# Agglomeration: A Dynamic Approach*

W. Walker Hanlon
UCLA and NBER

Antonio Miscio
Columbia University

***Preliminary Draft***

December 15, 2013

**Abstract**

This paper introduces a dynamic approach for estimating the sources of agglomeration economies. Our framework allows us to simultaneously estimate the importance of agglomeration forces due to (1) cross-industry spillovers, (2) within-industry spillovers, and (3) overall city-size on the growth of city-industries. This is done while controlling for fixed locational fundamentals, city-specific shocks, and national industry growth rates. We implement the approach using detailed new data describing the industry composition of English cities from 1851-1911. We find strong evidence that cross-industry connections can influence industry growth, particularly for industries with small firms and more skilled workers. Within-industry effects are not present for most industries, but may be important in a small number of sectors characterized by large firms and low-skilled workers. Once we control for the role of industry composition, we find a strong negative relationship between city size and city-industry growth rates.

*We thank Don Davis, Jonathan Dingel, Matt Kahn and participants in the Columbia International Trade Colloquium, UCLA's KALER seminar, and the Urban Economics Association annual conference for their helpful comments and suggestions. Funding for this project was provided by a grant from UCLA's Ziman Center for Real Estate.

# 1 Introduction

At the heart of many theories of urban agglomeration lies the idea that firms benefit from proximity to one other. These localized benefits could take many forms, such as input-output connections, the benefits of sharing common labor pools, or the exchange of knowledge spillovers (Marshall (1890)). These benefits may be shared by firms within the same industry, they may flow between particular industry pairs, or the benefits may be due to overall city size and diversity.[1] Balancing these localized benefits in leading urban models are the costs of firm proximity, through increased competition for local consumers or congestion in the use of local inputs. Recently researchers have made important advances in understanding some of these agglomeration forces.[2] Yet important questions remain unanswered. How important are spillovers occurring within industries relative to spillovers flowing across industries? How do these benefits compare to the cost of proximity arising through congestion or competition? How can we separate all of these features from the fixed locational advantages of cities?

This study attempts to shed new light on these questions by introducing a dynamic panel-data approach. We begin by constructing a new set of city-industry panel data. Using these data, we investigate the extent to which employment growth in any industry i is influenced by the initial size of that industry, by spillovers from other industries through channels such as input-output connections, and by the initial size of the city overall. We calculate these relationships while controlling for time-varying industry-specific shocks at the national level as well as time-varying city-specific shocks. This approach offers several advantages over the static approach taken in other recent studies. First, using a dynamic approach allows us to deal with concerns about fixed locational fundamentals that affect the size of a city-industry. Dealing with fixed locational fundamentals has been a key challenge in the agglomeration literature. Existing studies generally rely on constructing measures of key locational fundamentals that can be used as control variables. But with panel data we can deal flexibly with any fixed locational features. A second advantage of our approach is that we can look at multiple channels – within-industry spillovers, cross-

[1]Glaeser *et al.* (1992) refer to within-industry and across industry spillovers as, respectively, Marshall-Arrow-Romer spillovers and Jacobs spillovers.

[2]Major recent contributions include Ellison *et al.* (2010), Greenstone *et al.* (2010), both of which focus on identifying the channels that drive inter-industry spillovers.

industry spillovers, city-size effects – at the same time. In contrast, recent studies have generally focused on one, or at most two, of these stories at once. Moreover, we can be more flexible when analyzing cross-industry spillovers, by allowing spillovers from industry i to j to differ from spillovers from j to i.[3]

Our primary data come from Britain and cover 25 of the largest English cities (based on 1851 population) for the period 1851-1911 (currently being extended to later years). Within each city, we construct employment for 27 industry groups. These industries span nearly the entire private sector economy, including services, so we can go beyond an analysis of manufacturing alone. As we will discuss, these data are unique in providing a high level of detail over such a long time period and a wide set of cities and industries. A second advantage offered by this setting is that we are able to look at a system of cities that should be near equilibrium.[4] A third advantage is that government intervention was insignificant, due largely to the limited size of the central government and the strong laissez-faire ideology that was dominant in Britain over most of this period.

Our estimation strategy is motivated by a simple model of city-industry growth. In the model, productivity growth in a city-industry in a period is determined by within and across-industry spillovers which depend on the level of initial employment in the spillover-producing industries in that period. In any period, the model is in spatial equilibrium and trade is costless, so more rapid productivity growth is reflected in industry employment. This main contribution of this theory is that it disciplines the empirical specification.

In estimating the relationship between employment in industry j in a city and the subsequent growth of industry i in the city, we may be concerned that employment in industry j is affected by factors that also affect growth in industry i,, such as spillovers from i to j in the previous period. We deal with this concern by using an approach suggested by Bartik (1991).[5] We use industry growth at the national level multiplied by lagged industry employment to generate *predicted* employment in industry j in a period. This predicted employment level is then used as an instrument for actual employment in industry j. Put another way, we take advantage of the

[3] This is not possible in the coagglomeration approach offered by Ellison *et al.* (2010).

[4] This contrasts with the system of cities in the U.S., where the same period was characterized by huge shifts toward the West, as well as the integration of the South into the national labor market.

[5] The Bartik approach is commonly used in studies in this literature. One recent example is Diamond (2012).

national industry growth rate to generate predicted industry employment levels within a city that are plausibly exogenous to local spillovers in the previous period. We apply the same methodology when we estimate the relationship between employment in industry j in a city and the subsequent growth of that industry.

It is not possible to estimate the full matrix of inter-industry spillovers because of the large number of coefficients involved. Thus, it is necessary to look for channels through which information is likely to flow between industries, and then use these channels to parameterize the inter-industry spillover terms. Previous research has suggested that information flows between firms that buy or sell to one another (see, e.g., Javorcik (2004) or Kugler (2006)) or through worker flows (see, e.g., Poole (2013) and Balsvik (2011)). Thus, we use input-output connections and industry occupational similarity to parameterize our cross-industry spillover terms.[6]

Our main results can be divided into those related to cross-industry spillovers, within-industry spillovers, and city-size effects. For cross-industry spillovers, we find strong evidence of dynamic effects. These operate primarily through forward linkages; employment in an industry grows more rapidly in cities in which many of that industry's suppliers are present. We also find some evidence of positive labor market pooling benefits, i.e., employment in an industry tends to grow more rapidly in cities with other industries that employ similar worker types. We find no evidence of positive effects occurring through backward linkages and some evidence that these effects may actually be negative. In other words, industries seem to grow more rapidly when more of their buyers are located outside of their city. This may suggest that rapid growth requires industries to look beyond supplying local consumers. Unpacking these results, we show that cross-industry spillovers have larger employment effects in tradable industries, and that they are stronger in industries with smaller firms and more educated workers.

On average, we find little evidence of positive within-industry spillovers, though there is substantial heterogeneity across industries. In most cases, the larger the is an industry at the beginning of a period, the slower it will grow. This is consistent with local competition forces outweighing positive localized spillover benefits. In contrast, a small number of industries, such as shipbuilding and textiles, do show positive and statistically significant within-industry benefits. The industries exhibiting positive

[6]This approach is similar to that used in Ellison *et al.* (2010), but because we are looking at dynamic effects, the motivation is slightly different.

within-industry spillovers are characterized by large average firm sizes, a high degree of spatial agglomeration, and less educated workforces.

For city size, we find strong evidence that, all else equal, industries grow more slowly in larger cities. Thus, we find no support for (dynamic) urbanization economies and strong evidence of congestion forces operating at the city level. At the same time, city growth is unrelated to initial city size, consistent with Gibrat's Law. In short, the negative congestion force generated by city size is balanced by the positive localized spillovers operating within and across industries in cities.

While the results described above focus on effects occurring within cities, we can extend our analysis to consider cross-city effects adjusting for the the distance from each city to the other cities in our sample. Specifically, we focus on how city-industry growth is affected by (1) access to consumers in other nearby cities and (2) cross-industry spillovers from other nearby cities. We find some evidence that city-industry growth is positively related to market access to other nearby cities, calculated using the distance weighted population of those cities. This result is consistent with Hanson (2005). We also find some evidence of cross-city cross-industry spillovers occurring between industries with similar labor forces. However, these effects are an order of magnitude smaller than the effect of spillovers within a city.

This project is related to a substantial literature investigating agglomeration forces. One strand of this literature, motivated by the endogenous growth literature, focuses on estimating the dynamic agglomeration forces generated by city charcteristics, such as size and industrial specialization or diversity (Glaeser *et al.* (1992), Henderson *et al.* (1995)).[7] The standard approach is to compare long-differenced growth of industry employment in a selection of industries in U.S. cities to fairly course measures of city features, such as industry concentration or a city's industrial diversity, in the base year. A more recent contribution to this literature, Henderson (2003), extends this approach using firm-level panel data.

A second and more recent strand of the literature exploits data on the structure connections between industries, such as input-output flows, labor force similarity, and measures of technological spillovers, in order to identify the channels that generate these agglomeration forces (Rosenthal & Strange (2001), Ellison *et al.* (2010), Faggio *et al.* (2013)). These studies use cross-sectional data and therefor focused on levels,

[7]A related paper, Kim (1995), focuses on similar issues at the regional level.

rather than growth effects. A concern with this cross-sectional approach is the role of fixed locational feathers. To deal with this, current studies include a vector of control variables representing a variety of locational features.

A third strand of related literature focuses on identifying the causal impact of changes in local economic activity on either productivity or employment (Greenstone *et al.* (2010) and Hanlon (2013)). These studies compare similar outcomes in similar locations, where some locations receive a plausably exogenous shock to the level of local economic activity. As a result, these studies are able to more clearly identify causal relationships, but only in special conditions.

Our study sits at the intersection of these three strands. We are interested in identifying dynamic agglomeration forces while at the same time shedding light on the channels that generate these effects. In terms of identification, we view our panel data approach and use of Bartic-style instrumentation as an improvement over previous studies looking at the broad econonomy. Identification using this approach is likely to be less clean than in studies utilizing natural experiments, but at the same time our approach has broader aplicability. Thus, we view our approach as complementary to previous studies focused on identifying causal links in more specific contexts. Finally, we are particularly interested in heterogeneity across industries in the importance of differetn agglomeration forces, a feature highlighted in a number of the studies cited above.[8]

The next section presents a simple theory of city-industry growth with inter-industry spillovers that is used to motivate our empirical specification. Section 3 describes the data. Our primary analysis, based on the British city-industry data set, is in Section 4 and Section 5 presents our main results. Section 6 extends the analysis to mechanisms that may operate across, rather than within, cities. Section 7 concludes.

[8]Studies were heterogeneity in the role of agglomeration forces across industries plays an important role include, among others, Henderson *et al.* (1995), Audretsch & Feldman (1996), Henderson (2003) and Faggio *et al.* (2013).

# 2 Theory

In this section we build a simple model of city growth incorporating localized spillovers within and across industries. The model clarifies the key forces that we are interested in and generates the empirical specification that we will take to the data. Since most of the elements of the model drawn directly from the existing literature, our description will be brief.

The theory focuses on localized spillovers that affect industry technology, and thereby influence industry growth rates. These dynamic effects are our primary object of interest. In order to keep things simple, we abstract from other localized factors that will affect the level of industry employment, such as the availability of cheaper inputs or the benefits of a larger labor pool.

The model is dynamic in discrete time. The dynamics of the model are driven by spillovers within and across industries which depend on industry employment and a matrix of parameters reflecting the extent to which any industry benefits from learning generated by employment in other industries (i.e., learning-by-doing spillovers). Because these dynamic effects are external to firms, they will not influence the static allocation of economic activity across space that is obtained given a distribution of technology levels. Thus, we can begin by focusing on the allocation of economic activity across space in any particular period. We then consider how the allocation in one period affects the evolution of technology and thus, the allocation of employment across city-industries in the next period.

## 2.1 Allocation within a static period

We begin by describing how the model allocates population and economic activity across geographic space within a static period, taking technology levels as given. The model economy is composed of many locations indexed by $c = \{1, ...N\}$. There are many industries indexed by $i = \{1, ...I\}$. Each industry produces one final good so final goods are also indexed by $i$. For simplicity, we assume that there are no transport costs required to move goods between locations, so each good will have the same price in every location.

