

Chapter 8

Spatial Economics, Urban Informatics, and Transport Accessibility



Ying Jin

Abstract One central pillar in the development of urban science which is key to the development of simulation of models of urban structure is spatial econometrics. In this chapter, we outline the way in which ideas pertaining to accessibility which we define conventionally, as in transport economics, as the relative nearness and size of locations to one another, can be embedded in a wider econometric framework. We are thus able to explore how GDP (gross domestic product) of different locations is influenced by different spatial investments. To illustrate this, we first outline the intellectual context, followed by a review of the most relevant econometric models. We examine the data required for such models and look at various quantifications in terms of elasticities of business productivity with respect to transport accessibility, using ordinary least squares, time-series fixed effects, and a range of dynamic panel-data models which narrow down the valid range of estimates. We then show how the model is applied to Guangdong province (with its connections to Hong Kong and Macau), which is one of the three major mega-city regions and a leading adopter of new technologies in China.

8.1 Introduction

In a nutshell, the contributions of spatial economics to urban informatics relate to the measurement, design, and interpretation of urban data that supports economic, social, and technological decisions regarding the locations, distributions, and layouts of urban activities, buildings, and infrastructure. In past decades, research at the frontier between spatial economics and urban informatics has largely been commissioned by governments, major banks, and businesses. Since new civic groups are playing an increasingly prominent role in investigating alternative options for spatial development (for a recent example in the UK, see the UK2070 Commission 2019), a full range of societal stakeholders have now been actively engaging with this area of interdisciplinary research. Students of urban informatics need an understanding

Y. Jin (✉)

Martin Centre for Architectural and Urban Studies, University of Cambridge, Cambridge, UK
e-mail: yj242@cam.ac.uk

© The Author(s) 2021
W. Shi et al. (eds.), *Urban Informatics*, The Urban Book Series,
https://doi.org/10.1007/978-981-15-8983-6_8

115

of spatial economics if they wish to influence the real decisions underpinning the planning, designing, funding, regulating, and maintaining of these spaces in cities and their hinterlands.

Spatial economics has a historic root as deep as all other main branches of modern economics. In particular, it can be traced back to the seminal works of von Thünen (1826). Since then, spatial economics has grown into a vast field of learning, which is sometimes referred to as the new economic geography (although this latter name does not have the consent of all geographers). Comprehensive handbooks on spatial economics have been compiled, for instance, see Duranton et al. (1987, 2018), and the higher-level overview by Redding and Rossi-Hansberg (2017). Somewhat paradoxically, this vastness of learning has often become a formidable barrier for those who work in urban informatics and wish to understand more about how spatial economies actually work.

This chapter adopts an approach that is complementary to the handbooks such as those referred to above—it aims to give students of urban informatics a feel for how spatial economics must tackle one of the critical issues that often confront them, that is, the measurement and interpretation of the contribution of inter-city transport accessibility improvements to the economy. According to Lakshmanan (2011), this is one of the most persistent spatial-economic issues in urban and regional transport studies. This approach which is an introduction by example is meant to encourage students of urban informatics to start with the quantitative skills that they may already have (e.g., simple ordinary least square or OLS regression models) and then engage with a cross section of advanced spatial-economics literature that is cogent to the topic.

The quantification of the economic contribution of transport accessibility improvements is particularly important for infrastructure investment. Significant progress has been made in recent years in spatial economics (see, for example, comprehensive reviews by Rosenthal and Strange 2004; Melo et al. 2009, 2013; Laird and Venables 2017). Nevertheless, in contrast to the considerable volume of research on the relationship between transport investment and productivity in the OECD countries, there are to date very few quantifications in this regard in emerging economies which are suitable for investment and loan decisions.

The complex, slow-evolving, and cumulative nature of the transport infrastructure investment makes the quantification of its impact one of the most challenging. Econometric modeling is the mainstay in current quantification of such impacts. Different types of regression and modeling methods have been developed over the years in this field, which started with OLS and time-series models that tested solely the effects of transport investment, and progressed with the introduction of a series of control variables, instrumental variables, and extended functional forms which are better able to deal with the heterogeneity and endogeneity issues of cumulative causation. This progression has led to more robust econometric models for such analysis.

In econometrics and only until recently, models have tended to be used in isolation rather than jointly. The quantification exercise tends to be carried out using the most advanced functional forms each time and this applies to the transport-related studies. However, using the alternative models jointly can offer valuable new insights into the quantification results. Bond et al. (2001) and Brühlhart and Mathys (2008) point

out that a comparison of the results of the alternative models with the theoretical, prior expectations may serve as an important bound test. Melo et al. (2013) have recently highlighted the empirical differences of the alternative model forms across different studies through a comprehensive meta-analysis on the effects of investment in transport infrastructure.

In this chapter, we show how a new approach to spatial-economic quantification of the transport effects can be developed using a series of regression models in the assessment of inter-city transport improvements. The econometric models are not only examined on their individual functional forms and estimation diagnostics, but also through a comparison of the outturn coefficient values with the prior theoretical expectations. Through this method, we aim to identify more precisely the transport effects on the real economy, while not substantially increasing the analytical work for practical studies designed, for example, for loan-project assessment.

