Spillovers across industries and regions in China's regional economic diversification

Jian Gao, Bogang Jun, Alex 'Sandy' Pentland, Tao Zhou & César A. Hidalgo

View supplementary material

Published online: 31 Mar 2021.

Submit your article to this journal

Article views: 30

View related articles

View Crossmark data

Spillovers across industries and regions in China's regional economic diversification

Jian Gao[a] , Bogang Jun[b] , Alex 'Sandy' Pentland[c] , Tao Zhou[d] and César A. Hidalgo[e]

ABSTRACT
Industrial diversification depends on spillovers from related industries and nearby regions, yet their interaction remains largely unclear. We study economic diversification in China during the period 1990–2015 and present supportive evidence on both spillover channels. We add to the literature by showing that these two channels behave as substitutes when explaining new entries and exits, and by using acceleration campaigns of high-speed rail to address some endogeneity concerns with regional spillovers. Our findings confirm the role of relatedness and geographical distance in the diffusion of economic capabilities and support the idea that improvements in transportation can facilitate the diffusion of productive capabilities.

KEYWORDS
economic development; industrial structure; economic complexity; spillovers; high-speed rail

INTRODUCTION

The pace and scale of China's economic expansion has no historical precedent (Eichengreen et al., 2012; Felipe et al., 2013; Song et al., 2011; Zhu, 2012). Between 1990 and 2015, China's overall gross domestic product (GDP) grew by a factor of 30, from less than US$400 billion in 1990 to more than US$10 trillion in 2015, and its GDP per capita (purchasing power parity – PPP) increased by a factor of nearly 10 (from US$1516 in 1990 to more than US$13,400 in 2015). But what explains China's remarkable economic success? Here, the literature provides multiple answers, from the advantages of backwardness (Gerschenkron, 1962), to government policies (Amsden, 1992; Evans, 2012; Left-wich, 1995; Lin et al., 2003; Wade, 2004), economic

reforms (Lin et al., 2003), financial development (Laurenceson & Chai, 2003) and institutional change (Qian, 2001).

What is also evident from the data is that China's expansion relied on the export of products that were unusually sophisticated for its income level (Hidalgo & Hausmann, 2009; Lin, 2012). Rodrik (2006) estimated the level of sophistication of China's exports by calculating the average income per capita of countries exporting the same products. He showed that even in 1992, when China's GDP per capita at PPP and constant prices was US$1844, its exports were associated with an average level of income of about US$13,500 (China's GDP per capita in 2015). Rodrik argued that this high level of export sophistication drove China's great economic expansion, a finding that is supported by the literature on

CONTACT

[a] (Corresponding author) ✉ jian.gao1@kellogg.northwestern.edu
CompleX Lab, University of Electronic Science and Technology of China, Chengdu, China; and Kellogg School of Management, Northwestern University, Evanston, IL, USA.
[b] ✉ bogang.jun@inha.ac.kr
Department of Economics, Inha University, Incheon, South Korea.
[c] ✉ sandy@media.mit.edu
MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA.
[d] ✉ zhutou@ustc.edu
CompleX Lab, University of Electronic Science and Technology of China, Chengdu, China.
[e] (Corresponding author) ✉ cesar.hidalgo@univ-toulouse.fr
Center for Collective Learning, University of Toulouse, Toulouse, France; School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA; and Alliance Manchester Business School, University of Manchester, Manchester, UK.
Supplemental data for this article can be accessed at https://doi.org/10.1080/00343404.2021.1883191

economic complexity (Gao & Zhou, 2018; Hausmann et al., 2014; Hidalgo, 2015, 2021; Hidalgo & Hausmann, 2009; Tacchella et al., 2012; Poncet & De Waldemar, 2013; Stojkoski et al., 2016).

But how did China learn to export products of high sophistication? And how did it manage to expand these lessons throughout its vast territory? The literature provides two possible answers. One is the idea that economies leverage the capabilities embodied in related industries by exploiting inter-industry spillovers (Hidalgo et al., 2007, 2018). Economies that are good at producing shirts can easily learn to produce trousers. But spillovers are regional too, so a region's success at making shirts also should increase with the number of neighbouring regions that are already good performers in the shirt industry. Yet, despite ample evidence of spillovers among related industries (Boschma et al., 2015; Guevara et al., 2016; Hidalgo et al., 2007, 2018; Kogler et al., 2013; Neffke & Henning, 2013; Neffke et al., 2011) and nearby regions (Audretsch & Feldman, 1996; Bahar et al., 2014; Boschma et al., 2013; Jaffe et al., 1993; Jun et al., 2020), the literature has not explored the interaction between these two channels deeply, or exhausted the search for policy interventions that can accelerate them.

Here we leverage two data sets for firms in China to study its period of rapid economic expansion during the period 1990–2015 through the lens of relatedness. First, we show that the probability that a province enters a new industry (develops comparative advantage) increases with the number of related industries present in that province (in agreement with the principle of relatedness; Guo & He, 2017; He et al., 2017; Hidalgo et al., 2007, 2018), and with the number of neighbouring provinces in which that industry is already present (in agreement with regional spillover studies; Audretsch & Feldman, 1996, 2004; Bahar et al., 2014; Boschma et al., 2013; Jaffe, 1986; Jaffe et al., 1993; Zhu et al., 2019). Second, we show that the two spillover channels behave as substitutes, meaning that the marginal contribution of one channel (related industries or nearby regions) is reduced when the other is sufficiently active. As a further robustness check, we show the causality of one of the channels (regional spillovers) by using the introduction of high-speed rail (HSR) as an instrument. The results confirm the role of relatedness and geographical distance in the diffusion of economic capabilities and provide evidence supporting the idea that improvements in transportation – such as HSR – can facilitate the diffusion of economic capabilities among regions.

LITERATURE REVIEW

Economic development usually involves structural transformation, which implies growing the variety and sophistication of economic output (Felipe et al., 2012; Hidalgo et al., 2007; Saviotti & Frenken, 2008; Saviotti et al., 2021). But how does structural transformation take place?

Economic diversification and development requires the simultaneous presence of multiple specific factors (Porter,

2008), from institutional changes and technology (Mokyr, 1992), to modern management structures (Chandler, 1994; Rosenberg & Birdzell, 2008), research and development (R&D) activities (Freeman & Soete, 1997), and systems of training and education (Goldin & Katz, 2009). Yet, since identifying and estimating the presence of specific factors is not straightforward, recent efforts have focused on methods that allow scholars to estimate the presence of multiple factors simultaneously. In particular, these methods estimate the relatedness between locations (e.g., countries, regions, cities) and activities (e.g., products, industries, occupations) by looking at the presence of similar activities in a location or in nearby locations (Hidalgo et al., 2007). In the literature, these methods have been used to map and quantify the path dependencies that govern the structural transformation of economies (Lin, 2012).

Relatedness methods have helped show that economies leverage the capabilities embodied in related industries (Hidalgo et al., 2018). In fact, there is now plenty of evidence documenting the importance of relatedness in predicting diversification events at multiple scales and for a variety of activities (Boschma et al., 2015; Gao et al., 2019; Guevara et al., 2016; Hausmann et al., 2014; Hidalgo et al., 2007; Kogler et al., 2013; Neffke & Henning, 2013; Neffke et al., 2011), including evidence from China as well (Guo & He, 2017; He et al., 2017; Jarreau & Poncet, 2012; Zhu et al., 2019). For example, Hidalgo et al. (2007) and Hausmann et al. (2014) used export data to show that the probability that a country develops comparative advantage in a product grows with the number of related products that the country already exports. Similar findings have shown relatedness to predict the evolution of the research portfolio of countries (Guevara et al., 2016), the presence of industries in regions (Balland et al., 2019; Boschma, 2017; Boschma & Iammarino, 2009; Boschma et al., 2012; Delgado et al., 2016; Frenken et al., 2007), and the patent portfolio of cities (Boschma et al., 2015; Castaldi et al., 2015; Kogler et al., 2013).

The second source of spillovers is geographical neighbours (Audretsch & Feldman, 1996; Boschma, 2005; Jaffe et al., 1993). For example, the probability that a province succeeds at making shirts increases with the number of neighbouring regions in which the shirt industry is already present. At the product level, Bahar et al. (2014), Boschma et al. (2017) and Jun et al. (2020) showed that the probability that a region starts to export a product increases significantly if that region shares a border with a nearby region that is already a successful exporter of the product. At the regional level, Acemoglu et al. (2015) found that spillover effects are sizable, accounting for about 50% of the quantitative impact of an expansion in the local state capacity. By analysing China's custom data, He et al. (2019) found that knowledge spillovers play an important role in industrial diversification, and spillovers are more efficient if two regions have larger institutional proximity.

Despite ample evidence supporting spillovers among related industries and nearby regions, there is relatively little research looking at the interaction between these two

spillover channels. In fact, recent work has focused more on unpacking the relatedness channel than on comparing relatedness with geographical distance. For instance, Jara-Figueroa et al. (2018) compared the role of a worker's industry-, occupation- and location-specific experience in the growth and survival of pioneer firms. Similarly, Jun et al. (2020) split the relatedness of international trade into three channels: product relatedness, exporter relatedness and importer relatedness. Farinha et al. (2019) unpacked relatedness among occupations by looking at occupations that co-locate, share skills or are employed in the same industries, and Diodato et al. (2018) unpacked co-location relatedness for industries in the United States.

