change of non-agricultural population (*modernization*). In addition, the inter-regional spillover effects are found to be more sensitive to urban-transformation features than to intra-regional spillover effects. This section further elaborates on the practical implications of our findings in metro-core, metro-ring, and non-metro regions.

According to Table 1, the average share of non-agricultural population increases in the metro-ring (2.9%) and the non-metro (2.5%) regions, but declines in the metro-core regions (–2.6%). As for the value in neighboring jurisdictions, the group means are similar (3.1%–3.9%). Our estimates (see column 6 of Table 3) suggest that λ_1 and λ_2 will be respectively increased by 1.85 and 7.40 percentage points, following a 10% increase in local modernization rate (*modern*) and the rate in neighboring jurisdictions (*modern_n*). Judging by the mean difference in *modern* and *modern_n* (see Table 1), *modernization* explains, at most, 1% of the difference in intra-regional spillover²⁹, and 0.6% of the difference in inter-regional spillover among the three groups. Comparing these values to $\hat{\lambda}_1 = 0.340$ and $\hat{\lambda}_2 = 0.143$, obtained in the baseline model, we conclude that *modernization* explains little between-group variations in technological spillovers.

Mobility is highest in non-metro regions but lowest in metro-core regions (see Table 1 and Fig. 6c). According to our estimates (see column 6 of Table 4), λ_1 and λ_2 will be increased by 1.23 and 3.80 percentage points, respectively, following a 10% increase in *mobility* either locally or in neighboring jurisdictions. Therefore, the between-group

difference in *mobility* (see Table 1) explains up to 2% of the difference in intra-regional spillover and 5.6% of the difference in inter-regional spillover. Thus, *mobility* has a moderate impact on the between-group variation in inter-regional spillover.

Disparity is highest between metro-core regions and their neighbors (see Table 1 and Fig. 6c). The estimates (see column 6 of Table 5) suggest a strong impact from this index on λ_1 and λ_2 : They are expected to change by 49.8 and –71.1 percentage points if the index increases by one unit. The between-group variation in this variable is as high as 0.487 between the metro-core and non-metro (see Table 1), which results in a difference of 24.3% in intra-regional spillover and a difference of –34.6% in inter-regional spillover. Clearly, regional disparity in economic structure explains most between-group variations in both spillover effects.

This discussion indicates that various features of urban transformation have heterogeneous impacts on technological spillovers in different types of regions. The traditional city centers (i.e., metro-core regions) are cursed by their urban-transformation feature of high-level regional disparity, which discourages spillovers from neighbors. They are also adversely affected by low level of labor mobility, both locally and in their neighbors, but to a lesser extent. In comparison, metro-ring and non-metro regions benefit more from inter-regional technological spillovers because of their urban-transformation features of a lower level of regional disparity and higher level of labor mobility.

These results have rich implications for business practitioners, urban planners, and policymakers. First, we find large differences of economic structure between traditional central cities and their neighbors. Although this urban transformation enhances productivity spillovers within the metro-core, it strongly impedes inter-regional productivity spillovers. Easing regional disparity, in particular that of metro-core regions, could generate sizable benefits to firms located in these areas. Second, the current model of economic development, especially that in Eastern China, is partially based on using long-distance migrants as cheap labor. Our analysis suggests that technological spillovers significantly benefit from short-distance migration. Since non-metro regions have the highest average level of mobility, firms in non-metro regions

²⁹ Table 1 shows that the difference in the mean of *modern* is at most 0.051. Given that a 1% increase in local modernization rate boosts the intra-regional productivity spillover by 0.185%, 5.1% difference in local modernization rate could, at most, explain 1% ($0.185\% \times 5.1 \approx 1\%$) of the difference in intra-regional spillover among the three groups.

could benefit more from inter-regional spillover. Urban policies that facilitate short-distance migration, especially in metro-core regions, might create a further boost to inter-regional spillovers. Third, beyond the systematic difference between groups, adequate attention must be paid to within-group variations in urbanization, which is high for all three types. Hosting a metro-core or not, regions with high levels of regional disparity or low levels of mobility deserve special policy considerations.

7. Concluding remarks

Productivity growth and urban transformation have been at the core of debate on sustainable growth and development. We use a spatial Durbin model to estimate the intra-regional and inter-regional productivity spillovers, employing a rich firm-level dataset, including 23,206 firms located in 700 districts and counties from 1999 to 2007. We find that intra-regional productivity spillover effects are larger than inter-regional spillover effects. Firms in large regions, rural counties or sparsely populated jurisdictions benefit more from intra-regional spillovers than they do from inter-regional spillovers. Both types of productivity spillovers are sourced from firm-specific R&D, exporting, and region-specific public expenditures and agglomeration factors.