We report an econometric analyses for Guangdong province, one of the three major mega-city regions and a leading adopter of new technologies in China. The analyses include Hong Kong and Macau as appropriate for the regional economic activities. Although we first started working on this quantification because of World Bank loan projects, we soon realized that Guangdong may be among the best case-study locations for such an investigation. Although the province has contributed to the highest provincial share of national GDP in China for more than two decades, its economic development is polarized, with a prosperous center and an underdeveloped periphery; its ways of doing business are being widely emulated by other provinces in China, thus are likely to represent what is to come in the rest of the country; and its land boundaries consist primarily of mountain chains which makes it straightforward to delineate a study-area boundary. This is in stark contrast to the amorphous limits of the other two main mega-city regions centered upon Beijing and Shanghai.

The chapter is organized accordingly in seven sections: Sect. 8.2 outlines the intellectual context, which is followed by Sect. 8.3 on the alternative econometric models. Section 8.4 presents the data. Section 8.5 presents the various quantifications in terms of elasticities of business productivity with respect to transport accessibility, using ordinary least squares, time-series fixed-effects and various dynamic panel-data models to narrow down the valid range of estimates. Section 8.6 discusses the wider implications of the findings and the extent of corroborations. Section 8.7 concludes with a short summary and considerations for future research directions.

8.2 Intellectual Context

Recent years have seen a growing body of research on the relationship between transport investment and productivity. The arguments are primarily built upon the spatial-economics literature, which gives due recognition to (1) consumers' and producers' love of variety in their use of products and services, (2) increasing returns to scale in production, and (3) the importance of transport costs in shaping the economic landscape. This has led to theoretical models that identify reasons why modern firms

tend to be more productive when they either concentrate in or have low cost links to large markets. Empirical studies have so far built up a substantial body of evidence which suggests that production and income are correlated with spatial proximity in the way suggested by the theories. Ciccone and Hall (1996), Rosenthal and Strange (2004), Redding and Venables (2004) and Melo et al. (2009, 2013) provide systematic surveys of the empirical evidence.

Inter-regional and city-scale theoretical models emerged about a decade after the initial trade models (see Fujita et al. 1999). Empirical studies followed. Rice et al. (2006) outlined an analytical framework within which interactions between the different aspects of regional inequality in per-employee productivity can be investigated econometrically using aggregate data. Kopp (2007) used a panel-data model to address the issue of endogeneity and identified contribution from transport investment to productivity, showing that doubling road stock in a country will lead to about 10% growth in total factor productivity in Western Europe. Combes et al. (2008) developed a general framework to investigate, respectively, the sources and mechanisms that lead to wage disparities across regional labor markets through sorting and self-selection. Graham and Kim (2008) investigated the relationship between spatial proximity and productivity using a large sample of financial accounting information from individual firms in the UK.

For emerging economies, Deichmann et al. (2005) distinguished between natural advantage, including infrastructure endowments, wage rates, and natural resource endowments, and production externalities that arise from the co-location of firms in the same or complementary industries, in their examination of the aggregate and sectoral geographic concentration of manufacturing industries for Indonesia. Lall et al. (2010) differentiated local and national infrastructure supply in India, and found that a city's proximity to international ports and highways connecting large domestic markets has the largest effect on its attractiveness for private investment.

In China, there has been a growing volume of literature that associates productivity benefits with agglomeration in Chinese cities and city regions (e.g., IBRD 2006, p. 145; Lu et al. 2007, p. 163). Using two nation-wide Censuses of Establishments of 1996 and 2001, Lu (2010) outlined the spatial distribution of economic activities across China and found through multivariate analysis that, during that period, the micro-economic explanations of agglomeration do not work well with publicly owned institutions, although they do work well with non-publicly owned institutions. Roberts and Goh (2012) showed that distance has a significant role in determining spatial productivity disparities in Chongqing municipality. Roberts et al. (2012) used counterfactual analysis based on a general equilibrium model to show that China's national expressway network has brought sizeable aggregate benefits to the Chinese economy, although its impact on regional disparities may be contingent upon factors such as migration.

These studies have shed an important light both on the statistical relationship between spatial proximity and productivity, and on a variety of complex issues of empirical modeling. Nevertheless, the studies have also shown that such statistical relationships may be highly context-specific.

At the heart of the difficulties of empirical measurements is the very nature of agglomeration as a process of circular, cumulative causation, which has become known since the work of Gunnar Myrdal: agglomeration propels endogenous growth—higher productivity leads to higher wages, which attracts employees of a higher caliber, which in turn draws in new investment, more productive technologies and so on; these lead to a new round of productivity growth. Conventionally, instrumental variables are used to overcome endogeneity issues in regressions; but by its very nature, agglomeration studies rarely have good instrumental variables for dealing with cumulative causation (Redding 2010).