Individuals are identical in all locations and, within any period, they consume an index of final goods given by $D_t$. The corresponding price index is $P_t$. The consumption idex takes a CES form,

$$D_t = \left( \sum_i \gamma_{it} x_{it}^{\rho} \right)^{\frac{1}{\rho}} \qquad , \qquad P_t = \left( \sum_i \gamma_{it}^{\sigma} p_{it}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

where $x_i$ is the quantity of good $i$ consumed, $\gamma_{it}$ is a distribution parameter that determines how important are the different final goods to consumers, $p_{it}$ is the price of final good $i$, and $\sigma$ is the (constant) elasticity of substitution between final goods. Note that we allow for changes in these preference parameters over time. It follows that the overall demand for any final good is,

$$x_{it} = D_t P_t^{\sigma} p_{it}^{-\sigma} \gamma_{it}^{\sigma}. \tag{1}$$

Production is undertaken by many perfectly competitive firms in each industry, indexed by $f$. Output by firm $f$ in industry $i$ is given by,

$$x_{icft} = A_{ict} L_{icft}^{\alpha} K_{icft}^{1-\alpha}, \tag{2}$$

where $A_{ict}$ is technology, $L_{icft}$ is labor input, and $K_{icft}$ is another input which we call *resources*. One important thing to note in this equation is that technology is not specific to any particular firm but that it is specific to each industry-location. This represents the idea that within industry-locations, firms are able to monitor and copy their competitors relatively easily, while information flows more slowly across locations.

Labor can move costlesly across locations to achieve spatial equilibrium. This is a standard assumption in urban economics models and one that is likely satisfied over the long time periods that we consider. The overall labor supply of labor to the economy is allowed to vary subject only to the fixed outside option wage $\bar{w}_t$ available to potential migrants. This reflects the fact that more successful cities (and countries) will attract migrants, but also that the wage needed to attract immigrants may vary over time.[9]

We also want to incorporate city-specific factors into our framework. Here we have

[9] One can think of this either as the wage that must be offered to immigrants or to a wage that is sufficient to attract labor to move from rural areas into the city.

in mind city-wide congestions forces, such as the price of housing, city-wide amenities, as well as the quality of city institutions. We incorporate these features by including a term $\lambda_{ct} > 0$ that represents a location-specific factor that affects the firm's cost of employing labor. The effective wage rate paid by firms in location $c$ is then $\bar{w}\lambda_{ct}$. When $\lambda_{ct} < 1$, the location is providing benefits that (net of their cost) increase the desirability of living in the city and thus reduce the effective wage rate. In contrast, values of $\lambda_{ct} > 1$ represent inefficient cities that increase the effective wage rate paid by firms in that location. In practice, this term will capture any fixed or time-varying city amenities or disamenities that affect all industries in the city.

In contrast to labor, resources are fixed geographically. They are also industry-specific, so that in equilibrium $\sum_f K_{icft} = \bar{K}_{ic}$, where $\bar{K}_{ic}$ is fixed for each industry-location and does not vary across time, though the level of $\bar{K}_{ic}$ does vary across locations. These fixed resources will be important for generating an initial distribution of industries across cities in our model, and allowing multiple cities to compete in the same industries in any period.

Firms solve:

$$\max_{L_{icft}, K_{icft}} p_{it} A_{ict} L_{icft}^{\alpha} K_{icft}^{1-\alpha} - \bar{w}_t \lambda_{ct} L_{icft} - r_{ict} K_{icft}.$$

Using the first order conditions, and summing over all firms in the industry-location, we obtain the following expression for employment in industry i and location c:

$$L_{ict} = A_{ict}^{\frac{1}{1-\alpha}} p_{it}^{\frac{1}{1-\alpha}} \left( \frac{\alpha}{\bar{w}_t \lambda_{ct}} \right)^{\frac{1}{1-\alpha}} \bar{K}_{ic}. \tag{3}$$

This expression tells us that employment in any industry $i$ and location $c$ will depend on technology in that industry-location, the fixed resource endowment for that industry-location, factors that affect the industry in all locations ($p_{it}$), city-specific factors ($\lambda_{ct}$), and factors that affect the economy as a whole ($\bar{w}_t$).

To close the static model, we need only ensure that income in the economy is equal to expenditures. This occurs when,

$$D_t P_t + M_t = \bar{w}_t \sum_c \lambda_{ct} \sum_i L_{ict} + \sum_i \sum_c r_{ict} \bar{K}_{ic}.$$

where $M_t$ represents net expenditures on imports. For a closed economy model we can set $M_t$ to zero and then solve for the equilibrium price levels in the economy.[10] Alternatively, we can consider a (small) open economy case where prices are given and solve for $M_t$. We are agnostic between these two approaches.

## 2.2 Dynamics: Technology growth over time

Technological progress in the model occurs through localized learning-by-doing spillovers. It is important that these spillovers are external to firms, so that firms are not forward looking when making their employment decisions within any particular period. We believe that this is a reasonable assumption over the long time-line that we consider, and it serves to greatly simplify the model.

Following the existing literature, we write the growth rate in technology as,

$$\ln\left(\frac{A_{ict+1}}{A_{ict}}\right) = S_{ict} + \epsilon_{ict}, \tag{4}$$

where $S_{ict}$ represent the amount of spillovers available in a city-industry in period t.[11] Some of the factors that we might consider are:

$$S_{ict} = f\Big( \text{within-industry spillovers, cross-industry spillovers,} \\ \text{national industry technology growth, city-level aggregate spillovers}\Big).$$

We can use Equation 4 to translate the growth in (unobservable) city-industry technology into the growth of (observable) city-industry employment. Starting with

[10]To solve for the price levels in the closed economy case, we use the first order conditions from the firm's maximization problem and Equation 3 to obtain,

$$p_{it} = \left(\frac{\alpha}{\bar{w}_t}\right)^{\frac{\alpha}{\alpha\sigma-\alpha-\sigma}} \left(\sum_c A_{ict}^{\frac{1}{1-\alpha}} \bar{K}_{ic} \lambda_{ct}^{\frac{\alpha}{\alpha-1}}\right)^{\frac{1-\alpha}{\alpha\sigma-\alpha-\sigma}} (D_t P_t^{\sigma})^{\frac{\alpha-1}{\alpha\sigma-\alpha-\sigma}} \gamma_{it}^{\frac{\sigma(\alpha-1)}{\alpha\sigma-\alpha-\sigma}}.$$

This equation tells us that in the closed-economy case, changes in the price level for goods produced by industry $i$ will depend on both shifts in the level of demand for goods produced by industry $i$ represented by $\gamma_{it}$, as well as changes in the overall level of technology in that industry (adjusted for resource abundance), represented by the summation over $A_{ict}$ terms.

[11]This formulation exactly matches that used in Glaeser *et al.* (1992) (Equation 6).

Equation 3 for period $t+1$, taking logs, and plugging in 4, we obtain,

$$\begin{aligned}\ln(L_{ict+1}) &= \left(\frac{1}{1-\alpha}\right)\ln(A_{ict}) + \left(\frac{1}{1-\alpha}\right)S_{ict} + \left(\frac{1}{1-\alpha}\right)\ln(P_{it+1}) + \left(\frac{1}{1-\alpha}\right)\ln(\alpha) \\ &- \left(\frac{1}{1-\alpha}\right)\ln(\lambda_{ct+1}) - \left(\frac{1}{1-\alpha}\right)\ln(\bar{w}_{t+1}) + \ln(\bar{K}_{ic}) + \left(\frac{1}{1-\alpha}\right)\epsilon_{ict}.\end{aligned}$$

Next, we use Equation 3 for period $t$, in logs, to substitute out $A_{ict}$ in the expression above, in order to obtain,

$$\begin{aligned}\ln(L_{ict+1}) - \ln(L_{ict}) &= \left(\frac{1}{1-\alpha}\right)\Bigg[S_{ict} + \Big[\ln(P_{it+1}) - \ln(P_{it})\Big] \\ &+ \Big[\ln(\lambda_{ct+1}) - \ln(\lambda_{ct})\Big] + \Big[\ln(\bar{w}_{t+1}) - \ln(\bar{w}_t)\Big] + e_{ict}\Bigg].\end{aligned}$$

where $e_{ict} = \epsilon_{ict+1} - \epsilon_{ict}$ is the error term.

The last step towards obtaining a usable basis for our empirical specification involves placing more structure on the spillovers term. Existing empirical evidence provides little guidance on what form this function should take, so here we follow the existing literature and take a fairly simple approach in which technology growth is a linear function of the sum of spillovers, so that

$$S_{ict} = \sum_k \tau_{ki}\ln(L_{kct}) + \xi_{it} + \psi_{ct}$$

where each $\tau_{ki} \in (0,1)$ is a parameter that determines the level of spillovers from industry $k$ to industry $i$. While admittedly arbitrary, this functional form does incorporates a number of desirable features. First, if there is very little employment in industry $k$ in location $c$ (e.g., $L_{kct} = 1$), then that industry makes no contribution to technology growth in industry $i$. Similarly, if $\tau_{ki} = 0$ then industry $k$ makes no contribution to technology growth in industry $i$. This specification does require that all industries have positive employment levels, such that $L_{ict} \geq 1$. As this is the case in the data we consider, we do not view this as an issue. This functional form also rules out complementarity between technological spillovers from different industries.

While such complementarities may exist, they are beyond the scope of the current paper. One thing to note about Equation 4 is that it will exhibit scale effects. While this may be a concern in other types of models, it is a desirable feature in a model of agglomeration economies, where these positive scale effects will be balanced by offsetting congestion forces.

Plugging this in, we obtain,

$$\begin{aligned}\ln(L_{ict+1}) - \ln(L_{ict}) &= \left(\frac{1}{1-\alpha}\right)\Bigg[\tau_{ii}\ln(L_{ict}) + \sum_{k\neq i}\tau_{ki}\ln(L_{kct}) + \xi_{it} + \psi_{ct} \\ &+ \Big[\ln(P_{it+1}) - \ln(P_{it})\Big] + \Big[\ln(\lambda_{ct+1}) - \ln(\lambda_{ct})\Big] \\ &+ \Big[\ln(\bar{w}_{t+1}) - \ln(\bar{w}_t)\Big] + e_{ict}\Bigg].\end{aligned} \tag{5}$$

To highlight that this expression incorporates both within and cross-industry spillovers, we have pulled the within-industry spillover term out of the summation. This equation expresses the change in log employment in industry $i$ and location $c$ in terms of (1) within-industry spillovers generated by employment in industry $i$, (2) cross-industry spillovers from other industries, (3) national industry-specific factors that affect industry $i$ in all locations, (4) city-specific factors that affect all industries in a location, and (5) aggregate changes in the wage (and thus national labor supply) that affects all industries and locations.

This expression for city-industry growth will motivate our empirical specification. One feature that is worth noting here is that we have two factors, city-level aggregate spillovers ($\psi_{ct}$) and other time-varying city factors ($\lambda_{ct}$), both of which vary at the city-year level. Empirically we will not be able to separate these positive and negative effects and so we will only be able to identify their *net* impact. Similarly, we cannot separate positive and negative effects that vary at the industry-year level.

# 3 Data

## 3.1 The British city-industry database

Our main database is drawn from the British Census of Population summary reports which were prepared by the Census Office. These data were collected by trained registrars during a relatively short time period, usually a few days in April of each census year. As part of the census, individuals were asked to provide one or more occupation. The primary occupation listed on the individual forms were then tabulated by the Census and summary reports were provided in printed format. To build our database we digitized hundreds of pages of printed documents.[12]

The cities included in the database are those that had a population of 50,000 or more in the 1851 census within the municipal boundaries.[13] To this were added a set of slightly smaller towns in Lancashire and Yorkshire for which data were previously available from Hanlon (2013). These towns are Blackburn, Halifax and Huddersfield.[14] This means that our database is slightly oversampling industrial cities. While the population of London was sometimes reported separately by borough, London is treated as one metropolitan area in the database. The geographic extent of these cities does change over time as the cities grow, a feature that we view as desirable for the purposes of our study[15].

Table 1 provides a list of the cities included in our main database, as well as the 1851 population of each city, the number of workers in the city in 1851, and the number of workers in 1851 that are working in one of the industry groups used in our analysis. A map showing the location of these cities in England is available in the

[12]We are grateful to the UK Data Archive for providing the scanned copies of the printed census reports. Because of the quality of the original documents they had to be digitized by hand-entry using a double-entry procedure to reduce errors.

[13]An exception to this rule was made for Wolverhampton, Staffordshire, with a population 49,985. Also, Plymouth is excluded from our database because in early years Plymouth data includes nearby Devonport while in later years it does not, resulting in an inconsistent series.

[14]The 1851 populations of these towns were: Blackburn 46,536, Halifax 33,582 and Huddersfield 30,880.