We extend the baseline analysis considering impacts of urban transformation on productivity spillovers. Using the sub-prefecture level of population census data in 2000 and 2010, we measure three features of urban transformation, resulting from the demographic change (*modernization*), the influx of short-distance mobile workers (*mobility*), and the change of economic structure (*disparity*). Our empirical evidence indicates that both *modernization* and *mobility* enhance inter-regional productivity spillovers. Regional disparity stimulates intra-regional productivity spillovers, but weakens inter-regional spillovers. The results are robust to changes in specifications and controls for regional characteristics, such as land area, administrative hierarchy, or population density.

Further, we distinguish and discuss the impact of urban transformation on productivity spillovers in three types of regions. Our classification of metro-core, metro-ring, and non-metro regions is not only based on size, administrative hierarchy, and population density, but also on historical legacy and neighborhood relationship. This regional division is new, conceptually close to the city-center idea, and more practical for urban planners and policymakers to understand the heterogeneous impact of urban transformation on technological spillovers.

Among the three urban transformation measures, we find that *disparity* explains most of the variations in spillover effects, followed by *mobility*, and *modernization*. During China's rapid urban transformation, metro-core regions are characterized by higher levels of regional disparity and lower levels of labor mobility. As a result, production externalities are, to a larger extent, locked inside the metro-core regions, while inter-regional spillovers are much more obstructed, despite their high levels of economic development. Different from traditional city centers, non-core regions reap substantial benefits from inter-regional spillovers, since they attract a large percent of short-distance migrant workers, and the economic structures are similar to their neighbors. Our results highlight the importance of policy interventions in regions that have large disparity and lower mobility during rapid urban transformation. Particular attention should be paid to the removal of institutional and policy-induced migration barriers across regions, and the balance of employment structure across sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.econmod.2020.03.014>.