8.3 Econometric Models

The underlying empirical model can thus be presented in a general form:

$$y_i = f(M_i, X_i) \quad (8.1)$$

where y_i is a measure of per-worker income or productivity in zone i , and $f(M_i, X_i)$ is a measure of transport accessibility of zone i , denoted by M_i , and a set of control variables X_i that reflect other zone-specific characteristics that may affect per-worker income or productivity. We define accessibility as measured by an aggregate economic mass (EM) that is accessible from a given location:

$$M_i = \sum_j \left(\frac{P_j}{g_{ij}^\alpha} \right), \text{ for all zones } j \text{ including } j = i \quad (8.2)$$

where

- i Location of the ‘home’ zone, for which the EM is computed as measuring accessibility from this location.
- j All relevant zones in the study area for market access, including $j = i$.
- g_{ij} Cost of travel between i and j , which may include time and monetary costs.
- P_j A measure of economic activity in zone j .
- α A parameter that controls the distance-decay effect; e.g., it was set to 1 by Graham and Kim (2008) and UK DfT (2006).

It goes without saying that the EM of location i increases if there is an increase in the level of economic activity in i , or there are decreases in the generalized costs of travel between i and j (e.g., through some transport intervention). By the same token, increased level of traffic congestion or dispersion of economic activity around a zone will reduce its EM.

We note that with this measure, the calculation of EM includes the contribution from the home zone (i.e., for $j = i$). This is the average travel cost for journeys within each zone, such as defined in transport studies.

A second popular functional form for the EM uses an exponential function to represent the effects of travel costs, in line with travel demand models:

$$M_i = \sum_j (P_j e^{-\theta g_{ij}}) \quad (8.3)$$

where P , i , j , and g_{ij} are defined as previously, and θ is a parameter for the exponential function that controls the distance-decay effect. θ may be calibrated through observed travel demand, and empirically, for inter-city travel, θ tends to reduce in value as the economic cost of travel increases. Rice et al. (2006) tested a variation of this exponential function as well as the Hansen function in their analyses of productivity effects.

8.3.1 Isotropic Versus Hierarchical Market Linkages for Economic Mass (EM) Computation

The two EM functions above may be used to cover market access to all destinations, or only a subset of the destinations which are relevant to the home zone in question. In the former case, the measurement is said to be isotropic in the sense that economic linkages between any cities, towns, and so on are considered in an identical way. This has been a common approach in the wider New Economic Geography literature.

In developing economies with limited technical specialization across locations, a hierarchical approach to covering the true market area (as originally defined by Christaller 1933) may be more realistic. This means that the cities and towns are central places of different orders in a regional hierarchy, and the linkages between different orders often tend to be stronger than those among centers of the same order. This is particularly true for learning new skills and transferring technology.

This is not a criticism of the existing EM measures in the literature, because they have largely been defined for regions of developed countries where the inter-city and inter-regional transport networks today are so well connected that they enable nearby central places at the same level of hierarchy to specialize and cross-trade to an extent that was not seen in Christaller's time. Extensive analyses of inter-city and inter-regional travel in Europe and Australia during the 1960s and 1970s indicated that the spatial patterns of travel in that era still exhibited features of the central place hierarchies (Bullock 1980). Our field work in Guangdong has also shown that regional hierarchies are important when firms consider their suppliers, markets, and linkages for technology transfer.

8.3.2 *Control Variables*

Other than transport accessibility that is represented by the EM, per-employee earnings in a given zone are influenced by a range of factors such as the number of hours worked, capital investment, level of skills, industry composition, and so on. If workers in a given zone work longer hours (e.g., through routine overtime working), they get higher nominal total pay. All being equal, better capital endowment enables higher output. Higher-skilled workers are paid more, and a high proportion of skilled workers in zonal employment would raise the level of average earnings. Similarly, employees working in some industries, such as finance, business services, IT, and research and development are often seen to be paid more than in other industries. These influences on per-worker earnings must be tested, and if significant, controlled for.

Here, we control the effect of working hours by modeling the average hourly earnings per employee as the dependent variable, that is, the annual average per-employee earnings are divided by the average number of working weeks and the average working hours per week. Similarly, we control for employee skills using as a proxy the proportions of those who achieved college, university, and post-graduate qualifications among the employees. In addition, we include control variables to represent industry composition and capital investment.

The regression analyses have been conducted using time-series data for 1999–2008, consisting of assembled economic data at the county or urban-district level and the economic mass (EM) data estimated by the study team using car travel times at the inter-county or urban-district level and a real GDP, as discussed above.

8.3.3 *Representing Spatial Spillover Effects*

The spatial econometrics literature suggests that there can be significant spillover effects between neighboring counties or urban districts. A formal way to deal with such spillover effects is to construct a spatial-weights matrix such that the lagged dependent and independent variables of all the near and distant neighbors are tested as explanatory variables, in addition to the independent variables of each county or urban district. Given that the EM variable has by definition already accounted for spatial proximity to each employment center, a weights matrix containing the influences of both near and distant neighbors would make the regression model over-complicated if used simultaneously with the dynamic panel-data models. We have therefore adopted here a simplified approach of only including as additional control variables the nearest neighbor of each county or urban district for such spillover effects. As a rule, including the nearest neighbor in the spatial spillover, analysis should take account of 70–80% of the spillover effects (LaSage 2012).