Moreover, the literature is also thin when it comes to causal evidence. Jara-Figueroa et al. (2018) developed a Bartik-style instrument to show that diversification events were more successful when pioneer firms hired workers from related industries, but not necessarily related occupations. While Jara-Figueroa et al. (2018) examined the role of three relatedness channels, they showed causal evidence only for one of them. This is largely because instruments can be hard to come by since they need to satisfy some key conditions, such as having an effect on the explanatory variable without affecting the dependent variable directly. To meet these requirements, random events, such as policy shocks or natural disasters, are commonly used as instruments (Coscia et al., 2020). For example, Catalini et al. (2020) used the introduction of Southwest Airlines as an instrument to show that reductions in ticket prices of direct flights between US cities increased collaboration among scholars from universities in cities connected by cheaper flights.

In the context of China's regional development, the construction of a transportation system such as HSR is increasingly used as an instrument since it can be largely regarded as a quasi-random political event (Zheng & Kahn, 2013; Zhu et al., 2019). In particular, during China's great economic expansion, the commercial train service improved through several 'speed-up' campaigns, which increased the speed of trains from an average of about 48 km/h (in the 1990s) to more than 300 km/h in the best cases (Jiao et al., 2014). By 2015, over 90 cities in China were connected by HSR (Lin et al., 2021), and as of September 2016, China had the world's longest HSR network, with over 20,000 km of track – longer than the rest of the world's HSR tracks combined (Cao et al., 2013). The introduction of HSR reduces travel time between provinces, encourages face-to-face interactions and promotes spillovers (Zheng & Kahn, 2013). At the same time, it is staggered, adding temporal variations that we can use to study the effect of the differential introduction of HSR on the diffusion of economic capacities.

The recent literature has explored the influence of HSR on accessibility (Fan et al., 2019), labour employment (Lin, 2017), population mobility and agglomeration (Li & Xu, 2018), productivity (Yang et al., 2019), industrial development (Qin, 2017), and more (Chen & Vickerman, 2017; Xie et al., 2020). In fact, the construction of

HSR between provinces should be close to random (Qin, 2017), at least with respect to the variable of interest (e.g., industrial similarity among provinces) and to province-level characteristics (Bertrand et al., 2004; Besley & Case, 2000), because HSR construction was largely driven by political considerations (not primarily to connect provinces with similar industrial structures). For example, the 'Go West' plan connected the east coast with China's Far West. There is also the 'Silk Road Economic Belt' plan (Albalate & Bel, 2012; Rolland, 2015), and plans to connect China with Southeast Asia (Garver, 2006). Along this line, Qin (2017) pointed out that the introduction of HSR in China can be treated as a quasi-natural experiment because most initial HSR was implemented on existing railway lines.

In this paper, we contribute to the literature on relatedness and geographical spillovers by studying the interaction between two spillover channels, namely, relatedness among industries (inter-industry spillovers) and among nearby regions (interregional spillovers). Our hypothesis is that both channels have significant effects on China's regional economic development, but they act as substitutes when it comes to promoting industrial diversification. Moreover, we provide some causal evidence in support of the effects of regional spillovers on economic development by using the introduction of HSR as an instrument. Considering that HSR reduces travel time and facilitates face-to-face interactions, we hypothesize that HSR should play role in the acceleration of regional spillovers and have a positive effect on regional economic growth if there is enough factor mobility (Banerjee et al., 2020; Berger & Enflo, 2017).

DATA DESCRIPTION

We use two data sets in our analysis: publicly listed firm data from China's two major stock markets, and Chinese enterprise data from various statistical yearbooks. Both data sets provide useful information, but neither by itself is necessarily perfect. For example, the firm data cover a large time period, while it has potential biases towards large firms. The enterprise data cover more firms and enterprises, but they have a relatively short time period. Here, our strategy is to use China's firm data for the main results and Chinese enterprise data for additional robustness checks mainly because the firm data cover a longer time period that makes our further causal analysis possible. By leveraging these two data sets, we aim to weaken the potential limitations of each data set in our analysis.

China's firm data
Our primary firm data is from the RESSET Financial Research Database, which is maintained by the Beijing Gildata RESSET Data Tech Co., Ltd (http://www.resset.cn). The data set covers a period of 1990–2015 and provides some basic registration and financial information of publicly listed firms on China's stock exchanges, such as listing date, delisting date, registered address, industry category, yearly revenue and number of employees

(see the supplemental data online for details). Although the numbers of newly listed and delisted firms in each year fluctuate, the overall number of firms grows almost linearly over time (see Figure A1 in Appendix A in the supplemental data online). The registered addresses of firms cover 31 provinces in mainland China. Industries of firms are aggregated into two levels: 18 categories at the sector level and 70 subcategories at the subsector level. The aggregation is based on the *Guidelines for the Industry Classification of Listed Companies* issued by the China Securities Regulatory Commission (CSRC) in 2011.[1] The detailed CSRC category and subcategory codes as well as their associated industry names are provided in Figure A2 in Appendix A online.

Chinese enterprise data

The enterprise data are from the *China Statistical Yearbook* released by the National Bureau of Statistics of China (http://www.stats.gov.cn). The yearbook provides yearly census data of total labour and economic output of industries in all provinces. Chinese enterprise data used in this study cover about 84.67 million workers employed by industrial enterprises above a designated size during the period 1998–2008 and 31 provinces in mainland China. The classification of industries corresponds to China's Industrial Classification for National Economic Activities (Version 2002). The data cover 188 industrial classes, which can be further aggregated into 39 divisions and three sections according to the industry classification. In particular, the three sections in our data all belong to processing-related industries, namely, Mining (six divisions), Manufacturing (30 divisions) and Electricity (three divisions).

Distance and macroeconomic indicators of China

To study regional spillovers, we use geographical distance as a measure of physical proximity among regions. The geographical distance ($D_{i,j}$) between provinces i and j is defined as the distance between the capital cities of two provinces. From the *China Statistical Yearbook*, we collect macroeconomic data at the province level, including GDP per capita, resident population, total value of imports and exports, urban area, and total area. Moreover, we use the share of the urban area in a province as an urbanization metric. These macroeconomic indicators cover the period 1990–2015. For brief descriptions and summary statistics of these indicators, see Table B1 in Appendix B in the supplemental data online.

EMPIRICAL STRATEGY

Measuring inter-industry spillovers

Based on China's publicly listed firm data, we first explore inter-industry spillovers. Specifically, we analyse how the probability of an industry appearing in a province is affected by the number of related industries already present in it by constructing a network of industries, or 'industry space'.

We connect provinces and industries by building a 'province–industry' bipartite network, where the weight of link $x_{i,\alpha}$ is the number of firms in province i that operate in industry α (see Figure C1 in Appendix C in the supplemental data online). Using this network, we calculate the proximity $\phi_{\alpha,\beta}$ between industries α and β by using the cosine similarity between $x_{i,\alpha}$ and $x_{i,\beta}$ across all provinces. Following Hidalgo et al. (2007), we assume the co-location of industries to be an imperfect proxy of their similarity, since pairs of industries that tend to co-locate are more likely to require similar capabilities. Formally, we let $x_{i,\alpha,t}$ and $x_{i,\beta,t}$ be the number of firms in province i that operate in industries α and β, respectively, in year t. The proximity $\phi_{\alpha,\beta,t}$ is then given by:

$$\phi_{\alpha,\beta,t} = \frac{\sum_i x_{i,\alpha,t} x_{i,\beta,t}}{\sqrt{\sum_i (x_{i,\alpha,t})^2}\sqrt{\sum_i (x_{i,\beta,t})^2}}. \tag{1}$$

Figure 1(A) shows China's industry space for 2015 (see the supplemental data online for details on the visualization methods). We note that China's industry space exhibits both a core–periphery and a dumb-bell structure. There is a tightly knit core of manufacturing industries (on the left) and another tightly knit core of service and information-related activities (on the right). This dumb-bell structure is also visible when looking at the hierarchically clustered matrix of industrial proximity (see Figure C3 in Appendix C in the supplemental data online). In line with previous findings based on data of products (Hidalgo et al., 2007), we find that extractive industries and agriculture tend to occupy the periphery of the industry space.

We estimate the number of related industries in a region using the relatedness density indicator introduced by Hidalgo et al. (2007). First, we define the revealed comparative advantage $RCA_{i,\alpha,t}$ of province i in industry α in year t. $RCA_{i,\alpha,t}$ as the ratio between the observed number of firms operating in industry α in province i and the expected number of firms of that industry in that province (Balassa, 1965). Formally, $RCA_{i,\alpha,t}$ is given by:

$$RCA_{i,\alpha,t} = \frac{\dfrac{x_{i,\alpha,t}}{\sum_\alpha x_{i,\alpha,t}}}{\dfrac{\sum_i x_{i,\alpha,t}}{\sum_\alpha \sum_i x_{i,\alpha,t}}}, \tag{2}$$

where $x_{i,\alpha,t}$ is the number of firms in province i that operate in industry α in year t. We say industry α is significantly present in province i in year t if $RCA_{i,\alpha,t} \geq 1$. When considering the evolution of China's provincial industrial structure (e.g., see the four illustrative examples in Figure D1 in Appendix D in the supplemental data online), we find that new industries present in a province tend to be connected to other closely related industries that were already present in that province.

