Regional Concentration of Industry in China:

decentralised choices or a central plan?

Stephen Sheppard[1] and Dan Zhao[2]

December 7, 2016

[1]Williams College Department of Economics, 24 Hopkins Hall Drive, Williamstown, MA 01267
[2]Williams College Department of Economics, 24 Hopkins Hall Drive, Williamstown, MA 01267

Abstract

The economic growth and development of China during the past 35 years has been associated with profound impacts on the well-being of the Chinese people, on patterns of global trade and prices of manufactured goods, and on industrial location within China itself. Many would view China's government and its policies as the dominant, perhaps exclusive, force in determining location and concentration of Chinese industry. This raises the question: can a theoretical approach based on decentralised optimization and location choice provide insights concerning the ongoing changes in industrial concentration in China? We address this question, putting forward a simple model and testing it using Chinese data.

1 Introduction

The economic growth and development of China during the past 35 years has proceeded at a rapid pace. With it, by some measures, there have been significant increases in economic inequality between households and between regions. Both within its borders and beyond, as China's development proceeded, there have been large changes in the sources of global trade, the prices of manufactured goods, and in patterns of industrial location within China itself. Though service industries (what China refers to sometimes as 'tertiary industries') have grown as a component of China's growth and productive capacity, secondary industries (what China refers to as manufacturing industries) have played the significant, if not central, role in China's regional and national economic development through the bulk of its more recent history Jeon (2006). This is particularly true given China's focus on heavy industrialization during its development and even its current activities in mining and manufacturing.

With its Special Economic Zones, export bases, industrial parks, territorial development policies, continued improvements to its public and transport infrastructure, as well as other geographic experimentation relying on varying degrees of liberalization and linkages, China's policies have potentially significant impacts on the spatial structure of its economy. Its policies appear to be designed to try and channel development towards different regions and to affect the spatial concentration of production. These large-scale attempts to spatially restructure its economy make it an interesting subject to study.

With China's state-centric reputation and its centralised political and national policy-making apparatus, it is unclear whether standard economic models of industrial location serve as appropriate structures for studying the concentration of industry in China. Most analyses of China view government policy as the dominant, perhaps exclusive, force in economic activity, including determination of patterns of location and concentration of Chinese industry. This raises a key question: can a theoretical perspective driven by decentralised individual choice and optimization, in the context of uneven utility differentials between locations, provide explanation and understanding of the changes in China's industrial concentration and organization?

This paper investigates some of the factors that contribute to this changing industrial geography and its organization in China's secondary industries. We tackle this question by developing and expanding upon a simple, analytically tractable core-periphery model based on Pfluger (2004), with relatively simple forces driving patterns of agglomeration and dispersion, primarily through the channel of transport costs of produced goods across regions. From this model, we obtain specific predictions concerning the range and behavior of industrial location quotients and how this range evolves with other changes in the economy, such as per capita income and transportation costs.

Using data from 41 industries across 27 provinces over the past 13 years, we analyze the changes that have been observed in industrial concentration and examine the ability of these relatively simple models to provide an understanding of these changes. Through our empirical analysis, we show that the data are consistent with the predictions of this model. Mindful of the possible endogeneity associated with our different measures of transportation costs, we employ IV techniques using measures of physical geography, geographic topology of provincial terrain, and similar variables as instruments. We show that our results are robust to this correction for potential bias.

Here, we aim not to explain every single contingency or to model every economic mechanism, but instead to specifically elucidate a significant, specific set of phenomena. Though our analyses cannot definitively say that government policy has no impact on industrial location, it reveals the power of relatively simple theoretical insights in providing an understanding of a complex economy. Though it may be a partial, albeit strongly supported, explanation, natural economic forces alone, from our model of spatial optimization, are largely consistent and sufficient in explaining what we observe in China's changing industrial geography and agglomerations.

2 Background

Particular attention has been paid to China due to its accelerated economic growth, especially due to its reputation as a heavily state-managed economy, its nature as a centralised political state, and its spatial experimentation with regional economic growth via industrial parks and free economic zones throughout the years, as noted by Meng (2003). As a result, the political economy of China's industrial growth is an interesting aspect to consider. In the pre-reform era, China inherited a spatial structure characterised by an extremely dispersed industrial configuration under heavy state management, a slow pace of urbanization, sparse and scattered cities, low levels of regional specialization, and low factor mobility (Chan, Henderson & Tsui (2008)).

Although China's liberalizations have relaxed (and in many cases, replaced), state planning and management, China's political and state institutions, though slowly changing themselves, are still intact, preserving the key structures that continue to shape the spatial configuration and structure of resource allocation (Chan et al. (2008)). Though there is debate on whether the non-state sector or the domestic private-sector better accounts for China's free market activity (Haggard & Huang (2008)), both sectors together likely encompass and capture the full scope of economic dynamics and market incentives–especially given the blend of state and private characteristics which constitute China's institutions and its economic system. For example, Township and Village Enterprises (TVEs), though collectively owned and not purely private by nature, possessed features characterised by both central management and a free-market reign, such as the dual-track pricing systems introduced in the transition process. The overall trend remains: in the 21st century, China has largely adopted, in general, a more relaxed attitude towards day-to-day production decisions and resource allocations since its pre-reform era.

With the early 1980s, the state liberalised the domestic private sector while orienting state-owned banks toward supporting the domestic private sector (Haggard & Huang (2008)). Though the government has somewhat shifted course between 1988 and the early 1990s on this stance (before reverting back to liberalization in the late 1990s and 2000s), it continued to aggressively court Foreign Direct Investment throughout this period to attract foreign private firms. Over time, the range of industries subject to government management has declined, especially for most manufacturing industries, culminating into drastic change (Brandt, Rawski & Sutton (2008)). Initially, direct foreign investment and exports from foreign-linked firms clustered along the eastern seaboard, leading to coastal China's differential development in its emerging clusters–which, ultimately, also led to the differential development of inland provinces (Brandt et al. (2008)).

3 Previous Studies

Much of the spatial literature in economics has been devoted to explaining how space influences producer location choices and incentives. Specific phenomena such as agglomeration or dispersion of industrial activity can occur, depending on economic conditions. As a result of the concentration of certain production activities, these clusters may exert economic forces, such as encouraging further agglomeration, on their local or regional economies. Generally, the spatial literature focuses on one of three themes: either to document and study the behavior of existing clusters, to analyze the impacts, if any, from these concentrations on their local economy, or to study the factors which cause their formation or dissolution. Our paper predominantly address the first and third themes which, naturally, also touches upon the second. Specifically, we do so with China, a country that has undergone large, fundamental economic changes in a relatively short span of time.

Clusters and concentrations have become increasingly prominent in the spatial distribution of economic activity in economies world wide. As these clusters become more prevalent, they play significant roles in influencing the range of economic activities available, input productivity, and even international trade flows. Fundamentally, the organization of economic activity across space requires the transport of inputs and outputs across different locations. Final goods and services are transported to consumers, intermediate goods and inputs are delivered to different producers in the supply chain, and consumers migrate to areas where goods and services are offered. Naturally, since the economy is spatially heterogeneous, transportation infrastructure and the cost of transport become important factors to firms in their decisions on where to produce.

China is no exception. Over the past two decades, China has largely increased its investment in transportation infrastructure. China's National Trunk Highway System has involved the construction of around 21,747 miles of highways over a period of 15 years at an estimated construction cost of around $120 billion in current U.S. dollars (Faber (2014)). There is rhyme and reason to the investment. The economic benefits from transport infrastructure development are widely documented. For instance, in China, access and proximity to transportation networks and infrastructure have been shown to have a moderate positive effect on *per capita* GDP levels across industry sectors (Banerjee, Duflo & Qian (2012)). However, the effects of such investment on transport costs themselves and on industrial organization are not well-documented and remain somewhat unclear.

In an analysis over a similar period, using census data at the firm level in 1995 and 2004, Long & Zhang (2012) use various multi-dimensional measures on industrial clustering to show that China's rapid industrialization has been characterised by increasing spatial concentration within industries and within regions. They show that the growth in the number of firms is greater in clusters than in un-clustered regions, concluding that China's industrialization is a largely cluster-based phenomenon. Similar studies on a lesser scale have shown that the spatial concentration of manufacturing firms has increased from 2002 to 2008, using data on specific firm addresses and measuring concentration with the Duranton and Overman index (Brakman, Garretsen & Zhao (2014)).

Though China is primarily known for its manufacturing, Yang & Yeh (2013) also reveal a spatial concentration of producer services in some urban areas over the years, although to a much smaller extent. Most importantly, they note that most other studies looking at similar topics focus mainly on individual cities, leaving spatial behavior of production activities at, or across, the regional and national levels unexamined. Other kinds of concentrations have been noted as well; in the past twenty to thirty years, the initial rise of cultural clusters (i.e. clusters of cultural organizations, museums, etc.) in cities has, through local spillovers over time, further spurred the creation of new cultural clusters Ko & Mok (2014).

However, in light of China's complicated developmental history and complex economy, economists still struggle to provide an appropriate context for explaining and understanding the ongoing transformation of Chinese industrial locations. Fan & Scott (2003) note the emerging patterns of industrial concentration and argue that the concentration itself is driving economic development through agglomeration economies and external economies of scale. They draw particular attention to effective government policy-making as a mechanism for encouraging the emergence of industrial concentrations that have, in turn, helped drive economic growth. Wen (2004) notes that from 1953 to 1978, industrial location in China was determined centrally, out of concern for outcomes in potential military conflicts. Looking at data from 1980 to 1995, he notes a tendency towards increased concentration of industries and credits the government's development policies as the source of these changes.

Brakman et al. (2014) identify the potential for policy to have any influence whatsoever on location as an open question. They note that compared to other countries, the historical levels of concentration of Chinese industry seem low; the possibility that China's Hukou system, one that regulates where persons can live (or at least where they can have access to certain public services), may serve to limit industrial concentration, and that a liberalization of this government policy may permit increased concentration. Using micro-geographic data they discover several results in support of showing that localization tendencies are stronger for private and foreign-owned firms than for state-owned firms. However, as noted above, the distinction between private and public is largely a formal one given the mixed, overlapping nature of China's production economy and its firms/enterprises. More recently, Hsieh & Song (2015) note that the transformation of the Chinese economy and its productivity growth have largely taken place through the government's reforms of privatizing smaller state-owned firms while retaining state ownership or partially privatizing several large producers. Along with China's gradual liberalization approach in the reform era, these policies seem to reinforce the blended nature of ownership and operation. While not focused directly on industrial location, their finding is particularly interesting in light of the finding in Brakman et al. (2014) that it is exactly these smaller firms that are driving the changes in localization of Chinese industry.

The general consensus, regardless of the approach taken or metric used, has pointed to an increasing geographic concentration of industry that has coincided with China's recent and rapid economic rise. A historical comparison of the concentration levels between 1980, 1985 and 1995 suggests that manufacturing industries, along with others, have become more geographically concentrated in the past two decades or so. This is documented in Wen (2004). These industrial clusters have had a variety of effects on their local and regional economies. From human capital externalities generated by pooled labour markets (Liu (2014)) to widening regional gaps in economic development (Van Huffel, Luo & Catin (2005)), these effects, though varied in nature, have significant effects on regional economic growth, regional convergence (or divergence), the spatial distribution of growth across different regions of the country, as well as the economic geography of industrial production, per-capita income, pollution, and the overall quality of life in China's cities (Zheng, Sun, Qi & Kahn (2014)).

From the perspective of economic theory, the main determinant of the spatial structures underlying economic activity lies in the cost of transporting inputs and outputs between locations, be it maintaining a market, extending a market, or creating a market. Testing this perspective is difficult because transport costs depend on both traffic levels as well as provision of transport infrastructure and so are endogenously determined. Two approaches have been widely used to address the endogeneity problems that arise in estimating relationships between transport costs and economic outcomes: historical route instruments and planned route instruments (Redding & Turner (2014)). The first relies upon historical transportation networks as instruments, whose validity hinges upon the fact that, given control variables, factors that do not directly affect economic activity determine the configuration of these historical networks. As an example, Duranton & Turner (2012) use maps of historical transportation networks and historic routes of major expeditions used to explore the U.S. as sources of quasi-random variation in the U.S. interstate highway network when trying to predict MSA-level economic outcomes.

The second method uses previously planned routes as instruments, which are exogenously valid only if the original purpose behind the routes is uncorrelated with its modern transport network counterpart; for instance, Baum-Snow (2007) uses a 1947 plan for the interstate highway network as a source of quasi-random variation in the way the actual network was developed. Since the network plan was designed for military purposes, the validity of this instrument hinges on the extent to which military purposes are irrelevant to the needs of commuters and producers in peacetime. Similarly, in studying the impact of the Gold Quadrilateral project (GQ), a previous upgrading project of India's central highway network which now connects many manufacturing hubs that only came into existence relatively recently, Ghani, Goswami & Kerr (2016) employ an IV approach that uses straight line distances at most 10 kilometers away from the actual GQ highway network route. This contention rests on the assumption that the GQ planners and policy makers did not know about the growth potential of the regions that the GQ would connect and, therefore, did not attempt to aid their growth through the highway project. Put differently, this rests on the claim that the cities that are now the nodes of the GQ network were not established as a result of this transport network. As such, they resort to a what-if analysis by focusing on what the layout would have been if this network was instead established based upon minimal distances between nodes.

Regardless of the approach, the main obstacle in determining the causal effects of transport costs and infrastructure on the spatial organization of economic activity is that infrastructure is not randomly assigned to locations, but rather through the same unobserved location characteristics that affect economic activity. Given the nature of these variables, using some form of transport network or routes as instruments does not completely shield against criticism or endogeneity. Even so, Redding & Turner (2014) note, "while these approaches remain open to criticism and refinement, they are about as good as can be hoped for in an environment where experiments seem implausible."

In this paper, which builds on and extends Zhao (2014), we directly account for the importance of space in producers' location choices by acknowledging transportation costs in our theoretical and empirical frameworks. Our research makes two contributions. First, we use general infrastructure measures as well as more precise industry-specific measures of transport cost burdens in our analysis. Second, we not only employ a novel identification strategy using the topography and geographic features of provincial terrain, along with additional variables, but we do so with extensive data spanning 13 years of China's economic development across all provinces and industries.

4 Model

With the development of the Chinese economy, extensive resources have been devoted to improving transportation infrastructure and improving the technology with which transportation of goods and of people is achieved. As these improvements have taken place, the skill levels of workers have also improved and as a result the consumption demands of residents have increased and evolved. The patterns of trade – both the internal distribution of goods and services production as well as external trade – have changed. Given the magnitude of these changes, it is not surprising that the patterns of industrial location and concentration have also changed dramatically. In this section we present a relatively simple model that can help to organise our thinking about these changes, and provide some guidance concerning what we would **expect** to observe as the process of economic development moves forward.

Our goal here is to adapt a simple spatial model based on decentralised, individually rational location choices for entrepreneurs and, to derive from this model, the expected relationship between the location quotients, which we will use as an index of industrial concentration empirically, and other variables with particular emphasis on transport costs. We then bring these predictions to the data as a way of assessing the ability of such a model to help us to understand the nature of industrial concentration in China, and how it is affected by transport costs.

Our starting point is the model of regional development and location choice introduced in Pfluger (2004). There are two regions, I and J. There are two classes of household, and each is endowed with one unit of a factor of production that can be supplied to earn income. In the two regions combined there are $L_I + L_J$ labouring households who supply unskilled labour to the economy. There are total of $K_I + K_J$ entrepreneurial households that own and provide capital. We can think of entrepreneurial households as supplying both human capital or physical capital. The labouring households cannot move between the regions, but the entrepreneurial households are mobile and employ their capital in the region where they live.

As noted in Xin (1999), the period from 1980 to 2000 brought significant increases in mobility for educated persons, so that assuming free mobility for owners of human capital along with other types of capital seems consistent with at least some perspectives of contemporary China. The general restrictions of the household registration or *hukou* system as applied to the general population have seen relatively less change. As noted in Young (2013), in some form the *Hukou* system has been present in China for thousands of years. The modern system of mobility restrictions took shape in the early 1900s, with the contemporary system being implemented after the revolution. While many people relocate to seek employment without benefit of formal permission under *hukou*, the limits on access to health care and education for such persons greatly reduces general labour mobility, largely conforming to the mobility assumptions of our model.

There are two sectors of the economy. The first sector is F whose output is homogeneous and is produced with constant returns to scale and costlessly traded within and between regions I and J. Sector F obtains input only from the labouring households.

The second sector is X, which obtains inputs both from labouring and entrepreneurial households. Each producer requires one unit of capital from an entrepreneurial household in order to produce any output. After these fixed costs are covered, output is produced at constant returns to scale in amounts proportional to the amount of unskilled labour employed. The output of sector X can be transported between regions with transport costs being of the "iceberg" variety with a cost parameter $\tau > 1$. Thus if a quantity X of this sector's output is transported between regions, only $\frac{X}{\tau}$ arrives to be sold at the destination.