In line with our field-survey findings, in the main regression models, we have assumed a lag of up to three years for the EM, capital stock, and education level in

each county or urban district to take effect. This is implemented through producing composite independent variables for any year t through producing a moving average of the same variable for t , $t - 1$, and $t - 2$. For the spillover effects, the main regression models that use spatial-lag variables take variables of the nearest neighbor from one year earlier.

In terms of the regression models, we exploit what is known in theory about the nature of the OLS, fixed-effects (FE) panel-data models, and dynamic panel-data models, in terms of coefficient estimation bias when used with a dataset such as ours which is autoregressive in nature and has a relatively short time-span. On the one hand, the pooled OLS estimation is likely to bias the coefficient upwards, because of potential endogeneity of the EM variable: if there exist un-measured zonal features that impact on per-employee productivity that would attract the businesses and output and thus impact upon the EM variable over time. The corresponding FE model which is intended for use with a long time series will bias the coefficients downwards if the time series is fairly short, which is often the case with the panel-data series assembled for transport impact studies.

Since our aim is to identify causal effects that run from the economic mass to per-employee hourly earnings, we have to account for the fact that all explanatory variables may be potentially endogenous. In this context, the dynamic panel-data model based on a linearized generalized method of moments (GMM) technique (Arellano and Bond 1991; Arellano and Bover 1995; Blundell and Bond 1998) would in theory be more appropriate than the pooled OLS and FE methods above. The idea of the dynamic panel-data model is to use the past realizations of the model variables as internal instrument variables, based on the assumptions that (1) past levels of a variable may have an influence on its current change, but not the opposite, and (2) past changes of a variable may have an influence on its current level, but not the opposite. The method suits well our requirements because truly exogenous instrumental variables are hard to find in investigations of urban agglomeration effects.

In large samples and given some weak assumptions, GMM models can be free of some of the estimation bias inherent in the OLS and FE models. However, the two variants of the GMM methods, namely DIFF-GMM and SYS-GMM, have different properties when used with small samples. While the DIFF-GMM technique may be unreliable under small samples (Bond et al. 2001), the SYS-GMM technique is expected to yield considerable improvements in such situations (Blundell and Bond 1998). As a rule, data samples of transport impact analyses are unlikely to be very big ones, especially in developing economies. It is therefore necessary to test all the above models in order to clarify the robustness of the models. In turn, a comparison with the theoretical, prior expectations may also serve as a robustness test (Brühlhart and Mathys 2008).

8.4 Data

The bulk of the Guangdong economy consists of manufacturing and local commerce. Despite being one of the richest provinces in China, Guangdong had a per-capita GDP of US\$6500 in 2008, which in real terms is equivalent to the level of the US per-capita output in the 1930s. The primary and manufacturing industries, mostly low-tech and labor intensive, account for over 70% of the provincial output, and the high-end R&D and business services are a small, unknown fraction of the tertiary sector output. Empirical evidence for the developed economies may not therefore be transferrable to Guangdong or elsewhere in China.

Data from Guangdong are available at two different spatial scales: the province is first divided into 21 municipalities, and the municipalities are in turn subdivided into 67 counties or county-level cities and 21 urban districts of the municipalities (therefore, 88 county-level units in total). This is the most detailed spatial level currently reachable.

The earnings data are for fully employed staff and workers in urban establishments. This definition excludes farmers and other workers in rural areas. Compared with other employment and earning data available, these are the most suitable, as the employees in urban establishments are the most relevant to the agglomeration effects on productivity.

The data for calculating the economic mass (EM) consist of the level of economic activity and travel costs. For economic activity, we chose zonal GDP as the main variable, and retained the zonal size of employment as a sensitivity test. The travel costs and times are those of business travel, because these trips are most directly related to business linkages, technology transfer, commercial transactions, and negotiations. Because our regression models presuppose that the EM variable is correlated with the control variables and respective error terms (see choice of regression modeling strategy below), we have opted to using business travel time as the main travel-cost variable, while retaining travel cost and general travel cost as sensitivity tests.

Road construction data have been assembled over the period of 1999–2008 from a variety of provincial sources. Road links from the 2008 road network are then modified backward in time. For time-series analysis, a road network has been produced for each year of 1999–2008 within the GIS tool. The resulting travel distance, cost, and time matrices at the county or urban-district level for 1999–2008 are checked using our transport modeling experience. Up to 2008, the use of rail for business travel was minimal within the province, and thus, it is not necessary to include rail costs and times in the travel data.

In order to carry out comparisons of different EM measures, both the Hansen and exponential EM function forms are calculated for both the isotropic and hierarchical market areas. For the hierarchical market-area computation, we assume that (1) a county or urban district always interacts with itself, with constant business travel times through all years 1999–2008, and (2) a county or urban district interacts with all component counties or urban districts within its own municipality, as well as the provincial-level centers of Guangzhou, Shenzhen, Zhuhai, and Hong Kong. The only

exceptions are Guangzhou and Foshan, which are effectively coalesced into the same metropolitan area—the two urban areas are allowed to interact with each other.