To avoid confusion with the indicator that we will introduce below for neighbouring provinces, we call this
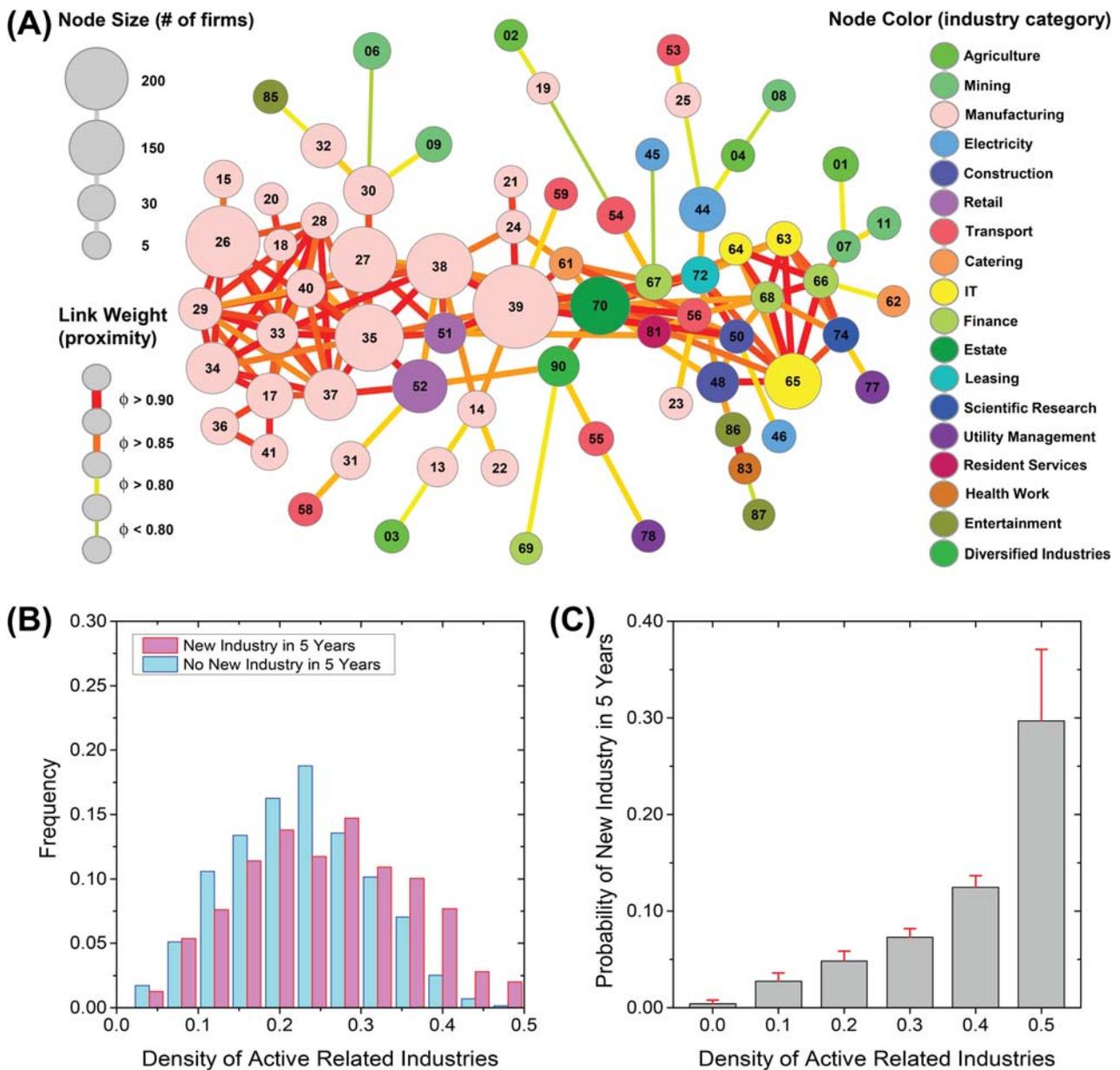
Figure 1. Inter-industry spillovers. (A) Network representation of China's industry space in 2015. Nodes represent industries, and links connect industries that have high proximity. Nodes are classified into 70 subcategories and coloured by 18 sectors. The size of nodes is proportional to the number of firms in that industry. The label of nodes corresponds to Figure A2 in Appendix A in the supplemental data online. The colour and weight of links correspond to the proximity ϕ. Readers of the print version can view the figure in colour online at https://doi.org/10.1080/00343404.2021.1883191. (B) Distribution of the density of related industries. The distributions are for the pairs of provinces and industries that succeed and failed in developing revealed comparative advantage (RCA) in the next five years, respectively. The mean of the pink distribution is significantly larger than that of the blue distribution. (C) Probability of a new industry appearing in a province as a function of the density of related industries (ω). Bars indicate averages and error bars indicate standard errors. Results are averaged for the period 2001–15 using a five-year interval. Source: Authors' calculations.

estimator the *density of related industries* (ω). Formally, the relatedness density $\omega_{i,\alpha,t}$ or industry α in province i in year t is given by:

$$\omega_{i,\alpha,t} = \frac{\sum_{\beta} \phi_{\alpha,\beta,t} M_{i,\beta,t}}{\sum_{\beta} \phi_{\alpha,\beta,t}}, \qquad (3)$$

where $M_{i,\beta,t}$ takes a value of 1 if province i has RCA in industry α in year t (i.e., $RCA_{i,\beta,t} \geq 1$), and 0 otherwise. Relatedness density is an indicator representing, for each

industry and location, the fraction of similar industries that are already present in the same location.

Using the above industry relatedness density measure defined in equation (3), we check the probability of a location i entering an industry α as a function of the density of related industries ($\omega_{i,\alpha,t}$) in that province. To reduce noise, we follow Bahar et al. (2014) and restrict the appearance of new industries to two conditions: a backward condition, which requires an industry to have RCA

< 1 during the two years before the beginning of the period; and a forward condition, which requires an industry to sustain RCA > 1 for the two years after the end of the period.

Figure 1(B) shows the frequency of densities for the industries that a province entered or did not enter in a five-year period. The distributions show that, on average, density is larger for the industries that provinces entered (analysis of variance (ANOVA) $p = 2.1 \times 10^{-40}$) than those they did not (see Figure D2 in Appendix D in the supplemental data online for additional robustness checks). Figure 1(C) shows the probability of a province entering an industry as a function of the density of related industries in that province five years ago. The increasing and convex relationship indicates that the probability of a province entering an industry increases strongly with density. To reduce noise, we visualize the data here using a fixed industry space ($\phi_{\alpha,\beta}$) in 2015 in equation (3). The results are robust when we use a time-varying industry space ($\phi_{\alpha,\beta,t}$), where industrial proximity is calculated using data only from previous years (see Figure D3 in Appendix D in the supplemental data online).

Measuring interregional spillovers
We then explore the regional spillovers through nearby regions. Figure 2(A) shows the spatial evolution of industries in provinces using data on RCA of four industries (Chemical products manufacturing, Pharmaceuticals, Electric machinery manufacturing, and Wholesale) between 1992 and 2015 (see Figure E3 in Appendix E in the supplemental data online for an equivalent chart using the number of firms). The saturation of the colour indicates the logarithm of RCA of that province in that industry. In these four examples, we observe that provinces that entered an industry tend to be neighbours of provinces that already had RCA in that industry, providing suggestive evidence of regional spillovers.

Similarly, we construct a density indicator for regional spillovers (Ω). For province i in industry α in year t, we define the density of neighbouring provinces by:

$$\Omega_{i,\alpha,t} = \frac{\sum_j \dfrac{M_{j,\alpha,t}}{D_{i,j}}}{\sum_j \dfrac{1}{D_{i,j}}}, \qquad (4)$$

where $D_{i,j}$ is the geographical distance between provinces i and j; and the binary variable $M_{j,\alpha,t}$ takes a value of 1 if $RCA_{j,\alpha,t} \geq 1$ when province j has industry α, and 0 otherwise. We calculate the density of neighbouring provinces (Ω) for each province and industry in a base year and identify new industries that appear in that province five years later. To reduce noise, we restrict the presence of new industries to a backward condition, requiring an industry to have RCA < 1 for two years before the beginning of the period; and a forward condition, requiring an industry to be present with RCA > 1 for two years after the end of the period.

Figure 2(B) compares the distribution of densities (Ω) for industry–province pairs that developed RCA in an industry in a five-year period and those that did not. We find that the average density of the province–industry pairs that developed RCA in a five-year period is significantly larger than the province–industry pairs that did not (ANOVA $p = 1.4 \times 10^{-37}$). Figure 2(C) shows the probability of a province developing RCA in an industry as a function of the density of active neighbouring provinces (Ω). Again, we find an increasing and convex relationship showing that the probability of a province developing RCA in an industry increases strongly with the fraction of active neighbouring provinces in that industry. These results are robust when using other distance metrics (see Figure E4 in Appendix E in the supplemental data online).

Econometric model
We use a multivariate probit model to estimate the probability of a province entering, or sustaining RCA in an industry, as a function of the density of related industries and the density of neighbouring provinces. We separate our data set into two subsets: one contains all province–industry pairs that did not have RCA (that could potentially be developed), and the other contains all pairs that had RCA (that could lose it). We then set up two probit regressions, one explaining the probability of a province without RCA in an industry developing RCA in that industry in the next five years, and the other explaining the probability of a province with RCA in an industry retaining RCA in that industry. In both regressions, we control for the size of the industry and the ubiquity of the industry.