The price of output from sector F is taken as *numeraire*. The price of unskilled labour is W and the price of capital is R. We think of sector X as representing production of manufactured goods and services, and sector F as representing agriculture and resource extraction. labouring households can switch between sectors, since both sectors use the labour they provide to produce output. Let L_I^F represent the amount of unskilled labour provided to sector F in region I, and F_I be the output of sector F in region I. We can choose units so that $L_I^F = F_I$, and this implies that $W = 1$. Since production is constant returns to scale we can without loss of generality think of F_I as the number of F-sector enterprises ("farms" and "mines") in region I. Similar observations can be made for region J.

Every household of either type has preferences represented by a quasi-linear utility function:

$$U = \alpha \ln C_X + C_F, \tag{1}$$

where:

$$C_X = \left(\int_0^{N_I} x_i^{\frac{\sigma-1}{\sigma}} di + \int_{N_I}^{N_J} x_j^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}} \text{ and } \alpha > 0,\ \sigma > 1. \tag{2}$$

C_X is an aggregate of manufactured goods and services of various types indexed by i and j, each produced by a separate monopolistically-competitive firm employing one unit of capital and an amount of labour c per unit of output. Those having indices i with $0 \leq i \leq N$ are made in region I and those having indices j with $N \leq j \leq N^*$ are produced in region J.

The budget constraint for each household in region I is given by:

$$P \cdot C_X + C_F = Y, \text{ where } P = \left(N_I \cdot P_i^{1-\sigma} + N_J \cdot (\tau P_j)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{3}$$

where $\tau > 1$ measures iceberg transportation costs as noted above, Y is household income, P is the *CES* price index for manufactured goods and services appropriate for the region, and N_I and N_J represent the number of good or service types that are produced in region I and J respectively.

This gives rise to demand for output of sector F of $C_F = Y - \alpha$, demand for the aggregate of manufactured goods and services of $C_X = \alpha \cdot P^{-1}$, and to indirect utility of $V = -\alpha \cdot \ln P + Y + (\alpha \cdot (\ln \alpha - 1))$.

With the output of sector F as *numeraire* and constant returns to scale in production in that sector, the equilibrium wage rate is $W = 1$. With the production of goods and services in sector X as described following equation 2, the fixed cost for production of each type is given by the cost of capital R and the monopolistically competitive industrial structure and zero profit equilibrium generates markup pricing applied to marginal cost c, with a scale of operation for a firm of:

$$x_i = \frac{R \cdot \sigma}{(\sigma - 1)} \tag{4}$$

This mirrors the setup of Pfluger (2004). If we limit the intensity of demand for produced goods and services by assuming that:

$$\alpha < \frac{\rho_I \cdot \sigma}{(2\rho_I + 1) \cdot (\sigma - 1)} \text{ with } \rho_I = \frac{L_I}{K_I + K_J} \tag{5}$$

then at least some labour will be employed in sector F. labour employment in sector F in region I is then given by:

$$L_I^F = L_I - N_I \cdot R \cdot (\sigma - 1) \tag{6}$$

A similar expression can be derived for region J so that both sectors are active in both regions. In the long run, households with capital migrate between regions seeking the highest possible level of utility. Equilibrium location occurs when the difference in utility is zero. The difference in utility of households with capital in regions I and J is given by:

$$\Gamma_{I,J} = \frac{\alpha}{1-\sigma} \cdot \ln\left(\frac{\lambda\phi + (1-\lambda)}{\lambda + \phi \cdot (1-\lambda)}\right) + \left(\frac{\alpha \cdot (1-\phi)}{\sigma}\right) \cdot \left(\frac{\rho_I + \lambda}{\lambda + \phi \cdot (1-\lambda)} - \frac{\rho_J + (1-\lambda)}{\lambda\phi + (1-\lambda)}\right), \tag{7}$$

where $\phi = \tau^{1-\sigma}$, the variable $\lambda = \frac{K_I}{K_I + K_J}$ gives the share of mobile households (and hence manufacturers of goods and services) who are located in region I, and other variables are as defined above. Given values for parameters ϕ, σ, α, ρ_I and ρ_J, the values of λ that result in $\Gamma_{I,J} = 0$ will identify spatial equilibrium arrangements of industry between the regions.

The value of λ is central to measurement of the concentration of production in this model. A widely-used measurement for this concentration is the location quotient LQ. Assuming each household that supplies labour to the sector operates as an independent firm, the firm location quotient for region I is given by:

$$LQ_I = \frac{\frac{K_I}{(L_I^F + K_I)}}{\frac{K_I + K_J}{(L_I^F + L_J^F + K_I + K_J)}} = \frac{\lambda}{\frac{L_I^F + K_I}{(L_I^F + L_J^F + K_I + K_J)}} \tag{8}$$

This is more easily examined if expressed as a logarithm:

$$\ln LQ_I = \ln \lambda - \ln \left(L_I^F + K_I\right) + \ln \left(L_I^F + L_J^F + K_I + K_J\right) \tag{9}$$

The second term on the right hand side of equation 9 is the logarithm of total producers in either sector in region I. The final term is the logarithm of total producers in both sectors in both regions. Organised in this way, the terms are ordered in decreasing order of variability with respect to transportation costs τ.

As transport costs increase, the values of λ that equilibrate the model change considerably, except for the value $\lambda = \frac{1}{2}$ which is always an equilibrium of the model, although at moderate and low values of τ it is an unstable equilibrium. At high values of τ it identifies the unique equilibrium of the economy. When $\lambda = \frac{1}{2}$ then firms are equally distributed between the regions and the location quotient $LQ_I = 1$ and $\ln LQ_I = 0$.

As τ increases, the number of sector X firms in region I can rise or fall, but the number of sector F firms moves in the opposite direction so that the change in the second term is modest. As τ changes it is possible that the total number of firms in both regions changes, but this change is even smaller so that the third term exhibits the least variation with respect to τ.

To illustrate the relationship that emerges between transportation costs and the logarithm of location quotients in equilibrium, figure 1 below illustrates an example. As indicated, equal division of firms between regions ($\lambda = 1/2$) is an equilibrium but at lower levels of transportation costs to the left of the figure it is not a stable equilibrium. At lower levels of τ, there are equilibria involving concentration of sector X, with one possible outcome being region I having a larger share of firms from the sector (and hence $\ln LQ_I > 0$ and an alternative outcome where region I has a smaller share of firms from sector X (and hence $\ln LQ_I < 0$).

As transport costs τ increase, the difference between the two possible equilibrium levels of industrial dispersion narrows. Eventually, transport costs become so high that it is not economically feasible to ship manufactured goods and services between the regions. The unique equilibrium of $\lambda = \frac{1}{2}$ and $\ln LQ_I = 0$ holds.

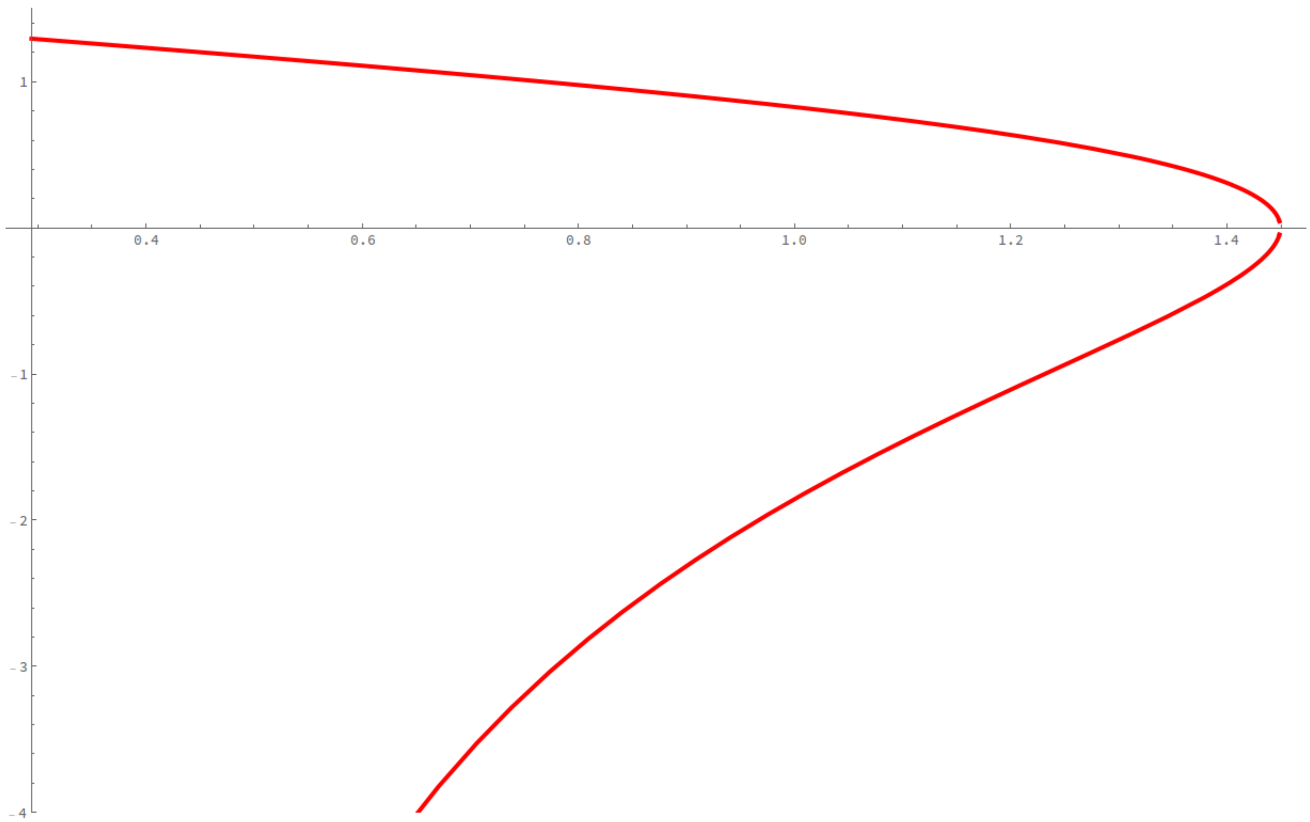

Figure 1: Equilibrium location quotients as a function of transport costs

This prediction of the relationship between transport costs and firm location quotients is derived from our simple model. We might wonder if there is any point in proceeding with empirical analysis to investigate whether this relationship is present in actual data from the Chinese economy. We discuss below the full extent of our analysis and the possible measures we use to represent transportation costs. For an initial indication of the model's potential and to inform our understanding of industrial concentration in China, Figure 2 juxtaposes the plots of potential equilibrium levels from the simple model and a scatter plot of the logarithm of firm output location quotients in 41 industries across all Chinese provinces calculated over the years 1998 through 2010, against the logarithm transport costs measured using the direct costs of transportation for each industry in the province for the appropriate year.

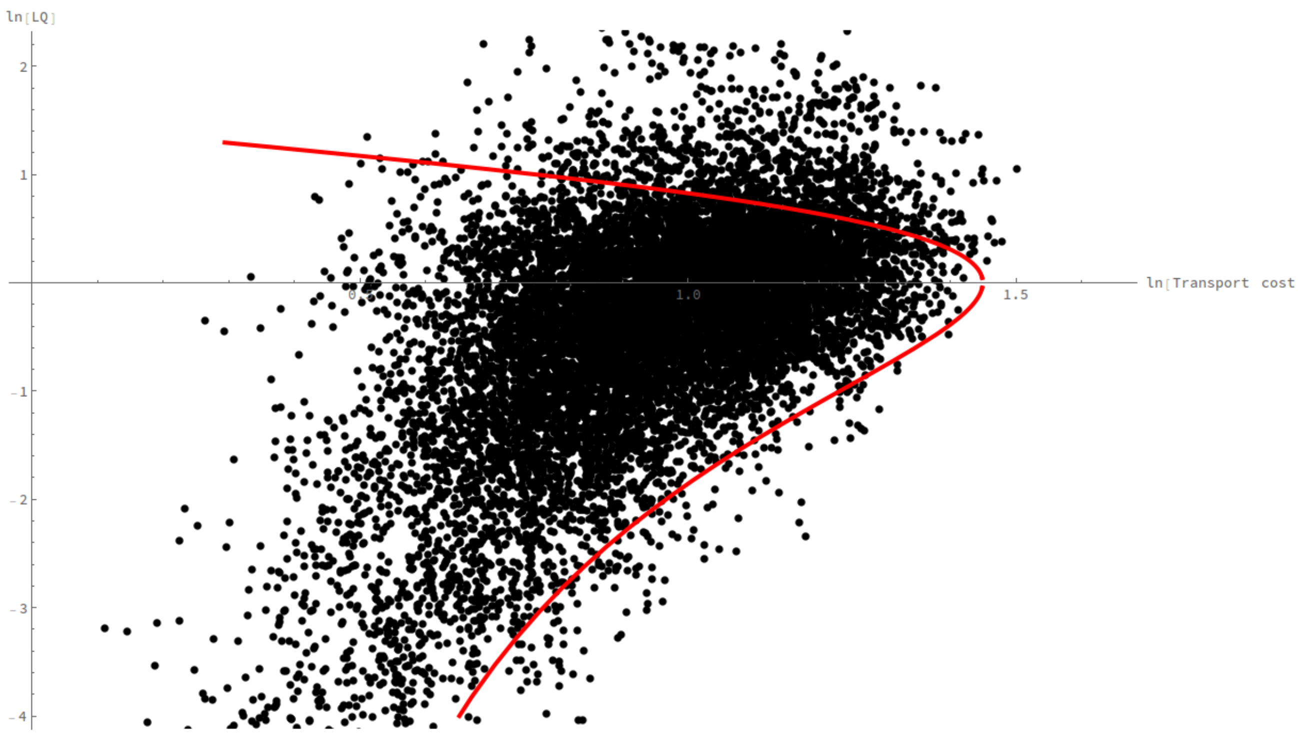

Figure 2: Location quotients and transport costs measured by direct costs from inter-industry data

This plot suggests that the actual data are at least generally similar to the pattern suggested by the theory. Not all data points lie along the locus of equilibrium location quotients, as expected of data drawn from a variety of industries facing different levels of demand and transport costs. Further refinements in measurement and inclusion of controls for other relevant variables can be expected to improve the fit, indicating that detailed data analysis will be of interest. We turn to such an analysis in the next section.

5 Data

Testing the model presented in section 4 requires data on industry-level location quotients, transportation costs, and other variables that may account for variation in relationship between the level of industrial concentration and transport costs. These latter controls would include factors that may capture variation in the intensity of demand for agricultural goods and the elasticity of substitution between different types of manufactured goods. Other factors such as the extent of state involvement in the economy or the importance of foreign trade may also be partly responsible for the dispersion apparent in the data illustrated in figure 2, and a test of the relationship between industrial concentration and transport costs may require control for these factors.

The data on location quotients are calculated from China's provincial and national statistical yearbooks, starting from 1998 and ending with 2010. Provincial yearbooks contain data specific to each province of China while the national yearbooks provide a general overview of the economic characteristics of China as a whole. Every year, each provincial yearbook records data on its industrial output, establishment count, and other statistics for each of the approximately 41 industries (for all firms with at least an annual revenue of 5 million RMB) as well as data on each province's transportation infrastructure, industry and other statistics. Analogously, the national yearbooks contain the same content but on an aggregated, national level.

We include the 27 provinces and equivalent administrative direct-controlled municipalities of China: Anhui, Beijing, Chongqing, Fujian, Gansu, Guangdong, Guizhou, Hainan, Hebei, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Qinghai, Shaanxi, Shandong, Shanghai, Shanxi, Sichuan, Tianjin, Yunan, and Zhejiang. We include Guangxi, an autonomous region, because its inclusion, together with the other provinces, forms a contiguous block of provinces for China as a whole. We exclude Taiwan, Hong Kong, Macau, Inner Mongolia, Ningxia Autonomous Region (bordering Inner Mongolia), Uyghur Autonomous Region in the far west, and Tibet, because of issues related to data comparability, data quality, and greater than usual limits on mobility of workers and producers. Selecting these provinces provides a contiguous whole covering the major economic production centers of China between which industrial relocation and internal migration are possible, while within the limits of central government regulation and the constraints of the Hukou system. The map presented in Figure 3 marks the provinces included in our analysis with cross-hatching, and shows them in the context of the entire nation.

Figure 3: Provinces included in analysis

While it is easy to speak of "transportation costs" in the abstract, obtaining reliable measures of these costs in different industries, provinces, and years is difficult. No Chinese data sources of which we are aware provide transport cost measures clearly or directly by industry or region. We make use of two broad approaches to construct a total of eight alternative measures of transport costs for each industry, province, and year and then test our theoretical approach using each of these measures. In the broadest sense, we have two measures: one group of measures based on the amounts of transportation infrastructure and another based on the flows of payments from each industrial sector to the transportation sector revealed in tables of inter-industry flows (input-output coefficients).

