Local Productivity Spillovers

By NATHANIEL BAUM-SNOW, NICOLAS GENDRON-CARRIER, AND RONNI PAVAN*

Draft: June 27, 2023

Using Canadian administrative data, this paper presents evidence of revenue and productivity spillovers across firms at fine spatial scales. Accounting for the endogenous sorting of firms across space, we estimate an average elasticity of firm revenue and productivity with respect to the average quality of other firms within 75 meters of 0.024. We find scant evidence that the average firm benefits from being surrounded by a greater amount of economic activity at this spatial scale. Sorting of higher quality firms into more productive locations and higher average and aggregate quality peer groups is salient in the data.

Considerable evidence quantifies the scale and nature of agglomeration economies at the regional and local labor market levels. Greenstone, Hornbeck and Moretti (2010), Ellison, Glaeser and Kerr (2010), Bloom, Schankerman and Reenen (2013), Faggio, Silva and Strange (2017), Hanlon and Miscio (2017), and others all provide evidence that firm and worker productivity are increasing in the prevalence of nearby firms to which they are connected, with connectivity measured through input-output relationships, patent citations or occupational similarity. There is also extensive evidence that firms and workers in larger cities are more productive on average, with about half of city size wage premia driven by greater returns to work experience in larger cities (Baum-Snow and Pavan, 2012; De la Roca and Puga, 2017). The natural implication is that city scale enhances firm and worker productivity, likely in part through spillovers that operate between firms and workers at microgeographic spatial scales. Despite this extensive evidence for broad regions, little empirical evidence exists about the magnitude and composition of productivity spillovers at the very local level within cities. Evidence in the literature at microgeographic spatial scales is primarily descriptive (Duranton and Overman, 2005; Kerr and Kominers, 2015) or specific to certain narrowly defined industries (Rosenthal and Strange, 2003; Arzaghi and Henderson, 2008).

* Baum-Snow: Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, ON, Canada, M5S3E6. e-mail: Nate.Baum-Snow@rotman.utoronto.ca. Gendron-Carrier: Department of Economics, McGill University, 855 Sherbrooke St. West, Montreal, QC, Canada, H3A2T7. e-mail: nicolas.gendron-carrier@mcgill.ca. Pavan: Department of Economics, University of Rochester, 280 Hutchinson Road, Rochester, NY, USA, 14627. e-mail: rpavan@ur.rochester.edu. Rolando Campusano, Guangbin Hong, Xiaoyang Li, and Kenwin Maung provided superb research assistance. We thank Gabriel Ahlfeldt, Kirill Borusyak, Jorge De la Roca, Florian Mayneris, Eleonora Patacchini, Stephen Ross, William Strange, and many seminar and conference participants for useful comments.

Using panel data on high-skilled services firms in three large Canadian cities, this paper provides the first causal estimates of revenue and productivity spillovers at fine spatial scales for a broad set of firms, quantifies the underlying mechanisms driving these spillovers, and characterizes sorting patterns of firms across peer groups and locations. We find strong evidence of revenue and productivity spillovers that operate between firms within 75 meter radius peer group areas. We estimate an average elasticity of firm revenue and productivity to the average quality of other firms within 75 meters of 0.024, measuring quality as static firm-specific components of these outcomes. This estimate indicates that going from the 10^{th} to 90^{th} percentile of peer groups in our data increases revenue by 7 percent. Conditional on these linear-in-means type spillovers, we find scant evidence that the average firm benefits from being surrounded by a greater amount of economic activity at spatial scales smaller than 500 meter radius areas. Linear-in-means spillovers are found to be very local in nature. Tests for mediation of spillovers through various industry connections suggest that learning or knowledge transfer between nearby firms is the primary mechanism driving spillovers at microgeographic spatial scales. In particular, we find greater spillovers to firms operating in industries that typically hire workers from peers' industries and to firms that have more peers in 2-digit industries other than their own. Moreover, about two-thirds of linear-in-means spillovers are estimated to accrue from firms in the top tercile of the local firm quality distribution.