References

- Aiello, F., Pupo, V., Ricotta, F., 2014. Explaining total factor productivity at firm level in Italy: does location matter? *Spatial Econ. Anal.* 9 (1), 51–70.
- Annez, P.C., Buckley, R.M., 2009. Urbanization and growth: setting the context. In: Spence, M., Annez, P.C., Buckley, R.M. (Eds.), *Urbanization and Growth*. The World Bank, Washington, DC, pp. 1–45.
- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Anwar, S., Sun, S., 2014. Heterogeneity and curvilinearity of FDI-related productivity spillovers in China's manufacturing sector. *Econ. Modell.* 41, 23–32.
- Au, C.-C., Henderson, J.V., 2006a. Are Chinese cities too small? *Rev. Econ. Stud.* 73 (3), 549–576.
- Au, C.-C., Henderson, J.V., 2006b. How migration restrictions limit agglomeration and productivity in China. *J. Dev. Econ.* 80 (2), 350–388.
- Baltagi, B.H., 2008. *Econometric Analysis of Panel Data*. John Wiley & Sons, London, UK.
- Baltagi, B.H., Egger, P.H., Kesina, M., 2015. Sources of productivity spillovers: panel data evidence from China. *J. Prod. Anal.* 43 (3), 389–402.
- Bao, Q., Huang, J., Wang, Y., 2015. Productivity and firms' sales destination: Chinese characteristics. *Rev. Int. Econ.* 23 (3), 620–637.
- Baum-Snow, N., 2007. Did highways cause suburbanization? *Q. J. Econ.* 122 (2), 775–805.
- Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., Zhang, Q., 2017. Roads, railroads, and decentralization of Chinese cities. *Rev. Econ. Stat.* 99 (3), 435–448.
- Berkowitz, D., Ma, H., Nishioka, S., 2017. Recasting the iron rice bowl: the reform of China's state-owned enterprises. *Rev. Econ. Stat.* 99 (4), 735–747.
- Boeing, P., Mueller, E., Sandner, P.G., 2016. China's R&D explosion—analyzing productivity effects across ownership types and over time. *Res. Pol.* 45 (1), 159–176.
- Bond, E., Riezman, R., 2016. *Urbanization and Economic Development: A Tale of Two Barriers*. CESifo Working Paper Series 5742.
- Bosker, M., Brakman, S., Garretsen, H., Schramm, M., 2012. Relaxing Hukou: increased labor mobility and China's economic geography. *J. Urban Econ.* 72 (2–3), 252–266.
- Bournakis, I., Christopoulos, D., Mallick, S., 2018. Knowledge spillovers and output per worker: an industry-level analysis for OECD countries. *Econ. Inq.* 56 (2), 1028–1046.
- Bournakis, I., Mallick, S., 2018. TFP estimation at firm level: the fiscal aspect of productivity convergence in the UK. *Econ. Modell.* 70, 579–590.
- Brakman, S., Garretsen, H., van Marrewijk, C., 2016. Urban development in China. *Camb. J. Reg. Econ. Soc.* 9 (3), 467–477.
- Brandt, L., Van Biesebroeck, J., Zhang, Y., 2012. Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. *J. Dev. Econ.* 97 (2), 339–351.
- Brandt, Loren, Van Biesebroeck, Johannes, Zhang, Yifan, 2014. Challenges of working with the Chinese NBS firm-level data. *China Econ. Rev.* 30 (9), 339–352.
- Cai, F., Du, Y., 2011. Wage increases, wage convergence, and the Lewis turning point in China. *China Econ. Rev.* 22 (4), 601–610.
- Cai, F., Wang, M., 2010. Growth and structural changes in employment in transition China. *J. Comp. Econ.* 38 (1), 71–81.
- Cao, Z., Zheng, X., Liu, Y., Li, Y., Chen, Y., 2018. Exploring the changing patterns of China's migration and its determinants using census data of 2000 and 2010. *Habitat Int.* 82, 72–82.
- Chan, K.W., 2010. Fundamentals of China's urbanization and policy. *China Rev.* 10 (1), 63–93.
- Chen, M., Guariglia, A., 2013. Internal financial constraints and firm productivity in China: do liquidity and export behavior make a difference? *J. Comp. Econ.* 41 (4), 1123–1140.
- Chen, M., Liu, W., Tao, X., 2013. Evolution and assessment on China's urbanization 1960–2010: under-urbanization or over-urbanization? *Habitat Int.* 38, 25–33.
- Chen, Q., Song, Z., 2014. Accounting for China's urbanization. *China Econ. Rev.* 30, 485–494.
- Chen, Z., Haynes, K.E., Zhou, Y., Dai, Z., 2019. *High Speed Rail and China's New Economic Geography: Impact Assessment from the Regional Science Perspective*. Edward Elgar Publishing, Northampton, MA.
- Combes, P.-P., 2000. Economic structure and local growth: France, 1984–1993. *J. Urban Econ.* 47 (3), 329–355.
- Demurger, S., 2001. Infrastructure development and economic growth: an explanation for regional disparities in China. *J. Comp. Econ.* 29 (1), 95–117.
- Deng, P.D., Jefferson, G.H., 2011. Explaining spatial convergence of China's industrial productivity. *Oxf. Bull. Econ. Stat.* 73 (6), 818–832.
- Ding, S., Guariglia, A., Harris, R., 2016. The determinants of productivity in Chinese large and medium-sized industrial firms, 1998–2007. *J. Prod. Anal.* 45 (2), 131–155.
- Duranton, G., 2008. From cities to productivity and growth in developing countries. *Can. J. Econ.* 41 (3), 689–736.
- Eapen, A., Yeo, J., Sasidharan, S., 2019. Finance constraints and technology spillovers from foreign to domestic firms. *Econ. Modell.* 76, 50–62.
- Elhorst, J.P., 2003. Specification and estimation of spatial panel data models. *Int. Reg. Sci. Rev.* 26 (3), 244–268.
- Elhorst, J.P., 2010. Spatial panel data models. In: Fischer, M.M., Getis, A. (Eds.), *Handbook of Applied Spatial Analysis*. Springer, Berlin, Heidelberg, pp. 377–407.
- Eriksson, R.H., 2011. Localized spillovers and knowledge flows: how does proximity influence the performance of plants. *Econ. Geogr.* 87 (2), 127–152.
- Faggian, A., Rajbhandari, I., Dotzel, K.R., 2017. The interregional migration of human capital and its regional consequences: a review. *Reg. Stud.* 51 (1), 128–143.
- Fu, Y., Gabriel, S.A., 2012. Labor migration, human capital agglomeration and regional development in China. *Reg. Sci. Urban Econ.* 42 (3), 473–484.