For the control variables, we use the percentage of workers with college degree and above as a proxy for labor skills from the statistical yearbooks at the county or urban-district level. The statistical yearbooks report the levels of fixed asset investment per year. The Economic Census of 2004 also reports the total capital stock for production purposes per municipality. We estimate the county or urban-district level capital stock through these sources and build up the yearly capital stock for the entire time series that incorporates a standard capital stock depreciation rate of 5% per year. Investment in residential properties is excluded. We divide the zonal total capital stock by the total of full-time workers and staff in that zone to obtain the per-employee capital endowment. According to the National Labour Statistics Yearbook 2009, finance, information technology, and R&D industries are ranked as the top three high-earning sectors in Guangdong Province. We use the number of employees by region in these three sectors to control for the effects that can potentially arise from such differences in industrial composition. Specifically, we construct the index of sectoral composition following the definition of location quotient (LQ).

8.5 Model Test Results

The regression analyses have been conducted using time-series data for 1999–2008, consisting of assembled economic data at the county or urban-district level and the economic mass (EM) data estimated by the study team using inter-county or urban-district level business car travel times and level of economic activity, as discussed above.

To recap, on the left-hand side of the regression equations, the dependent variable is a vector of zonal data representing per-employee productivity levels: the average nominal hourly earnings at the county or urban-district level is used as the main test variable, with per-employee average GDP as a sensitivity test variable. On the right-hand side of the equations, the list of independent zonal variables at the county or urban-district level includes the EM representing transport accessibility, a range of variables representing zonal capital investment, skills, and industrial composition, and spatial-lag variables from the nearest neighbor zones. The independent variables are tested as appropriate for each specific functional form. In addition, the GMM models use time-lagged independent variables as instruments as specified.

Through the regressions, we have tested different measures of productivity (i.e., hourly earnings and per-employee GDP), different EM terms (i.e., using distance, travel time, and generalized travel cost for isotropic and hierarchical market areas), and different measurements of capital endowment and labor skills. All regression models have returned consistent results, among which we have found that the equations using hourly nominal earnings, hierarchical EM using time to measure travel cost, accumulated and depreciated capital stock, and parentage of college and above graduates to measure labor skills, have an overall best fit. This is in line with our

field-survey findings. Both the Hansen-type and exponential functional forms of the EM variable are tested. Owing to the limit of space, we report the core estimation results in Table 8.1. The other tests are available upon request.

In Table 8.1, Model (1) is a pooled OLS model which returns an EM coefficient of 0.24, with the EM and the control variables (for capital stock and education level) being statistically significant and a relatively high R -squared = 0.69. However, we have good theoretical reasons to suspect that the coefficients are biased upwards and this model result embodies an absolute upper bound of the productivity elasticities.

By contrast, with Model (2) which is the time-series fixed-effect (FE) model, the EM coefficient drops to 0.115 when the period dummies (representing the period-specific effects) are included for the Hansen EM formulation. The EM coefficient further drops to 0.052 in Model (3) when the exponential EM variable is used. Our theoretical expectations are that these are biased downwards for respective EM functional forms, and thus could be considered as a lower bound to the EM coefficient.

This is reflected in the DIFF-GMM model in Column (4). The EM coefficient output from this model is at 0.151, between the upper and lower bounds as we expect, although the coefficients are not statistically significant. The SYS-GMM model (5) gives a similar EM coefficient at 0.141: Both the EM and the capital stock coefficients are now significant; note that this model includes additional explanatory variables that represent the spillover effects from the nearest neighbor zones in terms of capital stock endowment and education level of the employees.

The GMM-SYS Model (6) is a standard test to assess the robustness of the model by reducing the number of instrument variables (from 115 to 69), which has raised somewhat the significance of the education-level variable but has not altered the nature of the model results nor the magnitude of the coefficients. The standard tests of the GMM models suggest that there are no apparent misspecification problems. The Hansen test for over identification restrictions, and the difference Hansen tests for the validity of the GMM and IV instruments, indicate that the instruments are valid. The Arellano-Bond AR2 test suggests that no second-order residual auto-correlation is present.

Model (7) presents the SYS-GMM results for the exponential functional form of EM, which returns an EM coefficient of 0.087. The estimation diagnostics are similarly good. A test to reduce the number of instruments (from 103 to 75) has also been carried out as Model (8) and has confirmed that the instruments are valid.

Given that the exponential form of the EM variable embodies the distance-decay parameters that are consistent with the travel-behavior model calibrated in China, it would seem sensible to consider Model (7) as the preferred estimate of the productivity elasticity (i.e., 0.087 with a standard error of 0.03 and robust t statistic 2.89) with respect to transport accessibility.