Data on input-output (IO) coefficients of China's industries are in the form of the statistical input-output tables of China, released roughly every three years, since 1997, by the National Bureau of Statistics for China [1]. We make use of the IO coefficients that are calculated on the national level for different industries; province-specific IO coefficient tables are either inconsistent from year to year within a province or non-existent altogether. Data on the geographic and topological features of each province were obtained from SRTM digital elevation data and calculated for each province using ArcGIS. Our main transport cost measures focus on the lengths of highway and railroad infrastructure relative to active firms in the province, as well as the IO coefficients for each industry's transport cost share.

Chinese data sources divide the economy into what are characterised as "Primary", "Secondary" and "Tertiary" industrial sectors. These do not necessarily correspond to the more or less conventional interpretation of these terms. For example, the Secondary sector includes many mining and resource extraction activities, with the "Primary" sector restricted to agriculture and animal husbandry. The total value of produced output summed across these three sectors is equivalent to China's GDP. For our analysis, we focus on producers and industries active in one of 41 "Secondary" sector activities, whose individual sector descriptions are presented in the Appendix, Table 10. Collectively, the industries in the secondary sector accounted for approximately 47 percent of Chinese GDP in 2010, and have experienced about 12 percent average year on year growth during our sample period. Table 1 presents total economic output, output in these sectors and GDP *per capita* for 1998 through 2010.

In this sense, the "Primary" sector corresponds most naturally to sector F described in the model presented in section 4. We do not analyze production or agglomeration in what is called the "Tertiary" sector, which largely consists of professional services and accounted for 43 percent of GDP in 2010. As mentioned earlier, much of China's developmental history and industrialization up to even the 21st century centered largely on manufacturing, with a growing proportion of services materializing only later on.

In our overall provincial-level data set, there are approximately 41 industries that are consistently recorded throughout time and across provinces. The industries are listed along with other variables in table 10 in the Appendix. Two industries (one is *Other Mining & Dressing* and the other is *Weapons and Ammunition*

[1] See National Bureau of Statistics (2009), National Bureau of Statistics (2006), National Bureau of Statistics (2002) and National Bureau of Statistics (1999)

Table 1: Aggregate industrial output in China, 1995-2010
(100 million RMB, in current prices)

Year	Gross National Income	Gross Domestic Product	Primary Industry	Secondary Industry Total	Secondary Industry	Secondary Construction	Tertiary Industry	Per Capita GDP
1995	59810.5	60793.7	12135.8	28679.5	24950.6	3728.8	19978.5	5046
1996	70142.5	71176.6	14015.4	33835	29447.6	4387.4	23326.2	5846
1997	78060.9	78973	14441.9	37543	32921.4	4621.6	26988.1	6420
1998	83024.3	84402.3	14817.6	39004.2	34018.4	4985.8	30580.5	6796
1999	88479.2	89677.1	14770	41033.6	35861.5	5172.1	33873.4	7159
2000	98000.5	99214.6	14944.7	45555.9	40033.6	5522.3	38714	7858
2001	108068.2	109655.2	15781.3	49512.3	43580.6	5931.7	44361.6	8622
2002	119095.7	120332.7	16537	53896.8	47431.3	6465.5	49898.9	9398
2003	135174	135822.8	17381.7	62436.3	54945.5	7490.8	56004.7	10542
2004	159586.8	159878.3	21412.7	73904.3	65210	8694.3	64561.3	12336
2005	183618.5	184937.4	22420	87598.1	77230.8	10367.3	74919.3	14185
2006	215883.9	216314.4	24040	103719.5	91310.9	12408.6	88554.9	16500
2007	266411	265810.3	28627	125831.4	110534.9	15296.5	111351.9	20169
2008	315274.7	314045.4	33702	149003.4	130260.2	18743.2	131340	23708
2009	341401.5	340902.8	35226	157638.8	135239.9	22398.8	148038	25608
2010	403260	401202	40533.6	187581.4	160867	26714.4	173087	29992

Source: China National Yearbooks

Manufacturing) exhibit large swaths of missing values along the years and across provinces; however, our results are robust regardless of whether or not they were included or removed, so the remaining observations are included in the data.

Our focus on provincial level data for analysis owes to the stability and consistency of the province as a geographic unit. A Chinese province is analogous to a US state; its defining boundaries have remained largely stable and constant over the past few decades, at the very least. *Prefectures*, a similar geographic unit that are included under provinces, are not an official geographic unit or political division recognised by the Chinese constitution. Furthermore, they are subject to a more frequent redrawing of boundaries. As a result, the study of industrial agglomerations over the years can be interrupted and misplaced if some prefectures are absorbed, pared down, or converted. Provincial boundaries have remained largely static since the 17$^{\text{th}}$ century and remain the centerpiece of jurisdictional rulings, policies, and a cultural identity with which people and businesses identify as their native homes or customer base.

As suggested above in section 4 we focus on the location quotient to capture the geographic concentration of industry. For a given industry i in a province p during some year t, it is defined as

$$LQ_{ipt} = \frac{\beta_{ipt} / \sum_i \beta_{ipt}}{\sum_p \beta_{ipt} / \sum_i \sum_p \beta_{ipt}} \tag{10}$$

where β is some economic characteristic of some industry i in province p at time t. Examples of characteristics would be total employment, total output, or number of firms. Originating with Florence (1939), the location quotient (LQ) provides a measure of the concentration of an industry in a region relative to the total level of activity of the province relative to the same but on the national scale.

Other measures of industrial concentration exist such as the industrial location Gini as in Wen (2004), or the Duranton & Overman (2005) index, which are two popular measures with similar variants. However, the Gini's downward bias is particularly a concern without a rather large sample size in each period/unit of time, resulting in the possible exclusion of some regions and industries if implemented. Similarly, the latter requires massive amounts of data and the exact spatial address of every establishment. While it is easy to test the statistical significance of the Duranton and Overman index via Monte Carlo methods, such methods may make several results difficult to reproduce if not completely irreproducible. The large data requirements of many concentration indices are of particular concern given the difficulty in obtaining large amounts of consistent data, in terms of quality and granularity, for China.

The LQ has several attractive features as a measure of industrial concentration. Its primary strength lies in its simplicity, which allows for intuitive interpretation and consistent comparison across regions and industries. If an LQ is equal to 1, then an industry has the same share of total production in its province as it does in the nation. If the LQ is greater than 1, then the industry has a greater share of its provincial production than in the overall country, and vice-versa if an LQ falls below 1. Secondly, the LQ is typically unbiased and is generally more accurate for inference as the size of the relevant geographic unit increases (Billings & Johnson (2012)), making the provincial level data an attractive choice. Using U.S. County Business Patterns data from 2000, Billings and Johnson show that the accuracy of inference, specifically with confidence bands, increases with the level of the geographic unit. Finally, it is unit-less, so the measure of concentration is independent of the system of units used to calculate the LQ.

For our purposes, we use an industry's gross output as our economic characteristic in the LQ. Our provincial level data contain gross output statistics for all firms, across all industries, with an annual revenue of at least 5 million RMB. Doing so allows for the closest measurement of industrial clustering on the basis of actual production. Since each industry is different by nature of what they produce and how they produce, other characteristics such as the number of firms, employment, or value-added of each industry can be influenced by each industry's specific nature. For instance, power suppliers could have fewer establishments but have power grids to distribute their good, while pharmaceutical manufacturers may require many pharmacies and outlets to distribute; as such, establishment count or employment can be misleading in exploring the scale of industrial concentration. In some cases, in the the data, value-added is negative for several industries and provinces – or missing and unrecorded altogether in many cases, resulting in a nonsensical interpretation of a negative LQ. As such, using gross industrial output is more reliable in having a consistent measure and comparison of industrial production shares across the different industries.

Moreover, a focus on the provincial level not only lends some desirable statistical properties to the LQ as an estimator, but, more importantly, the provincial level offers an optimal balance between granularity and consistency in the data. Industry categories are largely consistent across the provinces and the years, while still remaining relatively large in number for meaningful study of different industrial concentrations. Small errors on the lower local levels within a province also tend to average out, especially in using a relative index such as the LQ. The data quality on lower levels of geographic hierarchy, such as for cities or counties, is poor, resulting in smaller sample sizes, less variation, or inconsistent and missing data. For instance, Ke, He, and Yuan (2013) study industry concentrations on the city level using only 10 general industry categories because of the inconsistent industry-sector breakdowns, data quality and continuity problems (across time and cities), and a lack of data on output and establishment count for each specific industry. The focus on the the provincial level also presents a safe, conservative approach to capture the extent of agglomeration activity: if clusters were measured on a lower geographic level but exist on a larger scale or level, using any unit below the province would neglect the full extent of agglomeration activity.

Table 2 presents descriptive statistics of the main covariates used in the model over the entire sample period; we see that they exhibit a large range that reflects the variety and diversity of provincial characteristics which make up China as whole. Tables 3 and 4 present descriptive statistics from the first and last year of the sample data, respectively. Over time, from the first year of our sample (1998) compared to the last (2010), living standards have largely increased, on average–especially per capita income. Population density has increased slightly while trade dependency has increased while maintaining the same spread, more or less, among the provinces.

Table 2: Descriptive statistics for selected covariates (All Provinces, 1998-2010)

Variable	μ	σ	Min	Max
Transport costs based on interindustry data				
Direct costs	65047	163889	0.0454	3342164
Direct costs per firm	559	2505	0.0454	108670
Total costs	240938	715663	0.1457	20100000
Total costs per firm	1673	6739	0.1457	245837
Transport costs based on infrastructure measures				
Industry firms per rail km	0.211	0.58	0.0002	11.06
Total firms per rail km	7.477	12.65	0.2618	90.52
Industry firms per highway km	0.006	0.02	0.0000	0.67
Total firms per highway km	0.215	0.47	0.0087	5.47
Other covariates				
Current income	9697	4115	4050	25973
Trade dependency	0.170	0.20	0.0016	0.91
Pop density	0.040	0.04	0.0007	0.22
State share GDP	0.502	0.21	0.0931	0.90

As detailed earlier, at the most basic level, we take two different approaches to constructing transportation cost measures. The first type of measure is based on the ratio between the number of active firms in the province relative to the length of transport infrastructure. We explore one measure based on the length of railroads operating in the province and another based on the length of roadway (and highway) in the province. As the number of firms producing output rises relative to the amount of transport infrastructure, it seems natural to take this as indicative of an increase in transportation costs due to their crowding capacity on transportation infrastructure.

The two infrastructure-based measures of transportation costs tell an interesting story in the descriptive statistics below. For each province, rail and highway are measured in km, Income is per capita, population density is in 10,000 persons per sq km, and trade dependency is the export value as a share of GDP. While the highway based measure has more than doubled since the beginning of our sample, the railway based measure has decreased on average, albeit by a very small amount. This could be due to the competing effects between an expansion in railway infrastructure and the increase in production activity through an increase in firm count. We also see that across the provinces, the range has shifted up for the highway measure as all provinces, on average, have increased their highway infrastructure relative to the number of firms; meanwhile, the spread of the rail based measure has increased as the range also shifts up slightly for all provinces.

Our other approach to measuring transportation costs is based on analysis of the matrix of input-output coefficients for the Chinese economy. Using the available inter-industry matrix that most nearly corresponds to the year under study, we examine the transportation inputs per currency unit of output for each of our industry groups. This can be done in two ways. We can simply consider the direct inputs from the transportation sector as a measure of transport costs relevant for the industry, or we can take a more expansive view and consider both the direct and indirect effects to capture the total transport requirements of the industrial sector. This approach basically asks: how much transportation production would have to expand if we increased the value of output in the particular industry by one unit? Naturally, the direct transport needs must be met, but so too must the indirect transport requirements of other industries providing intermediate inputs to the industry in question. In this way, we can construct a total transport cost measure for each industry that varies over time (with revisions of the matrix of inter-industry flows).

Table 3: Descriptive statistics for selected covariates (All Provinces, 1998)

Variable	μ	σ	Min	Max
Transport costs based on interindustry data				
Direct costs	11571	20379	0.9218	178261
Direct costs per firm	142	605	0.1783	13544
Total costs	34394	62612	2.8680	751597
Total costs per firm	391	1873	0.6703	43606
Transport costs based on infrastructure measures				
Industry firms per rail km	0.206	0.71	0.0002	11.06
Total firms per rail km	7.542	16.57	0.5209	90.52
Industry firms per highway km	0.011	0.04	0.0000	0.67
Total firms per highway km	0.410	1.01	0.0319	5.47
Other covariates				
Current income	5552	1426	4050	9002
Trade dependency	0.138	0.16	0.0016	0.73
Pop density	0.038	0.04	0.0007	0.21
State share GDP	0.630	0.19	0.2649	0.89

Table 4: Descriptive statistics for selected covariates (All Provinces, 2010)

Variable	μ	σ	Min	Max
Transport costs based on interindustry data				
Direct costs	165807	329401	16.8043	3342164
Direct costs per firm	1053	3812	0.2758	77403
Total costs	678859	1487272	60.0151	20100000
Total costs per firm	3609	11625	1.0749	233161
Transport costs based on infrastructure measures				
Industry firms per rail km	0.232	0.57	0.0003	5.23
Total firms per rail km	8.146	11.39	0.3362	40.30
Industry firms per highway km	0.005	0.01	0.000005	0.18
Total firms per highway km	0.169	0.28	0.0089	1.39
Other covariates				
Current income	15242	4187	9677	25973
Trade dependency	0.161	0.18	0.0069	0.71
Pop density	0.042	0.04	0.0008	0.22
State share GDP	0.376	0.17	0.1073	0.78

In Table 5 we see the range and spread of transport input-output coefficients compared with the range and spread of location quotients on a national level for a few selected industries. Since these statistics are for the national level and the location quotient compares a local economy to the national, the mean location quotient on a national level lacks meaning. However, we see that the observations are in line with our predictions. As the average transport coefficient τ_{avg} (cost) falls from the beginning of our sample to the end, we would expect the location quotients to be more spread out and vice versa – which is what we see here as the σ_{LQ} increases for these selected industries. Of course, we still need to control for other variables, which we do in our models later on.

To better see and illustrate the effects, we focus on a specific province, Guangdong, and its selected industries, which then allows for an average location quotient be meaningful. Table 6 lines up with what our model predicts: as the average transport coefficient (cost) increases from the beginning to the end of our sample, σ_{LQ} decreases as the spread narrows between the location quotients; consequentially, the mean location quotient falls as well. The opposite direction holds true as well when we examine a fall in the average transport coefficients (costs) for an industry in the province.

Table 5: Location Quotients & Transport Cost Spreads (Selected Industries, National Level)

	1998-2000				2008-2010			
Variable	σ_{LQ}	Min_{LQ}	Max_{LQ}	τ_{avg}	σ_{LQ}	Min_{LQ}	Max_{LQ}	τ_{avg}
Elec. & Heat Supply	0.515	0.434	2.821	0.031	0.712	0.563	3.986	0.012
Metal Products	0.438	0.197	1.626	0.029	0.417	0.129	1.679	0.021
Non-Metal Mining	0.983	0.002	5.036	0.041	0.793	0	2.915	0.064
Print. & Record Prod.	0.705	0.159	3.385	0.014	0.497	0.166	2.177	0.019
Tobacco	3.870	0	20.466	0.012	3.252	0	19.463	0.027

Before proceeding to examine the relationship between industrial concentration and transport costs while controlling for other relevant variables in a formal model, we should note the important role of the national state sector in manufacturing and production. Though we have noted earlier that to consider the full scope of either private or state production activities, both should be considered together given the overlapping, mixed nature, it is helpful to show the share of activity formally designated as one or the other. Well into the period of reform and rapid economic growth, and even to the present day, the state-controlled sector plays

Table 6: Location Quotients & Avg. Transport Costs (Selected Industries, Guangdong Province)

	1998-2000					2008-2010				
Variable	μ_{LQ}	σ_{LQ}	Min_{LQ}	Max_{LQ}	τ_{avg}	μ_{LQ}	σ_{LQ}	Min_{LQ}	Max_{LQ}	τ_{avg}
Agricul. & Food Process.	0.616	0.041	0.570	0.649	0.011	0.447	0.033	0.422	0.485	0.017
Beverage Production	0.768	0.010	0.761	0.780	0.012	0.611	0.036	0.578	0.650	0.027
Clothes, Shoes, & Hats	1.850	0.130	1.729	1.987	0.010	1.476	0.053	1.418	1.521	0.021
Comms. & Elec. Equip.	2.227	0.053	2.169	2.272	0.019	2.799	0.072	2.716	2.847	0.014
Instrum., Meters, & Office	2.522	0.249	2.359	2.809	0.014	1.916	0.167	1.782	2.103	0.019

a significant role in the Chinese economy. Table 7 presents some descriptive statistics on the significance of the state-owned sector, indicating that it accounted for nearly two thirds of the value of goods and services produced in 1998. This has declined during the first decade of the 21st century, but even in 2010 state owned and state controlled enterprises accounted for about one third of total output by value.