Specifically, our econometric model for exploring the inter-industry spillover effects is:

$$M_{i,\alpha,t+5} = \beta_0 + \beta_1 \omega_{i,\alpha,t} + \beta_2 U_{\alpha,t} + \beta_3 N_{i,t} + \mu_t + \varepsilon_{i,\alpha,t}, \qquad (5)$$

where $M_{i,\alpha,t+5}$ ($M_{i,\alpha,t}$) takes a value of 1 if $RCA_{i,\beta,t+5} \geq 1$ ($RCA_{i,\beta,t} \geq 1$), and 0 otherwise; $\omega_{i,\alpha,t}$ is the density of active related industries for industry α in province i in year t; $U_{\alpha,t} = \sum^i M_{i,\alpha,t}$ is the number of provinces in which that industry has RCA (its ubiquity); $N_{i,t} = \sum^\alpha M_{i,\alpha,t}$ is the number of industries with RCA in that province (its diversity); and $\varepsilon_{i,\alpha,t}$ is the error term. Here we include $U_{\alpha,t}$ and $N_{i,t}$ to account for heterogeneity for provinces to develop RCA in industries. For example, provinces that already have RCA in many industries are more likely to develop a new industry. The regression equation includes year-fixed effects μ_t to control for any time-varying characteristics of provinces and industries. Likewise, our econometric model for exploring the regional spillover effects is:

$$M_{i,\alpha,t+5} = \beta_0 + \beta_1 \Omega_{i,\alpha,t} + \beta_2 N_{i,t} + \beta_3 U_{\alpha,t} + \mu_t + \varepsilon_{i,\alpha,t}, \qquad (6)$$

where $\Omega_{i,\alpha,t}$ is the density of active neighbouring provinces
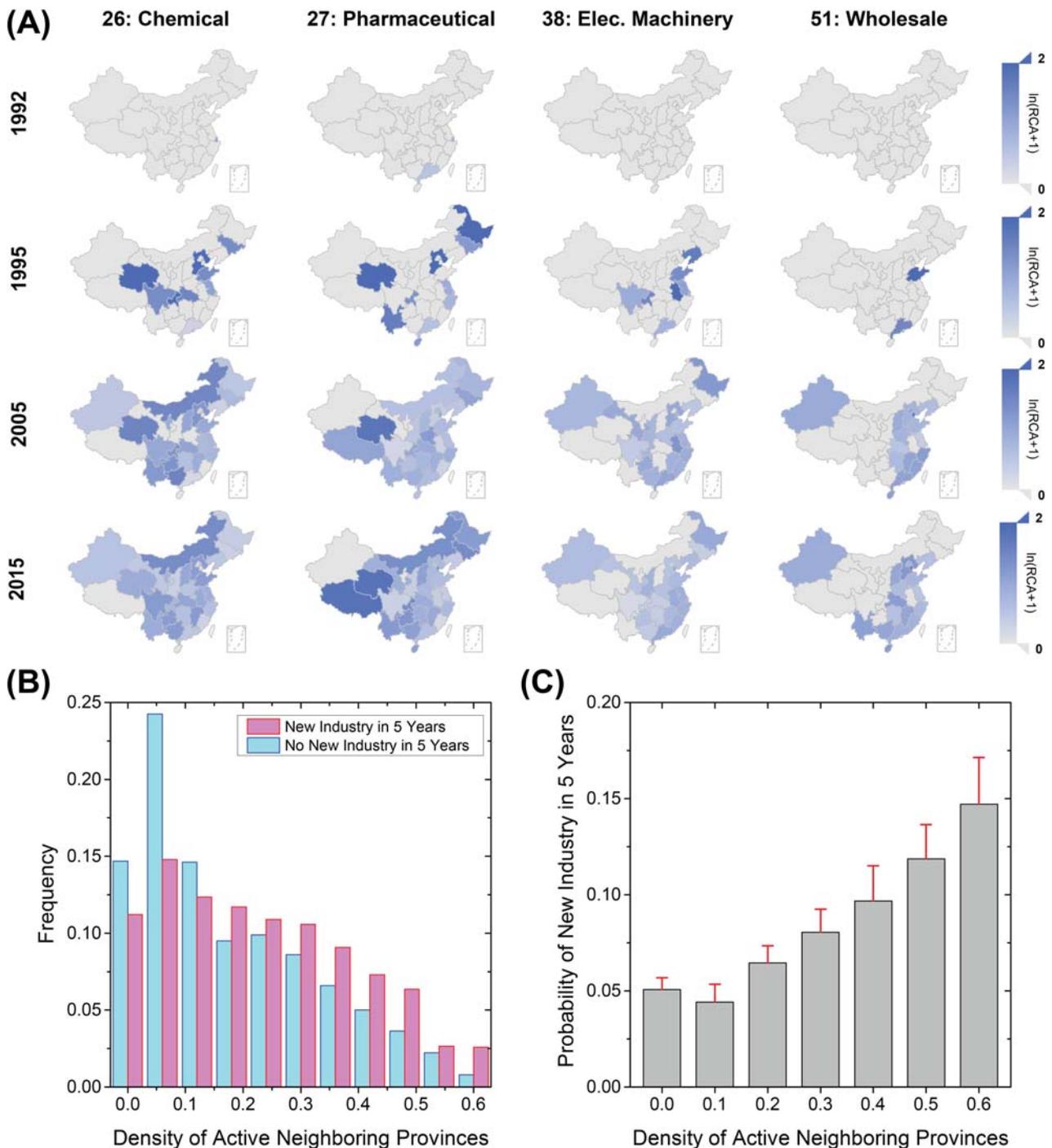
Figure 2. Interregional spillovers. (A) Evolution of revealed comparative advantage (RCA) in China's provinces between 1992 and 2015. The four illustrated industries include: 26, Chemical Products Manufacturing Industry; 27, Pharmaceutical Industry; 38, Electric Machinery Manufacturing Industry; and 51, Wholesale Industry. The keys of labels correspond to Figure A2 in Appendix A in the supplemental data online. The colour saturation indicates the $\ln(RCA + 1)$ value. (B) Distribution of the density of active neighbouring provinces. The distributions are for pairs of provinces and industries that developed RCA in the next five years. They also show the pairs of provinces and industries that succeeded and failed in developing RCA in the next five years, respectively. The mean of the pink distribution is significantly larger than that of the blue distribution. (C) The probability of a province developing comparative advantage in an industry as a function of the density of active neighbouring provinces five years ago. Bars indicate averages and error bars indicate standard errors. Results are averaged for the period 2001–15 using a five-year interval. Source: Authors' calculations.

for industry α and province i in year t. All other variables are defined as the same as those used in equation (5).

In the previous two models, we explore inter-industry and regional spillovers in China's economic development separately. However, do inter-industry and regional spillovers work together, or are they substitutes? To answer this question, we further explore the interaction between the two spillover channels by using a probit model,

where the dependent variable $J_{i,\alpha,t+5}$ indicates whether province i successfully develops industry α in a five-year period. Again, we consider a backward and forward condition to reduce noise. Formally, $J_{i,\alpha,t+5} = 1$ (i.e., province i develops industry α) if $M_{i,\alpha,t} \,\&\, M_{i,\alpha,t-1} \,\&\, M_{i,\alpha,t-2} = 0$ (two-year backward condition), and $M_{i,\alpha,t+5} \,\&\, M_{i,\alpha,t+6} \,\&\, M_{i,\alpha,t+7} = 1$ (two-year forward condition). Our empirical specification for estimating the joint spillover effects is:

$$J_{i,\alpha,t+5} = \beta_0 + \beta_1 \Omega_{i,\alpha,t} + \beta_2 \omega_{i,\alpha,t} + \beta_3 \Omega_{i,\alpha,t}\omega_{i,\alpha,t}$$
$$+ \mu_t + \varepsilon_{i,\alpha,t}, \tag{7}$$

where $\Omega_{i,\alpha,t}$ is the density of active neighbouring provinces; $\omega_{i,\alpha,t}$ is the density of related industries; $\Omega_{i,\alpha,t}\omega_{i,\alpha,t}$ is the interaction term of the two densities; μ_t are the year-fixed effects; and $\varepsilon_{i,\alpha,t}$ is the error term.

EMPIRICAL RESULTS

Spillovers among related industries and nearby regions

In our empirical analysis, we use the subset of data that contains all pairs of provinces and industries that do not have RCA to predict those that will develop RCA. Meanwhile, we use the subset of data that contains all pairs of provinces and industries that have RCA to predict those that can sustain RCA. The results of the regressions for both inter-industry and regional spillovers are given below.

Table 1 summarizes the results of probit regressions based on equation (5) exploring the effects of spillovers among related industries. Columns 1–3 show the effects of spillovers among related industries on regions entering an industry. Columns 4–6 show the effect of relatedness on the preservation of an existing industry in a region. We find that the density of related industries is a strong,

positive and significant predictor of the probability of entering a new industry and of maintaining comparative advantage in an industry in all cases and including all controls. The fact that the effects are robust even with control variables implies that our findings do not just reflect the industrial diversity of a province or the ubiquity of an industry.

Table 2 summarizes the results of probit regressions based on equation (6) for exploring the effects of regional spillovers. Columns 1–3 show that the density of active related industries is a positive and significant predictor of the industries that a province will enter in the future, suggesting that provinces are more likely to enter an industry when they have neighbours that are competitive in that industry. As shown in columns 4–6, however, the effects of active neighbouring provinces on sustaining RCA in an industry are not as clear. The bivariate effects are negative, but become positive after controls. We interpret this as evidence of tension between competition and spillovers, since an active neighbouring province is a source of spillovers when that province does not have an industry, but is also a source of competition when that province has developed that industry.

Interaction between industrial and regional spillovers

To explore the joint effects of industrial and regional spillovers, we first calculate the joint probability of a new industry emerging in a province as a function of both the density of active neighbouring provinces (Ω) and the density of active related industries (ω). Figure 3 shows that the probability of an industry entering a province in a five-year period increases with the density of active neighbouring provinces (Ω on the horizontal axis) and the density of related industries (ω on the vertical axis). This observation suggests that the two spillover channels work jointly to develop new industries. The results are robust when

Table 1. Probit regressions for inter-industry spillovers.