[15]The alternative is working with fixed geographic units. While that may be preferred for some types of work, given the growth that characterizes most of the cities in our sample, using fixed geographic units would mean either that the early observations would include a substantial portion of rural land surrounding the city, or that a substantial portion of city growth would not be part of our sample in the later years. Either of these options is undesirable. Other studies in the same vein, such as Michaels *et al.* (2013), also use metropolitan boundaries that expand over time.

Table 1: Cities in the primary analysis database

| City | Population in 1851 | Working population in 1851 | Workers in analysis industries in 1851 |
|---|---|---|---|
| Bath | 54,240 | 28,302 | 22,805 |
| Birmingham | 232,841 | 112,523 | 93,238 |
| Blackburn | 46,536 | 26,281 | 24,248 |
| Bolton | 61,171 | 31,291 | 28,617 |
| Bradford | 103,778 | 58,565 | 54,613 |
| Brighton | 69,673 | 33,521 | 27,129 |
| Bristol | 137,328 | 64,824 | 53,110 |
| Halifax | 33,582 | 18,159 | 16,162 |
| Huddersfield | 30,880 | 13,984 | 12,092 |
| Kingston-upon-Hull | 84,690 | 37,390 | 30,456 |
| Leeds | 172,270 | 83,980 | 73,480 |
| Leicester | 60,584 | 31,317 | 28,051 |
| Liverpool | 375,955 | 166,184 | 135,068 |
| London | 2,362,236 | 1,096,384 | 866,640 |
| Manchester | 367,232 | 205,314 | 180,839 |
| Newcastle-upon-Tyne | 87,784 | 38,804 | 32,133 |
| Norwich | 68,195 | 34,369 | 28,879 |
| Nottingham | 57,407 | 34,104 | 30,526 |
| Oldham | 52,820 | 38,932 | 35,690 |
| Portsmouth | 72,096 | 31,571 | 18,538 |
| Preston | 69,542 | 36,998 | 32,601 |
| Sheffield | 135,310 | 58,775 | 50,860 |
| Stockport | 53,835 | 30,209 | 27,632 |
| Sunderland | 63,897 | 24,978 | 21,253 |
| Wolverhampton | 49,985 | 22,844 | 19,423 |

Appendix. In general, our analysis industries cover most of the working population of the cities. Much of the remaining working population is employed by the government or in agricultural work.[16]

The occupations listed in the census reports closely correspond to industries, an important feature for our purposes. Examples from 1851 include “Banker”, “Glass Manufacture” or “Cotton manufacture”. The database does include a few occupations that do not directly correspond to industries, such as “Labourer”, “Mechanic”, or “Gentleman”, but these are a relatively small share of the population. These cat-

[16]For example, in Portsmouth, the large gap between working population and workers in the analysis industries is due to the fact that this was a major base for the Royal Navy.

egories are not included in the analysis. A major challenge faced in using these data is that the occupational categories listed in the census reports varied over time. To deal with this issue we combined multiple industries in order to construct consistent industry groupings over the study period. Individual categories in the years were combined into industry groups based on (1) the census' occupation classes, and (2) the name of the occupation. This process generates 27 consistent private sector occupation categories. Of these, 23 can be matched to input-output categories and used in the analysis.[17] Table 2 describes the industries included in the database.

It is worth pausing to highlight the differences between our British city-industry database and other potential alternatives. Focusing on U.S. data, there are two likely alternatives. The first is to use occupation and industry data from the U.S. Population Census (IPUMS). These data have been used in papers such as Katz & Margo (2013) and Michaels *et al.* (2013), but always in a more aggregated way than we need. The main issue with using these data for our purposes is that they are based on 1% or 5% samples, rather than the full census. One consequence is that measurement error becomes a major concern when disaggregating the data to city-industries, even when we focus on large municipalities. A second, more promising alternative, is to use County Business Patterns data. These data are publicly available starting in the 1970's, and in recent work, Duranton *et al.* (2013) have digitized earlier data dating to 1956. While the time-period covered by these data is more limited than the period available in our primary data, these data do offer the necessary level of detail. One additional concern with these data is that some employment is censored for some county-industries, which is not a concern in the British data.

[17] Another issue faced when using these data is that in some years occupation was reported for only some age groups. Specifically, the data for 1851-1861 are available divided into workers over 19 and workers under 20, by occupation, but in 1871, data by occupation are available only for workers over 19, while in 1881-1891 data are only available for all workers. It is, therefore, necessary to estimate values for 1871 employees under 20, as this was an important fraction of the labor force at this time. This is done by calculating the average ratio of all employees to employees over 20, in each industry and location, in 1851 and 1861, using data from the towns in Lancashire and Yorkshire available from Hanlon (2013). Only Lancashire and Yorkshire towns are used here because age-specific data have not been entered for the other towns. This value is then multiplied by the number of employees in each industry and location in 1871 in order to obtain 1871 values that are consistent with the other years. For the years 1901 and 1911 the data are available for workers over 10 years of age. It is assumed that this encompasses the entire workforce starting in 1901. The database includes both male and female workers.

Table 2: Industries in the primary analysis database with 1851 employment

| | | | |
|---|---|---|---|
| **Manufacturing** | | **Services and Professional** | |
| Chemicals & drugs | 17,814 | Engineers & Surveyors* | 2,288 |
| Dress | 320,613 | Clerks* | 27,108 |
| Instruments & jewelry* | 31,462 | General services | 454,825 |
| Earthenware & bricks | 18,247 | Merchant, agent, accountant, etc. | 30,492 |
| Leather & hair goods | 26,214 | Messenger, porter, etc. | 71,645 |
| Metal & Machines | 161,615 | Shopkeeper, salesmen, etc. | 26,570 |
| Oil, soap, etc. | 12,063 | | |
| Paper and publishing | 41,805 | **Transportation services** | |
| Shipbuilding | 13,962 | Railway transport | 9,878 |
| Textiles | 308,984 | Road transport | 34,771 |
| Vehicles | 8,609 | Sea & canal transport | 63,569 |
| Wood & furniture | 68,587 | | |
| | | **Food, etc.** | |
| **Others industries** | | Food processing | 111,316 |
| Building | 134,643 | Spiritous drinks, etc. | 7,892 |
| Mining | 22,920 | Tobacconists* | 3,224 |
| Water & gas services | 3,847 | | |

Industries marked with a * are available in the database but are not used in the baseline analysis because they cannot be linked to categories in the 1907 British input-output table.

## 3.2 Industry economic proximity measures

The second necessary piece of data for our analysis is a set of matrices measuring the pattern of connections between industries. Ideally we would like historical measures of the pattern of input-output connections, labor force similarity, and technological similarity, for each pair of industries in our data. However, the availability of such data for the period we study is necessarily sparse.

Our best measures are those reflecting input-output connections. Our main measure here is based on an input-output table constructed by Thomas (1987) based on the 1907 British Census of Production (Britain's first industrial census). This matrix is divided into 41 industry groups. We construct two variables: $IOin_{ij}$, which gives the share of industry i's intermediate inputs that are sourced from industry j, and $IOout_{ij}$ which gives the share of industry i's sales of intermediate goods that are purchased by industry j.

To measure labor force similarity between industry pairs, we take advantage of details about the characteristics of workers in each occupation available from the census data. For each occupation, we have information on the share of male and female workers, as well as the share of workers under 20. While rough, these divisions reflect important industry features in the 19th century that varied widely across sectors. For example, textile industries employed substantial amounts of female and child labor, while metal and heavy machinery industries employed few female or young workers. Our main measure of labor force similarity, $EMP_{ij}$, simply divides workers in each industry into these four available bins (male/female and over20/under20) and calculates the correlation in shares across industries. We also construct a measures based only on gender ($EMPgender_{ij}$) and ($EMPage_{ij}$). These measures are based on the absolute value of the difference between the gender or age shares for any pair of industries. Note that, unlike $EMP_{ij}$, these are *dissimilarity* measures.

The most difficult of the Marshallian connections to measure is technology spillovers. Currently, no matrix is available to measure the technological similarity of industries in the historical context we study, though we are working to construct one.

### 3.3 Additional data sets

We also collect data on a wide variety of other industry and city characteristics. The 1851 Census of Population was particularly detailed, and provides information on firm sizes in each industry and education levels in each city. From the 1907 Census of Production, Britain's first industrial census, we have collected data on the share of salaried workers in each industry as well as the industries coal usage. From the 1907 input-output table, we have measures of the share of industry output that is sold directly to households, as well as the share exported abroad. Finally, we collect data on the distance between cities from Google Maps, which we will use when considering cross-city effects in Section 6.

## 4 Empirical approach

The starting point for our analysis is based on Equation 5, which represents the growth rate of a city-industry as a function of the learning spillovers as well as time-varying city-specific and national industry-specific factors. Rewriting this as a regression equation we have,

$$\triangle \ln(L_{ict+1}) = \tilde{\tau}_{ii} \ln(L_{ict}) + \sum_{k \neq i} \tilde{\tau}_{ki} \ln(L_{kct}) + \theta_{ct} + \phi_{it} + e_{ict} \tag{6}$$

where $\triangle$ is the first difference operator, $\tilde{\tau}_{ii}$ and $\tilde{\tau}_{ki}$ include the coefficient $\left(\frac{1}{1-\alpha}\right)$, $\theta_{ct}$ is a full set of city-year effects and $\phi_{it}$ is a full set of industry-year effects. The first term on the right hand side represents within-industry spillovers, while the second term represents cross-industry spillovers. We purposely omitted the last term of Equation 5, namely $\triangle \ln(\bar{w}_{t+1})$, because although it could be estimated as a year-specific constant, it would be collinear with both the (summation of) industry-year and city-year effects. Moreover, in any given year we also need to drop one of the city or industry dummies in order to avoid collinearity. We chose to drop in all specifications the industry-year dummies associated with the "General services" sector.

One issue with Equation 6 is that there are too many parameters for us to credibly estimate given the available data.[18] In order to reduce the number of parameters,

[18] Our main dataset consists of $C = 25$ cities, $I = 27$ industries and $T + 1 = 7$ decades worth of

we need to put additional structure on the spillover terms. To do this, we look for measures reflecting the channels through which ideas may flow between industries. One channel suggested by the literature is that firms may share information with their customers or suppliers. For example, Javorcik (2004) and Kugler (2006) provide evidence that the presence of foreign firms (FDI) affects the productivity of upstream and downstream domestic firms. To reflect this channel, we will look at how growth in a city-industry is affected by the presence of upstream and downstream industries in the city. Another channel for knowledge flow may be the movement of workers, who may carry ideas between industries. Research by Poole (2013) and Balsvik (2011), using data from Brazil and Norway, respectively, has highlighted this channel. To reflect this, we will look at how growth in a city-industry is affected by the presence of other industries in the city that share similar labor pools, which is likely to influence the ability of worker to move between industries.

We parametrize the spillovers across industries, using the available input-output and occupational similarity matrices, in following way:

$$\tilde{\tau}_{ki} = \beta_0 + \beta_1 IOin_{ki} + \beta_2 IOout_{ki} + \beta_3 EMP_{ki} \quad \forall\, i,\, k$$

Substituting this into 6 we obtain:

$$\begin{aligned}\triangle \ln(L_{ict+1}) &= \tilde{\tau}_{ii}\ln(L_{ict}) + \beta_0 \sum_{k\neq i}\ln(L_{kct}) + \beta_1 \sum_{k\neq i} IOin_{ki}\ln(L_{kct}) + \beta_2 \sum_{k\neq i} IOout_{ki}\ln(L_{kct}) \\ &+ \beta_3 \sum_{k\neq i} EMP_{ki}\ln(L_{kct}) + \theta_{ct} + \phi_{it} + e_{ict} \qquad (7)\end{aligned}$$

Instead of a large number of parameters measuring spillovers across industry, Equation 7 now contains only three parameters multiplying three (weighted) summations of log employment. The first summation, where log employment in all industries $k \neq i$ are equally weighted, is very close to the log of total employment in the city and therefore we will interpret $\beta_0$ as the effect of city size on growth. The remaining summation terms use instead weights measuring input and output connections, and

data, thus we have 4,050 data points. Without imposing further structure, Equation 6 requires us to estimate 729 spillover coefficients and 306 fixed effects.

labor pooling. This will serve as the baseline specification in the analysis.