We see extensive evidence of non-random sorting of firms across peer groups and locations. Specifically, using estimates of firm quality, we show that higher quality firms tend to be located in peer groups of greater average and aggregate quality. Locations with better fundamentals also attract higher quality firms on average. Each of these patterns is more pronounced for above median quality firms. Externalities that increase in levels with both own firm quality and average peer quality incentivize the non-random sorting of larger and higher quality firms into better peer groups and locations. A positive equilibrium relationship between average and aggregate peer group quality ensues. Because the spillover process is linear-in-means, however, there are only small aggregate gains associated with the observed peer group composition relative to a random allocation of firms across locations. Absent consideration of potential general equilibrium effects and assuming homogeneous treatment effects, counterfactual allocations that randomly assign firms to peer groups reduce aggregate firm revenue by 0.27-0.74 percent through reductions in linear-in-means type spillovers, mostly because the highest quality firms experience smaller spillovers in this environment.

The use of restricted access administrative tax data on the universe of firms in Canada is central to this analysis. We use information on sales, inputs, factor prices, and postal codes for over 55,000 firms in more than 3,500 peer group locations for each year 2001-2012. We focus on the densest areas in Montreal, Toronto, and Vancouver, where postal codes are less than 75 meters in radius. As in De Loecker (2011), reasonable assumptions about the data generating process for revenue that accommodate variation in factor intensity and market power across sectors allow us to recover estimates of total factor productivity (TFP) in addition to revenue spillovers. We find that sizes and attributes of TFP and revenue spillovers are not statistically different.

Our empirical analysis adopts and extends a common specification in the peer effects literature into the context of interactions between firms, a context that has not been considered before in the literature in this way. In our empirical model, a firm's log revenue depends on a fixed firm-specific component and a weighted aggregate of this object for other firms in the peer group conditional on local area-year and industry-year fixed effects. Our key parameter of interest is the coefficient on this peer group aggregate. Arcidiacono et al. (2012) (AFGK) show how to estimate peer effects with panel data in analogous environments in which children may sort across classrooms on fixed unobserved attributes. We extend their setup to distinguish between the relative importance of aggregate versus linear-in-means type spillovers and to assess the relative importance of different types of industry connectivity weights. Through specification of the weights that aggregate peer attributes, we can measure each of these types of spillovers. Extension of the AFGK model to estimate the impacts of multiple types of spillovers simultaneously facilitates this analysis. Such "horse race" type specifications have not been explored much in the peer effects literature but are essential to recovering these important insights.[1]

Our fundamental source of identifying variation comes from changes in the composition of firms over time within small areas. We use this sort of variation to separately identify spillovers from location fundamentals or "contextual effects" of neighborhoods. In addition to selection on time-invariant unobserved attributes, one may be additionally concerned that firm location choices may depend on localized productivity, infrastructure, or worker amenity shocks. If neighborhoods with improving business environments attract higher quality new arrivals and those with deteriorating business environments see departures of higher quality firms, our spillover estimates would be overstated. On the other hand, if deaths of low quality firms disproportionately occur in poor business environments, our estimates would be understated. As examples of such neighborhood attributes that may matter, a refurbished road, a new transit station, or upgraded internet service may both promote improved outcomes for existing firms nearby and draw in new more productive firms. As such, the main threat to identification is that the quality of arriving or departing firms may be correlated with unobserved trends in neighborhood fundamentals.

[1] Conley et al. (2015) and Liu, Patacchini and Zenou (2014), which estimate spillover parameters in analyses of peer effects on studying effort and participation in school sport activities, are exceptions.

To account for the possibility that firms select locations in a way that is correlated with such location-specific shocks, our primary identification strategy takes advantage of the spatial granularity in our data and includes 500 meter radius area fixed effects interacted with year. Identifying variation comes from a combination of cross-sectional differences in firm composition in nearby 75 meter radius regions and differential changes in firm composition over time in these same peer groups when compared within larger 500 meter radius regions. The inclusion of neighborhood-year fixed effects coupled with changes over time in firm composition within peer groups allows us to identify peer effects separately from changes in location fundamentals. Controlling for firm fixed effects fully accounts for sorting across peer groups and locations on levels of firm quality.