	Probit model					
	Entering (developing RCA in the next five years)			Not exiting (keeping RCA in the next five years)		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
Density of active related industries	3.8844***	4.2084***	11.510***	−0.5753**	−1.7392***	15.266***
	(0.1622)	(0.1661)	(0.3826)	(0.2266)	(0.2665)	(1.1658)
Number of active provinces in an industry		0.0559***	0.0624***		−0.0740***	−0.0834***
		(0.0028)	(0.0029)		(0.0059)	(0.0062)
Number of active industries in a province			−0.1348***			−0.3101***
			(0.0063)			(0.0193)
Observations	25,713	25,713	25,713	6837	6837	6837
Pseudo-R^2	0.0626	0.0924	0.1217	0.0119	0.0497	0.1397

Notes: Probit regressions modelling the probability of developing a new industry, or keeping an industry, in a province are a function of the density of active related industries in a province, the number of provinces active in an industry and the number of industries active in a province. Data are for the 2001–15 period. Probit regressions include year-fixed effects. Robust standard errors are reported in parentheses. RCA, revealed comparative advantage.
Significance level: *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.
Source: Authors' calculations.

Table 2. Probit regressions for interregional spillovers.

	Probit model					
	Entering (developing RCA in a five-year period)			Not exiting (keeping RCA in a five-year period)		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
Density of active neighbouring provinces	1.5393***	1.5621***	1.6969***	−1.4079***	−1.7160***	0.5836*
	(0.0781)	(0.0782)	(0.2116)	(0.1317)	(0.1332)	(0.3311)
Number of active industries in a province		0.0404***	0.0402***		−0.0555***	−0.0660***
		(0.0025)	(0.0025)		(0.0049)	(0.0051)
Number of active provinces in an industry			−0.0053			−0.1045***
			(0.0075)			(0.0132)
Observations	25,713	25,713	25,713	6837	6837	6837
Pseudo-R^2	0.0473	0.0648	0.0648	0.0342	0.0636	0.0793

Notes: Probit regressions modelling the probability of developing a new industry, or keeping an industry, in a province are a function of the density of active neighbouring provinces in an industry, the number of industries active in a province and the number of provinces active in an industry. Data are for the 2001–15 period. Probit regressions include year-fixed effects. Robust standard errors are reported in parentheses. RCA, revealed comparative advantage.
Significance level: *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.
Source: Authors' calculations.

using other density measures (see Figure F1 in Appendix F in the supplemental data online).

We then formalize this observation using regression models. Table 3 summarizes the results of probit regressions based on equation (7) for exploring the joint effects of inter-industry and regional spillovers on the development of a new industry and the preservation of an existing industry (see Table F1 in Appendix F in the supplemental data online for summary statistics of variables). Column 1 shows the basic regression considering the density of active neighbouring provinces (Ω) and the density of related industries (ω), where we find both effects are jointly significant. After adding an interaction term ($\Omega\omega$) between the two densities in column 2, the individual coefficients for both densities (Ω and ω) increase. The interaction term is negative and significant, indicating that the two spillover channels behave as substitutes.

We further consider alternative definitions for both densities. In columns 3–4, we repeat our analysis using the ratio of active neighbouring provinces and related industries as independent variables (see the supplemental data online for details). This is equivalent to calculating both densities using simple proportions instead of weighted averages. Again, we find that both effects are significant and the two spillover channels act as substitutes. Column 5 presents a negative control, where the model uses the number of neighbouring provinces and related industries, regardless of whether they are active or not. In this case, the model loses almost all its explanatory power, and the effects are relatively small. These observations suggest that the source of our results is the presence of active neighbouring provinces and active related industries rather than from having many neighbouring provinces or many industries.

HSR and regional spillovers

Finally, we use the introduction of HSR as a means to address some endogeneity concerns for regional spillovers. We explore the role of HSR in the acceleration of regional spillovers by using a difference-in-differences (DID) analysis. Here we use the entry of HSR as an intervention. Also, we note that similarity among industries is less likely to be affected by new HSR connections. This makes the introduction of HSR a means by which to isolate the effects of regional spillovers from related industry spillovers.

In our DID analysis, pairs of provinces belong to the treatment group if they are connected by HSR in 2015, and otherwise they belong to the control group. Although there are many rounds of 'speed-up' campaigns of HSR in China, we consider only the period between 2004 and
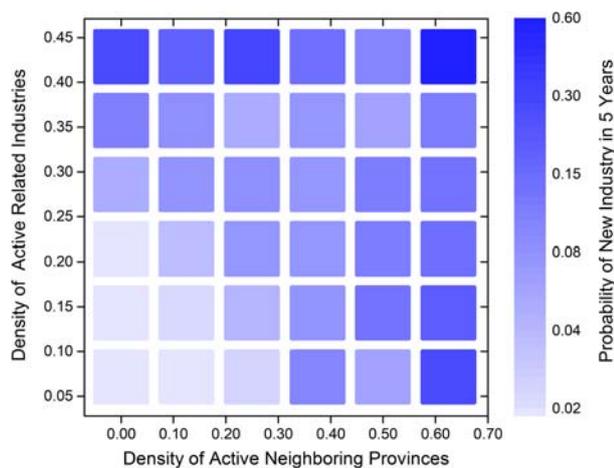
Figure 3. Joint probability of developing new industries in the next five years. The joint probability of a province developing a new industry in a five-year period, given the density of active neighbouring provinces (Ω), is shown on the horizontal axis, and the density of related industries (ω) is shown on the vertical axis.
Source: Authors' calculations.

Table 3. Interaction between inter-industry and regional spillovers.

Variables	Probit model: Industry entries in a five-year period				
	(1)	(2)	(3)	(4)	(5)
Density of active neighbouring provinces	1.3092***	4.3405***			
	(0.0807)	(0.2421)			
Density of active related industries	3.7163***	6.2435***			
	(0.1713)	(0.2616)			
Interaction term 1		−11.8437***			
		(0.9136)			
Ratio of active neighbouring provinces			0.5474***	0.6643***	
			(0.0499)	(0.0678)	
Ratio of active related industries			0.7802***	0.8502***	
			(0.0374)	(0.0472)	
Interaction term 2				−0.3701**	
				(0.1572)	
Number of neighbouring provinces					0.0049
					(0.0103)
Number of related industries					0.0178**
					(0.0090)
Interaction term 3					0.0010
					(0.0019)
Observations	25,713	25,713	25,713	25,713	25,713
Pseudo-R^2	0.0819	0.0974	0.0653	0.0658	0.0234

Notes: Regressions consider the effects of both regional and inter-industry spillovers by using densities and their alternative definitions. Data are for the 2001–15 period. The probit regressions include the year-fixed effects. Robust standard errors are reported in parentheses.
Significance level: *$p < 0.1$, **$p < 0.01$, and ***$p < 0.01$.
Source: Authors' calculations.
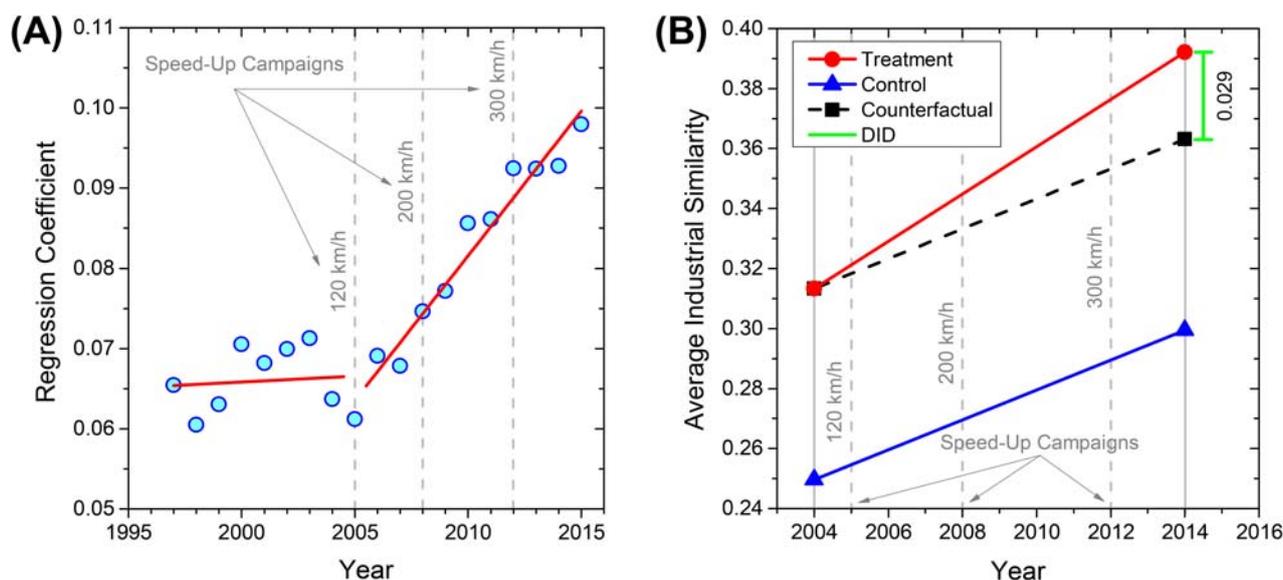
Figure 4. Differences-in-differences (DID) analysis on industrial similarity. (A) Results of the event study. The y-axis shows the regression coefficient β_k in equation (8) as a function of year, after regressing the industrial similarity of province pairs that were connected by high-speed rail (HSR) against the HSR entry. Lines are linear fits for the periods 1997–2005 and 2005–15. (B) Results of the DID analysis. The y-axis is the average industrial similarity of all province pairs connected by HSR (circles) or not (triangles). The value of DID is 0.029 and statistically significant. Vertical dash lines correspond to the years after rail speed-up campaigns, next to which the approximate average speeds of HSR are shown.
Source: Authors' calculations.