There are two issues to address at this point, both of which could violate the exogeneity restriction needed to estimate this equation. First, there could be a measurement error in $L_{ict}$. Since this variable appears both on the left and right hand side, this would mechanically generate an attenuation bias in our within-industry spillover estimates. Moreover, since $L_{ict}$ is correlated with the other explanatory variables, such measurement error would also bias the remaining estimates. We deal with measurement error in $L_{ict}$ on the right hand side by instrumenting it with what we will call henceforth a Bartik instrument, following an approach similar to Bartik (1991). Under the assumption that the measurement error in any given city-industry pair is *iid* across cities and time, our instrument is $L_{ict}^{Bart} = L_{ict-1} \times g_{i-ct}$, where $L_{ict-1}$ is the lag of $L_{ict}$ and $g_{i-ct}$ is the decennial growth rate in industry $i$ computed using employment levels in all cities *except* city $c$.

Second, even after addressing a potential measurement error in $L_{ict}$, the exogeneity restriction may still fail due to a simultaneity bias. For instance, if there is some factor not included in our model which causes growth in two industries $i$ and $k \neq i$ in the same city, a naive estimation would impute such growth to the spillover effect from $k$ to $i$, thus biasing the estimated spillover upward. We can control for this potential simultaneity bias also using a Bartik approach, whereby all the summation terms in Equation 7 such as $\sum_{k \neq i} IOin_{ki} \ln(L_{kct})$ will be instrumented with $\sum_{k \neq i} IOin_{ki} \ln(L_{kct}^{Bart})$, where $L_{kct}^{Bart}$ is computed as described above.

In addition, we need to discuss whether the presence of city-year effects $\theta_{ct}$ or industry-year effects $\phi_{it}$ raises concerns in the estimation of Equation 7. First, one may wonder whether the presence of a lagged dependent variable in a fixed effects estimation invalidates our estimation strategy. Notice that we would run into the type of endogeneity studied by Arellano & Bond (1991) if we had time-invariant city-industry effects. We do not have such effects because in the model itself these were captured by the term $\bar{K}_{ic}$ and were differenced out when we derived our estimating equation. Second, notice that the first two terms in Equation 7, namely $\ln(L_{ict})$ and $\sum_{k \neq i} \ln(L_{kct})$, are jointly collinear with the city-year effect $\theta_{ct}$. In our primary regression specifications, we drop $\sum_{k \neq i} \ln(L_{kct})$ and let $\theta_{ct}$ absorb all the effect of city size. In some alternative specifications we instead use a time-invariant city effect $\theta_c$ in order to retain the term $\sum_{k \neq i} \ln(L_{kct})$.

The estimation is performed using OLS and two-step GMM under a variety of assumptions for the standard errors, namely arbitrarily heteroskedastic and clustered by city. The different assumptions made on the standard errors do not matter for the point estimates obtained under OLS but they can affect the GMM results. To simplify the exposition, we will hereafter collectively refer to the set of regressors $\ln(L_{ict}), \quad i = 1...I$ as the *within* variables. Similarly, with a small abuse of notation the term $\sum_{k \neq i} \ln(L_{kct})$ will be referred to as *employment*, the term $\sum_{k \neq i} IOin_{ki} \ln(L_{kct})$ as $IOin$, and so on for $IOout$ and $EMP$. Finally, we will collectively refer to the latter terms as the *between* regressors since they are the parametrized counterpart of the spillovers across industries.

# 5 Main results

Our main regression results are based on the specification described in Equation 7. Regressions based on this specification generate results that can tell us about cross-industry spillovers, within-industry spillovers, city-wide factors, and industry-specific factors. In the following subsections, we will discuss results related to each of these forces in turn, but it is important to keep in mind that these results are coming out of regressions in which all of these factors are present. We begin by considering the pattern of spillovers across industries.

## 5.1 Cross-industry spillovers

Our estimation strategy involves using three proxies for the pattern of cross-industry spillovers: forward input-output linkages, backward input-output linkages, and occupational similarity. We begin our analysis, in Table 3 by looking at results that include only one of these proxies at a time. Columns 1-2 include only the forward input-output linkages. First we calculate results using the largest sample for which the IO measure is available. Then, for comparability to later results, we confine the sample to the set of observations for which measures of all three channels are available. The same exercise is done for backward input-output linkages in columns 3-4 and occupational similarity in columns 5-6. We find evidence of strong positive spillovers through forward input-output connections and some evidence of benefits operating through the occupational similarity channel. There is little evidence of benefits occurring through backward input-output linkages. In Table 4 we focus on the restricted sample and use Bartik instruments to address the potential endogeneity of the *within* and *between* variables. The results are qualitatively unchanged with the exception of the coefficient on $IOout$, which changes sign but remains statistically indistinguishable from zero.

Table 3: Regressions including only one spillover path at a time

| VARIABLES | (1)<br>lhs | (2)<br>lhs | (3)<br>lhs | (4)<br>lhs | (5)<br>lhs | (6)<br>lhs |
|---|---|---|---|---|---|---|
| IOin | 0.0848*** | 0.0662*** | | | | |
| | (0.0237) | (0.0093) | | | | |
| IOout | | | 0.0041 | 0.0041 | | |
| | | | (0.0085) | (0.0085) | | |
| EMP | | | | | 0.0033*** | 0.0021** |
| | | | | | (0.0013) | (0.0009) |
| Observations | 3,750 | 3,450 | 3,450 | 3,450 | 4,050 | 3,450 |
| estimation | ols | ols | ols | ols | ols | ols |
| FE1 | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year |
| FE2 | City*Year | City*Year | City*Year | City*Year | City*Year | City*Year |
| instruments | none | none | none | none | none | none |
| instrumented | none | none | none | none | none | none |
| SE | clustered | clustered | clustered | clustered | clustered | clustered |
| sample | whole | restrict | whole | restrict | whole | restrict |

Standard errors clustered by city. Regressors *within* and fixed effects included in all regressions but not displayed. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Next, consider all three channels together, while applying Bartik instrumentation to deal with endogeneity in industry employment. Our baseline results are displayed in Table 5. The first two columns present OLS results with either robust standard errors or standard errors clustered at the city level. In columns 3 and 4, we use the Bartik approach described above to instrument for the within variables, represented by $\ln(L_{ict})$ in Equation 7. Because this instrumentation approach requires the use of a lag, these regressions do not include observations from 1851. In columns 5-6, we use the Bartik approach to instrument for both the within and cross-industry spillover terms.

Across all specifications, we consistently find a positive and statistically significant coefficient on the $IOin$ coefficient. This suggests that industries grow more rapidly in

Table 4: Instrumented regressions including only one spillover path at a time

| VARIABLES | (1)<br>lhs | (2)<br>lhs | (3)<br>lhs | (4)<br>lhs | (5)<br>lhs | (6)<br>lhs |
|---|---|---|---|---|---|---|
| IOin | 0.0531*** | | | 0.0548*** | | |
| | (0.0103) | | | (0.0108) | | |
| IOout | | -0.0070 | | | -0.0086 | |
| | | (0.0083) | | | (0.0081) | |
| EMP | | | 0.0032*** | | | 0.0030** |
| | | | (0.0012) | | | (0.0012) |
| Observations | 2,871 | 2,871 | 2,871 | 2,871 | 2,871 | 2,871 |
| estimation | gmm | gmm | gmm | gmm | gmm | gmm |
| FE1 | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year |
| FE2 | City*Year | City*Year | City*Year | City*Year | City*Year | City*Year |
| instruments | Bartik | Bartik | Bartik | Bartik | Bartik | Bartik |
| instrumented | wtn | wtn | wtn | wtn,btn | wtn,btn | wtn,btn |
| SE | clustered | clustered | clustered | clustered | clustered | clustered |
| sample | restrict | restrict | restrict | restrict | restrict | restrict |

Standard errors clustered by city. Regressors *within* and fixed effects included in all regressions but not displayed. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Note that the number of observations falls for the instrumented regressions because the instruments require a lagged employment term. Thus, data from 1851 are not available for these regressions. Acronyms: wtn = *within*, btn = *between*.

cities with more employment in their supplier firms. In contrast, the *IOout* coefficients are uniformly negative and generally statistically significant, suggesting that firms do not grow faster in cities with a greater number of their customer industries. This may reflect that city-industries achieve fast growth by serving markets outside of their own city, rather than by focusing on local customers. However, comparing these results to Table 3 suggests that some of this effect may be due to correlations between our measures of different spillover channels. Thus, we do not interpret this as strong evidence of a negative effect. Finally, we find evidence that firms grow more rapidly in cities where there are more other firms employing a similar workforce. This result strengthens when we use the Bartic instrument. One potential explanation for this is that having more industries pulling from a similar local labor pool could have a negative short-run effect, through competition for workers, but a positive effect in the long-run. In that case, the Bartik instrumentation would deal with the short-term endogeneity, helping us identify the positive long-run effect. Additional results, available in Appendix A.2.1, show that we obtain similar results when London is excluded from the analysis.

Table 5: Baseline results – with all spillover channels

| VARIABLES | (1)<br>lhs | (2)<br>lhs | (3)<br>lhs | (4)<br>lhs | (5)<br>lhs | (6)<br>lhs |
|---|---|---|---|---|---|---|
| IOin | 0.0732*** | 0.0732*** | 0.0637*** | 0.0669*** | 0.0679*** | 0.0699*** |
| | (0.0094) | (0.0105) | (0.0099) | (0.0106) | (0.0102) | (0.0114) |
| IOout | -0.0152* | -0.0152* | -0.0262*** | -0.0247*** | -0.0286*** | -0.0273*** |
| | (0.0079) | (0.0084) | (0.0087) | (0.0090) | (0.0084) | (0.0089) |
| EMP | 0.0014 | 0.0014 | 0.0031*** | 0.0023*** | 0.0027** | 0.0022*** |
| | (0.0009) | (0.0010) | (0.0010) | (0.0005) | (0.0010) | (0.0005) |
| Observations | 3,450 | 3,450 | 2,871 | 2,871 | 2,871 | 2,871 |
| estimation | ols | ols | gmm | gmm | gmm | gmm |
| FE1 | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year |
| FE2 | City*Year | City*Year | City*Year | City*Year | City*Year | City*Year |
| instruments | none | none | Bartik | Bartik | Bartik | Bartik |
| instrumented | none | none | wtn | wtn | wtn,btn | wtn,btn |
| SE | robust | clustered | robust | clustered | robust | clustered |

Standard errors clustered by city. Regressors *within* and fixed effects included in all regressions but not displayed. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Note that the number of observations falls for the instrumented regressions in columns 3-6 because the instruments require a lagged employment term. Thus, data from 1851 are not available for these regressions. Acronyms: wtn = *within*, btn = *between*.

One potential concern with the results generated above is that some of our industries are non-traded. In this case, growth in local employment may not reflect growth in industry productivity. In contrast, this will not be a concern for tradable goods when labor is mobile, particularly given the long time-frame we study. To address this concern, Table 6 uses data on each industry's share of sales that are exported to divide industries into those above or below the median in terms of tradability. We then analyze industries with more and less traded good separately. We find stronger results for more-traded industries and weaker results for less-traded industries, though less traded industries still show positive and statistically significant growth effects from the local presence of supplier firms.

The results above reveal average patterns across all industries. We can further unpack these effects by estimating industry-specific coefficients for each of the spillover channels. Specifically, we replace $\beta_1$, $\beta_2$, and $\beta_3$ in Equation 7, with industry-specific coefficients $\beta_1^i$, $\beta_2^i$, and $\beta_3^i$. The estimated industry-specific coefficients are presented

Table 6: Baseline results – by tradability

| | More traded industries | | | Less traded industries | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| IOin | 0.0496*** | 0.0369** | 0.0484*** | 0.0505*** | 0.0606*** | 0.0524*** |
| | (0.0137) | (0.0164) | (0.0174) | (0.0191) | (0.0167) | (0.0178) |
| IOout | -0.0186** | -0.0291*** | -0.0324*** | -0.0093 | -0.0166 | -0.0144 |
| | (0.0089) | (0.0095) | (0.0096) | (0.0149) | (0.0135) | (0.0136) |
| EMP | 0.0027*** | 0.0038*** | 0.0035*** | -0.0154*** | -0.0041 | -0.0067 |
| | (0.0010) | (0.0013) | (0.0012) | (0.0045) | (0.0041) | (0.0043) |
| Observations | 1,800 | 1,496 | 1,496 | 1,650 | 1,375 | 1,375 |
| estimation | ols | gmm | gmm | ols | gmm | gmm |
| FE1 | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year |
| FE2 | City*Year | City*Year | City*Year | City*Year | City*Year | City*Year |
| instruments | none | Bartik | Bartik | none | Bartik | Bartik |
| instrumented | none | wtn | wtn,btn | none | wtn | wtn,btn |
| SE | clustered | clustered | clustered | clustered | clustered | clustered |

The median export share is used to split the sample in tradable vs non-tradable. Since we have an odd number of industries, the tradable sample contains one industry more than the non-tradable sample. Standard errors clustered by city. Regressors *within* and fixed effects included in all regressions but not displayed. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Acronyms: wtn = *within*, btn = *between*.

in the Appendix. We can compare these industry-specific cross-industry spillover coefficients to available information on industry characteristics, in order to identify the features of industries where each type of cross-industry spillover is important.