The existence of frictions in commercial real estate markets in the central business district areas of large cities and our focus on high skilled service industries support our identification strategy. In order to hedge against business cycle risk, landlords typically rent out space on a rolling basis with 5-10 year commercial leases (Rosenthal, Strange and Urrego, 2021), generating smooth variation in tenant turnover and making it difficult for firms to coordinate on location. As a result, there are typically few options available for new commercial space within a 500 meter radius area in any given year. Therefore, the opportunity for firms to sort on changes in fundamentals at small spatial scales is very limited after controlling for neighborhood-year fixed effects. Bayer, Ross and Topa (2008) employ a similar strategy in the residential housing market context to quantify the extent to which neighbors provide each other with job referrals. Data from dense locations provides identifying variation while simultaneously making it unlikely that changes in firm location choices could be correlated with annual shocks to small area fundamentals. We perform a number of post-estimation identification checks using our estimates of firm quality. Among other things, we show that firm revenue residuals are not correlated with various attributes of future or past peer quality that could reflect trends in location fundamentals that operate at spatial scales smaller than 500 meter radius areas or selective migration. Our focus on high-skilled services that are traded beyond local neighborhoods reduces the possibility that very local shocks to demand conditions and associated changes in local output prices at spatial scales smaller than a 500 meter radius area may be driving results. Robustness checks that use model structure to account for endogenous price responses corroborate our more reduced form estimates.

One key goal of the analysis is to distinguish between linear-in-means and aggregate forms of spillovers. This distinction is important, as greater aggregate gains are typically available through the internalization of agglomeration type spillovers relative to the internalization of linear-in-means type spillovers. Many urban economic geography models that incorporate local agglomeration, from

Fujita and Ogawa (1982) to Ahlfeldt et al. (2015), abstract away from firm heterogeneity. Instead, they consider aggregate production functions for (implicitly) identical firms with constant returns to scale production. Rather than indexing TFP by firm, TFP is indexed by location and is typically an increasing function of nearby employment. This assumption about the form of agglomeration economies shapes a related empirical literature that focuses on finding scale effects using aggregate rather than firm level data. In contrast, the peer effects literature focuses primarily on estimating linear-in-means type spillovers between individuals and does not consider aggregate type spillovers (e.g., Guryan, Kroft and Notowidigdo, 2009; Cornelissen, Dustmann and Schonberg, 2017). As mean and aggregate peer firm quality are positively correlated in our context, credible estimates of each type of spillover requires considering both simultaneously in estimation. Otherwise, it is easy to confuse one type of spillover for the other. We hope our evidence on the relative importance of linear-in-means type spillovers sparks innovation in urban economic geography modeling to accommodate such essential firm heterogeneity.

At first blush, it might appear that our evidence that linear-in-means type spillovers dominate simple aggregation (agglomeration) spillovers is at odds with observed productivity and wage premia that are associated with city size. However, coupled with our evidence that higher quality firms experience larger spillovers in dollar terms from peer groups of the same quality than do lower quality firms, our baseline results indicate an important interaction between sorting and firm externalities that generates aggregate increasing returns at the city level. That is, the existence of larger and more productive firms in larger cities itself can generate agglomeration economies. All of this is consistent with Combes et al. (2012)'s evidence that static firm TFP distributions have higher means and more right dilation in larger cities. The "Plant Size-Place Effect" of larger firms in larger cities (Manning, 2009) also means there will be larger firm-to-firm spillovers in larger cities, resulting in higher aggregate productivity. This is the firm-level counterpart to Baum-Snow and Pavan (2012) and De la Roca and Puga (2017)'s evidence that workers' returns to experience are greater in larger cities, and that this profile is increasing in worker ability.

Methodologically, our investigation is similar to a number of papers in the peer effects literature. Perhaps most closely related, Cornelissen, Dustmann and Schonberg (2017) formulate a similar empirical model to ours, in which a worker's wage depends in part on spillovers from components of coworkers' wages that are fixed over time. Using administrative data from the Munich region in Germany, they estimate wage elasticities to averages of their peers amongst those working routine tasks within firms of about 0.05. In contrast to our results, they find smaller spillovers for more skilled occupations, indicating a very different process for human capital spillovers within than between firms. Our very localized evidence is in line with Moretti (2004), Kantor and Whalley (2014), and Serafinelli (2019)'s more macro evidence on knowledge flows that operate between firms.