Table 4. Difference-in-differences (DID) regressions using the introduction of high-speed rail (HSR) as an instrument.

	DID regressions using an ordinary least squares (OLS) model: Industrial similarity		
Variables	(1)	(2)	(3)
HSR entry	0.0290*	0.0266*	0.0268*
	(0.0152)	(0.0150)	(0.0152)
Treatment group	0.0637***	0.0565***	0.0588***
	(0.0107)	(0.0110)	(0.0108)
After entry	0.0498***	0.0466***	0.0506***
	(0.0091)	(0.0091)	(0.0090)
Δ Population		−0.0204***	
(log)		(0.0049)	
Δ GDP per capita		−0.0207**	
(log)		(0.0081)	
Δ Urbanization			0.0160***
			(0.0127)
Δ Trade (log)			−0.0068***
			(0.0024)
Observations	930	930	930
Robust R^2	0.1628	0.1833	0.1689
RMSE	0.1109	0.1097	0.1106

Notes: DID regressions use the OLS model considering the effects of HSR entry on the industrial similarity after controlling macroeconomic indicators such as population, gross domestic product (GDP) per capita, urbanization and trade. Data are for 2004 (before HSR entry) and 2014 (after HSR entry). Robust standard errors are reported in parentheses. RMSE, root mean square error.
Significance level: *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.
Source: Authors' calculations.

2014, since this allows us to capture the construction of numerous railroads (and, hence, to obtain a larger sample). The introduction of HSR was identified by using the Google Maps API in 2015 considering the accessibility between capital cities of provinces through HSR passenger trains (see the supplemental data online for details).

An important way to validate our choice of instrumental variable is to study the pre-trend of the dependent variable of the control and treatment groups. Our data satisfy this condition before 2005, as shown in Figure 4(A). Specifically, we perform an event study using the following ordinary least squares (OLS) regression model based on the data between 1997 and 2015, where we predict the industrial similarity between provinces i and j for each year:

$$\varphi_{i,j,t} = \beta_0 + \sum_{k=1997}^{2015} \beta_k(Treat_{i,j}*1\{t=k\}) + \varepsilon_{i,j}, \quad (8)$$

where $Treat_{i,j}$ is a dummy variable denoting whether provinces i and j are connected by HSR; and $1\{t=k\}$ is an event time indicator with a value of 1 for the year when the province pair was connected by HSR. In other words, equation (8) regresses industrial similarity between province pairs considering whether they are connected by HSR. Larger coefficients (β_k) suggest that the industrial similarity of province pairs connected by HSR increases compared with those that remain unconnected. Figure 4 (A) shows the graphic results of equation (8). We can see that before the entry of HSR (1997–2005), there is no trend in β_k. After the introduction of HSR (2005–15), the effect of the treatment begins to increase significantly, meaning that the treated provinces grew more similar after HSR was introduced (see Figure G1 in Appendix G in the supplemental data online for additional robustness checks).

Next, we explore the effects of HSR entry on interregional spillovers. Specifically, we use a DID analysis of the following form:

$$\varphi_{i,j,t} = \beta_0 + \beta_1(Treat_{i,j}*After_t) + \beta_2 Treat_{i,j}$$
$$+ \beta_3 After_t + AX' + \varepsilon_{i,j}, \quad (9)$$

where $\varphi_{i,j,t}$ is the industrial similarity between provinces i and j in year t; and $\varepsilon_{i,j}$ is the error term. $Treat_{i,j}*After_t$ is the DID term, where the dummy $Treat_{i,j}$ denotes whether provinces i and j are affected by HSR entry. $After_t$ denotes

Table 5. Probit regressions for inter-industry and regional spillovers based on Chinese enterprise data.

	Probit model: Developing new industries in two years			
Independent variables	(1)	(2)	(3)	(4)
Density of active related industries	0.186***		0.157***	0.163***
	(0.013)		(0.014)	(0.015)
Density of active neighbouring regions		0.172***	0.141***	0.152***
		(0.014)	(0.015)	(0.015)
Interaction term				−0.042**
				(0.017)
Observations	28,518	28,518	28,518	28,518
Pseudo-R^2	0.028	0.027	0.039	0.040

Notes: Probit regressions modelling the probability of developing a new industry in a region are a function of the density of active related industries and the density of active neighbouring regions. Probit regressions include year-fixed effects. Robust standard errors are reported in parentheses.
Significance level: *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.
Source: Authors' calculations.

Table 6. Probit regressions for industrial and regional spillovers based on Brazilian labour data.

Independent variables	Probit model: Developing new industries in two years			
	(1)	(2)	(3)	(4)
Density of active related industries	0.240***		0.193***	0.234***
	(0.002)		(0.002)	(0.003)
Density of active neighbouring regions		0.255***	0.225***	0.261***
		(0.002)	(0.002)	(0.002)
Interaction term				−0.071**
				(0.002)
Observations	1,301,335	1,301,335	1,301,335	1,301,335
Pseudo-R^2	0.040	0.062	0.086	0.092

Notes: Probit regressions modelling the probability of developing a new industry in a region are a function of the density of active related industries and the density of active neighbouring regions. Probit regressions include year-fixed effects. Robust standard errors are reported in parentheses.
Significance level: *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.
Source: Authors' calculations.

whether the time is after HSR entry for year t. The vector X denotes the inclusion of gravity considerations among province pairs, such as the differences between populations, GDP per capita, urbanization and trade.

Figure 4(B) summarizes the results of the DID analysis exploring the effect of HSR on industrial similarity. The DID between the treatment group (circles) and the expected trend from the control group (dashed black line) is 0.029, indicating that pairs of provinces became more industrially similar after the introduction of HSR. Table 4 further summarizes the full regression results. Columns 2–3 show the results of DID regressions with additional controls for the differences in the level of population, GDP per capita, urbanization and trade among province pairs (see Table G1 in Appendix G in the supplemental data online for summary statistics of variables). The regression coefficient (β_1) of the interaction term ($Treat_{i,j} * After_t$) is positive and significant, and it is robust to additional controls (see Table G2 in Appendix G online). These results suggest that the introduction of HSR promoted the increase in industrial similarity among connected provinces, showing some causal evidence on the role of regional spillovers in facilitating economic development.

Robustness checks using Chinese and Brazilian data

We use China's publicly listed firm data in our main analysis, which unfortunately has limitations in its coverage and geographical resolution. One concern is that our results could be biased toward large firms because China's publicly listed firms are relatively large and well-established enterprises. Also, the provincial scale makes our study more similar to those at the country level, since provinces in China are relatively large administrative units. We try to overcome some of these limitations by bringing in two additional data sets: one from China, including many more firms, and the other from Brazil, which we use to reproduce some of our findings at a much finer spatial resolution. More specifically, to check the robustness of our main results, we repeat our analysis using the Chinese

enterprise data, which cover about 84.67 million workers employed by industrial enterprises above the designated size during the period 1998–2008, and the Brazilian labour data set (Gao, 2017; Jara-Figueroa et al., 2018), which covers about 76.62 million workers in Brazil's formal labour market during the period 2006–13.

The Chinese enterprise data cover 188 industries at the class level (further aggregated into 39 divisions and three sections) for all 31 provinces in mainland China. These additional Chinese enterprise data are of relatively large scale and cover more industrial enterprises than the publicly listed firm data. Table 5 summarizes the result of probit regressions constructed using the Chinese enterprise data. In column 1, the density of active related industries has significant and positive effects on the presence of a new industry in provinces in two years. Similarly, as shown in column 2, the effects of the density of active neighbouring provinces are also significant and positive. We further check the robustness of the joint effects of the two spillover channels. In columns 3–4, we find that the two channels have significantly positive effects, and the interaction term has a negative regression coefficient. These results show that our main results remain robust when using an alternative Chinese enterprise data set.

Similarly, Table 6 repeats this analysis using the Brazilian labour data, which covers 588 micro-regions (similar to metro areas, but covering the whole territory) and 669 industries at the class level (further aggregated into 87 divisions and 21 sections according to the National Classification of Economic Activities). Once again, we find the same pattern, showing that our observations are not due to the particularity of China's RESSET firm data set.

DISCUSSION AND CONCLUSIONS

In this study we explored the role of industrial relatedness and geographical proximity in China's economic diversification during a period of its rapid economic expansion. We confirmed previous findings by showing that the probability that an industry enters (exits) a province increases (decreases) with the presence (absence) of related industries. We also

confirmed similar regional spillover effects. In addition to previous works, we explored the interaction between these two spillover channels, finding that they behave as substitutes, and then confirmed this finding using a second data set form China and a granular data set from Brazil. To further support the robustness of our findings, we provided some causal evidence on role of regional spillovers in facilitating economic development. Specifically, we showed that the introduction of HSR significantly increased the industrial similarity of China's provinces that were connected, supporting the idea that improvements in transportation accelerate regional spillovers.

Our findings add to the literature on relatedness and geographical spillovers, but at the same time they should be interpreted in the light of some limitations. One possible improvement would be to consider the issue of multi-locational firms. Investments by the same large firm in different locations (e.g., the establishment of new subsidiaries) may result in the diversification of industries in regions (Elekes et al., 2019), which represent within firm spillovers. Moreover, making a distinction between the effects of the same industry or related industries in neighbouring regions would be helpful for better isolating the effects of regional spillovers. The former may have a negative impact on regional diversification and can take up a competition effect. Echoing this point, our current results are more likely to be a conservative estimation, since we observe the effects of regional spillovers even when there may be potential competition. Nevertheless, our analyses could be further strengthened by using large-scale and granular data that have a better firm-level coverage and a higher geographical resolution.