Table 7: Output from state and non-state enterprises (100 millions of Yuan)

Year	State-owned Enterprises	Private Industries	Macao, HK, Taiwan, & Foreign	State Total ratio	Private State ratio
1998	33621	2083	16758	0.64	0.06
1999	35571	3245	18954	0.62	0.09
2000	40554	5220	23465	0.59	0.13
2001	42408	8761	27221	0.54	0.21
2002	45179	12951	32459	0.50	0.29
2003	53408	20980	44358	0.45	0.39
2004	70229	35141	65995	0.41	0.50
2005	83750	47778	79860	0.40	0.57
2006	98910	67240	100077	0.37	0.68
2007	119686	94023	127629	0.35	0.79
2008	143950	136340	149794	0.33	0.95
2009	146630	162026	152687	0.32	1.11
2010	185861	213339	189917	0.32	1.15

This presents central government policy makers with multiple opportunities to pursue political objectives in selecting locations for industrial concentration and development. The location patterns that best attain these objectives would not necessarily align with the location patterns consistent with individual optimization by firms and mobile households. For that reason there is doubt about how useful a model, such as that developed in section 4, can be in helping us to understand and predict the spatial concentration of industry in China. We turn in the next section to some tests of the model by examining the relationship between measures of industrial concentration and transportation costs or other covariates.

6 Estimates

For a more complete assessment of the model predictions applied to recent Chinese economic development, we estimate quantile models that can capture the upper limit of location quotients for those industries and provinces experiencing concentration of manufactured goods and services and the lower limit of location quotients for provinces where there is less concentration of particular industries. Our theoretical model predicts that as transport costs increase, the inter-quantile differences between these limits should be reduced.

We include in our models controls for income, population density, trade dependency and the share of production taking place in the state sector. In addition, we control for year fixed effects (relative to the datum of 2010) as a way of adjusting for nationwide and cross-industry trends including changes in price levels. We also control for fixed effects associated with selected extraction industries that may be tied to specific geological features and mineral deposits; these industries are less responsive to scale economies that would otherwise encourage concentration if transportation costs fall because of their dependency on fixed locations.

Table 8 presents a series of estimates for inter-quantile models using the complete data from 1998 through 2010 for all 41 industries in all 27 provinces. After column one containing variable names, the estimates presented in columns two through four use varying sets of controls and measure transport costs using a proxy developed to capture the total impact on transport demand of increasing the output of the industry, as discussed in section 5 above. The estimates presented in columns five through seven present models based on the direct use of transport costs in the industry. In every case, the dependent variable is the spread between the 10^{th} and the 90^{th} percentile of the logarithm of the location quotient. All continuously variable right hand side variables are measured as logarithms.

As predicted by the theory presented in section 4 above, each of the models estimated confirms the central prediction: an increase in transport costs is associated with a decrease in the spread of location quotients. As transport costs increase, industry becomes less concentrated as predicted by a model based on utility and profit maximization, with free mobility by owners of physical and human capital.

Table 8 presents estimates for the models using these controls. The expected sign for parameters associated with controls for extraction and mining-oriented firms depends on the degree to which the resources being processed are concentrated in the Chinese landscape. If the resource is relatively concentrated, we may expect this concentration to lead to greater dispersion between provincial location quotients than for more footloose industries, and a significant and positive estimated coefficient would be the result. Petroleum and coal might be reasonable examples. The estimates presented in Table 8 mostly match these expectations.

Table 8: Interquantile estimates

	Direct costs	(Dir costs / firms)	Total costs	(Tot costs / firms)	(Ind firms / highway km)	(Ind firms / rail km)	(Total firms / highway km)	(Total firms / rail km)
Transport costs	-0.3760***	-0.2311***	-0.3667***	-0.2377***	-0.5412***	-0.5033***	-0.3655***	-0.1400***
σ	*0.0106*	*0.0232*	*0.0113*	*0.0262*	*0.0171*	*0.0176*	*0.0463*	*0.0342*
Income	0.2307*	1.7101***	0.3051***	1.5390***	2.4343***	2.1737***	2.1244***	1.6813***
σ	*0.1189*	*0.1944*	*0.1081*	*0.1993*	*0.1285*	*0.1471*	*0.2180*	*0.2197*
Pop density	-0.1514***	-0.5956***	-0.1663***	-0.5839***	-0.2745***	-0.1764***	-0.4855***	-0.6265***
σ	*0.0304*	*0.0441*	*0.0273*	*0.0440*	*0.0332*	*0.0398*	*0.0563*	*0.0631*
State share	-0.3602***	0.2971***	-0.3571***	0.2679***	0.5541***	0.2380***	0.6904***	0.6491***
σ	*0.0456*	*0.0669*	*0.0351*	*0.0875*	*0.0424*	*0.0428*	*0.0496*	*0.0633*
Trade	0.0751***	-0.1011***	0.0440**	-0.1136***	0.0651***	-0.0847***	-0.0641	-0.1090***
σ	*0.0227*	*0.0351*	*0.0201*	*0.0398*	*0.0238*	*0.0284*	*0.0409*	*0.0382*
Other mine	-1.1715**	0.6546	-1.5481**	0.4231	-1.4033***	-1.5158***	1.4163*	1.3783*
σ	*0.5315*	*0.8252*	*0.6110*	*0.8378*	*0.4749*	*0.4603*	*0.8030*	*0.7684*
Coal mine	0.4064***	2.1742***	0.1414	2.0501***	0.2494*	0.3445	2.0209***	1.9542***
σ	*0.1286*	*0.1791*	*0.1432*	*0.2738*	*0.1431*	*0.2197*	*0.3087*	*0.3719*
Ferrous mine	0.1308	0.8975***	-0.0866	0.8710***	-0.5972***	-0.5571***	1.0109***	1.0400***
σ	*0.1312*	*0.2908*	*0.1277*	*0.3263*	*0.1585*	*0.1512*	*0.2347*	*0.2465*
Non-ferr mine	0.2424***	1.2080***	-0.1138	1.0918***	-0.1586	0.0605	1.5603***	1.6267***
σ	*0.0725*	*0.1909*	*0.1002*	*0.1643*	*0.1875*	*0.1721*	*0.2830*	*0.2309*
Non-met mine	-0.5887***	-0.2061	-0.8782***	-0.3934**	-0.6758***	-0.6430***	0.1576	0.1141
σ	*0.0652*	*0.1847*	*0.0684*	*0.1798*	*0.1010*	*0.0966*	*0.1529*	*0.1675*
Petrol & gas	1.0118***	3.2106***	0.8605***	3.1076***	-0.1529	0.3712	3.4861***	3.4432***
σ	*0.1254*	*0.1925*	*0.1196*	*0.1771*	*0.2668*	*0.3032*	*0.1759*	*0.1772*
Y_{1998}	-0.6324***	0.6851***	-0.6460***	0.4172*	2.6071***	1.9783***	1.8179***	1.0608***
σ	*0.1264*	*0.2470*	*0.1190*	*0.2426*	*0.1496*	*0.1437*	*0.2765*	*0.2825*
Y_{1999}	-0.6331***	0.5734**	-0.6414***	0.3288	2.3298***	1.8039***	1.5737***	0.9226***
σ	*0.1220*	*0.2313*	*0.1125*	*0.2238*	*0.1306*	*0.1492*	*0.2213*	*0.2605*
Y_{2000}	-0.3174**	0.6703***	-0.3996***	0.4558**	2.1279***	1.6556***	1.4171***	0.7961***
σ	*0.1398*	*0.2231*	*0.1019*	*0.2242*	*0.1222*	*0.1515*	*0.2397*	*0.2308*

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
Y_{2001}	-0.2885***	0.6306***	-0.3855***	0.3593*	1.8152***	1.4436***	1.2036***	0.7041***
σ	*0.1125*	*0.1952*	*0.1097*	*0.1994*	*0.1299*	*0.1337*	*0.2308*	*0.2283*
Y_{2002}	-0.3119***	0.5931***	-0.3505***	0.3765*	1.5591***	1.2024***	1.0475***	0.6902***
σ	*0.0927*	*0.2066*	*0.1040*	*0.2047*	*0.1211*	*0.1206*	*0.2023*	*0.2085*
Y_{2003}	-0.2944***	0.6101***	-0.3127***	0.3934**	1.4319***	1.1022***	1.1506***	0.7722***
σ	*0.1030*	*0.1692*	*0.0987*	*0.1812*	*0.1137*	*0.1034*	*0.2067*	*0.1961*
Y_{2004}	-0.1980*	0.6116***	-0.2415**	0.4388**	1.3986***	1.1361***	1.1835***	0.8646***
σ	*0.1022*	*0.1783*	*0.0951*	*0.1733*	*0.1167*	*0.1110*	*0.1804*	*0.1847*
Y_{2005}	-0.1132	0.5745***	-0.0985	0.4873***	1.1017***	0.9085***	0.8913***	0.5882***
σ	*0.0895*	*0.1587*	*0.0961*	*0.1743*	*0.0913*	*0.1042*	*0.1606*	*0.1748*
Y_{2006}	-0.1070	0.4635***	-0.0978	0.3729**	0.6095***	0.7192***	0.5476***	0.4641***
σ	*0.0735*	*0.1621*	*0.0871*	*0.1679*	*0.0881*	*0.0826*	*0.1386*	*0.1415*
Y_{2007}	-0.1475*	0.2492*	-0.1650**	0.1578	0.4176***	0.5124***	0.3737***	0.3263**
σ	*0.0775*	*0.1373*	*0.0698*	*0.1527*	*0.0875*	*0.0895*	*0.1339*	*0.1321*
Y_{2008}	-0.1084	0.1567	-0.1088	0.1307	0.3895***	0.4264***	0.3818***	0.3391**
σ	*0.0765*	*0.1428*	*0.0675*	*0.1405*	*0.0898*	*0.0955*	*0.1443*	*0.1360*
Y_{2009}	-0.1106	-0.0456	-0.1074*	-0.1050	0.1402*	0.1194	0.1422	0.1263
σ	*0.0680*	*0.1261*	*0.0645*	*0.1321*	*0.0835*	*0.0823*	*0.1483*	*0.1304*
Constant	2.8409**	-14.3493***	2.5329**	-12.2878***	-25.1599***	-20.4610***	-19.8596***	-15.0492***
σ	*1.2158*	*1.9426*	*1.0993*	*2.0210*	*1.3104*	*1.4753*	*2.1934*	*2.2292*
Observations	12021	11793	12021	11793	12231	12231	12360	12360
$R^2_{0.9}$	0.1522	0.0995	0.1527	0.1015	0.0693	0.072	0.0456	0.0446
$R^2_{0.1}$	0.5032	0.1879	0.499	0.1896	0.3696	0.3505	0.1168	0.1133

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Apart from the usual challenges associated with the estimation of quantile models, another major hurdle that confronts our analysis relates to the possible endogeneity of transport costs. The infrastructure investments that are capable of producing changes in transportation costs do not rain down randomly on Chinese provinces. They are the outcome of human decisions. If the decision to invest is correlated with errors in the estimated relationship between location quotients, transportation costs, and other controls then the potential for endogeneity exists, and uncorrected estimates of the impact may be biased. We address this issue by considering a group of instruments for transportation costs (and for selected other controls) that are based on geographic and topographic features of the province that are exogenous to random error in location quotients.

It must be acknowledged that there are potential excludability concerns related to some of these instruments for some industries. For example, some production processes may operate at greater efficiency at low altitudes and therefore elevation will directly affect the concentration of industry as well as providing a useful instrument for transport costs. We argue, however, that the wide range of industries and services to which we apply our analysis includes many production processes that are immune to such concerns, and that the potential advantage of reducing bias in estimation for these industries makes such procedures worthwhile.

Our first approach to this problem is to produce IV estimates of an inter-quantile spread model by using a collection of instruments in a first stage to produce estimated values of both inter-industry and infrastructure based measures of transport costs associated with each year, industry and province. The first-stage estimates are presented in Tables 11 through 15 below. As indicated in section 5, the instruments are based on physical characteristics of the landscape and topography in each province. The results presented suggest that collectively these instruments perform reasonably well to produce estimated transport costs.

We use these first-stage results to estimate inter-quantile spread models, for which we bootstrap the standard errors. The results for both total and direct transport costs and different combinations of controls are presented in table 9. The IV estimates increase the absolute value of the estimated impact of transport costs on the inter-quantile spread by somewhere between two and three standard errors, and the estimates all remain negative and statistically significant. Additionally, an expected increase in the standard error of the estimates renders the estimated impact of provincial per capita income levels insignificant, although the sign remains unchanged.

Table 9 presents estimates of the models using these instrumented values for endogenous variables along with controls for year and selected industry that are assumed to be exogenous. The results are remarkably consistent across different approaches to measurement of transportation costs. An increase in transportation costs is associated with a decrease in industrial concentration, as predicted by the model discussed in section 4. Despite the varieties of noise and confounding influences in the data, the results are estimated with sufficient precision that each of these impacts is significant at the 1% level or better.

Other parameter estimates mostly match our expectations. Increases in *per capita* income are associated with increases in industrial concentration, and increases in population density, *ceteris paribus* enable more diverse local production and decrease concentration. Controls for mining and dressing industries are more varied but most tend to be positive, suggesting higher levels of concentration (near fixed ore deposits). As noted above, this is expected when the resource being extracted and processed is concentrated in particular provinces due their fixed geographic locations, so that the spread between location quotients for provinces with and without the resource is increased.

Overall, we argue that the estimates from Table 9, which we regard as the central results of this analysis, contribute to our confidence in the ability of a simple model, based on individual optimization and decentralised choice, to provide an analytic framework for understanding specific changes in the concentration of industry in China.

Table 9: Interquantile IV estimates with bootstrap standard errors

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
Transport costs	-0.4108***	-0.9372**	-0.4111***	-0.9334**	-0.6676***	-0.6740***	-0.4440***	-0.2142**
σ	*0.0707*	*0.4353*	*0.0732*	*0.4518*	*0.1060*	*0.1388*	*0.1135*	*0.1059*
Income	0.7180**	1.3341***	0.7246**	1.3459***	3.2162***	3.2037***	2.9113***	2.5293***
σ	*0.3663*	*0.4846*	*0.3542*	*0.4701*	*0.3756*	*0.3691*	*0.3970*	*0.3551*
Pop density	-0.5011***	-0.4592***	-0.5019***	-0.4628***	-0.3805***	-0.3892***	-0.4574***	-0.5706***
σ	*0.0559*	*0.1115*	*0.0572*	*0.1152*	*0.0680*	*0.0796*	*0.0786*	*0.0732*
State share	0.1507	0.6688***	0.1508	0.6695***	0.8855***	0.3403**	0.9398***	0.8730***
σ	*0.1730*	*0.2371*	*0.1834*	*0.2417*	*0.0760*	*0.1649*	*0.0837*	*0.1504*
Trade	0.0889	-0.0697	0.0887	-0.0709	-0.0624	-0.1961***	-0.0789	-0.1437***
σ	*0.0663*	*0.0672*	*0.0652*	*0.0667*	*0.0627*	*0.0607*	*0.0575*	*0.0553*
Other mine	-0.8636	0.6065	-1.1129	0.0407	-1.7949*	-1.8874*	1.0216	1.0277
σ	*0.7427*	*0.7735*	*0.8041*	*0.8572*	*0.9181*	*0.9896*	*0.7474*	*0.7774*
Coal mine	2.1312***	2.5330***	1.9407***	2.1148***	1.8431***	1.8398***	1.8858***	1.8570***
σ	*0.3076*	*0.4790*	*0.2906*	*0.3331*	*0.2756*	*0.2879*	*0.3056*	*0.3631*
Ferrous mine	0.8484***	1.5116***	0.6788**	1.1306***	0.2504	0.3476	1.0776***	1.1400***
σ	*0.3149*	*0.2883*	*0.2890*	*0.2952*	*0.2689*	*0.2893*	*0.2717*	*0.2597*
Non-ferr mine	1.3020***	2.0359***	1.1630***	1.4707***	0.8252***	0.9316***	1.4407***	1.5040***
σ	*0.2069*	*0.2822*	*0.2162*	*0.2121*	*0.2798*	*0.2328*	*0.2293*	*0.2662*
Non-met mine	-0.0604	0.2729**	-0.3070*	-0.2884	-0.1680	-0.1852	0.1430	0.1836
σ	*0.1475*	*0.1356*	*0.1696*	*0.2272*	*0.1187*	*0.1414*	*0.1311*	*0.1211*
Petrol & gas	3.2950***	6.1946***	3.2242***	6.0226***	1.4515***	1.3949***	3.5371***	3.4509***
σ	*0.1978*	*1.3218*	*0.2069*	*1.2605*	*0.3484*	*0.4234*	*0.2135*	*0.2265*
Y_{1998}	-0.7394*	-1.3872	-0.8231*	-1.5746	2.8385***	2.5112***	2.2991***	1.6558***
σ	*0.4377*	*1.1724*	*0.4354*	*1.2989*	*0.4339*	*0.4052*	*0.4552*	*0.3998*
Y_{1999}	-0.7832*	-1.4782	-0.8680**	-1.6654	2.5328***	2.2111***	2.0125***	1.4154***
σ	*0.4063*	*1.1115*	*0.4172*	*1.2564*	*0.3936*	*0.3786*	*0.4204*	*0.3804*
Y_{2000}	-0.5794	-0.9905	-0.6637*	-1.1802	2.2559***	2.0590***	1.8385***	1.3117***
σ	*0.3754*	*0.8770*	*0.3712*	*0.9763*	*0.3673*	*0.3620*	*0.3946*	*0.3597*