In summary, the econometric results show that transport accessibility as represented by the EM is statistically significant after controlling for control-variable endogeneity and spatial spillover effects. Our preferred estimate comes from Model (7) in Table 8.1, which adopts a SYS-GMM formulation and exponential EM formula and returns a productivity elasticity of 0.087, with a robust standard error of 0.030. The

Table 8.1 Time-series model results

Dependent variable: ln(hourly earnings [t])	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pooled OLS (Hansen)	Fixed effects (Hansen)	Fixed effects (Exp)	DIF-GMM (Hansen)	SYS-GMM (Hansen)	SYS-GMM (Hansen)	SYS-GMM (Exp)	SYS-GMM (Exp)
ln (economic mass)	0.240*** [0.016]	0.115*** [0.043]	0.052 [0.017]	0.151 [0.152]	0.141*** [0.067]	0.162*** [0.077]	0.087*** [0.030]	0.101*** [0.037]
ln (per-worker capital stock)	0.155*** [0.013]	0.007 [0.011]	0.007 [0.011]	0.026 [0.035]	0.133*** [0.032]	0.151*** [0.032]	0.127*** [0.035]	0.129*** [0.035]
ln (% of college graduates)	0.193*** [0.011]	-0.021 [0.016]	-0.026 [0.016]	-0.046 [0.048]	0.031 [0.049]	0.062 [0.040]	0.119*** [0.039]	0.110*** [0.037]
Period dummies ln (per-worker capital stock)—nearest county	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
					0.008		0.017	
					[0.033]		[0.028]	
ln (% of college graduates)—nearest county					0.071		0.006	
					[0.041]		[0.041]	
Constant	-0.569*** [0.127]	1.975 [0.181]			0.170 [0.478]	0.104 [0.364]	0.387 [0.423]	0.491 0.407
Observations	704	704		616	704	704	616	616
R-squared	0.69	0.57						
Number of instruments				64	115	69	103	75
AR(2) test				0.430	0.170	0.790	0.370	0.411
Hansen test of over identification restrictions				0.274	0.980	0.352	0.818	0.245

(continued)

Table 8.1 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: ln(hourly earnings [t])	Pooled OLS (Hansen)	Fixed effects (Hansen)	Fixed effects (Exp)	DIF-GMM (Hansen)	SYS-GMM (Hansen)	SYS-GMM (Hansen)	SYS-GMM (Exp)	SYS-GMM (Exp)
Difference in Hansen tests of exogeneity of instrument subsets:								
For GMM instruments (<i>p</i> -value)					0.713	0.367	0.516	0.249
For IV instruments (<i>p</i> -value)				0.115	0.909	0.365	0.744	0.330

Notes (1) For the GMM models, the variables for the economic mass and average percentage of college graduates are treated as GMM instrument variables while the year dummies and average per-worker capital stock are treated as IV variables. For GMM models they are based on two-step Windmeijer finite-sample adjustment. For the FE and GMM models, the fixed effects are modeled but not reported here. (2) Robust standard errors are in square parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

model diagnostics suggest that all the SYS-GMM model results are robust. Furthermore, the GMM model results fit our prior expectations regarding the upper bounds established by the pooled OLS models and the lower bounds by the time-series fixed-effects models.

8.6 Discussions

An extensive series of regression model tests show a consistent pattern for a statistically robust relationship between transport accessibility and business productivity. In particular:

- (a) As expected, the pooled OLS regressions produced high elasticity estimates while the time-series fixed-effects (FE) regressions produced low estimates. The dynamic panel-data models using the linearized generalized method of moments (GMM) tend to return intermediate elasticity values.
- (b) Our understanding of the regression models and the development process in Guangdong, China gives grounds to prefer the GMM model estimates (particularly the SYS variant which corrects for relatively small samples). This is because the SYS-GMM models are capable of making a sound use of the short panel dataset.
- (c) Our preferred estimate of the productivity elasticity with respect to transport accessibility is 0.087 (with robust standard error 0.03 and t statistic 2.89). This comes from the SYS-GMM model which uses the exponential formula in measuring transport accessibility. This positive relationship remains robust after controlling for a range of control variables, endogeneity, and nearest neighbor spillover effects. The robustness of this estimate is confirmed through both the regression diagnostics and a comparison with results from the alternative models.

This central productivity elasticity estimate of 0.087 implies that a 10% improvement in transport accessibility would give rise to an increase of per-worker productivity of 0.83% (i.e., $(1 + 10\%)^{0.087} - 1 = 0.0083$), and a doubling in transport accessibility would imply an increase of per-worker productivity of 6.2% (i.e., $(1 + 100\%)^{0.087} - 1 = 0.0622$). This is well within the consensus range of productivity elasticities from a comprehensive review of such evidence in predominantly developed economies that “doubling city size seems to increase productivity by an amount that ranges from... roughly 5–8%” (Rosenthal and Strange 2004), and is comparable with the elasticity range from the latest meta-analysis of productivity elasticities published by Melo et al. (2013), who suggests the central elasticity value is around 0.05.

In assessing the estimates we may also compare them with our prior expectations: transport accessibility and agglomeration are thought to play an important role in knowledge spillover and technological improvements in China (IBRD 2006).