While we provide evidence in support of inter-industry and regional spillovers, we still lack a micro-channel for them. Are these spillovers the result of spin-off companies (Klepper & Sleeper, 2005), migrant workers (Jara-Figueroa et al., 2018), supply and demand externalities (Hidalgo et al., 2020), labour market pooling or other channels? The observed spillovers could be due to the mechanism of learning and other factors, such as knowledge spillovers, transportation infrastructure, labour market economies and basic factor conditions. These micro-level explanations are important for understanding the role of relatedness and regional spillovers in economic development, but are beyond the scope of this study. Future research could assess the impact of interregional linkages on diversification and spillovers (Miguelez & Moreno, 2018), such as use of interregional scientific collaboration (Acosta et al., 2011), co-patenting (Von Proff & Brenner, 2014) and transportation infrastructure (Catalini et al., 2020; Dong et al., 2020). Also, it would be interesting to dig further into the intersection among industrial relatedness, geographical proximity and new transportation tools in facilitating regional development and economic diversification.

In addition, our work helps expand the body of literature supporting the idea that economic development is a path-dependent process, as it is affected by the presence of related industries and the industrial development of neighbours. This finding represents good news for developing countries looking to modernize their planning and economic development efforts (Gao, 2017; Zhu et al., 2017). For example, an optimal economic diversification strategy should balance the development of related and unrelated activities (Alshamsi et al., 2018), and the development of related activities in nearby regions can be a rationale for upgrading transportation infrastructure (Catalini et al., 2020; Dong et al., 2020). Our analysis also indicates that improvements in transportation can facilitate the diffusion of economic capabilities, suggesting HSR as an important policy implication on the local economies (Diao, 2018; Vickerman, 2018; Yu et al., 2019). As a summary, we hope that this study helps to stimulate new research on the intersection of relatedness and transportation.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

NOTE

1. See http://www.csrc.gov.cn/pub/csrc_en/laws/overRule/Announcement/201302/W020130225570141407159.doc.

ORCID

Jian Gao http://orcid.org/0000-0001-6659-5770
Bogang Jun http://orcid.org/0000-0003-4994-702X
Alex 'Sandy' Pentland http://orcid.org/0000-0002-8053-9983
Tao Zhou http://orcid.org/0000-0003-0561-2316
César A. Hidalgo http://orcid.org/0000-0002-6031-5982

REFERENCES

Acemoglu, D., García-Jimeno, C., & Robinson, J. A. (2015). State capacity and economic development: A network approach. *American Economic Review*, *105*(8), 2364–2409. https://doi.org/10.1257/aer.20140044

Acosta, M., Coronado, D., Ferrándiz, E., & León, M. D. (2011). Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance. *Scientometrics*, *87*(1), 63–74. https://doi.org/10.1007/s11192-010-0305-6

Albalate, D., & Bel, G. (2012). High-speed rail: Lessons for policy makers from experiences abroad. *Public Administration Review*, *72*(3), 336–349. https://doi.org/10.1111/j.1540-6210.2011.02492.x

Alshamsi, A., Pinheiro, F. L., & Hidalgo, C. A. (2018). Optimal diversification strategies in the networks of related products and of related research areas. *Nature Communications*, *9*(1), 1328. https://doi.org/10.1038/s41467-018-03740-9

Amsden, A. H. (1992). *Asia's next giant: South Korea and late industrialization*. Oxford University Press.

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review*, *86*(3), 630–640. www.jstor.org/stable/2118216

Audretsch, D. B., & Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. *Handbook of Regional and Urban Economics*, *4*, 2713–2739. https://doi.org/10.1016/S1574-0080(04)80018-X

Bahar, D., Hausmann, R., & Hidalgo, C. A. (2014). Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics*, *92*(1), 111–123. https://doi.org/10.1016/j.jinteco.2013.11.001

Balassa, B. (1965). Trade liberalisation and 'revealed' comparative advantage. *The Manchester School*, *33*(2), 99–123. https://doi.org/10.1111/j.1467-9957.1965.tb00050.x

Balland, P. A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European Union: Relatedness, knowledge complexity and regional diversification. *Regional Studies*, *53*(9), 1252–1268. https://doi.org/10.1080/00343404.2018.1437900

Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, *145*, 102442. https://doi.org/10.1016/j.jdeveco.2020.102442

Berger, T., & Enflo, K. (2017). Locomotives of local growth: The short- and long-term impact of railroads in Sweden. *Journal of Urban Economics*, *98*, 124–138. https://doi.org/10.1016/j.jue.2015.09.001

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275. https://doi.org/10.1162/003355304772839588

Besley, T., & Case, A. (2000). Unnatural experiments? Estimating the incidence of endogenous policies. *The Economic Journal*, *110*(467), 672–694. https://doi.org/10.1111/1468-0297.00578

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies*, *51*(3), 351–364. https://doi.org/10.1080/00343404.2016.1254767

Boschma, R., Balland, P. A., & Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, *24*(1), 223–250. https://doi.org/10.1093/icc/dtu012

Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, *85*(3), 289–311. https://doi.org/10.1111/j.1944-8287.2009.01034.x

Boschma, R., Martín, V., & Minondo, A. (2017). Neighbour regions as the source of new industries. *Papers in Regional Science*, *96*(2), 227–245. https://doi.org/10.1111/pirs.12215

Boschma, R., Minondo, A., & Navarro, M. (2012). Related variety and regional growth in Spain. *Papers in Regional Science*, *91*(2), 241–256. https://doi.org/10.1111/j.1435-5957.2011.00387.x

Boschma, R., Minondo, A., & Navarro, M. (2013). The emergence of new industries at the regional level in Spain: A proximity approach based on product relatedness. *Economic Geography*, *89*(1), 29–51. https://doi.org/10.1111/j.1944-8287.2012.01170.x

Cao, J., Liu, X. C., Wang, Y., & Li, Q. (2013). Accessibility impacts of China's high-speed rail network. *Journal of Transport Geography*, *28*, 12–21. https://doi.org/10.1016/j.jtrangeo.2012.10.008

Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: An analysis of US state-level patenting. *Regional Studies*, *49*(5), 767–781. https://doi.org/10.1080/00343404.2014.940305

Catalini, C., Fons-Rosen, C., & Gaulé, P. (2020). How do travel costs shape collaboration? *Management Science*, *66*(8), 3340–3360. https://doi.org/10.1287/mnsc.2019.3381

Chandler, A. D. Jr. 1994. *Scale and scope: The dynamics of industrial capitalism*. Belknap.

Chen, C. L., & Vickerman, R. (2017). Can transport infrastructure change regions' economic fortunes? Some evidence from Europe and China. *Regional Studies*, *51*(1), 144–160. https://doi.org/10.1080/00343404.2016.1262017

Coscia, M., Neffke, F. M., & Hausmann, R. (2020). Knowledge diffusion in the network of international business travel. *Nature Human Behaviour*, *4*(10), 1011–1020. https://doi.org/10.1038/s41562-020-0922-x

Delgado, M., Porter, M. E., & Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, *16*(1), 1–38. https://doi.org/10.1093/jeg/lbv017

Diao, M. (2018). Does growth follow the rail? The potential impact of high-speed rail on the economic geography of China. *Transportation Research Part A: Policy and Practice*, *113*, 279–290. https://doi.org/10.1016/j.tra.2018.04.024

Diodato, D., Neffke, R., & Clery, N. (2018). Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics*, *106*, 1–26. https://doi.org/10.1016/j.jue.2018.05.002

Dong, X., Zheng, S., & Kahn, M. E. (2020). The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics*, *115*, 103212. https://doi.org/10.1016/j.jue.2019.103212

Eichengreen, B., Park, D., & Shin, K. (2012). When fast-growing economies slow down: International evidence and implications for China. *Asian Economic Papers*, *11*(1), 42–87. https://doi.org/10.1162/ASEP_a_00118

Elekes, Z., Boschma, R., & Lengyel, B. (2019). Foreign-owned firms as agents of structural change in regions. *Regional Studies*, *53*(11), 1603–1613. https://doi.org/10.1080/00343404.2019.1596254

Evans, P. B. (2012). *Embedded autonomy: States and industrial transformation*. Princeton University Press.

Fan, J., Li, Y., Zhang, Y., Luo, X., & Ma, C. (2019). Connectivity and accessibility of the railway network in China: Guidance for spatial balanced development. *Sustainability*, *11*(24), 7099. https://doi.org/10.3390/su11247099

Farinha, T., Balland, P.-A., Morrison, A., & Boschma, R. (2019). What drives the geography of jobs in the US? Unpacking

relatedness. *Industry and Innovation*, 26(9), 988–1022. https://doi.org/10.1080/13662716.2019.1591940

Felipe, J., Kumar, U., Abdon, A., & Bacate, M. (2012). Product complexity and economic development. *Structural Change and Economic Dynamics*, 23(1), 36–68. https://doi.org/10.1016/j.strueco.2011.08.003

Felipe, J., Kumar, U., Usui, N., & Abdon, A. (2013). Why has China succeeded? And why it will continue to do so. *Cambridge Journal of Economics*, 37(4), 791–818. https://doi.org/10.1093/cje/bes058

Freeman, C., & Soete, L. (1997). *The economics of industrial innovation*. Psychology Press.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5), 685–697. https://doi.org/10.1080/00343400601120296

Gao, J. (2017, December 15–17). Maximizing the collective learning effects in regional economic development. In *2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing* (pp. 337–341). IEEE.