We focus on three features of industries for which data are available. First, we consider two measures of firm size in the industry in 1851. Second, we consider the share of industry output that goes to exports or that are sold directly to households. Third, we consider the skill level in the industry, using data on the share of workers paid on salary. In each case we run a simple univariate regression where the dependent variable is our industry-specific cross-industry spillover coefficient and the independent variable is one of the five industry characteristics.[19] The goal of these regressions is simply to identify the characteristics of industries that benefit from cross-industry spillovers; these results should not be interpreted as offering causal evidence.

Table 7 lists the coefficients generated by univariate regressions. Further details on the regression specifications used in this table, and full regression results, are available

[19]Univariate regressions are used because we are working with a relatively small number of observations.

in the Appendix. In rows 1-2, we see evidence that small firm size in an industry is associated with more cross-industry spillover benefits. Rows 3 and 4 show a weaker a relationship between who an industry sells to and the level of cross-industry spillover benefits, though there is some evidence that firms supplying more final goods to households may experience slightly larger cross-industry spillovers. In row 5, we see that high-skill industries – this with a larger share of salaried workers – are more likely to benefit from cross-industry spillovers.

Table 7: Features of industries that benefit from each type of cross-industry spillover – coefficients from univariate regressions

| | DV: Estimated industry-specific cross-industry spillover coefficients | | |
|---|---|---|---|
| Industry | Spillovers channel: | | |
| features: | IO-in | IO-out | EMP |
| Average firm size | -3.500*** | -3.548 | -2.739* |
| | (1.217) | (9.756) | (1.475) |
| Median worker's firm size | -0.466*** | -0.766 | -0.358* |
| | (0.134) | (1.138) | (0.174) |
| Share of industry output exported abroad | -0.726 | -1.131 | -0.877 |
| | (0.430) | (3.048) | (0.548) |
| Share of industry output sold to households | 0.320* | -1.172 | 0.384* |
| | (0.175) | (1.418) | (0.224) |
| Salaried worker share | 3.737*** | 6.041 | 2.560* |
| | (0.884) | (9.549) | (1.240) |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variables are the estimated cross-industry spillover coefficients for each industry and each spillover channel. More details are available in Appendix A.2.1. Firm size data comes from the 1851 Census of Population. The share of industry output exported or sold to households and is from the 1907 Input-Output matrix. The share of salaried workers in each industry is from the 1907 Census of Production

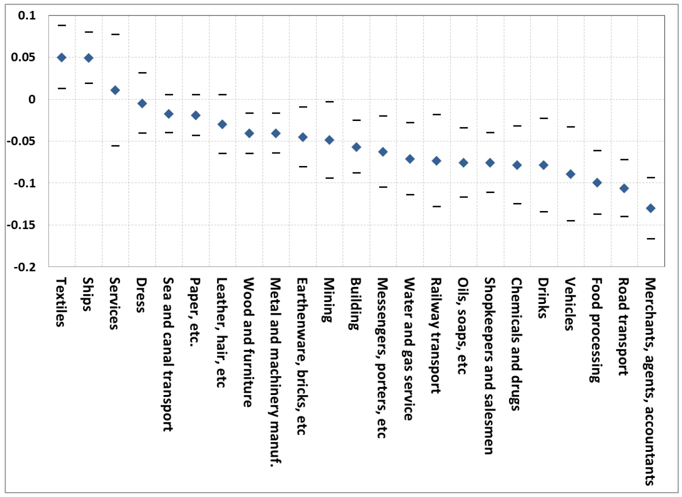
## 5.2 Within-industry spillovers

Our analysis can also help us understand the strength of within-industry spillovers.[20] These spillovers are reflected in the $\ln(L_{ict})$ term in Equation 6, which is an instrumented variable. Figure 1 presents the within-industry coefficients and 95% confidence intervals for regression specifications corresponding to columns 5 and 6 of Table 5. The top panel presents results from a specification with heteroskedasticity-robust standard errors, while the bottom panel uses standard errors that are clustered at the city level. We show both because, when using clustered standard errors, a few of the terms are dropped tue to colinearity. Both approaches display a similar pattern; localization economies appear to be an important positive factor in only a small number of industries, such as textile production and shipbuilding. These industries are characterized by increasing returns and strong patterns of geographic concentration. In contrast, we observe statistically significant negative coefficients for a number of industries, suggesting that there may be negative effects operating through channels such as competition for local customers or local inputs. Many of the industries showing negative coefficients are primarily non-tradable producers, such as "Water and Gas Service" to "Merchants, Agents, and Accountants".

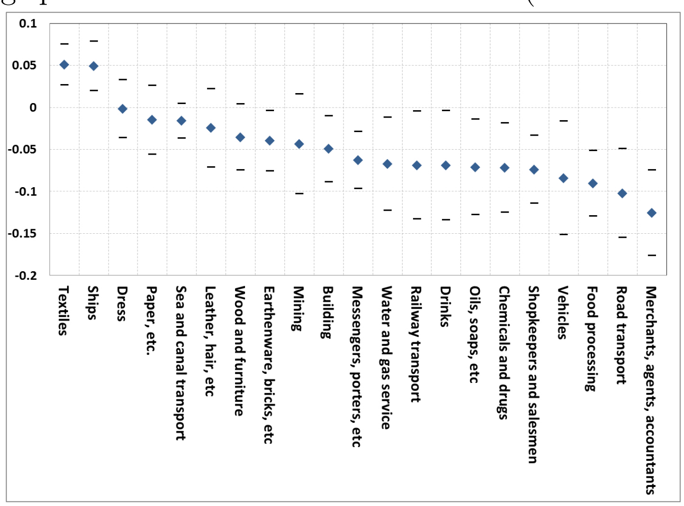
[20] In a static context these are often referred to as localization economies.

Figure 1: Strength of localization economies by industry

Using specification in column 5 of Table 5 (robust S.E.s)




Using specification in column 6 of Table 5 (clustered S.E.s)




Results are based on regression in columns 5 and 6 of Table 5. These regressions include a full set of city-year and industry-year terms, and both the within and between terms are instrumented using the Bartik approach. Estimation is done using a GMM approach.

In Table 8 we use additional data to consider some of the industry characteristics that may be generating the range of different within-industry spillover estimates we observe. We view these results as merely suggestive, since they are based on few observations. Columns 1-2 focus on the role of firm size using two different measures. We observe a strong positive relationship between firm size in an industry and the strength of within-industry spillovers.[21] The third and fourth columns look at the buyers served by each industry. We find no relationship between within-industry spillovers and the importance of exports. However, within-industry spillovers are associated with a lower share of industry output going directly to households, though this relationship is statistically weak. In the last column, we show that there is a negative relationship between the share of high-skilled workers in an industry, proxied by the share of workers paid on salary, and the level of within industry spillovers. Overall, we find that within-industry spillovers were positive and substantial only for a small number of industries. These industries were characterized by large manufacturing firms employing relatively low-skilled workers.

Table 8: Correlates of within-industry spillovers

| | DV: Estimated industry-specific within-industry spillover coefficients | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Average firm size | 0.517** | | | | |
| | (0.227) | | | | |
| Median worker's firm size | | 0.0653** | | | |
| | | (0.0260) | | | |
| Share of industry output exported abroad | | | 0.116 | | |
| | | | (0.0743) | | |
| Share of industry output sold to households | | | | -0.0639* | |
| | | | | (0.0329) | |
| Salaried worker share | | | | | -0.442** |
| | | | | | (0.178) |
| Constant | -0.0815*** | -0.0622*** | -0.0659*** | -0.0233 | 0.00403 |
| | (0.0182) | (0.0114) | (0.0139) | (0.0163) | (0.0209) |
| hline Observations | 20 | 20 | 23 | 23 | 20 |
| R-squared | 0.224 | 0.260 | 0.105 | 0.152 | 0.255 |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The number of observations varies because the explanatory variables are drawn from different sources and are not available for all industries. The within coefficients come from the specification used in column 6 of Table 5 (clustered S.E.s). Firm size data comes from the 1851 Census of Population. The export's and household's share of industry output come from the input-output table. The share of industry workers paid on salary comes from the 1907 Census of Production.

[21]More data on firm size by industry are available in the Appendix.

### 5.3 City-wide effects

Next, we want to look for effects operating at the city level. In particular, we are interested in the effect of city size on city-industry growth. City size may reduce city growth through congestion forces, but may also increase city-industry growth if there are substantial agglomeration benefits from being in a large city. Such aggregate city-size agglomeration forces play a role in existing theories, such as Davis & Dingel (2012). In addition to city size, we will also investigate the importance of a city's industrial diversity, which has been considered in several previous studies. The motivation for that interest comes from work by Jacobs (1969), who suggests that industries may benefit from locating near a diverse set of other industries.

We begin by focusing on the effect of city size in 1851 on city growth during the 1851-1911 period. In the left-hand panel of Figure 2, we plot the relationship between each city's working population in 1851 and the average growth rate of city's working population over 1851-1911.[22] There is no evidence of a strong relationship between city size and city growth, consistent with Gibrat's law (Gabaix (1999)). However, this relationship may be influenced by the type of industries found in a city, and the connections between them.

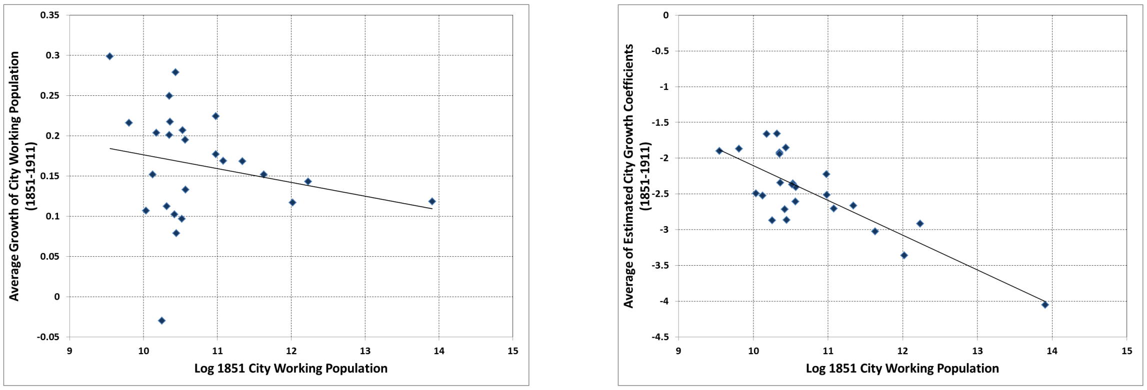
Our methodology allows us to isolate city-size effects from other factors related to a city's industrial composition. These city-size effects are reflected in the estimated coefficients on the city-year indicator variables in our analysis. In the right-hand panel of Figure 2, we plot these estimated coefficients, averaged at the city level over the 1851-1911 period, against the log of city working population in 1851. We now see evidence of a strong negative relationship between the average of estimated city-year coefficients and initial city size. This supports the idea that industry growth can face a substantial drag from congestion forces related to overall city size, but that this congestion drag is offset by localized industry spillovers.[23]

[22]This relationship is essentially unchanged if we focus on total city population rather than the city's working population.

[23]A similar relationship emerges if we focus on overall city size, rather than a city's working population. These results are also robust to excluding London, which is an outlier in terms of initial city size.

Figure 2: The effect of city size on city growth

City size vs. actual city growth | City size vs. estimated city growth coefficients




| | **DV: Average growth of city working population** |
|---|---|
| Log city working population in 1851 | -0.017 |
| | (0.016) |
| Constant | 0.348* |
| | (0.169) |
| R-squared | 0.05 |
| Observations | 25 |

| | **DV: Average of estimated city growth coefficients** |
|---|---|
| Log city working population in 1851 | -0.486*** |
| | (0.075) |
| Constant | 2.758*** |
| | (0.815) |
| R-squared | 0.64 |
| Observations | 25 |

The left-hand graph shows the relationship between city growth during the 1851-1911 period and each city's working population in 1851. The right-hand graph shows the relationship between an average over our estimated city-year growth coefficients for 1851-1911 and each city's working population in 1851. The city-year growth coefficients are estimated using the specification described in Column 3 of Table 5. The tables below each graph show corresponding OLS regression results for the fitted line. *** p<0.01, ** p<0.05, * p<0.1.