We emphasize that while our analysis faces a number of identification challenges, we formulate our empirical model such that it is not subject to the reflection problem. Given the considerable empirical challenges associated with credible identification of "endogenous effects", in which a firm's outcome directly impacts peers' outcomes (Manski, 1993; Angrist, 2014), we do not attempt to isolate this component of our spillover estimates. Instead, we follow Gibbons, Overman and Patacchini (2015)'s advice and focus on estimating spillovers from exogenous attributes of nearby firms, as captured in their estimated fixed effects. Indeed, we think our setting is unlikely to generate much in the way of endogenous effects, as nearby firms in most industries have little reason to try to coordinate on revenue. Moreover, as we discuss further below, our empirical model and identification strategy are explicitly formulated to focus on the recovery of exogenous effects only.[2] Absent any endogenous effects, our elasticity estimates can be interpreted as the ratio of the impact of the aggregated exogenous attributes of peers to those of the firm's own exogenous attributes.

This paper proceeds as follows. In Section I, we develop a theoretical framework that justifies and interprets our use of revenue as the main outcome variable of interest. Section II describes our empirical model, identification, and estimation. Section III describes the data and sample. Section IV discusses the main results and identification checks. Section V presents counterfactuals oriented toward isolating the impacts of firm sorting. Finally, section VI concludes.

I. Theoretical Framework

In this section, we lay out a conceptual framework that delivers empirical specifications describing the operation of productivity spillovers between firms at microgeographic spatial scales. Beginning with a standard profit maximization problem, we derive an estimation equation in which a firm's log revenue (sales) depends on its own fixed effect and a weighted aggregate of the fixed effects of its peers. The key parameter of interest to be estimated is the elasticity of a firm's log revenue with respect to the weighted aggregate of its peers' fixed effects. We show that under certain conditions this parameter measures the average TFP spillover between firms within each peer group.

Our main estimation equation accommodates both perfectly and imperfectly competitive environments. If output prices are exogenous, time-differencing log revenue reveals that revenue innovations induced by changes in peer group composition must be related to changes in firm TFP, with an

[2] Credible evidence of endogenous productivity spillovers has used supply chain network structure for identification, as in Bazzi et al. (2017).

adjustment for the variable input share. If output prices are endogenous and specific to the firm, firm re-optimization in response to a positive TFP shock (and associated reduced marginal cost) results in a reduced firm-specific output price. The magnitude of this endogenous price response depends on both the size of the increase in TFP and the elasticity of demand faced by the firm. We derive an additional adjustment to account for this endogenous price response, allowing us to recover measures of TFP spillovers under imperfect competition as well with some modeling assumptions and parameter calibration.

A. Basic Setup

Each year, each firm chooses its variable input quantity L conditional on location. Because of commercial real estate market frictions, firms can change locations but cannot choose the exact block b in which to locate, only the broader neighborhood $B(b)$. Each block is associated with a fixed amount of space. The only way a firm can adjust its space input is to move to a different block. In the empirical work, we vary the size of the block by aggregating postal codes to areas of 75 to 250 meter radii within 500 meter radius broader neighborhoods.

The resulting short-run profit of firm i in block b and industry k at time t is

$$\pi_{i,b,k,t} = p_{i,b,k,t} A_{i,b,k,t} L_{i,b,k,t}^{\theta_k} - w_{B(b),k,t} L_{i,b,k,t} - F_{i,b,k,t}.$$