Gao, J., Zhang, Y.-C., & Zhou, T. (2019). Computational socioeconomics. *Physics Reports*, 817, 1–104. https://doi.org/10.1016/j.physrep.2019.05.002

Gao, J., & Zhou, T. (2018). Quantifying China's regional economic complexity. *Physica A: Statistical Mechanics and its Applications*, 492, 1591–1603. https://doi.org/10.1016/j.physa.2017.11.084

Garver, J. W. (2006). Development of China's overland transportation links with Central, South-West and South Asia. *The China Quarterly*, 185, 1–22. https://doi.org/10.1017/S0305741006000026

Gerschenkron, A. (1962). *Economic backwardness in historical perspective: A book of essays*. Belknap/Harvard University Press.

Goldin, C., & Katz, L. F. (2009). *The race between education and technology*. Harvard University Press.

Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., & Hidalgo, C. A. (2016). The research space: Using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics*, 109(3), 1695–1709. https://doi.org/10.1007/s11192-016-2125-9

Guo, Q., & He, C. (2017). Production space and regional industrial evolution in China. *GeoJournal*, 82(2), 379–396. https://doi.org/10.1007/s10708-015-9689-4

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press.

He, C., Zhu, S., Hu, X., & Li, Y. (2019). Proximity matters: Interregional knowledge spillovers and regional industrial diversification in China. *Tijdschrift voor Economische en Sociale Geografie*, 110(2), 173–190. https://doi.org/10.1111/tesg.12345

He, C., Zhu, S., & Yang, X. (2017). What matters for regional industrial dynamics in a transitional economy? *Area Development and Policy*, 2(1), 71–90. https://doi.org/10.1080/23792949.2016.1264867

Hidalgo, C. A. (2015). *Why information grows: The evolution of order, from atoms to economies*. Basic Books.

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, 3, 92–113. https://doi.org/10.1038/s42254-020-00275-1

Hidalgo, C. A., Balland, P. A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., & Neffke, F. (2018, July 22–27). The principle of relatedness. In *International Conference on Complex Systems* (pp. 451–457). Springer.

Hidalgo, C. A., Castañer, E., & Sevtsuk, A. (2020). The amenity mix of urban neighborhoods. *Habitat International*, 106, 102205. https://doi.org/10.1016/j.habitatint.2020.102205

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences, USA*, 106(26), 10570–10575. https://doi.org/10.1073/pnas.0900943106

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837), 482–487. https://doi.org/10.1126/science.1144581

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. *American Economic Review*, 76(5), 984–1001. https://www.jstor.org/stable/1816464

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3), 577–598. https://doi.org/10.2307/2118401

Jara-Figueroa, C., Jun, B., Glaeser, E. L., & Hidalgo, C. A. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences, USA*, 115(50), 12646–12653. https://doi.org/10.1073/pnas.1800475115

Jarreau, J., & Poncet, S. (2012). Export sophistication and economic growth: Evidence from China. *Journal of Development Economics*, 97(2), 281–292. https://doi.org/10.1016/j.jdeveco.2011.04.001

Jiao, J., Wang, J., Jin, F., & Dunford, M. (2014). Impacts on accessibility of China's present and future HSR network. *Journal of Transport Geography*, 40, 123–132. https://doi.org/10.1016/j.jtrangeo.2014.07.004

Jun, B., Alshamsi, A., Gao, J., & Hidalgo, C. A. (2020). Bilateral relatedness: Knowledge diffusion and the evolution of bilateral trade. *Journal of Evolutionary Economics*, 30(2), 247–277. https://doi.org/10.1007/s00191-019-00638-7

Klepper, S., & Sleeper, S. (2005). Entry by spinoffs. *Management Science*, 51(8), 1291–1306. https://doi.org/10.1287/mnsc.1050.0411

Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping knowledge space and technological relatedness in US cities. *European Planning Studies*, 21(9), 1374–1391. https://doi.org/10.1080/09654313.2012.755832

Laurenceson, J., & Chai, J. C. (2003). *Financial reform and economic development in China*. Edward Elgar Publishing.

Leftwich, A. (1995). Bringing politics back in: Towards a model of the developmental state. *Journal of Development Studies*, 31(3), 400–427. https://doi.org/10.1080/00220389508422370

Li, Z., & Xu, H. (2018). Highspeed railroads and economic geography: Evidence from Japan. *Journal of Regional Science*, 58(4), 705–727. https://doi.org/10.1111/jors.12384

Lin, J. Y. (2012). *New structural economics: A framework for rethinking development and policy*. World Bank.

Lin, J. Y., Cai, F., & Li, Z. (2003). *The China miracle: Development strategy and economic reform*. Chinese University Press.

Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from China's high speed railway system. *Journal of Urban Economics*, 98, 98–123. https://doi.org/10.1016/j.jue.2016.11.002

Lin, Y., Qin, Y., & Xie, Z. (2021). Does foreign technology transfer spur domestic innovation? Evidence from the high-speed rail sector in China. *Journal of Comparative Economics*, 49(1), 212–229. https://doi.org/10.1016/j.jce.2020.08.004

Miguelez, E., & Moreno, R. (2018). Relatedness, external linkages and regional innovation in Europe. *Regional Studies*, 52(5), 688–701. https://doi.org/10.1080/00343404.2017.1360478

Mokyr, J. (1992). *The lever of riches: Technological creativity and economic progress*. Oxford University Press.

Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3), 297–316. https://doi.org/10.1002/smj.2014

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, *87*(3), 237–265. https://doi.org/10.1111/j.1944-8287.2011.01121.x

Poncet, S., & De Waldemar, F. S. (2013). Export upgrading and growth: The prerequisite of domestic embeddedness. *World Development*, *51*, 104–118. https://doi.org/10.1016/j.worlddev.2013.05.010

Porter, M. E. (2008). *On competition*. Harvard Business Press.

Qian, Y. (2001). Government control in corporate governance as a transitional institution: Lessons from China. In J. E. Stiglitz & S. Yusuf (Eds.), *Rethinking the East Asian miracle* (pp. 295–322). World Bank.

Qin, Y. (2017). No county left behind? The distributional impact of high-speed rail upgrade in China. *Journal of Economic Geography*, *17*(3), 489–520. https://doi.org/10.1093/jeg/lbw013

Rodrik, D. (2006). What's so special about China's exports? *China & World Economy*, *14*(5), 1–19. https://doi.org/10.1111/j.1749-124X.2006.00038.x

Rolland, N. (2015). *China's new silk road*. National Bureau of Asian Research. http://www.nbr.org/research/activity.aspx?id=531

Rosenberg, N., Birdzell, L. E. Jr. (2008). *How the west grew rich: The economic transformation of the industrial world*. Basic Books.

Saviotti, P. P., & Frenken, K. (2008). Export variety and the economic performance of countries. *Journal of Evolutionary Economics*, *18*(2), 201–218. https://doi.org/10.1007/s00191-007-0081-5

Saviotti, P.-P., Pyka, A., & Jun, B. (2021). Diversification, structural change, and economic development. *Journal of Evolutionary Economics*, *30*, 1301–1335. https://doi.org/10.1007/s00191-020-00672-w

Song, Z., Storesletten, K., & Zilibotti, F. (2011). Growing like China. *American Economic Review*, *101*(1), 196–233. https://doi.org/10.1257/aer.101.1.196

Stojkoski, V., Utkovski, Z., & Kocarev, L. (2016). The impact of services on economic complexity: Service sophistication as route for economic growth. *PLoS ONE*, *11*(8), e0161633. https://doi.org/10.1371/journal.pone.0161633

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., & Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific Reports*, *2*(1), 723. https://doi.org/10.1038/srep00723

Vickerman, R. (2018). Can high-speed rail have a transformative effect on the economy?. *Transport Policy*, *62*, 31–37. https://doi.org/10.1016/j.tranpol.2017.03.008

Von Proff, S., & Brenner, T. (2014). The dynamics of inter-regional collaboration: An analysis of co-patenting. *The Annals of Regional Science*, *52*(1), 41–64. https://doi.org/10.1007/s00168-013-0573-1

Wade, R. (2004). *Governing the market: Economic theory and the role of government in East Asian industrialization*. Princeton University Press.

Xie, M., Bai, W., Wu, Q., Gao, J., & Zhou, T. (2020). Impact of high-speed railway on economic development. *Journal of University of Electronic Science and Technology of China*, *49*(6), 891–904. https://doi.org/10.12178/1001-0548.2020229

Yang, X., Lin, S., Zhang, J., & He, M. (2019). Does high-speed rail promote enterprises productivity? Evidence from China. *Journal of Advanced Transportation*, *2019*, 1279489. https://doi.org/10.1155/2019/1279489

Yu, F., Lin, F., Tang, Y., & Zhong, C. (2019). High-speed railway to success? The effects of high-speed rail connection on regional economic development in China. *Journal of Regional Science*, *59*(4), 723–742. https://doi.org/10.1111/jors.12420

Zheng, S., & Kahn, M. E. (2013). China's bullet trains facilitate market integration and mitigate the cost of megacity growth. *Proceedings of the National Academy of Sciences, USA*, *110*(14), E1248–E1253. https://doi.org/10.1073/pnas.1209247110

Zhu, S., He, C., & Luo, Q. (2019). Good neighbors, bad neighbors: Local knowledge spillovers, regional institutions and firm performance in China. *Small Business Economics*, *52*(3), 617–632. https://doi.org/10.1007/s11187-017-9975-2

Zhu, S., He, C., & Zhou, Y. (2017). How to jump further and catch up? Path-breaking in an uneven industry space. *Journal of Economic Geography*, *17*(3), 521–545. https://doi.org/10.1093/jeg/lbw047

Zhu, X. (2012). Understanding China's growth: Past, present, and future. *Journal of Economic Perspectives*, *26*(4), 103–124. https://doi.org/10.1257/jep.26.4.103