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
Y_{2001}	-0.5935*	-0.8985	-0.6750*	-1.0810	1.9876***	1.7594***	1.5950***	1.1785***
σ	*0.3499*	*0.8108*	*0.3593*	*0.9092*	*0.3216*	*0.3068*	*0.3478*	*0.3098*
Y_{2002}	-0.4640	-0.5811	-0.5542*	-0.7861	1.6848***	1.4914***	1.3405***	0.9758***
σ	*0.3078*	*0.5919*	*0.3044*	*0.7058*	*0.2822*	*0.2791*	*0.3077*	*0.2825*
Y_{2003}	-0.2617	-0.3061	-0.3514	-0.5086	1.6864***	1.4786***	1.4156***	1.0868***
σ	*0.2841*	*0.5109*	*0.2865*	*0.6292*	*0.2567*	*0.2421*	*0.2776*	*0.2609*
Y_{2004}	-0.1392	-0.1308	-0.2304	-0.3382	1.7536***	1.4415***	1.4939***	1.1233***
σ	*0.2572*	*0.4911*	*0.2713*	*0.5897*	*0.2560*	*0.2310*	*0.2665*	*0.2179*
Y_{2005}	-0.1248	-0.1502	-0.1539	-0.2185	1.3206***	1.1955***	1.0966***	0.8218***
σ	*0.2068*	*0.3505*	*0.2078*	*0.3940*	*0.2241*	*0.2186*	*0.2287*	*0.2213*
Y_{2006}	-0.0551	-0.0866	-0.0846	-0.1546	0.8335***	0.9549***	0.7110***	0.6523***
σ	*0.1813*	*0.2654*	*0.1735*	*0.3011*	*0.1736*	*0.1799*	*0.1770*	*0.1729*
Y_{2007}	-0.2806*	-0.3414	-0.3003*	-0.3844	0.4805***	0.5963***	0.4210***	0.4304***
σ	*0.1635*	*0.3088*	*0.1611*	*0.3332*	*0.1405*	*0.1528*	*0.1497*	*0.1533*
Y_{2008}	-0.1208	-0.1933	-0.1408	-0.2351	0.4933***	0.5757***	0.4327***	0.4442***
σ	*0.1439*	*0.2627*	*0.1493*	*0.2981*	*0.1304*	*0.1205*	*0.1348*	*0.1360*
Y_{2009}	-0.0602	-0.3374	-0.0820	-0.3845*	0.2256*	0.1857	0.1153	0.1132
σ	*0.1137*	*0.2067*	*0.1168*	*0.2233*	*0.1208*	*0.1214*	*0.1183*	*0.1212*
Constant	-1.5918	-5.7076	-1.0885	-4.5676	-33.2222***	-31.1467***	-27.1405***	-22.5765***
σ	*3.9634*	*6.8344*	*4.0027*	*7.3461*	*3.9236*	*3.7647*	*3.8935*	*3.3519*

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Our final approach to testing the predictions of our simple model in the Chinese context is to use an approach described in detail in Chetverikov, Larsen & Palmer (2016). This approach produces instrumental variable estimates for each specified quantile quickly. Here we display the results in graphical form in Figures 4 and 5.[2]

The results are presented not as an estimated impact of variables and controls on the inter-quantile spread, but rather on the estimated individual quantiles themselves. Referring to our model developed in section 4 above, and the predicted relationship illustrated in figure 9, we expect that for the lowest quantile (range of location quotients) an increase in transportation costs should increase the estimated quantile. Conversely, for the highest quantiles (the highest location quotients with the most concentrated industrial clusters), increases in transportation costs should reduce the estimated quantile. This generates the result that an increase in transportation costs moves all locations towards production for local consumption only.

In Figures 4 and 5 we see this effect pretty clearly. As expected, lower quantiles exhibit a positive association with transport costs (after making adjustments for treating endogeneity) and higher quantiles exhibit a negative association. Using this method, the precision of the estimated effects depends on the measure of transportation costs used. Measures of transport costs based on inter-industry coefficients (either direct costs or total costs) have a significant positive impact on 15^{th} and 20^{th} percentiles, and a significantly negative impact on the 95^{th}. The case for significant impacts is better for rail transport infrastructure than for highway. In all cases, however, the general tendency of the point estimates is supportive of our theoretical perspective.

Not only is the qualitative relationship as predicted by the model presented in section 4, but the estimates even reflect the predicted relationship that the association is numerically stronger for the lower quantiles (more rapid increase) than it is for the upper quantiles.

While the magnitudes of estimated coefficients differ, along with the precision of those estimates, each of these confirms the basic results presented above. Transport costs are always statistically significant in determining the provincial concentration of industrial production, and the sign of this impact is consistent with the model presented above.

[2]We thank Chris Palmer for making the STATA code available to us.

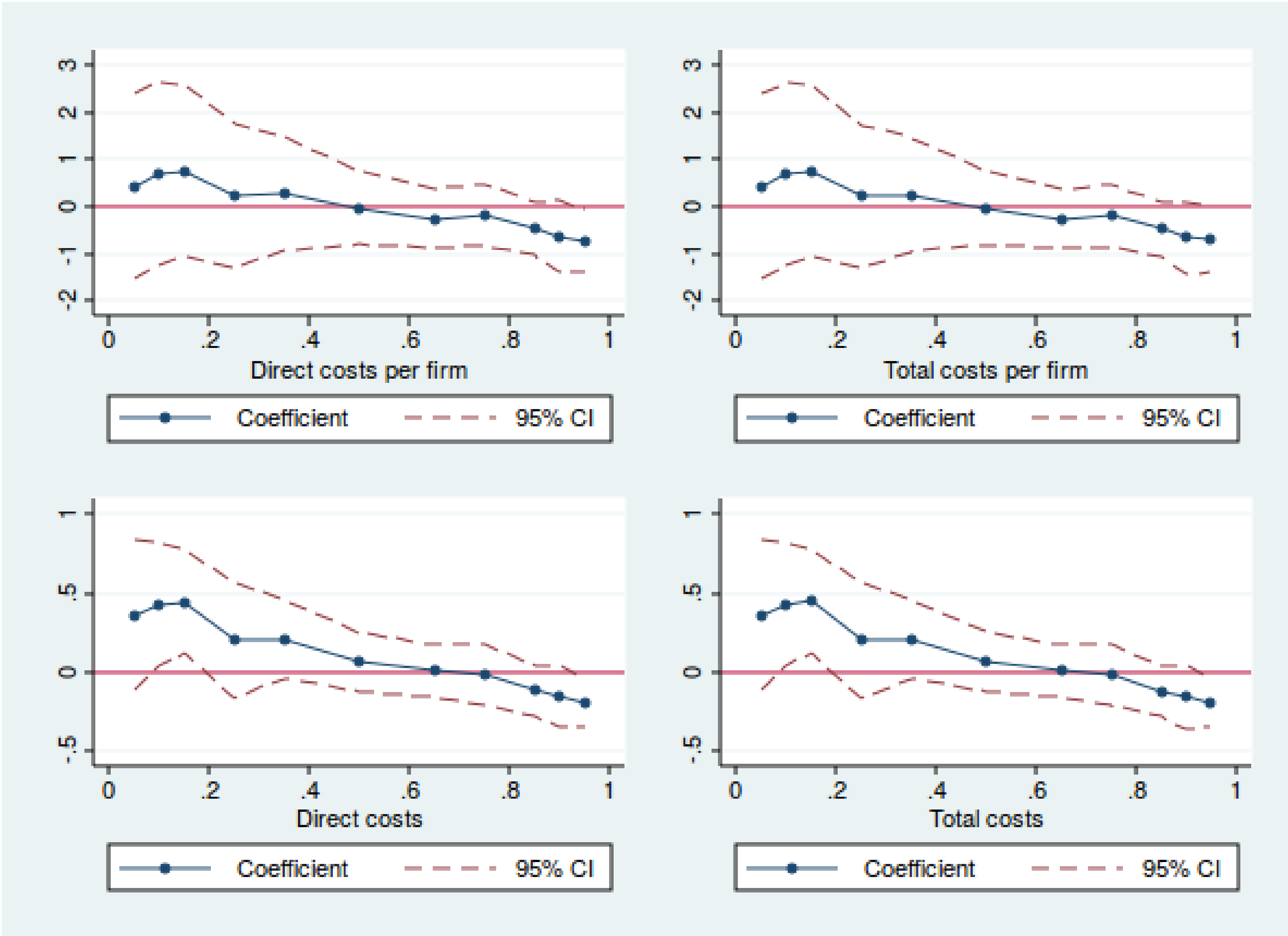

Figure 4: Impacts of transport costs measured using inter-industry flows

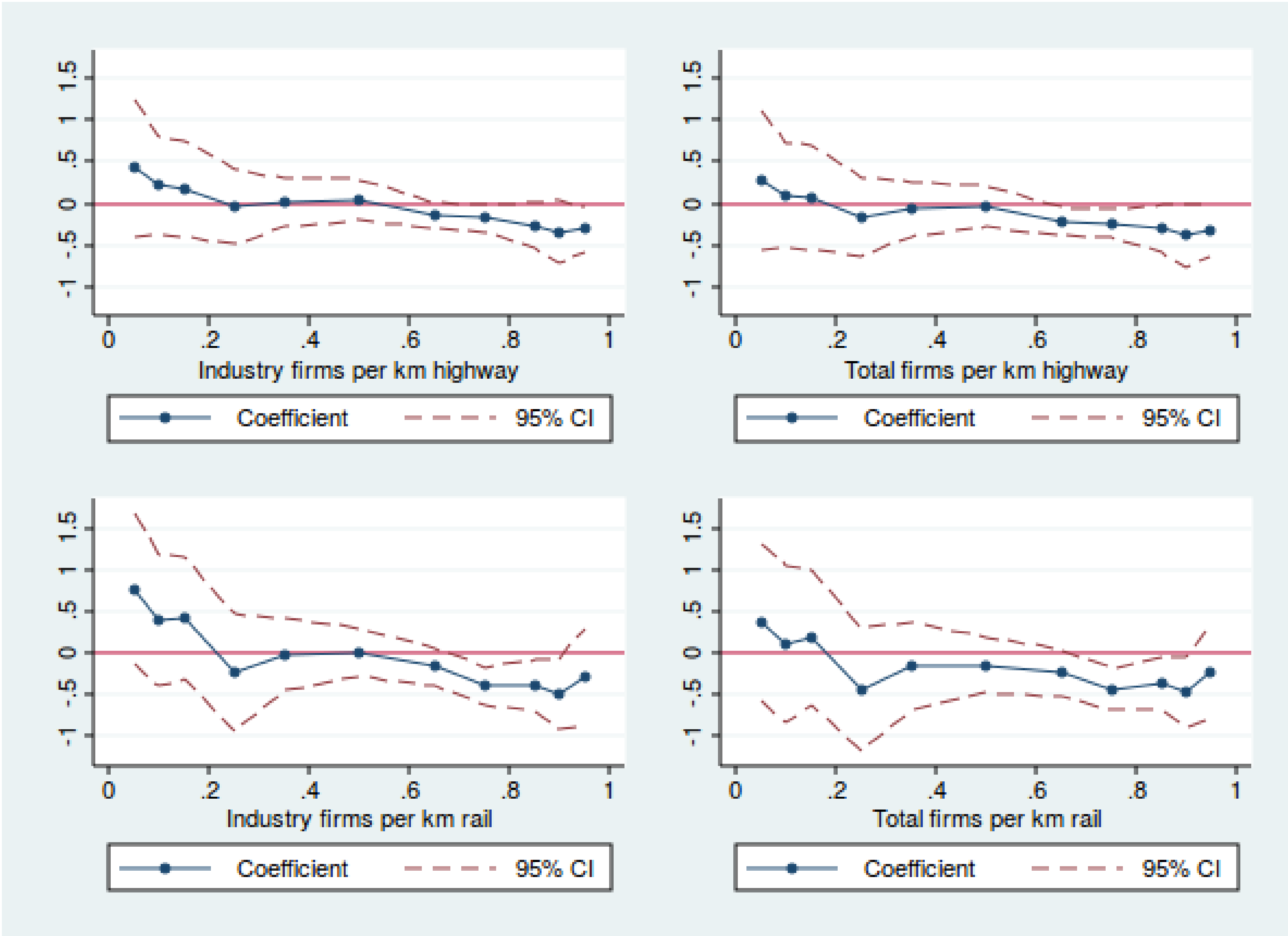

Figure 5: Impacts of transport costs measured using firms per unit infrastructure

7 Conclusion

The rapidly growing Chinese economy presents a variety of challenges in economic study and analysis. Output and the spatial concentration of industry have changed side-by-side along the significant growth in China's industries. While the role of the state-owned and state-controlled sectors has somewhat diminished as a share of total industrial output, it remains large. The central government continues to require residence permits from its citizens in order for them to access many local public services and housing, constraining labour mobility in ways not experienced in most market economies. In the context of China's industrial and manufacturing economy, we ask whether a model of decentralised choices and spatial optimization can provide an explanation and accurate prediction for the changes observed in China over the past decade. From our results, we arrive at an unambiguously clear answer: yes. Analysis using three different empirical approaches and eight different measures of transportation costs all produce results that are consistent with the predictions of our model.

The extensive involvement of state actors in the economy raises the question of how to understand the emerging patterns of industrial location. Are these emerging patterns the product of political decisions, so that some type of political-economy model is required? Further research might reveal that a more complex model of this nature could add to our understanding of the rapidly-evolving Chinese economy, but the analysis above suggests that a simple and more parsimonious model will often be sufficient. Our analysis suggests that a tractable "New Economic Geography" type model in which owners of physical or human capital are mobile to different degrees depending on the type and amount of capital they own, and where industries are monopolistically competitive with profit-maximizing firms, the data and the predictions of the model concerning industrial concentration are largely consistent with each other. The location and concentration of economic activity emerge as an equilibrium outcome from utility maximizing households and firm optimization. The result is a specific prediction of the relationship between transportation costs and the pattern of location quotients, which provide a measure of industrial concentration in the model.

We have examined the correspondence between the predictions of this model and the patterns of industrial concentration and transport costs across Chinese provinces. This required construction of possible measures of transport costs in each province, for which we have no direct measurements, as well as consideration of procedures to adjust for potential endogeneity of transport costs and other co-variates. While criticism can always be raised about the role of particular instruments for endogenous variables or about particular estimation methods, we argue that the results of our various estimates make a persuasive case for the model.

Each of the models we estimate indicates that industrial concentration increases as transport costs fall, and we can easily reject the hypothesis that the true magnitude of this association is zero. This is true for all of the possible measures of transportation costs that we consider and it continues to hold when we employ an instrumental variables approach to adjust for endogeneity. Furthermore, for the estimation approach we present that both adjusts for endogeneity and looks separately at the upper and lower quantiles, the association of these quantiles with transportation costs conforms to the pattern predicted by our simple analytic model.

It is of course possible that government policies are imposing significant costs on the economy, and even interfering in ways that result in significant departures from other aspects of behavior we would expect to see in an economy with unconstrained location choice and optimization. Alternatively, it may be the case that while Chinese public policy is sometimes presented as a type of omnipotent control over many aspects of the economy and decisions of households, this is simply not the case. It could be that the state's policies are only effective over certain things, such as flows of unskilled or semi-skilled labour, or that, beyond the short-run, their effects are short-lived in longer-run equilibrium. The central government's intervention may only be concentrated in specific industries (e.g. weapons and ammunition manufacturing) that are strategically important, but account for a small portion of total production capacity; or, the government only intervenes during crisis periods and intermittently while maintaining observation for the rest of the time. Another possibility is that the effects of central government policies, for the large part, may "cancel" out provincial government policies over time.

Perhaps through one mechanism or another, given China's geographic and provincial diversity capital owners figure out how to deploy their assets in the locations that provide them with the greatest returns, firms seek profits and enter regions and industries until the incentive for them to do so is eliminated, and households purchase bundles of local and non-local goods that maximise their well-being so that the economic environment is basically in accord with that assumed in our simple model. This would be an expected result consistent with China's "Grasping the Large and Releasing the Small" reforms which started in the mid-1990's where the central government scaled back the state sector, privatised state-owned firms across its economy, but retained control over the largest state-owned enterprises. Even then, large state-owned enterprises were corporatised to an extent. From 1998 to 2007, while capital productivity of state-owned firms remained lower than that of the private sector, labour productivity of state-owned firms converged to that of private firms and total factor productivity growth of state-owned firms was faster than that of private firms (Hsieh & Song (2015).) If this is indeed the case, then this new understanding of the nature of China's economy carries new, profound policy and welfare implications for its government and its people.

Whichever of these perspectives is adopted, we can say that the observed patterns of industrial concentration, and in particular the correspondence between transportation costs and this concentration, conforms well with a relatively standard model that assumes mobility of capital and individual optimization. Provisionally, it seems that such models can be usefully applied to help us understand spatial concentration of industry in China.