The key object of interest in this expression is TFP $A_{i,b,k,t}$, which is firm-year specific, and is influenced by location fundamentals, industry, and fixed attributes of neighboring firms. The variable input quantity is $L_{i,b,k,t}$, which we think of mostly as labor. For small adjustments in $L_{i,b,k,t}$, which may occur year to year in response to changes in $p_{i,b,k,t}$, $A_{i,b,k,t}$, and $w_{B(b),k,t}$, the short-run production technology is decreasing returns to scale. We allow the variable input share $\theta_k < 1$ to differ across industries. The input price $w_{B(b),k,t}$ is determined at a broader level of spatial aggregation $B(b)$ than the block and thus can be controlled for with local area and industry fixed effects interacted with time. If firms are price takers, the output price $p_{i,b,k,t} = p_{B(b),k,t}$ can also be controlled for with local area and industry fixed effects interacted with time. Empirically, we focus on the high-skilled services sector. As a result, output prices are likely to be determined at a broader level of spatial aggregation than the block, with no local price competition. With market power, output prices differ across firms as developed in Section I.D below and in Appendix A.A1. The fixed cost $F_{i,b,k,t}$ captures real estate and capital inputs. These are fixed in the short run but their implicit prices can vary over time and space.

Under perfect competition, firm log revenue in block b is

$$\ln R_{i,b,k,t} = \ln p_{B(b),k,t} + \ln A_{i,b,k,t} + \theta_k \ln L^*_{i,b,k,t}, \tag{1}$$

where $L^*_{i,b,k,t}$ is the variable input demand function. Substitution of the input demand function into equation (1) yields the following reduced form expression for log revenue:

$$\ln R_{i,b,k,t} = \frac{\theta_k}{1-\theta_k}\ln\theta_k + \frac{1}{1-\theta_k}\ln p_{B(b),k,t} + \frac{1}{1-\theta_k}\ln A_{i,b,k,t} - \frac{\theta_k}{1-\theta_k}\ln w_{B(b),k,t}. \tag{2}$$

The structural responses of the variable factor input to TFP and output price shocks are identical to those for revenue shown in equation (2).

The goal of the empirical work is to isolate revenue and productivity spillovers between firms from variation in peer composition and in firms' own log revenue or variable inputs. Doing so requires holding constant location-specific attributes of wages and output prices, which we control for with various fixed effects and adjustments described below. Conditional on output prices and wages, equation (2) indicates the extent to which shocks to log revenue that spill over from nearby firms fully reflect log TFP spillovers between firms. With a variable input share of 0.7, an observed 10 percent change in revenue would reflect a 3 percent change in TFP conditional on the output price and variable input cost.

B. TFP Spillovers

To complete the structural representation of our estimation equation, we specify the process through which we conceptualize TFP propagates between nearby firms. We allow firm i's TFP in year t to depend on a firm-specific component that is fixed over time α_i^A, spillovers from a weighted aggregate of this same object in all other firms j in block b at time t, and area-industry-time fixed effects. Put together, we have the following data generating process for firm i's TFP at time t:

$$\ln A_{i,b,k,t} = \alpha_i^A + \phi^A_{B(b),k,t} + \gamma^A\left[\sum_{j\in M_{b,t},\neq i}\omega_{ij}(M_{b,t})\alpha_j^A\right] + \varepsilon^A_{i,b,k,t}. \tag{3}$$

γ^A is the key object in this equation that we aim to estimate. It denotes the elasticity of firm i's TFP with respect to an aggregation of the firm-specific component of TFP that is fixed over time across other firms in firm i's peer group. $M_{b,t}$ is the set of firms in peer group location b in year t. Weights $\omega_{ij}(M_{b,t})$ are equal across peers and sum to one in the "linear-in-means" (LIM) specifications and sum to $|M_{b,t}|-1$ in "agglomeration" (Agg) specifications. Local area-industry-year fixed effects

$\phi^A_{B(b),k,t}$ capture a combination of location fundamentals and industry level TFP shocks.[3]

We follow much of the literature in conceptualizing firm quality α^A_i as being a fixed firm-specific component of TFP. We model spillovers to depend on aggregations of α^A_j only, with the idea that immutable firm attributes like management expertise and competence are more likely to influence nearby firms than transitory TFP shocks would. To the extent that TFP growth rates differ systematically across firms within local area, industry, and year fixed effects $\phi^A_{B(b),k,t}$, such growth rates will be positively correlated with average firm TFP. Even if higher firm quality promotes more rapid TFP growth, our specification will still attribute greater firm quality to more rapidly growing firms.[4]