The empirical findings in this chapter are to an extent supported by emerging estimates for China, although our estimates are considerably lower. For instance, Au and Henderson (2006), using data of 1990 and 1997 from 205 Chinese cities, suggested that there are significant urban agglomeration benefits: for example, moving from a city of 635,000 to one of 1.27 m increases the real output per worker by 14%, after controlling a range of other influences. More recently, Zhang's analysis (Zhang 2008) using the 1993–2004 data put the mean elasticity value at 0.106 in China after controlling for spatial spillover effects.

Our field studies in Guangdong (see EASCS 2014a, b) have also started to investigate the actual mechanisms through which businesses benefit from transport accessibility improvements in terms of employee productivity. It indicates that the agglomeration benefits accrued by transport improvements are well understood by the businesses and individuals, and the extent to which they exploit such benefits is comparable with those observed in developed economies. This provides a degree of corroboration at the micro-level. Of course, further work is still needed to quantify such effects at the level of individual businesses and employees.

8.7 Conclusions

This chapter aims to introduce the theories and methods of spatial economics through one specific example of quantifying the economic contribution of transport accessibility improvements, which may well be a research question that often confronts the students of urban informatics. The chapter starts with simple OLS regression models that are commonly used in urban-informatics research and then extends the models step by step using a cross section of spatial analytical and economic theories. The resulting models reach the current frontier of the field, and they serve to fill a gap in current literature. In developing the models, there is also an ethos of developing a methodology which is theoretically rigorous but can be made operational with a level of data availability that is generally achievable in the emerging economies. In the low- and middle-income developing countries such as China, such empirical evidence for spatial-economic effects of transport is currently poor and the practical needs for them are urgent, for example for assessing major investment initiatives.

Of course, the current econometric models may not yet fully control for other differences between zones, for example, the spatial self-selection and sorting of employees within and among the counties and urban districts. Clearly, spatial proximity resulting from transport improvements plays an enabling role in spatial self-selection and sorting. Nevertheless, it is yet difficult to discern the precise contribution of transport improvements to such mechanisms within the available data sources.

Also, it is not for econometric studies alone to establish causality between transport accessibility and productivity where there is a process of significant cumulative causation; that task should be supported by an in-depth understanding of the actual mechanisms at work, for example through field studies as discussed above.

Additional future work may further improve the robustness of the findings presented here; the list below would serve to indicate the scope of further research on this topic:

First, it may be possible to expand the time series under consideration both in years covered and the range of explanatory variables, which is likely to make the model more robust and improve the precision of the coefficient estimates.

Secondly, similar econometric models can be estimated for the economically less-developed regions in China (e.g., inland regions such as Sichuan), as well as other affluent regions along the Eastern Coast (e.g., the Yangtze River Delta centered upon Shanghai and the Bohai Bay Metropolitan Area centered upon Beijing). This would clarify whether there are significant differences among regions of different levels of development.

Thirdly, if and when the disaggregate Economic Census data become available from the Chinese statistics bureaux, enterprise-level production functions (e.g., of the Translog type) can be estimated, which would provide more precise estimates of the agglomeration effects including possible spatial sorting effects. The Economic Census data were collected by enterprise, although so far they have not been released for use in research in China.

Fourthly, micro-level case studies of firms and institutions will help us understand how firms actually respond to transport improvements, and through what mechanisms they gain from agglomeration effects or otherwise.

The cumulative evidence through the above could eventually provide a fuller understanding of economic development in terms of dynamic general equilibrium processes, for example as suggested by Au and Henderson (2006) and Lakshmanan (2011). Such understanding would in turn enable us to better plan transport projects, particularly to promote shared prosperity and poverty alleviation in under-developed regions.

Acknowledgements and Disclaimer The example reported in this chapter is primarily based on a working paper sponsored by the Technical Assistance of the World Bank project “Regional economic impact analysis of high-speed rail in China”. The funding support from the World Bank is gratefully acknowledged. The author also acknowledges additional funding support from the UK EPSRC Centre for Smart Infrastructure and Construction Phase 2 (EP/N021614/1) on the theoretical extensions and testing of the quantification methods. The contributions of World Bank colleagues during the project, particularly John Scales, Gerald Ollivier, Richard G. Bullock, and Wanli Fang, are gratefully acknowledged. The views and any remaining errors are the responsibility of the author alone.