- Gao, H., Van Biesebroeck, J., 2014. Effects of deregulation and vertical unbundling on the performance of China's electricity generation sector. *J. Ind. Econ.* 62 (1), 41–76.
- Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in cities. *J. Polit. Econ.* 100 (6), 1126–1152.
- Golley, J., Meng, X., 2011. Has China run out of surplus labour? *China Econ. Rev.* 22 (4), 555–572.
- Haskel, J.E., Pereira, S.C., Slaughter, M.J., 2007. Does inward foreign direct investment boost the productivity of domestic firms? *Rev. Econ. Stat.* 89 (3), 482–496.
- He, Z., Shi, X., Wang, X., Xu, Y., 2017. Urbanisation and the geographic concentration of industrial SO₂ emissions in China. *Urban Stud.* 54 (15), 3579–3596.
- Henderson, J.V., 2003a. Marshall's scale economies. *J. Urban Econ.* 53 (1), 1–28.
- Henderson, J.V., 2003b. The urbanization process and economic growth: the so-what question. *J. Econ. Growth* 8 (1), 47–71.
- Howell, A., He, C., Yang, R., Fan, C., 2016. Technological relatedness and asymmetrical firm productivity gains under market reforms in China. *Camb. J. Reg. Econ. Soc.* 9 (3), 499–515.
- Hsieh, C.-T., Klenow, P.J., 2009. Misallocation and manufacturing TFP in China and India. *Q. J. Econ.* 124 (4), 1403–1448.
- Hsieh, C.-T., Song, Z.M., 2015. Grasp the large, let go of the small: the transformation of the state sector in China. *Brookings Pap. Econ. Activ.* 2015 (Spring), 295–346.
- Hu, C., Xu, Z., Yashiro, N., 2015. Agglomeration and productivity in China: firm level evidence. *China Econ. Rev.* 33 (4), 50–66.
- Hu, Y., Fisher-Vanden, K., & Su, B. (in press). Technological spillover through industrial and regional linkages: firm-level evidence from China. *Economic Modelling*.
- Jefferson, G.H., Hu, A.G.Z., Su, J., 2006. The sources and sustainability of China's economic growth. *Brookings Pap. Econ. Activ.* 37 (2), 1–60.
- Jefferson, G.H., Rawski, T.G., Zhang, Y., 2008. Productivity growth and convergence across China's industrial economy. *J. Chin. Econ. Bus. Stud.* 6 (2), 121–140.
- Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure of estimating a spatial autoregressive model with autoregressive disturbances. *J. R. Estate Finance Econ.* 17 (1), 99–121.
- Kelejian, H.H., Prucha, I.R., Yuzefovich, Y., 2004. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results. In: Lesage, J.P., Pace, R.K. (Eds.), *Advances in Econometrics*, vol. 18. Emerald Group Publishing Limited, London, UK, pp. 163–198.
- Knight, J., Deng, Q., Li, S., 2011. The puzzle of migrant labour shortage and rural labour surplus in China. *China Econ. Rev.* 22 (4), 585–600.
- Krugman, P., 1997. *The Age of Diminished Expectations: US Economic Policy in the 1990s*. MIT Press, Cambridge, MA.
- Kumar, A., Kober, B., 2012. Urbanization, human capital, and cross-country productivity differences. *Econ. Lett.* 117 (1), 14–17.
- Lee, L.-F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. *J. Econom.* 154 (2), 165–185.
- Lemoine, F., Poncet, S., Ünal, D., 2015. Spatial rebalancing and industrial convergence in China. *China Econ. Rev.* 34, 39–63.
- LeSage, J., Pace, R.K., 2009. *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL.
- Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. *Rev. Econ. Stud.* 70 (2), 317–341.
- Li, H., Wei, Y., Ning, Y., 2016. Spatial and temporal evolution of urban systems in China during rapid urbanization. *Sustainability* 8 (7), 1–17.
- Liang, Z., 2001. The age of migration in China. *Popul. Dev. Rev.* 27 (3), 499–524.
- Lin, H.-L., Li, H.-Y., Yang, C.-H., 2011. Agglomeration and productivity: firm-level evidence from China's textile industry. *China Econ. Rev.* 22 (3), 313–329.
- Liu, T., Qi, Y., Cao, G., Liu, H., 2015. Spatial patterns, driving forces, and urbanization effects of China's internal migration: county-level analysis based on the 2000 and 2010 censuses. *J. Geogr. Sci.* 25 (2), 236–256.