To generate more rigorous results, we use a regression specification,

$$\hat{\theta}_{ct} = \alpha_0 + \alpha_1 \ln(CitySize1851_c) + \mu_t + u_{it}$$

where $\hat{\theta}_{ct}$ is the estimated city-time fixed effect from our baseline regressions (Equation 7), $\ln(CitySize1851_c)$ is the log of the working population of the city in 1851 and $\mu_t$ is a set of year fixed effects. Column 1 of Table 9 presents the relationship between log initial city size and *actual* city growth rates. The coefficient is small and not statistically significant, consistent with the results described above. Column 4 presents the relationship between log initial city size and the *estimated* city-time growth coefficients, i.e., the relationship while controlling for factors related to a city's industrial composition. Here we observe a clear statistically significant negative relationship between log city size and city growth. Given that the log of initial city size in our database has a mean of 10.75 and standard deviation of 0.9, our results suggest that a one standard deviation in log initial size size lowers city growth by about 9% over a decade. Moreover, we can see that city size now explain a large fraction of the variation in city growth.[24]

Columns 2 and 4 conduct a similar exercise with a city's industrial diversity. Once we control for each city's industrial composition and spillovers between industries, there is a clear positive relationship between the initial concentration of city industries and the average of estimated city growth coefficients for 1851-1911. The mean and standard deviation of the city Herfendahl variable are, respectively, 0.23 and 0.11. Thus, a one standard deviation increase in a city's Herfendahl index increase city growth by about 6%. We can only speculate on the channels that may be generating this relationship. One possible channel is that cities with a more concentrated industry mix are better able to coordinate on growth-enhancing policies or public goods investments which improve growth across all industries in the city. Columns 3 and 6 combine the two effects. The negative city-size and positive city-diversity effects remain when cotrolling for a city's industrial composition.

An alternative approach for highlighting the role of city size is to alter Equation 7 by replacing the city-year effects $\theta_{ct}$ with city fixed effects $\theta_c$ and including a city size term $\beta_0 \sum_{k \neq i} \ln(L_{kct})$. The estimated $\beta_0$ coefficient then describes the relationship between employment in all city-industries other than industry $i$ and the growth rate of industry $i$. Regression results generated using this approach, which are available in Table 15 in Appendix A.2.2, show a negative and statistically significant relationship between city size and industry employment growth.

Two lessons can be drawn from these results. First, the impact of a city's size and industrial diversity on city-industry growth become much clearer once we control for other city-industry effects, and these two factors can explain a large fraction of the remaining variation in city growth. This has important implications for how we study these city-level factors. Second, we can see that city size exerts a strong negative effect on city growth, consistent with an important role for congestion operating at

[24]This result is not driven by the inclusion of year effects in the regression. If we do not include these effects the estimated coefficients are essentially unchanged and the R-squared statistics rises from 0.013 in the specification corresponding to column 1 to 0.235 in the specification corresponding to column 4.

the aggregate city level. However, we observe no clear relationship between initial city size and *actual* city growth, suggesting that these negative forces just balance the positive spillover benefits experienced by industries, such that Gibrat's Law holds.

Table 9: Influence of city size and diversity on city growth

| Dependent variable: | Actual city pop. growth rate | | | Estimated city growth coefficient | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Log city working population in 1851 | -0.0208 | | -0.0242 | -0.0894*** | | -0.0765*** |
| | (0.0161) | | (0.0169) | (0.0131) | | (0.0132) |
| Herfendahl of city-industry concentration | | -0.0323 | -0.0911 | | 0.539*** | 0.353*** |
| | | (0.133) | (0.139) | | (0.116) | (0.108) |
| Constant | 0.434** | 0.218*** | 0.491** | 0.670*** | -0.417*** | 0.450*** |
| | (0.177) | (0.0449) | (0.197) | (0.143) | (0.0393) | (0.153) |
| Year effects | X | X | X | X | X | X |
| Observations | 125 | 125 | 125 | 125 | 125 | 125 |
| R-squared | 0.062 | 0.049 | 0.065 | 0.400 | 0.293 | 0.450 |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The actual city population growth refers only to the city's working population. The estimated city growth coefficients are the city-year effects coefficient estimates corresponding to the regression shown in Column 6 of Table 5, where we instrument for both within and cross-industry terms. All estimates use observations from 1861-1911, with 1851 data used to generate the Bartik instrument for 1861 values.

# 6 Extensions

In this section, we extend our analysis to consider the possibility that city-industry growth may be influenced not just by factors within the city, but also through the influence of other nearby cities. We consider two potential channels for this supra-city effects. First, industries may benefit from proximity to consumers in nearby cities. This *market potential* effect is motivated by research by Hanson (2005), who finds that regional demand linkages play an important role in generating spatial agglomeration using modern U.S. data. Second, industries may benefit from spillovers from other industries in nearby towns, through any of the channels that we have identified.

There is substantial variation in the proximity of cities in our database to other nearby cities (see the Appendix for a map). Some cities, particularly those in Lancashire, west Yorkshire, and the North Midlands, are located in close proximity to a number of other nearby cities. Others, such as Norwich, Hull, and Portsmouth are located a relatively long distance from other cities.

We begin our analysis by collecting data on the distance (as a crow flies) between each of the cities in our database, which we call $distance_{ij}$. Using these, we construct two measures for the remoteness of one city from another, $d1_{ij} = 1/distance_{ij}$ and $d2_{ij} = exp(-distance_{ij})$.[25] Our measures of market potential for each city are then,

$$MP1_{ct} = \sum_{j \neq c} POP_{jt} * d1_{cj} \qquad MP2_{ct} = \sum_{j \neq c} POP_{jt} * d2_{cj}.$$

where $POP_{jt}$ is the population of city $j$. This differs slightly from Hanson's approach, which uses income in a city instead of population. Unfortunately, we do not have data allowing us to calculate income in a city.

We also want to measure the potential for cross-industry spillovers occurring across industries. We measure proximity to an industry $i$ in other cities as the distance weighted sum of log employment in that industry across all other cities. Our full regression specification, including both cross-city market potential and spillover effects, is then,

$$\begin{aligned}
\triangle \ln(L_{ict+1}) &= \tilde{\tau}_{ii} \ln(L_{ict}) + \beta_0 \sum_{k \neq i} \ln(L_{kct}) \\
&+ \beta_1 \sum_{k \neq i} IOin_{ki} \ln(L_{kct}) + \beta_2 \sum_{k \neq i} IOout_{ki} \ln(L_{kct}) + \beta_3 \sum_{k \neq i} EMP_{ki} \ln(L_{kct}) \\
&+ \beta_4 \sum_{k \neq i} IOin_{ki} \sum_{j \neq c} d1_{jc} * \ln(L_{kjt}) + \beta_5 \sum_{k \neq i} IOout_{ki} \sum_{j \neq c} d1_{jc} * \ln(L_{kjt}) \qquad (8) \\
&+ \beta_6 \sum_{k \neq i} EMP_{ki} \sum_{j \neq c} d1_{jc} * \ln(L_{kjt}) \\
&+ MP1_{ct} + \theta_c + \phi_{it} + \epsilon_{ict}.
\end{aligned}$$

When estimating this equation, note that we include city fixed effects ($\theta_c$) in place of city-year effects. This is necessary because city-year effects would be perfectly correlated with our market potential measure. To simplify the exposition and in analogy with the previous section, we will refer to the cross-city term $\sum_{k \neq i} IOin_{ki} \sum_{j \neq c} d1_{jc} *$

[25] The first of these is attractive because of its simplicity, while the second is motivated by Hanson (2005).

$\ln(L_{kjt})$ as $IOin * d1$, and similarly for the other cross-city terms $IOout * d1$ and $EMP * d1$.

The results generated using this specification are shown in Table 10. The first thing to take away from this table is that our baseline results are essentially unchanged when we include these additional variables, with the exception of the effects reflected in the occupational similarity measure, which appear to weaken. The coefficients on the market potential measure are always positive, and sometimes statistically significant. This is consistent with the idea that a city's market access can contribute positively to city-industry growth. The results provide no evidence of cross-city spillovers occurring through either of the input-output channels. There is, however, some evidence of cross-city spillovers between industries employing similar labor forces. These cross-city effects are an order of magnitude smaller than the estimated within-city coefficients in our preferred specifications (columns 4 or 6). Thus, we find some evidence that industries benefit from cross-industry spillovers at a supra-city level, but these effects appear to be much smaller in magnitude than those occurring within a city.

Table 10: Regression results with cross-city variables

| VARIABLES | (1)<br>lhs | (2)<br>lhs | (3)<br>lhs | (4)<br>lhs | (5)<br>lhs | (6)<br>lhs |
|---|---|---|---|---|---|---|
| IOin | 0.0651*** | 0.0638*** | 0.0672*** | 0.0734*** | 0.0651*** | 0.0678*** |
| | (0.0107) | (0.0107) | (0.0118) | (0.0114) | (0.0119) | (0.0125) |
| IOout | -0.0303*** | -0.0305*** | -0.0299*** | -0.0294*** | -0.0304*** | -0.0300*** |
| | (0.0090) | (0.0090) | (0.0091) | (0.0091) | (0.0091) | (0.0091) |
| EMP | 0.0026** | 0.0028*** | 0.0018* | 0.0017 | 0.0023** | 0.0020 |
| | (0.0011) | (0.0011) | (0.0011) | (0.0012) | (0.0011) | (0.0012) |
| employment | -0.0141*** | -0.0143*** | -0.0122*** | -0.0135*** | -0.0138*** | -0.0140*** |
| | (0.0023) | (0.0022) | (0.0019) | (0.0022) | (0.0023) | (0.0023) |
| MP1 | 0.1625* | | | | 0.1762* | |
| | (0.0947) | | | | (0.0972) | |
| MP2 | | 0.2572*** | | | | 0.3034 |
| | | (0.0950) | | | | (0.1913) |
| IOin*d1 | | | -0.0089 | | -0.0118 | |
| | | | (0.0234) | | (0.0236) | |
| IOout*d1 | | | 0.0011 | | 0.0015 | |
| | | | (0.0102) | | (0.0102) | |
| EMP*d1 | | | 0.0008 | | 0.0007 | |
| | | | (0.0008) | | (0.0008) | |
| IOin*d2 | | | | 0.0031* | | 0.0007 |
| | | | | (0.0018) | | (0.0028) |
| IOout*d2 | | | | -0.0010 | | -0.0006 |
| | | | | (0.0010) | | (0.0011) |
| EMP*d2 | | | | 0.0002* | | 0.0002* |
| | | | | (0.0001) | | (0.0001) |
| Observations | 2,871 | 2,871 | 2,871 | 2,871 | 2,871 | 2,871 |
| estimation | gmm | gmm | gmm | gmm | gmm | gmm |
| FE1 | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year |
| FE2 | City | City | City | City | City | City |
| instruments | Bartik | Bartik | Bartik | Bartik | Bartik | Bartik |
| instrumented | wtn,btn | wtn,btn | wtn,btn | wtn,btn | wtn,btn | wtn,btn |
| SE | robust | robust | robust | robust | robust | robust |

Standard errors are clustered by city to deal with serial correlation concerns. Regressors *within* and fixed effects included in all regressions but not displayed. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Acronyms: wtn = *within*, btn = *between*.

# 7 Conclusion

This paper introduces a dynamic approach to studying agglomeration forces and uses it to generate a rich set of results describing the nature and importance of agglomeration forces in generating employment growth in Britain from 1851-1911. Our framework allows us to investigate a range of potential agglomeration and dispersion forces – including within and across-industry spillovers, city-size congestion effects, market access, and supra-city spillovers – in a single unified framework. To our knowledge, this is the first study to analyze such a wide range of agglomeration theories in a unified framework.

Our results reveal substantial evidence of spillovers across industries. These spillovers operate through both input-output linkages and labor force similarity. However, we add to these results by showing that these linkages drive industry growth. Moreover, for input-output linkages, we show that it is the presence of suppliers, not customers, that plays the crucial role. These findings are consistent with, but also add to, the findings of Ellison *et al.* (2010) and Greenstone *et al.* (2010). We also find that spillovers within industries play little positive role in industry growth for most industries, but that they can have a positive effect for a few large-scale manufacturing industries such as textiles or shipbuilding. Our analysis cannot differentiate spillovers across firms in the same industry from within-firm increasing returns, so this finding may reflect either of these factors. Our overall finding that within-industry spillovers are relatively rare fits previous findings from the FDI literature suggesting that spillovers from FDI firms occurred primarily across industries rather than within them (Aitken & Harrison (1999), Javorcik (2004), Kugler (2006), et al.).