References

Banerjee, A., Duflo, E. & Qian, N. (2012), On the road: Access to transportation infrastructure and economic growth in China.

Baum-Snow, N. (2007), 'Did highways cause suburbanization?.', *Quarterly Journal of Economics* **122**(2), 775 – 805.

Billings, S. B. & Johnson, E. B. (2012), 'The location quotient as an estimator of industrial concentration', *Regional Science and Urban Economics* **42**, 642–647.

Brakman, S., Garretsen, H. & Zhao, Z. (2014), Spatial concentration of manufacturing firms in China.

Brandt, L., Rawski, T. G. & Sutton, J. (2008), China's industrial development., *in* L. Brandt & T. G. Rawski, eds, 'China's Great Economic Transformation', Cambridge University Press, Cambridge and New York, pp. 569 – 632.

Chan, K. W., Henderson, J. V. & Tsui, K. Y. (2008), Spatial dimensions of Chinese economic development., *in* L. Brandt & T. G. Rawski, eds, 'China's Great Economic Transformation', Cambridge University Press, Cambridge and New York, pp. 776 – 828.

Chetverikov, D., Larsen, B. & Palmer, C. (2016), 'IV quantile regression for group-level treatments, with an application to the distributional effects of trade.', *Econometrica* **84**(2), 809–833.

Duranton, G. & Overman, H. G. (2005), 'Testing for localization using micro-geographic data.', *Review of Economic Studies* **72**(4), 1077 – 1106.

Duranton, G. & Turner, M. A. (2012), 'Urban growth and transportation.', *Review of Economic Studies* **79**(4), 1407 – 1440.

Faber, B. (2014), 'Trade integration, market size, and industrialization: Evidence from China's national trunk highway system.', *Review of Economic Studies* **81**(3), 1046 – 1070.

Fan, C. C. & Scott, A. J. (2003), 'Industrial agglomeration and development: A survey of spatial economic issues in East Asia and a statistical analysis of Chinese regions.', *Economic Geography* **79**(3), 295 – 319.

Florence, S. (1939), *Report on the Location of Industry in Great Britain*, Political and Economic Planning, London.

Ghani, E., Goswami, A. G. & Kerr, W. R. (2016), 'Highway to success: The impact of the golden quadrilateral project for the location and performance of Indian manufacturing', *The Economic Journal* **126**(591), 317–357.

Haggard, S. & Huang, Y. (2008), The political economy of private-sector development in China., *in* L. Brandt & T. G. Rawski, eds, 'China's Great Economic Transformation', Cambridge University Press, Cambridge and New York, pp. 337 – 374.

Hsieh, C.-T. & Song, Z. M. (2015), 'Grasp the large, let go of the small: The transformation of the state sector in China.', *NBER, Working Paper Series* (21006).

Jeon, Y. (2006), 'Manufacturing, increasing returns and economic development in china, 1979-2004: A kaldorian approach', *Working Paper Series, Department of Economics, University of Utah* (2006-08).

Ko, K. W. & Mok, K. W. P. (2014), 'Clustering of cultural industries in Chinese cities: A spatial panel approach.', *Economics of Transition* **22**(2), 365 – 395.

Liu, Z. (2014), 'Human capital externalities in cities: Evidence from Chinese manufacturing firms.', *Journal of Economic Geography* **14**(3), 621 – 649.

Long, C. & Zhang, X. (2012), 'Patterns of China's industrialization: Concentration, specialization, and clustering.', *China Economic Review* **23**(3), 593 – 612.

Meng, G. (2003), *The Theory and Practice of Free Economic Zones: A Case Study of Tianjin, Peoples Republic of China*, Peter Lang - International Academic Publishers, Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien.

National Bureau of Statistics (1999), *1997 Input-Output Table of China*, China Statistics Press, Beijing.

National Bureau of Statistics (2002), *2000 Input-Output Table of China*, China Statistics Press, Beijing.

National Bureau of Statistics (2006), *2002 Input-Output Table of China*, China Statistics Press, Beijing.

National Bureau of Statistics (2009), *2007 Input-Output Table of China*, China Statistics Press, Beijing.

Pfluger, M. (2004), 'A simple, analytically solvable, Chamberlinian agglomeration model.', *Regional Science and Urban Economics* **34**(5), 565 – 573.

Redding, S. J. & Turner, M. A. (2014), Transportation costs and the spatial organization of economic activity.

Van Huffel, C., Luo, X. & Catin, M. (2005), Openness, industrialization, and geographic concentration of activities in China.

Wen, M. (2004), 'Relocation and agglomeration of Chinese industry.', *Journal of Development Economics* **73**(1), 329 – 347.

Xin, G. (1999), Revitalizing Chinese society: Institutional transformation and social change, *in* W. Guangwu & J. Wong, eds, 'China: Two Decades of Reform and Change', Singapore University Press, World Scientific Publishing Company, Inc., Singapore, chapter 3, pp. 67 – 100.

Yang, F. F. & Yeh, A. G. O. (2013), 'Spatial development of producer services in the Chinese urban system.', *Environment and Planning A* **45**(1), 159 – 179.

Young, J. (2013), *China's Hukou System: Markets, Migrants and Institutional Change*, Palgrave Macmillan, Houndmills, Basingstoke.

Zhao, D. (2014), The new economic geography of industrial agglomerations in China. Available at `http://library.williams.edu/theses/pdf.php?id=701`.

Zheng, S., Sun, C., Qi, Y. & Kahn, M. E. (2014), 'The evolving geography of China's industrial production: Implications for pollution dynamics and urban quality of life.', *Journal of Economic Surveys* **28**(4), 709 – 724.

8 Appendix

Material not necessary for publication.

Table 10: Variables used in analysis

How Used	Variable name	Description
Dependent	loglqoutput	Ln(Location Quotient for sector output in province and year)
Measures of transportation cost, one used in each model, estimated as endogenous controls in IV models		
Endogenous	logtranspdircost	Ln(IO for sector and storage x sector output in province during year)
Endogenous	logtranspDCperf	Ln(IO for transport and storage x sector output in province during year / firms in sector)
Endogenous	logtransptotcost	Ln(Cumulative total IO for transport and storage in industries supplying sector x sector output in province during year)
Endogenous	logtranspTCperf	Ln(Cumulative total IO for transport and storage in industries supplying sector x sector output in province during year / firms in sector)
Endogenous	ltfirmsperrail	Ln(Total firms per km of rail in province during year)
Endogenous	ltfirmsperhigh	Ln(Total firms per km of highway in province during year)
Endogenous	lifirmsperrail	Ln(Sector firms per km of rail in province during year)
Endogenous	lifirmsperhigh	Ln(Sector firms per km of highway in province during year)
Variables used in all models, estimated as endogenous controls in IV models		
Endogenous	loginc	Ln(Income per capita)
Endogenous	logpopden	LN(Total population / Area of province)
Endogenous	lstateshareoutput	Ln(Share of province GDP produced in state sector)
Endogenous	logtrade	Ln(Trade dependency index)
Variables used as exogenous controls		
Exogenous control	Industry$_9$	Coal Mining & Dressing
Exogenous control	Industry$_{15}$	Ferrous Metal Mining & Dressing
Exogenous control	Industry$_{23}$	Non-ferrous Metal Ores Mining & Dressing
Exogenous control	Industry$_{24}$	Non-metal Ores Mining & Dressing
Exogenous control	Industry$_2$	Other Mining & Dressing

Continued on next page

Continued from previous page

How Used	Variable name	Description
Exogenous control	Industry_{27}	Petroleum and Natural Gas Extraction
Exogenous controls	Year_{1998} to Year_{2009}	Indicator for year (2010 is excluded category)
		Variables used as instruments for IV models
Instrument	slopesharp	Tan(mean slope) to capture sharpness of incline
Instrument	slopesharprange	tangent of rangeslope (range of slope)
Instrument	varslope	stdevslope/meanslope (coeffcient of variation for slope)
Instrument	varelev	stedevelev/meanelev (coefficient of variation for elevation)
Instrument	geo	Mean slope divided by mean elevation for province
Instrument	meanstdelev	meanelev/stdelev (mean elevation standardised by standard deviation of elevation)
Instrument	meanstdslope	meanslope/stdslope (mean slope standardised by standard deviation of slope)
Instrument	tanmeanstdslope	tangent of meanstdslope
Instrument	medstdelev	median of elevation/stdelev
Instrument	slopetoelev	meanstdslope/meanstdelev
Instrument	numparks	number of parks (probably designated by province) annual data
Instrument	pubgreenarea	public green areas (km squared) public rights of access
Instrument	areacovdisas	area that the disaster covers/hits (km squared)
	Industrial sectors used for data calculation, not used as controls	
Sector	Industry_{1}	Nonmetal Mineral Products
Sector	Industry_{3}	Transport of Timber and Bamboo
Sector	Industry_{4}	Weapons and Ammunition Manufacturing
Sector	Industry_{5}	Agricultural and Sideline Foods Processing
Sector	Industry_{6}	Beverage Production
Sector	Industry_{7}	Chemical Fiber
Sector	Industry_{8}	Clothes Shoes and Hat Manufacture
Sector	Industry_{10}	Communications Equipment Computer and Other Electronic Equipment Manufacturing

Continued on next page

Continued from previous page

How Used	**Variable name**	**Description**
Sector	Industry_{11}	Craftwork and Other Manufactures
Sector	Industry_{12}	Cultural Educational and Sports Articles Production
Sector	Industry_{13}	Electric Machines and Apparatuses Manufacturing
Sector	Industry_{14}	Electricity and Heating Production and Supply
Sector	Industry_{16}	Food Production
Sector	Industry_{17}	Fuel Gas Production and Supply
Sector	Industry_{18}	Furniture Manufacturing
Sector	Industry_{19}	Instruments Meters Cultural and Office Machinery Manufacture
Sector	Industry_{20}	Leather Furs Down and Related Products
Sector	Industry_{21}	Medical and Pharmaceutical Products
Sector	Industry_{22}	Metal Products
Sector	Industry_{25}	Ordinary Machinery Manufacturing
Sector	Industry_{26}	Papermaking and Paper Products
Sector	Industry_{27}	Petroleum Processing Coking and Nuclear Fuel Processing
Sector	Industry_{29}	Plastic Products
Sector	Industry_{30}	Printing and Record Medium Reproduction
Sector	Industry_{31}	Raw Chemical Material & Chemical Products
Sector	Industry_{32}	Rubber Products
Sector	Industry_{33}	Smelting & Pressing of Ferrous Metals
Sector	Industry_{34}	Smelting & Pressing of Non-ferrous Metals
Sector	Industry_{35}	Special Equipment Manufacturing
Sector	Industry_{36}	Textile Industry
Sector	Industry_{37}	Timber Processing Bamboo Cane Palm Fiber and Straw Products
Sector	Industry_{38}	Tobacco Products Processing
Sector	Industry_{39}	Transport Equipment Manufacturing

Continued on next page

Continued from previous page

How Used	**Variable name**	**Description**
Sector	Industry_{40}	Waste Resources and Old Material Recycling and Processing
Sector	Industry_{41}	Water Production and Supply

Table 11: First stage estimates of endogenous transport costs

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
TanSlope	-0.0004	0.0009**	-0.0004	0.0009**	-0.0009**	-0.0006	-0.0006***	-0.0003**
σ	*0.0005*	*0.0004*	*0.0005*	*0.0004*	*0.0004*	*0.0004*	*0.0002*	*0.0002*
TanSlRange	0.0248***	0.0036***	0.0248***	0.0036***	0.0157***	0.0270***	0.0158***	0.0270***
σ	*0.0018*	*0.0013*	*0.0017*	*0.0012*	*0.0014*	*0.0014*	*0.0005*	*0.0006*
CVSlope	-0.0163	-0.0817	-0.0087	-0.0743	-2.8150***	1.1685***	-2.3903***	1.6235***
σ	*0.2769*	*0.2021*	*0.2734*	*0.1928*	*0.2201*	*0.2213*	*0.0854*	*0.0906*
CVElev	-0.3123*	0.1053	-0.3230*	0.0941	1.3086***	-1.3647***	1.2734***	-1.4204***
σ	*0.1862*	*0.1378*	*0.1839*	*0.1315*	*0.1528*	*0.1537*	*0.0590*	*0.0625*
$\left(\frac{\text{Slope}}{\text{Elev}}\right)$	14.9700***	4.8756***	14.9706***	4.8750***	23.9522***	25.0220***	26.0294***	27.0912***
σ	*0.8545*	*0.6209*	*0.8438*	*0.5922*	*0.6964*	*0.7003*	*0.2719*	*0.2884*
$\left(\frac{\text{Elev}}{\text{SDElev}}\right)$	-3.6134***	-1.0758***	-3.6146***	-1.0766***	-3.1077***	-1.7469***	-2.7222***	-1.3513***
σ	*0.1964*	*0.1454*	*0.1940*	*0.1387*	*0.1608*	*0.1617*	*0.0627*	*0.0665*
$\left(\frac{\text{Slope}}{\text{SDSlope}}\right)$	-0.2412	-0.0540	-0.2476	-0.0612	-2.3236***	-0.1479	-1.6880***	0.5131***
σ	*0.1746*	*0.1273*	*0.1724*	*0.1214*	*0.1391*	*0.1399*	*0.0539*	*0.0572*
$\left(\frac{\text{TanSlope}}{\text{SDSlope}}\right)$	-0.0256***	-0.0048**	-0.0256***	-0.0048**	-0.0261***	-0.0124***	-0.0230***	-0.0092***
σ	*0.0025*	*0.0020*	*0.0025*	*0.0019*	*0.0022*	*0.0022*	*0.0009*	*0.0009*
$\left(\frac{\text{MedElev}}{\text{SDElev}}\right)$	2.5719***	0.7031***	2.5747***	0.7055***	2.2819***	1.2150***	1.8872***	0.8117***
σ	*0.1642*	*0.1218*	*0.1622*	*0.1162*	*0.1349*	*0.1356*	*0.0525*	*0.0557*
$\left(\frac{\text{CVElev}}{\text{CVSlope}}\right)$	-0.8413***	-0.8153***	-0.8276***	-0.8007***	-2.5230***	1.1798***	-2.7834***	0.9296***
σ	*0.2963*	*0.2163*	*0.2926*	*0.2063*	*0.2400*	*0.2413*	*0.0921*	*0.0978*

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
NumParks	0.000350***	0.000232**	0.000353***	0.000235***	0.000221**	0.000591***	0.000413***	0.000781***
σ	*0.000126*	*0.000091*	*0.000124*	*0.000087*	*0.000102*	*0.000102*	*0.000039*	*0.000042*
GreenArea	0.000066***	0.000006	0.000066***	0.000006	0.000031***	0.000021***	0.000030***	0.000021***
σ	*0.000005*	*0.000004*	*0.000005*	*0.000003*	*0.000004*	*0.000004*	*0.000002*	*0.000002*
DisastArea	0.000217***	0.000042***	0.000216***	0.000041***	-0.000072***	-0.000022	-0.000085***	-0.000034***
σ	*0.000017*	*0.000012*	*0.000017*	*0.000012*	*0.000014*	*0.000014*	*0.000005*	*0.000006*
Other mine	-4.3637***	-0.4299***	-4.9665***	-1.0353***	-4.1593***	-4.2442***	0.0851**	0.0494
σ	*0.1907*	*0.1412*	*0.1883*	*0.1346*	*0.1435*	*0.1443*	*0.0360*	*0.0382*
Coal mine	0.6126***	0.7154***	0.1494	0.2555***	-0.0501	-0.0377	0.0015	0.0016
σ	*0.0992*	*0.0726*	*0.0979*	*0.0693*	*0.0810*	*0.0815*	*0.0313*	*0.0332*
Ferrous mine	-0.8811***	0.4240***	-1.2882***	0.0186	-1.2074***	-1.2062***	-0.0005	-0.0043
σ	*0.0982*	*0.0721*	*0.0970*	*0.0687*	*0.0804*	*0.0809*	*0.0318*	*0.0337*
Non-ferr mine	-0.3239***	0.5765***	-0.9262***	-0.0249	-0.8396***	-0.8178***	-0.0005	-0.0043
σ	*0.0992*	*0.0725*	*0.0979*	*0.0692*	*0.0809*	*0.0813*	*0.0318*	*0.0337*
Non-met mine	-0.3575***	0.1602**	-0.9597***	-0.4405***	-0.4503***	-0.4522***	0.0015	0.0016
σ	*0.0960*	*0.0702*	*0.0947*	*0.0670*	*0.0785*	*0.0789*	*0.0313*	*0.0332*
Petrol & gas	-0.2469**	2.9134***	-0.4210***	2.7402***	-3.1008***	-3.0608***	0.0507	0.0673*
σ	*0.1162*	*0.0848*	*0.1148*	*0.0809*	*0.0935*	*0.0941*	*0.0340*	*0.0360*
Y_{1998}	-1.7820***	-1.9892***	-2.0036***	-2.2102***	1.0403***	0.1812**	0.9914***	0.1349***
σ	*0.0857*	*0.0637*	*0.0847*	*0.0608*	*0.0710*	*0.0714*	*0.0275*	*0.0292*
Y_{1999}	-1.7427***	-1.9373***	-1.9647***	-2.1585***	0.9619***	0.1613**	0.9026***	0.1054***
σ	*0.0848*	*0.0630*	*0.0837*	*0.0601*	*0.0702*	*0.0706*	*0.0272*	*0.0289*
Y_{2000}	-1.3097***	-1.3956***	-1.5213***	-1.6071***	0.8473***	0.1525**	0.7850***	0.0917***
σ	*0.0846*	*0.0629*	*0.0835*	*0.0599*	*0.0702*	*0.0706*	*0.0271*	*0.0288*