- Lu, H., Xie, H., Yao, G., 2019. Impact of land fragmentation on marginal productivity of agricultural labor and non-agricultural labor supply: a case study of Jiangsu, China. *Habitat Int.* 83, 65–72.
- Lu, M., Chen, Z., 2006. Urbanization, urban-biased policies, and urban-rural inequality in China, 1987–2001. *Chin. Econ.* 39 (3), 42–63.
- Melo, P.C., Graham, D.J., Brage-Ardao, R., 2013. The productivity of transport infrastructure investment: a meta-analysis of empirical evidence. *Reg. Sci. Urban Econ.* 43 (5), 695–706.
- Pannell, C.W., 2002. China's continuing urban transition. *Environ. Plann. A* 34 (9), 1571–1589.
- Rosenthal, S.S., Strange, W.C., 2008. The attenuation of human capital spillovers. *J. Urban Econ.* 64 (2), 373–389.
- Shang, J., Li, P., Li, L., Chen, Y., 2018. The relationship between population growth and capital allocation in urbanization. *Technol. Forecast. Soc. Change* 135, 249–256.
- Shen, J., 2013. Increasing internal migration in China from 1985 to 2005: institutional versus economic drivers. *Habitat Int.* 39 (1), 1–7.
- Sheng, Y., Song, L., 2013. Re-estimation of firms' total factor productivity in China's iron and steel industry. *China Econ. Rev.* 24 (3), 177–188.
- Song, Y., 2014. What should economists know about the current Chinese hukou system? *China Econ. Rev.* 29, 200–212.
- Song, Z., Storesletten, K., Zilibotti, F., 2011. Growing like China. *Am. Econ. Rev.* 101 (1), 196–233.
- Soo, K.T., 2014. Zipf, Gibrat and geography: evidence from China, India and Brazil. *Pap. Reg. Sci.* 93 (1), 159–181.
- Su, Y., Tesfazion, P., Zhao, Z., 2018. Where are the migrants from? Inter-vs. intra-provincial rural-urban migration in China. *China Econ. Rev.* 47, 142–155.
- Syverson, C., 2011. What determines productivity? *J. Econ. Liter.* 49 (2), 326–365.
- Vijverberg, W.P.M., Fu, F.-C., Vijverberg, C.-P.C., 2011. Public infrastructure as a determinant of productive performance in China. *J. Prod. Anal.* 36 (1), 91–111.
- Walheer, B., He, M., 2020. Technical efficiency and technology gap of the manufacturing industry in China: does firm ownership matter? *World Dev.* 127.
- Wang, X.-R., Hui, E.C.-M., Choguill, C., Jia, S.-H., 2015. The new urbanization policy in China: which way forward? *Habitat Int.* 47, 279–284.
- Wei, Y., Liu, X., 2006. Productivity spillovers from R&D, exports and FDI in China's manufacturing sector. *J. Int. Bus. Stud.* 37 (4), 544–557.
- Wu, D., Rao, P., 2017. Urbanization and income inequality in China: an empirical investigation at provincial level. *Soc. Indic. Res.* 131 (1), 189–214.
- Yang, R., He, C., 2014. The productivity puzzle of Chinese exporters: perspectives of local protection and spillover effects. *Pap. Reg. Sci.* 93 (2), 367–384.
- Ye, X., Xie, Y., 2012. Re-examination of Zipf's law and urban dynamic in China: a regional approach. *Ann. Reg. Sci.* 49 (1), 135–156.
- Yu, M., 2015. Processing trade, tariff reductions and firm productivity: evidence from Chinese firms. *Econ. J.* 125 (585), 943–988.
- Zhang, L., LeGates, R., Zhao, M., 2016. *Understanding China's Urbanization: the Great Demographic, Spatial, Economic, and Social Transformation*. Edward Elgar Publishing, Northampton, MA.
- Zhang, W., Bao, S., 2015. Created unequal: China's regional pay inequality and its relationship with mega-trend urbanization. *Appl. Geogr.* 61, 81–93.
- Zhang, Y., Shao, T., Dong, Q., 2018. Reassessing the Lewis turning point in China: evidence from 70,000 rural households. *China World Econ.* 26 (1), 4–17.
- Zhou, Y., 2009. Urbanization, urban-rural income gap and overall income inequality in China: an empirical test of the inverse-U hypothesis. *China Econ. Quart.* 8 (4), 1239–1256.