We find substantial heterogeneity across industries in the importance of both within and across-industry spillovers. Specifically, industries that benefit more from cross-industry spillovers are characterized by smaller firms and higher-skilled workers, while within-industry spillovers are more important in industries with large firms and less-skilled workers. We are not the first to highlight the potential for such heterogeneity; this point appears in Henderson (2003) and, more recently, in Faggio *et al.* (2013).

We show that, once we control for the role of industries, city size has a strong negative effect on city growth. Thus, we reject a positive role of urbanization economies operating outside of industry spillovers in favor of congestion forces. This finding fits well with many leading urban models, where city-level congestion effects push against industry agglomeration forces. It is particularly interesting that the positive effect of industry spillovers balances the negative city size effects such that Gibrat's law holds.

How exactly this balance is achieved is an interesting topic for further research.

# References

Aitken, Brian J, & Harrison, Ann E. 1999. Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. *American Economic Review* , **89**(3), 605–618.

Arellano, Manuel, & Bond, Stephen. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, **58**(2), 277–297.

Audretsch, DB, & Feldman, MP. 1996. R&D spillovers and the geography of innovation and production. *American Economics Review*, **86**(3), 630–640.

Balsvik, Ragnhild. 2011. Is labor mobility a channel for spillovers from multinationals? Evidence from Norwegian manufacturing. *Review of Economics and Statistics*, **93**(1), 285–297.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Davis, Donald R, & Dingel, Jonathan I. 2012 (June). *A Spatial Knowledge Economy.*

Diamond, Rebecca. 2012 (November). *The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000.*

Duranton, Gilles, Morrow, Peter M, & Turner, Matthew A. 2013 (July). *Roads and Trade: Evidence from the U.S.* Working Paper.

Ellison, G., Glaeser, E., & Kerr, W. 2010. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, **100**(3), pp. 1195–1213.

Faggio, Guilia, Silva, Olmo, & Strange, William. 2013 (October). *Heterogeneous Agglomeration.* Working Paper.

Gabaix, X. 1999. Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, **114**(3), 739–767.

Glaeser, Edward L, Kallal, Hedi D, Scheinkman, Jose A, & Shleifer, Andrei. 1992. Growth in Cities. *Journal of Political Economy*, **100**(6), 1126–1152.

Greenstone, Michael, Hornbeck, Richard, & Moretti, Enrico. 2010. Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants. *Journal of Political Economy*, **118**(3), pp. 536–598.

Hanlon, WW. 2013 (July). *Industry Connections and the Geographic Location of Economic Activity.* UCLA Ziman Center Working Paper No. 2013-11.

Hanson, Gordon H. 2005. Market potential, increasing returns and geographic concentration. *Journal of International Economics*, **67**(1), 1 – 24.

Henderson, J Vernon. 2003. Marshall's Scale Economies. *Journal of Urban Economics* , **53**(1), 1–28.

Henderson, V, Kuncoro, A, & Turner, M. 1995. Industrial Development in Cities. *Journal of Political Economy*, **103**(5), 1067–1090.

Jacobs, Jane. 1969. *The Economy of Cities*. New York: Vintage Books.

Javorcik, Beata S. 2004. Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages. *The American Economic Review*, **94**(3), 605–627.

Katz, Lawrence F., & Margo, Robert A. 2013 (February). *Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective*. NBER Working Paper 18752.

Kim, Sukkoo. 1995. Expansion of Markets and the Geographic Distribution of Economic Activities - The Trends in US Regional Manufacturing Structure, 1860-1987. *Quarterly Journal of Economics* , **110**(4), 881–908.

Kugler, Maurice. 2006. Spillovers from foreign direct investment: Within or between industries? *Journal of Development Economics*, **80**(2), 444–477.

Marshall, Alfred. 1890. *Principles of Economics*. New York: Macmillan and Co.

Michaels, Guy, Rauch, Ferdinand, & Redding, Stephen. 2013 (January). *Task Specialization in U.S. Cities from 1880-2000*. NBER Working Paper No. 18715.

Poole, Jennifer P. 2013. Knowledge Transfers from Multinational to Domestic Firms: Evidence from Worker Mobility. *The Review of Economics and Statistics*, **95**(2).

Rosenthal, Stuart S, & Strange, William C. 2001. The Determinants of Agglomeration. *Journal of Urban Economics* , **50**(2), 191–229.

Thomas, Mark. 1987. *An Input-Output Approach to the British Economy, 1890-1914*. Ph.D. thesis, Oxford University.

# A Appendix

## A.1 Data appendix

Table 11: Map showing the location of cities in the analysis database

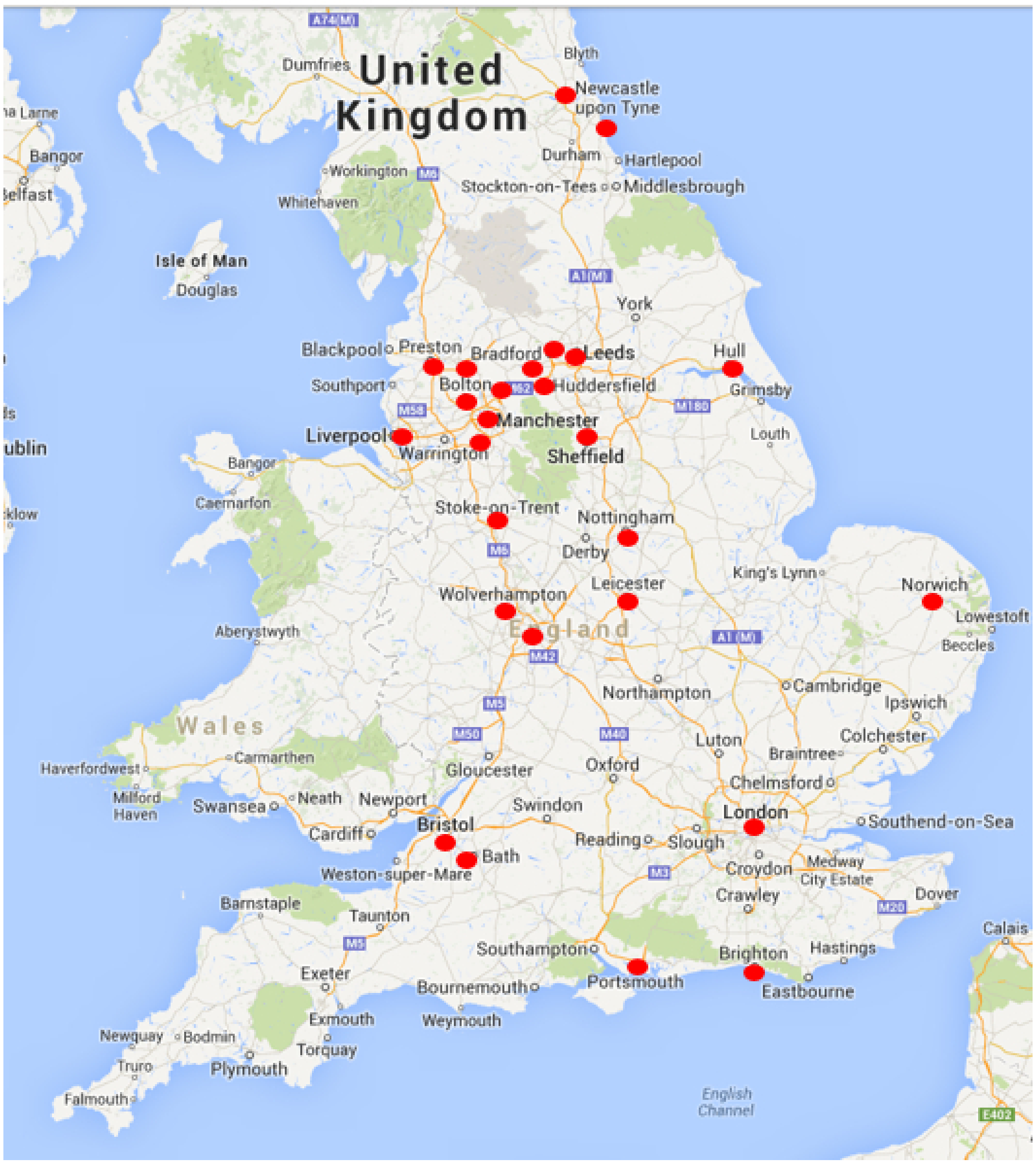


Table 12: Industry firm size data from 1851 Census of Population

| Industry | Median firm size | Median worker's firm size | Average firm size | Median firm size* | Median worker's firm size* | Average firm size* |
|---|---|---|---|---|---|---|
| Chemicals and drugs | 1 | 7 | 5 | 2 | 10 | 6 |
| Services | 1 | 1 | 2 | 1 | 2 | 3 |
| Dress | 1 | 5 | 4 | 2 | 6 | 6 |
| Merchants, agents, accountants | 2 | 9 | 6 | 3 | 10 | 7 |
| Shopkeepers and salesmen | 1 | 3 | 3 | 3 | 4 | 5 |
| Road transport | 1 | 3 | 3 | 3 | 5 | 5 |
| Sea and canal transport | 3 | 20 | 8 | 5 | 20 | 10 |
| Engineers and surveyors | 7 | 20 | 11 | 7 | 20 | 11 |
| Vehicles | 4 | 10 | 7 | 5 | 10 | 9 |
| Shipbuilding | 3 | 40 | 11 | 5 | 40 | 14 |
| Building | 2 | 7 | 5 | 3 | 8 | 7 |
| Food processing | 1 | 2 | 3 | 2 | 3 | 4 |
| Oils, soaps, etc | 1 | 5 | 4 | 2 | 6 | 5 |
| Leather, hair, etc | 2 | 8 | 5 | 3 | 9 | 7 |
| Drinks | 2 | 8 | 6 | 3 | 9 | 7 |
| Tobacco | 1 | 20 | 7 | 5 | 20 | 11 |
| Wood and furniture | 2 | 6 | 5 | 3 | 7 | 6 |
| Textiles | 3 | 150 | 17 | 5 | 150 | 22 |
| Paper, etc. | 2 | 20 | 8 | 4 | 20 | 11 |
| Mining | 3 | 50 | 13 | 4 | 50 | 15 |
| Earthenware, bricks, etc | 4 | 100 | 16 | 6 | 100 | 19 |
| Instruments and jewelry | 1 | 7 | 5 | 3 | 9 | 7 |
| Metal and machinery manuf. | 1 | 20 | 6 | 2 | 20 | 8 |

* Values in columns 3-6 are calculated dropping entries by masters reporting zero employees or not reporting the number of employees.

## A.2 Results appendix

### A.2.1 Cross-industry spillovers results appendix

We begin by showing results generated without including London. Estimates corresponding to our baseline results (Table 5) are shown in Table 13. Estimates differentiating tradable from non-tradable, corresponding to Table 6 in the main text, are shown in Table 14.

Table 13: Regression results without London

| VARIABLES | (1) lhs | (2) lhs | (3) lhs | (4) lhs | (5) lhs | (6) lhs |
|---|---|---|---|---|---|---|
| IOin | 0.0658*** | 0.0658*** | 0.0580*** | 0.0535*** | 0.0619*** | 0.0541*** |
| | (0.0097) | (0.0109) | (0.0103) | (0.0115) | (0.0107) | (0.0122) |
| IOout | -0.0089 | -0.0089 | -0.0188* | -0.0125 | -0.0217** | -0.0119 |
| | (0.0090) | (0.0093) | (0.0102) | (0.0093) | (0.0099) | (0.0102) |
| EMP | 0.0014 | 0.0014 | 0.0031*** | 0.0021*** | 0.0026** | 0.0019*** |
| | (0.0009) | (0.0011) | (0.0010) | (0.0005) | (0.0011) | (0.0006) |
| Observations | 3,312 | 3,312 | 2,756 | 2,756 | 2,756 | 2,756 |
| estimation | ols | ols | gmm | gmm | gmm | gmm |
| FE1 | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year |
| FE2 | City*Year | City*Year | City*Year | City*Year | City*Year | City*Year |
| instruments | none | none | Bartik | Bartik | Bartik | Bartik |
| instrumented | none | none | wtn | wtn | wtn,btn | wtn,btn |
| SE | robust | clustered | robust | clustered | robust | clustered |

These regressions replicate the baseline results from 5 but are based on a sample that excludes London. Regressors *within* and fixed effects included in all regressions but not displayed. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Acronyms: wtn = *within*, btn = *between*.