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
Y_{2001}	-1.2839***	-1.2748***	-1.4970***	-1.4881***	0.6514***	0.0565	0.5502***	-0.0453
σ	*0.0844*	*0.0627*	*0.0833*	*0.0598*	*0.0701*	*0.0705*	*0.0270*	*0.0287*
Y_{2002}	-0.9097***	-0.8746***	-1.1388***	-1.1036***	0.5892***	0.0451	0.5006***	-0.0428
σ	*0.0826*	*0.0614*	*0.0816*	*0.0586*	*0.0687*	*0.0691*	*0.0265*	*0.0281*
Y_{2003}	-0.8865***	-0.7496***	-1.1152***	-0.9780***	0.5392***	-0.0124	0.4912***	-0.0571**
σ	*0.0823*	*0.0613*	*0.0813*	*0.0585*	*0.0682*	*0.0686*	*0.0263*	*0.0279*
Y_{2004}	-0.6047***	-0.6397***	-0.8343***	-0.8691***	0.5537***	-0.0031	0.4954***	-0.0584**
σ	*0.0811*	*0.0604*	*0.0801*	*0.0576*	*0.0669*	*0.0672*	*0.0258*	*0.0274*
Y_{2005}	-0.4568***	-0.4487***	-0.5356***	-0.5274***	0.4956***	0.0918	0.4657***	0.0645**
σ	*0.0820*	*0.0610*	*0.0810*	*0.0582*	*0.0661*	*0.0665*	*0.0256*	*0.0271*
Y_{2006}	-0.2917***	-0.2931***	-0.3709***	-0.3723***	0.0260	0.1284*	0.0253	0.1254***
σ	*0.0815*	*0.0606*	*0.0805*	*0.0578*	*0.0661*	*0.0665*	*0.0254*	*0.0269*
Y_{2007}	-0.5801***	-0.4730***	-0.6319***	-0.5247***	0.0087	0.0524	0.0128	0.0537**
σ	*0.0811*	*0.0602*	*0.0801*	*0.0574*	*0.0660*	*0.0664*	*0.0254*	*0.0269*
Y_{2008}	-0.3661***	-0.3873***	-0.4175***	-0.4384***	0.0497	0.1110*	0.0459*	0.1068***
σ	*0.0823*	*0.0617*	*0.0813*	*0.0589*	*0.0667*	*0.0671*	*0.0256*	*0.0272*
Y_{2009}	-0.3389***	-0.3595***	-0.3899***	-0.4104***	0.0722	0.0960	0.0698***	0.0937***
σ	*0.0794*	*0.0601*	*0.0784*	*0.0573*	*0.0659*	*0.0663*	*0.0253*	*0.0269*
Constant	12.2809***	6.8866***	13.6583***	8.2642***	0.3383	-3.7468***	3.5248***	-0.6211***
σ	*0.5556*	*0.4051*	*0.5486*	*0.3864*	*0.4404*	*0.4429*	*0.1707*	*0.1811*
Obs	12055	11860	12055	11860	12679	12679	13625	13625
F	270.92	166.61	306.58	202.5	250.26	271.16	1302.71	1372.72
$\overline{R^2}$	0.4097	0.3021	0.4401	0.345	0.3787	0.3978	0.7476	0.7574
$\sqrt{MSE}$	1.7036	1.2375	1.6822	1.1803	1.4168	1.4248	0.56923	0.6039

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Table 12: First stage estimates of endogenous income

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
TanSlope	-0.00008**	-0.00008**	-0.00008**	-0.00008**	-0.00008**	-0.00008**	-0.00008**	-0.00008**
σ	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004
TanSlRange	-0.0015***	-0.0015***	-0.0015***	-0.0015***	-0.0015***	-0.0015***	-0.0015***	-0.0015***
σ	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
CVSlope	-0.0482**	-0.0482**	-0.0482**	-0.0482**	-0.0482**	-0.0482**	-0.0503***	-0.0503***
σ	0.0197	0.0197	0.0197	0.0197	0.0197	0.0197	0.0196	0.0196
CVElev	-0.0389***	-0.0389***	-0.0389***	-0.0389***	-0.0389***	-0.0389***	-0.0373***	-0.0373***
σ	0.0134	0.0134	0.0134	0.0134	0.0134	0.0134	0.0134	0.0134
$\left(\frac{\text{Slope}}{\text{Elev}}\right)$	4.9549***	4.9549***	4.9549***	4.9549***	4.9549***	4.9549***	4.9535***	4.9535***
σ	0.0631	0.0631	0.0631	0.0631	0.0631	0.0631	0.0627	0.0627
$\left(\frac{\text{Elev}}{\text{SDElev}}\right)$	-0.1493***	-0.1493***	-0.1493***	-0.1493***	-0.1493***	-0.1493***	-0.1497***	-0.1497***
σ	0.0143	0.0143	0.0143	0.0143	0.0143	0.0143	0.0142	0.0142
$\left(\frac{\text{Slope}}{\text{SDSlope}}\right)$	-0.0081	-0.0081	-0.0081	-0.0081	-0.0081	-0.0081	-0.0088	-0.0088
σ	0.0124	0.0124	0.0124	0.0124	0.0124	0.0124	0.0123	0.0123
$\left(\frac{\text{TanSlope}}{\text{SDSlope}}\right)$	-0.0019***	-0.0019***	-0.0019***	-0.0019***	-0.0019***	-0.0019***	-0.0019***	-0.0019***
σ	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
$\left(\frac{\text{MedElev}}{\text{SDElev}}\right)$	0.1036***	0.1036***	0.1036***	0.1036***	0.1036***	0.1036***	0.1038***	0.1038***
σ	0.0119	0.0119	0.0119	0.0119	0.0119	0.0119	0.0119	0.0119
$\left(\frac{\text{CVElev}}{\text{CVSlope}}\right)$	0.1351***	0.1351***	0.1351***	0.1351***	0.1351***	0.1351***	0.1326***	0.1326***

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
σ	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0212	0.0212
NumParks	0.000110***	0.000110***	0.000110***	0.000110***	0.000110***	0.000110***	0.000108***	0.000108***
σ	0.000009	0.000009	0.000009	0.000009	0.000009	0.000009	0.000009	0.000009
GreenArea	0.0000056***	0.0000056***	0.0000056***	0.0000056***	0.0000056***	0.0000056***	0.0000056***	0.0000056***
σ	0.0000004	0.0000004	0.0000004	0.0000004	0.0000004	0.0000004	0.0000004	0.0000004
DisastArea	-0.000059***	-0.000059***	-0.000059***	-0.000059***	-0.000059***	-0.000059***	-0.000059***	-0.000059***
σ	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001
Other mine	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0077	0.0077
σ	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083
Coal mine	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
σ	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072
Ferrous mine	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005
σ	0.0073	0.0073	0.0073	0.0073	0.0073	0.0073	0.0073	0.0073
Non-ferr mine	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005
σ	0.0073	0.0073	0.0073	0.0073	0.0073	0.0073	0.0073	0.0073
Non-met mine	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
σ	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072
Petrol & gas	0.0230***	0.0230***	0.0230***	0.0230***	0.0230***	0.0230***	0.0231***	0.0231***
σ	0.0078	0.0078	0.0078	0.0078	0.0078	0.0078	0.0078	0.0078
Y_{1998}	-0.8887***	-0.8887***	-0.8887***	-0.8887***	-0.8887***	-0.8887***	-0.8887***	-0.8887***
σ	0.0063	0.0063	0.0063	0.0063	0.0063	0.0063	0.0062	0.0062
Y_{1999}	-0.8084***	-0.8084***	-0.8084***	-0.8084***	-0.8084***	-0.8084***	-0.8084***	-0.8084***
σ	0.0062	0.0062	0.0062	0.0062	0.0062	0.0062	0.0061	0.0061
Y_{2000}	-0.7338***	-0.7338***	-0.7338***	-0.7338***	-0.7338***	-0.7338***	-0.7337***	-0.7337***
σ	0.0062	0.0062	0.0062	0.0062	0.0062	0.0062	0.0061	0.0061

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
Y_{2001}	-0.6646***	-0.6646***	-0.6646***	-0.6646***	-0.6646***	-0.6646***	-0.6646***	-0.6646***
σ	0.0061	0.0061	0.0061	0.0061	0.0061	0.0061	0.0061	0.0061
Y_{2002}	-0.5723***	-0.5723***	-0.5723***	-0.5723***	-0.5723***	-0.5723***	-0.5723***	-0.5723***
σ	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060
Y_{2003}	-0.4867***	-0.4867***	-0.4867***	-0.4867***	-0.4867***	-0.4867***	-0.4867***	-0.4867***
σ	0.0060	0.0060	0.0060	0.0060	0.0060	0.0060	0.0059	0.0059
Y_{2004}	-0.4690***	-0.4690***	-0.4690***	-0.4690***	-0.4690***	-0.4690***	-0.4691***	-0.4691***
σ	0.0059	0.0059	0.0059	0.0059	0.0059	0.0059	0.0058	0.0058
Y_{2005}	-0.3835***	-0.3835***	-0.3835***	-0.3835***	-0.3835***	-0.3835***	-0.3835***	-0.3835***
σ	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058
Y_{2006}	-0.2889***	-0.2889***	-0.2889***	-0.2889***	-0.2889***	-0.2889***	-0.2890***	-0.2890***
σ	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0057	0.0057
Y_{2007}	-0.1852***	-0.1852***	-0.1852***	-0.1852***	-0.1852***	-0.1852***	-0.1851***	-0.1851***
σ	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0057	0.0057
Y_{2008}	-0.1435***	-0.1435***	-0.1435***	-0.1435***	-0.1435***	-0.1435***	-0.1435***	-0.1435***
σ	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0057	0.0057
Y_{2009}	-0.0413***	-0.0413***	-0.0413***	-0.0413***	-0.0413***	-0.0413***	-0.0413***	-0.0413***
σ	0.0057	0.0057	0.0057	0.0057	0.0057	0.0057	0.0056	0.0056
Constant	9.4955***	9.4955***	9.4955***	9.4955***	9.4955***	9.4955***	9.5001***	9.5001***
σ	0.0394	0.0394	0.0394	0.0394	0.0394	0.0394	0.0392	0.0392
Obs	13739	13739	13739	13739	13739	13739	13871	13871
F	3555.39	3555.39	3555.39	3555.39	3555.39	3555.39	3598.31	3598.31
$\overline{R^2}$	0.8891	0.8891	0.8891	0.8891	0.8891	0.8891	0.8894	0.8894
$\sqrt{MSE}$	0.13144	0.13144	0.13144	0.13144	0.13144	0.13144	0.13145	0.13145

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Table 13: First stage estimates of endogenous population density

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
TanSlope	0.0052***	0.0052***	0.0052***	0.0052***	0.0052***	0.0052***	0.0052***	0.0052***
σ	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
TanSlRange	0.0193***	0.0193***	0.0193***	0.0193***	0.0193***	0.0193***	0.0193***	0.0193***
σ	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
CVSlope	-2.2008***	-2.2008***	-2.2008***	-2.2008***	-2.2008***	-2.2008***	-2.2040***	-2.2040***
σ	0.0554	0.0554	0.0554	0.0554	0.0554	0.0554	0.0551	0.0551
CVElev	2.1200***	2.1200***	2.1200***	2.1200***	2.1200***	2.1200***	2.1220***	2.1220***
σ	0.0378	0.0378	0.0378	0.0378	0.0378	0.0378	0.0376	0.0376
$\left(\frac{\text{Slope}}{\text{Elev}}\right)$	15.8022***	15.8022***	15.8022***	15.8022***	15.8022***	15.8022***	15.7922***	15.7922***
σ	0.1776	0.1776	0.1776	0.1776	0.1776	0.1776	0.1767	0.1767
$\left(\frac{\text{Elev}}{\text{SDElev}}\right)$	-0.6168***	-0.6168***	-0.6168***	-0.6168***	-0.6168***	-0.6168***	-0.6169***	-0.6169***
σ	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402	0.0400	0.0400
$\left(\frac{\text{Slope}}{\text{SDSlope}}\right)$	0.7040***	0.7040***	0.7040***	0.7040***	0.7040***	0.7040***	0.7031***	0.7031***
σ	0.0348	0.0348	0.0348	0.0348	0.0348	0.0348	0.0346	0.0346
$\left(\frac{\text{TanSlope}}{\text{SDSlope}}\right)$	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002
σ	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
$\left(\frac{\text{MedElev}}{\text{SDElev}}\right)$	-0.5601***	-0.5601***	-0.5601***	-0.5601***	-0.5601***	-0.5601***	-0.5594***	-0.5594***
σ	0.0336	0.0336	0.0336	0.0336	0.0336	0.0336	0.0334	0.0334
$\left(\frac{\text{CVElev}}{\text{CVSlope}}\right)$	-4.7487***	-4.7487***	-4.7487***	-4.7487***	-4.7487***	-4.7487***	-4.7473***	-4.7473***
σ	0.0599	0.0599	0.0599	0.0599	0.0599	0.0599	0.0596	0.0596

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
NumParks	0.000067***	0.000067***	0.000067***	0.000067***	0.000067***	0.000067***	0.000060**	0.000060**
σ	0.000026	0.000026	0.000026	0.000026	0.000026	0.000026	0.000025	0.000025
GreenArea	0.000016***	0.000016***	0.000016***	0.000016***	0.000016***	0.000016***	0.000016***	0.000016***
σ	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001
DisastArea	-0.000033***	-0.000033***	-0.000033***	-0.000033***	-0.000033***	-0.000033***	-0.000033***	-0.000033***
σ	0.000004	0.000004	0.000004	0.000004	0.000004	0.000004	0.000003	0.000003
Other mine	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007	0.0008	0.0008
σ	0.0232	0.0232	0.0232	0.0232	0.0232	0.0232	0.0232	0.0232
Coal mine	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
σ	0.0202	0.0202	0.0202	0.0202	0.0202	0.0202	0.0202	0.0202
Ferrous mine	-0.0022	-0.0022	-0.0022	-0.0022	-0.0022	-0.0022	-0.0022	-0.0022
σ	0.0205	0.0205	0.0205	0.0205	0.0205	0.0205	0.0205	0.0205
Non-ferr mine	-0.0022	-0.0022	-0.0022	-0.0022	-0.0022	-0.0022	-0.0022	-0.0022
σ	0.0205	0.0205	0.0205	0.0205	0.0205	0.0205	0.0205	0.0205
Non-met mine	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
σ	0.0202	0.0202	0.0202	0.0202	0.0202	0.0202	0.0202	0.0202
Petrol & gas	0.0120	0.0120	0.0120	0.0120	0.0120	0.0120	0.0120	0.0120
σ	0.0219	0.0219	0.0219	0.0219	0.0219	0.0219	0.0219	0.0219
Y_{1998}	0.1148***	0.1148***	0.1148***	0.1148***	0.1148***	0.1148***	0.1159***	0.1159***
σ	0.0176	0.0176	0.0176	0.0176	0.0176	0.0176	0.0175	0.0175
Y_{1999}	0.1200***	0.1200***	0.1200***	0.1200***	0.1200***	0.1200***	0.1210***	0.1210***
σ	0.0174	0.0174	0.0174	0.0174	0.0174	0.0174	0.0173	0.0173
Y_{2000}	0.1289***	0.1289***	0.1289***	0.1289***	0.1289***	0.1289***	0.1300***	0.1300***
σ	0.0174	0.0174	0.0174	0.0174	0.0174	0.0174	0.0173	0.0173
Y_{2001}	0.1144***	0.1144***	0.1144***	0.1144***	0.1144***	0.1144***	0.1154***	0.1154***