References

- Arellano M, Bover O (1995) Another look at the instrumental variable estimation of error-components models. *J Economet* 68:29–51
- Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev Econ Stud* 58:277–297

- Au CC, Henderson JV (2006) Are Chinese cities too small? *Rev Econ Stud* 73:549–576. <https://doi.org/10.1111/j.1467-937X.2006.00387.x>
- Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *J Economet* 87:115–143
- Bond S, Hoeffler A, Temple J (2001) GMM estimation of empirical growth models. CEPR Discussion Paper 3048
- Brühlhart M, Mathys NA (2008) Sectoral agglomeration economies in a panel of European regions. *Reg Sci Urban Econ* 38:348–362
- Bullock RG (1980) Regional travel patterns: review and empirical model estimation. Mimeo
- Christaller (1933) Central places in Southern Germany. English edition 1966. Prentice-Hall, Englewood Cliffs, NJ
- Ciccone A, Hall R (1996) Productivity and the density of economic activity. *Am Econ Rev* 86:54–70
- Combes P, Duranton G, Gobillon L (2008) Spatial wage disparities: sorting matters! *J Urban Econ* 63:723–742
- Deichmann U, Kaiser K, Lall SV, Shalizi Z (2005) Agglomeration, transport, and regional development in Indonesia. World Bank Policy Research Working Paper No. 3477. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=647682
- Duranton G et al (1987–2018) Handbook of urban and regional economics, 1–5. <https://www.sciencedirect.com.ezp.lib.cam.ac.uk/journal/handbook-of-regional-and-urban-economics/issues>
- EASCS (2014a) Regional economic impact analysis of high speed rail in China: main report. Report No.: ACS9734. The World Bank, Washington DC
- EASCS (2014b) Regional economic impact analysis of high speed rail in China: background Working Paper 2. Report No: ACS9734. The World Bank, Washington DC
- Fujita M, Krugman P, Venables AJ (1999) The spatial economy: cities, regions and international trade. MIT Press, Cambridge, MA
- Graham DJ, Kim HY (2008) An empirical analytical framework for agglomeration economies. *Ann Reg Sci* 42:267–289
- IBRD (2006) Governance, investment climate, and harmonious society: competitiveness enhancements for 120 cities in China. The World Bank and IBRD, Washington DC
- Kopp A (2007) Aggregate productivity effects of road investment: a reassessment for Western Europe. In: Karlsson C, Anderson WP, Johansson B, Kobayashi K (eds) The management and measurement of infrastructure. Edward Elgar, Cheltenham
- Laird JJ, Venables AJ (2017) Transport investment and economic performance: a framework for project appraisal. *Transp Policy* 56:1–11
- Lakshmanan TR (2011) The broader economic consequences of transport infrastructure investments. *J Transp Geogr* 19:1–12
- Lall SV, Gun WH, Deichmann U (2010) Infrastructure and city competitiveness in India. UNU-WIDER Working Paper W, World Institute for Development Economic Research
- LaSage J (2012) Understanding spill-overs. ERSA 2012 Congress keynote speech, Bratislava, August
- Lu D, Yao S, Liu H, Gao X, Li G, Duan J et al (2007) China regional development report: the urbanisation process and spatial expansion. The Commercial Press, Beijing ([In Chinese])
- Lu J (2010). Agglomeration of economic activities in China: evidence from Establishment Censuses. *Regional Stud* 44:281–297
- Melo PC, Graham DJ, Noland RB (2009) A meta-analysis of estimates of urban agglomeration economies. *Reg Sci Urban Econ* 39:332–342
- Melo P, Graham DJ, Brage-Ardao R (2013) The productivity of transport infrastructure investment: a meta-analysis of empirical evidence. *Reg Sci Urban Econ* 43:695–706
- Redding SJ (2010) The empirics of new economic geography. *J Reg Sci* 50:297–311
- Redding SJ, Venables A (2004) Economic geography and international inequality. *J Int Econ* 62:53–82
- Redding SJ, Rossi-Hansberg E (2017) Quantitative spatial economics. <https://www.princeton.edu/~erossi/QSE.pdf>

- Rice PG, Venables AJ, Patacchini E (2006) Spatial determinants of productivity: analysis for the regions of Great Britain. *Reg Sci Urban Econ* 36:727–752
- Roberts M, Goh C (2012) Density, distance and division: the case of Chongqing municipality, China. *Cambridge J Reg Econ Soc* 4:175–188
- Roberts M, Deichmann U, Fingleton B, Shi T (2012) Evaluating China's road to prosperity: a new economic geography approach. *Reg Sci Urban Econ* 42:580–594
- Rosenthal S, Strange WC (2004) Evidence on the nature and sources of agglomeration. *Rev Econ Stat* 85:377–393
- UK DfT (2006) Transport, wider economic benefits, and impacts on GDP. UK Department for Transport, London. <https://www.dft.gov.uk/pgr/economics/rdg/webia/transportwidereconomicbenefi3137>
- UK2070 Commission (2019) Fairer and stronger: first report of the UK2070 Commission. <https://uk2070.org.uk/publications/>
- von Thünen JH (1826) *Der isolierte staat in beziehung auf landwirtschaft und nationalökonomie*. Perthes, Hamburg. English translation: Pergamon Press, Oxford
- Zhang X (2008) Transport infrastructure, spatial spillover and economic growth: evidence from China. *Front Econ China* 3:585–597



Ying Jin is Director of the Martin Centre for Architectural and Urban Studies at University of Cambridge and a Fellow of Robinson College, Cambridge. He was elected a Fellow of the Academy of Urbanism, London in 2011, and his research interests are investigating the causes of urban change through data science and interdisciplinary models.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