Table 14: Regression results without London, by tradability

| | Tradeable industries | | | Non-traded industries | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| IOin | 0.0415*** | 0.0293* | 0.0402** | 0.0586*** | 0.0687*** | 0.0611*** |
| | (0.0138) | (0.0166) | (0.0175) | (0.0194) | (0.0177) | (0.0186) |
| IOout | -0.0131 | -0.0232* | -0.0271** | -0.0071 | -0.0125 | -0.0103 |
| | (0.0100) | (0.0120) | (0.0124) | (0.0164) | (0.0150) | (0.0151) |
| EMP | 0.0027** | 0.0037*** | 0.0035*** | -0.0143*** | -0.0031 | -0.0060 |
| | (0.0010) | (0.0012) | (0.0011) | (0.0050) | (0.0045) | (0.0049) |
| Observations | 1,728 | 1,436 | 1,436 | 1,584 | 1,320 | 1,320 |
| estimation | ols | gmm | gmm | ols | gmm | gmm |
| FE1 | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year |
| FE2 | City*Year | City*Year | City*Year | City*Year | City*Year | City*Year |
| instruments | none | Bartik | Bartik | none | Bartik | Bartik |
| instrumented | none | wtn | wtn,btn | none | wtn | wtn,btn |
| SE | clustered | clustered | clustered | clustered | clustered | clustered |
| tradable | yes | yes | yes | no | no | no |

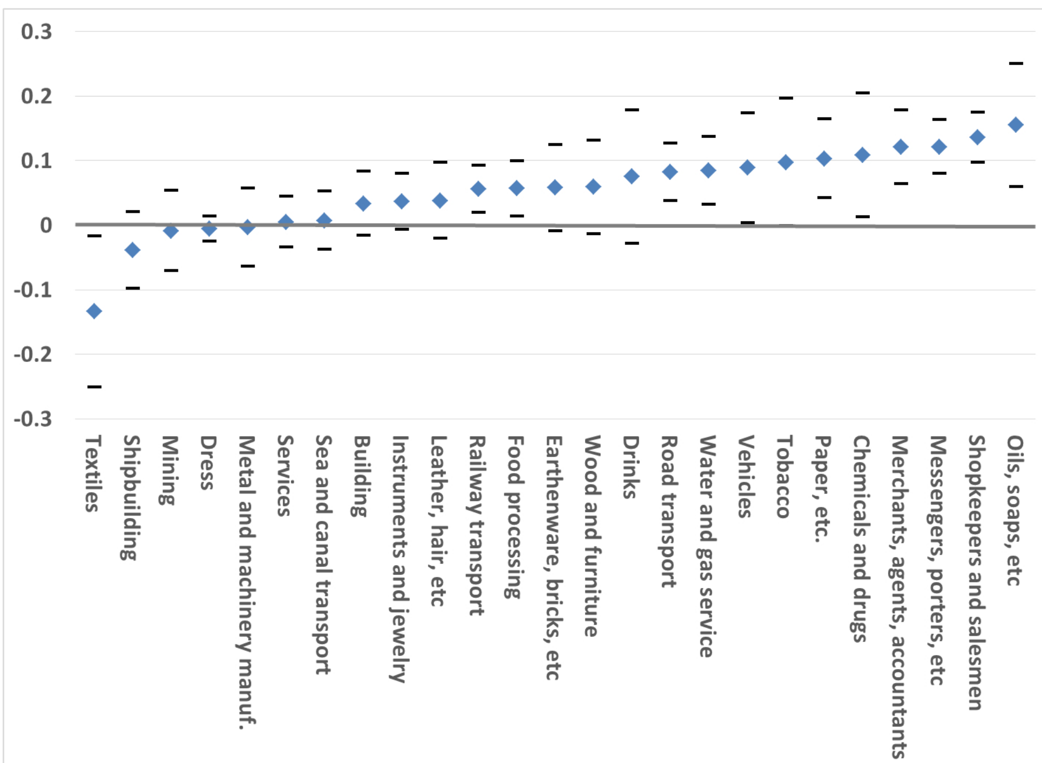
These regressions replicate the baseline results from 6 but are based on a sample that excludes London. Regressors *within* and fixed effects included in all regressions but not displayed. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Acronyms: wtn = *within*, btn = *between*.

Figures 3-5 present the estimated industry-specific cross-industry spillover coefficients for each of the spillover channel measures. Regressions are run with only one channel at a time to keep the number of estimated parameters manageable. Thus, the estimating equation for the first of these results is,

$$\triangle \ln(L_{ict+1}) = \tilde{\tau}_{ii} \ln(L_{ict}) + \beta_1^i \sum_{k \neq i} IOin_{ki} \ln(L_{kct}) + \theta_{ct} + \phi_{it} + \epsilon_{ict}.$$

The results are plotted in Figure 3 with 95% confidence intervals. These estimates are the dependent variable in the regressions shown in column 1 of Table 7.

Figure 3: Industry-specific cross-industry spillover coefficient estimates – IO in channel




A similar estimating equation is used for the IOout and EMP spillover channels. The result are shown in Figures 4 and 5, respectively. These estimates provide the dependent variables for, respectively, columns 2 and 3 of Table 7.

Figure 4: Industry-specific cross-industry spillover coefficient estimates – IO out channel

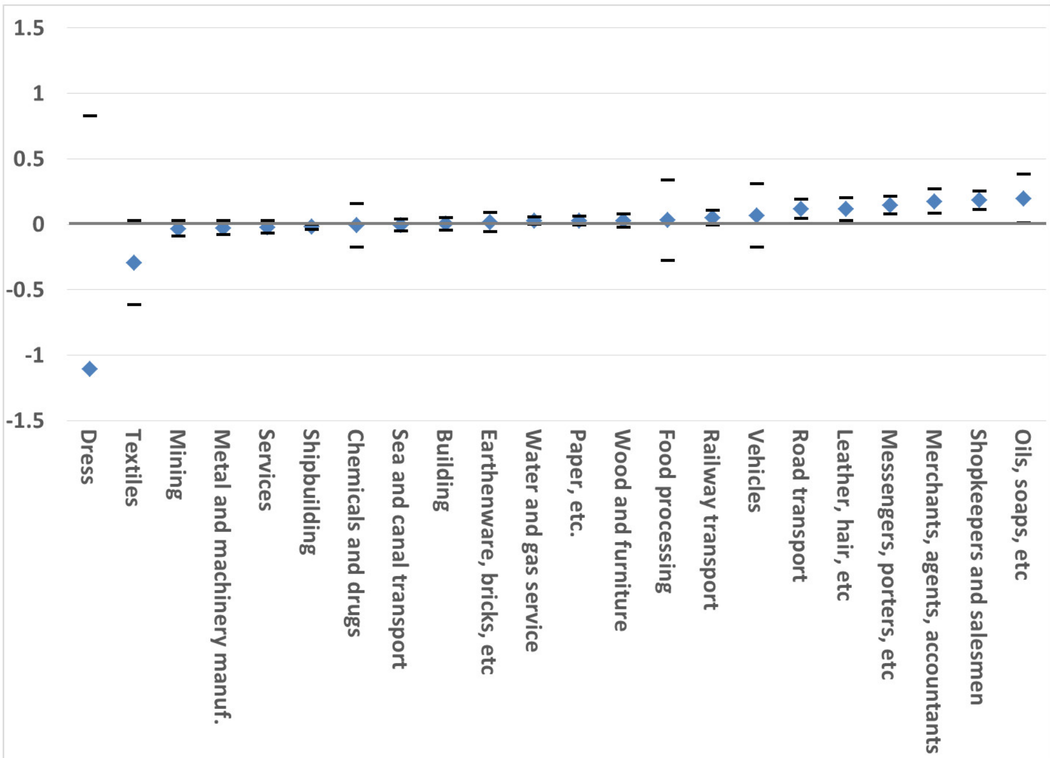
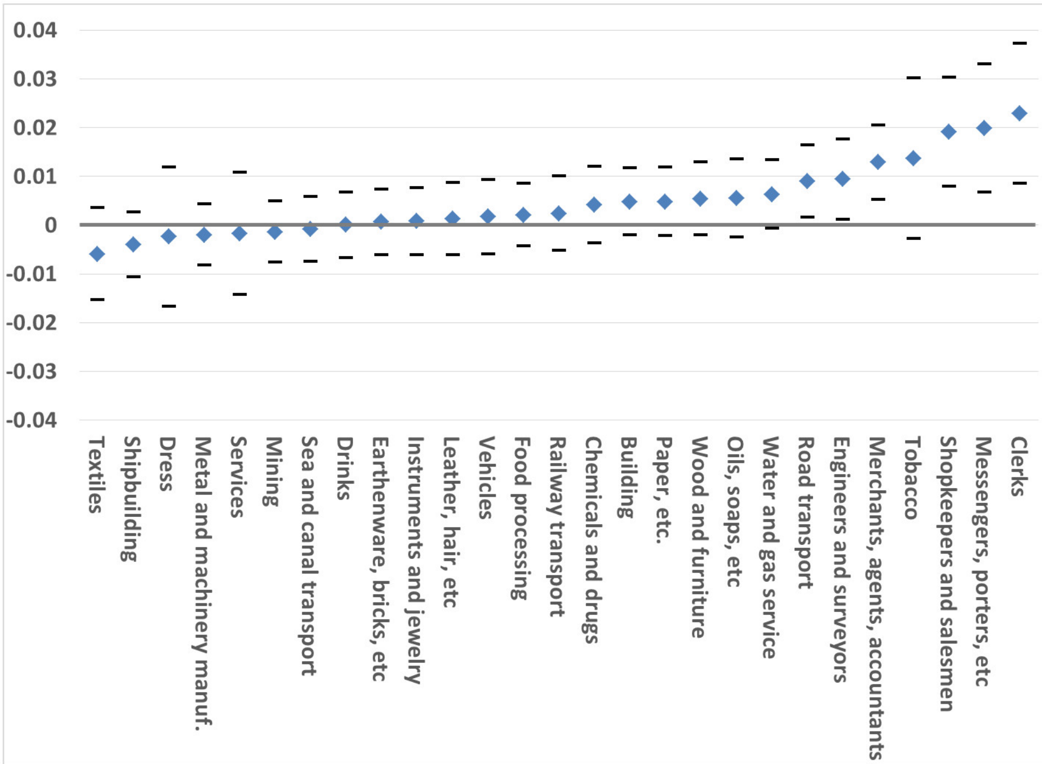


Figure 5: Industry-specific cross-industry spillover coefficient estimates – Employment similarity channel



### A.2.2 City-size results appendix

Table 15: Regression results with city size term and city fixed effects

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| VARIABLES | lhs | lhs | lhs | lhs | lhs | lhs |
| employment | -0.0098*** | -0.0098*** | -0.0154*** | -0.0151*** | -0.0125*** | -0.0121*** |
| | (0.0012) | (0.0017) | (0.0015) | (0.0028) | (0.0018) | (0.0030) |
| IOin | 0.0716*** | 0.0716*** | 0.0676*** | 0.0554*** | 0.0667*** | 0.0537*** |
| | (0.0097) | (0.0113) | (0.0103) | (0.0112) | (0.0106) | (0.0111) |
| IOout | -0.0155* | -0.0155* | -0.0260*** | -0.0173** | -0.0299*** | -0.0200*** |
| | (0.0086) | (0.0089) | (0.0093) | (0.0079) | (0.0090) | (0.0077) |
| EMP | 0.0011 | 0.0011 | 0.0019* | 0.0014*** | 0.0021** | 0.0015*** |
| | (0.0009) | (0.0011) | (0.0010) | (0.0004) | (0.0010) | (0.0004) |
| Observations | 3,450 | 3,450 | 2,871 | 2,871 | 2,871 | 2,871 |
| estimation | ols | ols | gmm | gmm | gmm | gmm |
| FE1 | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year | Ind*Year |
| FE2 | City | City | City | City | City | City |
| instruments | none | none | Bartik | Bartik | Bartik | Bartik |
| instrumented | none | none | wtn | wtn | wtn,btn | wtn,btn |
| SE | robust | clustered | robust | clustered | robust | clustered |

Employment represents the sum of employment in all industries in the city other than industry $i$, when looking at growth in industry $i$. Clustered standard errors are clustered by city to deal with serial correlation concerns. Regressors *within* and fixed effects included in all regressions but not displayed. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Note that the number of observations falls for the instrumented regressions in columns 3-6 because the instruments require a lagged employment term. Thus, data from 1851 are not available for these regressions. Acronyms: wtn = *within*, btn = *between*.