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
σ	0.0173	0.0173	0.0173	0.0173	0.0173	0.0173	0.0172	0.0172
Y_{2002}	0.1050***	0.1050***	0.1050***	0.1050***	0.1050***	0.1050***	0.1058***	0.1058***
σ	0.0169	0.0169	0.0169	0.0169	0.0169	0.0169	0.0168	0.0168
Y_{2003}	0.1006***	0.1006***	0.1006***	0.1006***	0.1006***	0.1006***	0.1015***	0.1015***
σ	0.0168	0.0168	0.0168	0.0168	0.0168	0.0168	0.0167	0.0167
Y_{2004}	0.0663***	0.0663***	0.0663***	0.0663***	0.0663***	0.0663***	0.0668***	0.0668***
σ	0.0165	0.0165	0.0165	0.0165	0.0165	0.0165	0.0164	0.0164
Y_{2005}	0.0560***	0.0560***	0.0560***	0.0560***	0.0560***	0.0560***	0.0565***	0.0565***
σ	0.0164	0.0164	0.0164	0.0164	0.0164	0.0164	0.0163	0.0163
Y_{2006}	0.0479***	0.0479***	0.0479***	0.0479***	0.0479***	0.0479***	0.0482***	0.0482***
σ	0.0164	0.0164	0.0164	0.0164	0.0164	0.0164	0.0162	0.0162
Y_{2007}	0.0509***	0.0509***	0.0509***	0.0509***	0.0509***	0.0509***	0.0514***	0.0514***
σ	0.0164	0.0164	0.0164	0.0164	0.0164	0.0164	0.0162	0.0162
Y_{2008}	0.0288*	0.0288*	0.0288*	0.0288*	0.0288*	0.0288*	0.0291*	0.0291*
σ	0.0163	0.0163	0.0163	0.0163	0.0163	0.0163	0.0161	0.0161
Y_{2009}	0.0219	0.0219	0.0219	0.0219	0.0219	0.0219	0.0221	0.0221
σ	0.0160	0.0160	0.0160	0.0160	0.0160	0.0160	0.0158	0.0158
Constant	1.1502***	1.1502***	1.1502***	1.1502***	1.1502***	1.1502***	1.1508***	1.1508***
σ	0.1110	0.1110	0.1110	0.1110	0.1110	0.1110	0.1105	0.1105
Obs	13739	13739	13739	13739	13739	13739	13871	13871
F	2865.17	2865.17	2865.17	2865.17	2865.17	2865.17	2893.67	2893.67
$\overline{R^2}$	0.866	0.866	0.866	0.866	0.866	0.866	0.866	0.866
$\sqrt{MSE}$	0.37009	0.37009	0.37009	0.37009	0.37009	0.37009	0.37007	0.37007

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Table 14: First stage estimates of endogenous state share of the economy

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
TanSlope	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00001	0.00001
σ	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
TanSlRange	-0.0065***	-0.0065***	-0.0065***	-0.0065***	-0.0065***	-0.0065***	-0.0065***	-0.0065***
σ	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
CVSlope	-2.1557***	-2.1557***	-2.1557***	-2.1557***	-2.1557***	-2.1557***	-2.1622***	-2.1622***
σ	0.0345	0.0345	0.0345	0.0345	0.0345	0.0345	0.0343	0.0343
CVElev	1.6189***	1.6189***	1.6189***	1.6189***	1.6189***	1.6189***	1.6256***	1.6256***
σ	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235	0.0234	0.0234
$\left(\frac{\text{Slope}}{\text{Elev}}\right)$	-4.7913***	-4.7913***	-4.7913***	-4.7913***	-4.7913***	-4.7913***	-4.7863***	-4.7863***
σ	0.1105	0.1105	0.1105	0.1105	0.1105	0.1105	0.1099	0.1099
$\left(\frac{\text{Elev}}{\text{SDElev}}\right)$	0.7076***	0.7076***	0.7076***	0.7076***	0.7076***	0.7076***	0.7118***	0.7118***
σ	0.0250	0.0250	0.0250	0.0250	0.0250	0.0250	0.0249	0.0249
$\left(\frac{\text{Slope}}{\text{SDSlope}}\right)$	-0.4513***	-0.4513***	-0.4513***	-0.4513***	-0.4513***	-0.4513***	-0.4510***	-0.4510***
σ	0.0217	0.0217	0.0217	0.0217	0.0217	0.0217	0.0216	0.0216
$\left(\frac{\text{TanSlope}}{\text{SDSlope}}\right)$	0.0114***	0.0114***	0.0114***	0.0114***	0.0114***	0.0114***	0.0114***	0.0114***
σ	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
$\left(\frac{\text{MedElev}}{\text{SDElev}}\right)$	-0.6833***	-0.6833***	-0.6833***	-0.6833***	-0.6833***	-0.6833***	-0.6868***	-0.6868***
σ	0.0209	0.0209	0.0209	0.0209	0.0209	0.0209	0.0208	0.0208
$\left(\frac{\text{CVElev}}{\text{CVSlope}}\right)$	-2.4245***	-2.4245***	-2.4245***	-2.4245***	-2.4245***	-2.4245***	-2.4326***	-2.4326***
σ	0.0372	0.0372	0.0372	0.0372	0.0372	0.0372	0.0371	0.0371

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
NumParks	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***
σ	0.000016	0.000016	0.000016	0.000016	0.000016	0.000016	0.000016	0.000016
GreenArea	-0.000001*	-0.000001*	-0.000001*	-0.000001*	-0.000001*	-0.000001*	-0.000001*	-0.000001*
σ	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001	0.000001
DisastArea	0.000016***	0.000016***	0.000016***	0.000016***	0.000016***	0.000016***	0.000016***	0.000016***
σ	0.000002	0.000002	0.000002	0.000002	0.000002	0.000002	0.000002	0.000002
Other mine	-0.0122	-0.0122	-0.0122	-0.0122	-0.0122	-0.0122	-0.0122	-0.0122
σ	0.0145	0.0145	0.0145	0.0145	0.0145	0.0145	0.0145	0.0145
Coal mine	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005
σ	0.0126	0.0126	0.0126	0.0126	0.0126	0.0126	0.0126	0.0126
Ferrous mine	-0.0021	-0.0021	-0.0021	-0.0021	-0.0021	-0.0021	-0.0021	-0.0021
σ	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127
Non-ferr mine	-0.0021	-0.0021	-0.0021	-0.0021	-0.0021	-0.0021	-0.0021	-0.0021
σ	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127
Non-met mine	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005
σ	0.0126	0.0126	0.0126	0.0126	0.0126	0.0126	0.0126	0.0126
Petrol & gas	-0.0280**	-0.0280**	-0.0280**	-0.0280**	-0.0280**	-0.0280**	-0.0281**	-0.0281**
σ	0.0136	0.0136	0.0136	0.0136	0.0136	0.0136	0.0136	0.0136
Y_{1998}	0.4535***	0.4535***	0.4535***	0.4535***	0.4535***	0.4535***	0.4538***	0.4538***
σ	0.0110	0.0110	0.0110	0.0110	0.0110	0.0110	0.0109	0.0109
Y_{1999}	0.4453***	0.4453***	0.4453***	0.4453***	0.4453***	0.4453***	0.4455***	0.4455***
σ	0.0108	0.0108	0.0108	0.0108	0.0108	0.0108	0.0108	0.0108
Y_{2000}	0.4189***	0.4189***	0.4189***	0.4189***	0.4189***	0.4189***	0.4191***	0.4191***
σ	0.0108	0.0108	0.0108	0.0108	0.0108	0.0108	0.0107	0.0107
Y_{2001}	0.3800***	0.3800***	0.3800***	0.3800***	0.3800***	0.3800***	0.3803***	0.3803***

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
σ	0.0108	0.0108	0.0108	0.0108	0.0108	0.0108	0.0107	0.0107
Y_{2002}	0.3269***	0.3269***	0.3269***	0.3269***	0.3269***	0.3269***	0.3271***	0.3271***
σ	0.0105	0.0105	0.0105	0.0105	0.0105	0.0105	0.0105	0.0105
Y_{2003}	0.2774***	0.2774***	0.2774***	0.2774***	0.2774***	0.2774***	0.2775***	0.2775***
σ	0.0105	0.0105	0.0105	0.0105	0.0105	0.0105	0.0104	0.0104
Y_{2004}	0.1251***	0.1251***	0.1251***	0.1251***	0.1251***	0.1251***	0.1253***	0.1253***
σ	0.0103	0.0103	0.0103	0.0103	0.0103	0.0103	0.0102	0.0102
Y_{2005}	0.2102***	0.2102***	0.2102***	0.2102***	0.2102***	0.2102***	0.2103***	0.2103***
σ	0.0102	0.0102	0.0102	0.0102	0.0102	0.0102	0.0101	0.0101
Y_{2006}	0.1462***	0.1462***	0.1462***	0.1462***	0.1462***	0.1462***	0.1464***	0.1464***
σ	0.0102	0.0102	0.0102	0.0102	0.0102	0.0102	0.0101	0.0101
Y_{2007}	0.0903***	0.0903***	0.0903***	0.0903***	0.0903***	0.0903***	0.0905***	0.0905***
σ	0.0102	0.0102	0.0102	0.0102	0.0102	0.0102	0.0101	0.0101
Y_{2008}	0.0350***	0.0350***	0.0350***	0.0350***	0.0350***	0.0350***	0.0351***	0.0351***
σ	0.0101	0.0101	0.0101	0.0101	0.0101	0.0101	0.0100	0.0100
Y_{2009}	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
σ	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0099	0.0099
Constant	2.3045***	2.3045***	2.3045***	2.3045***	2.3045***	2.3045***	2.3088***	2.3088***
σ	0.0691	0.0691	0.0691	0.0691	0.0691	0.0691	0.0687	0.0687
Obs	13739	13739	13739	13739	13739	13739	13871	13871
F	1771.16	1771.16	1771.16	1771.16	1771.16	1771.16	1792.59	1792.59
$\overline{R^2}$	0.7998	0.7998	0.7998	0.7998	0.7998	0.7998	0.8002	0.8002
$\sqrt{MSE}$	0.23027	0.23027	0.23027	0.23027	0.23027	0.23027	0.23027	0.23027

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Table 15: First stage estimates of endogenous trade dependency

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
TanSlope	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***
σ	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
TanSlRange	0.0109***	0.0109***	0.0109***	0.0109***	0.0109***	0.0109***	0.0108***	0.0108***
σ	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
CVSlope	-5.3784***	-5.3784***	-5.3784***	-5.3784***	-5.3784***	-5.3784***	-5.3704***	-5.3704***
σ	0.0967	0.0967	0.0967	0.0967	0.0967	0.0967	0.0962	0.0962
CVElev	3.3934***	3.3934***	3.3934***	3.3934***	3.3934***	3.3934***	3.3870***	3.3870***
σ	0.0659	0.0659	0.0659	0.0659	0.0659	0.0659	0.0656	0.0656
$\left(\frac{\text{Slope}}{\text{Elev}}\right)$	12.6840***	12.6840***	12.6840***	12.6840***	12.6840***	12.6840***	12.7137***	12.7137***
σ	0.3099	0.3099	0.3099	0.3099	0.3099	0.3099	0.3081	0.3081
$\left(\frac{\text{Elev}}{\text{SDElev}}\right)$	-3.0182***	-3.0182***	-3.0182***	-3.0182***	-3.0182***	-3.0182***	-3.0241***	-3.0241***
σ	0.0701	0.0701	0.0701	0.0701	0.0701	0.0701	0.0698	0.0698
$\left(\frac{\text{Slope}}{\text{SDSlope}}\right)$	-2.0529***	-2.0529***	-2.0529***	-2.0529***	-2.0529***	-2.0529***	-2.0509***	-2.0509***
σ	0.0607	0.0607	0.0607	0.0607	0.0607	0.0607	0.0604	0.0604
$\left(\frac{\text{TanSlope}}{\text{SDSlope}}\right)$	-0.0458***	-0.0458***	-0.0458***	-0.0458***	-0.0458***	-0.0458***	-0.0457***	-0.0457***
σ	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009
$\left(\frac{\text{MedElev}}{\text{SDElev}}\right)$	1.6077***	1.6077***	1.6077***	1.6077***	1.6077***	1.6077***	1.6140***	1.6140***
σ	0.0586	0.0586	0.0586	0.0586	0.0586	0.0586	0.0583	0.0583
$\left(\frac{\text{CVElev}}{\text{CVSlope}}\right)$	-6.5528***	-6.5528***	-6.5528***	-6.5528***	-6.5528***	-6.5528***	-6.5425***	-6.5425***
σ	0.1045	0.1045	0.1045	0.1045	0.1045	0.1045	0.1039	0.1039

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	(Dir costs / firms)	Total costs	(Tot costs / firms)	(Ind firms / highway km)	(Ind firms / rail km)	(Total firms / highway km)	(Total firms / rail km)
NumParks	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***
σ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GreenArea	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
σ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DisastArea	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
σ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Other mine	0.0145	0.0145	0.0145	0.0145	0.0145	0.0145	0.0148	0.0148
σ	0.0405	0.0405	0.0405	0.0405	0.0405	0.0405	0.0405	0.0405
Coal mine	0.0022	0.0022	0.0022	0.0022	0.0022	0.0022	0.0021	0.0021
σ	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352
Ferrous mine	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027
σ	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357
Non-ferr mine	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027
σ	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357
Non-met mine	0.0022	0.0022	0.0022	0.0022	0.0022	0.0022	0.0021	0.0021
σ	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352	0.0352
Petrol & gas	0.1373***	0.1373***	0.1373***	0.1373***	0.1373***	0.1373***	0.1374***	0.1374***
σ	0.0381	0.0381	0.0381	0.0381	0.0381	0.0381	0.0381	0.0381
Y_{1998}	0.3270***	0.3270***	0.3270***	0.3270***	0.3270***	0.3270***	0.3249***	0.3249***
σ	0.0308	0.0308	0.0308	0.0308	0.0308	0.0308	0.0305	0.0305
Y_{1999}	0.2341***	0.2341***	0.2341***	0.2341***	0.2341***	0.2341***	0.2321***	0.2321***
σ	0.0304	0.0304	0.0304	0.0304	0.0304	0.0304	0.0302	0.0302
Y_{2000}	0.4199***	0.4199***	0.4199***	0.4199***	0.4199***	0.4199***	0.4179***	0.4179***
σ	0.0303	0.0303	0.0303	0.0303	0.0303	0.0303	0.0301	0.0301
Y_{2001}	0.3352***	0.3352***	0.3352***	0.3352***	0.3352***	0.3352***	0.3335***	0.3335***

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*

Continued on next page

	Direct costs	$\left(\frac{\text{Dir costs}}{\text{firms}}\right)$	Total costs	$\left(\frac{\text{Tot costs}}{\text{firms}}\right)$	$\left(\frac{\text{Ind firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Ind firms}}{\text{rail km}}\right)$	$\left(\frac{\text{Total firms}}{\text{highway km}}\right)$	$\left(\frac{\text{Total firms}}{\text{rail km}}\right)$
σ	0.0302	0.0302	0.0302	0.0302	0.0302	0.0302	0.0300	0.0300
Y_{2002}	0.3104***	0.3104***	0.3104***	0.3104***	0.3104***	0.3104***	0.3087***	0.3087***
σ	0.0296	0.0296	0.0296	0.0296	0.0296	0.0296	0.0294	0.0294
Y_{2003}	0.4576***	0.4576***	0.4576***	0.4576***	0.4576***	0.4576***	0.4560***	0.4560***
σ	0.0293	0.0293	0.0293	0.0293	0.0293	0.0293	0.0291	0.0291
Y_{2004}	0.3832***	0.3832***	0.3832***	0.3832***	0.3832***	0.3832***	0.3820***	0.3820***
σ	0.0288	0.0288	0.0288	0.0288	0.0288	0.0288	0.0286	0.0286
Y_{2005}	0.3758***	0.3758***	0.3758***	0.3758***	0.3758***	0.3758***	0.3748***	0.3748***
σ	0.0285	0.0285	0.0285	0.0285	0.0285	0.0285	0.0284	0.0284
Y_{2006}	0.4458***	0.4458***	0.4458***	0.4458***	0.4458***	0.4458***	0.4449***	0.4449***
σ	0.0285	0.0285	0.0285	0.0285	0.0285	0.0285	0.0282	0.0282
Y_{2007}	0.4758***	0.4758***	0.4758***	0.4758***	0.4758***	0.4758***	0.4750***	0.4750***
σ	0.0286	0.0286	0.0286	0.0286	0.0286	0.0286	0.0282	0.0282
Y_{2008}	0.2879***	0.2879***	0.2879***	0.2879***	0.2879***	0.2879***	0.2875***	0.2875***
σ	0.0284	0.0284	0.0284	0.0284	0.0284	0.0284	0.0280	0.0280
Y_{2009}	-0.0303	-0.0303	-0.0303	-0.0303	-0.0303	-0.0303	-0.0307	-0.0307
σ	0.0280	0.0280	0.0280	0.0280	0.0280	0.0280	0.0276	0.0276
Constant	12.0204***	12.0204***	12.0204***	12.0204***	12.0204***	12.0204***	12.0093***	12.0093***
σ	0.1937	0.1937	0.1937	0.1937	0.1937	0.1937	0.1927	0.1927
Obs	13739	13739	13739	13739	13739	13739	13871	13871
F	955.99	955.99	955.99	955.99	955.99	955.99	966.06	966.06
$\overline{R^2}$	0.683	0.683	0.683	0.683	0.683	0.683	0.6832	0.6832
$\sqrt{MSE}$	0.64581	0.64581	0.64581	0.64581	0.64581	0.64581	0.64545	0.64545

**** - significant at 1%, ** - significant at 5%, * - significant at 10%*