In order to distinguish between mechanisms driving spillovers at a microgeographic scale, some of our empirical work jointly estimates linear-in-means and agglomeration type spillovers. In these cases, equation (3) becomes

$$\ln A_{i,b,k,t} = \alpha^A_i + \phi^A_{B(b),k,t} + \gamma^A_1 \left[\sum_{j \in M_{b,t}, \neq i} \omega^1_{ij}(M_{b,t})\alpha^A_j \right] + \gamma^A_2 \left[\sum_{j \in M_{b,t}, \neq i} \omega^2_{ij}(M_{b,t})\alpha^A_j \right] + \varepsilon^A_{i,b,k,t}.$$

All of the theoretical development in this section extends to such horse race model specifications.[5]

C. *Structural Interpretation of Revenue Spillovers*

The primary specification of our empirical model relates an aggregation of peers' fixed components of log revenue to a firm's own log revenue in year t, taking the same form as in equation (3). Our baseline estimation equation takes the following form, closely following Arcidiacono et al. (2012):

$$\ln R_{i,b,k,t} = \alpha^R_i + \phi^R_{B(b),k,t} + \gamma^R \left[\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})\alpha^R_j \right] + \varepsilon^R_{i,b,k,t}. \tag{4}$$

The framework in Sections I.A and I.B shows how to assign structural interpretations to each empirical model parameter in equation (4) and clarifies the conditions under which the reduced form parameter γ^R identifies the structural parameter γ^A. Inserting equation (3) into equation (2) delivers the structural interpretation of each parameter in equation (4).

We first consider the interpretation of local area-industry-year fixed effects $\phi^R_{B(b),k,t}$. Once these

[3]We conceptualize no role for endogenous effects, as TFP is unlikely to be chosen strategically in response to peers' choices.
[4]Such dynamics are more relevant for young firms, which tend to grow fastest. Firms in their first 5 years of existence make up 25% of our estimation sample.
[5]For computational reasons, we limit spillover comparisons to be between only two different peer group compositions at a time.

are understood, it is more straightforward to see what firm-specific factors remain. The primary empirical specification uses combinations of 500 meter radius area fixed effects, year fixed effects, and 2-digit industry fixed effects to control for contextual effects $\phi^R_{B(b),k,t}$. Under perfect competition, the structural interpretation of the fixed effects in equation (4) is

$$\phi^R_{B(b),k,t} = \frac{\theta_k}{1-\theta_k}\ln\theta_k + \frac{1}{1-\theta_k}\ln p_{B(b),k,t} - \frac{\theta_k}{1-\theta_k}\ln w_{B(b),k,t} + \frac{1}{1-\theta_k}\phi^A_{B(b),k,t}.$$

These fixed effects capture location and industry fundamentals, spatial variation in variable input prices, and industry-specific output demand conditions.

The remaining terms in equation (4) can be simplified with a rescaling of the structural fixed effect α_i^A. The relationship between the remaining terms in the reduced form estimation equation and the structural equation is:

$$\alpha_i^R + \gamma^R \sum_{j\in M_{b,t},\neq i}\left[\omega_{ij}(M_{b,t})\alpha_j^R\right] + \varepsilon^R_{i,b,k,t} = \frac{1}{1-\theta_{k(i)}}[\alpha_i^A + \gamma^A \sum_{j\in M_{b,t},\neq i}\left[\omega_{ij}(M_{b,t})\alpha_j^A\right] + \varepsilon^A_{i,b,k,t}].$$

Setting the firm-specific fixed effect α_i^R equal to $\alpha_i^A\frac{1}{1-\theta_{k(i)}}$, we can see that revenue spillovers γ^R directly measure TFP spillovers γ^A if all firms in firm i's peer group have the same variable input share. In the perfect competition case, the theory suggests that using $(1-\theta_{k(i)})\ln R_{i,b,k,t}$ as an outcome instead of $\ln R_{i,b,k,t}$ allows for recovery of the structural parameter of interest γ^A if other firms in firm i's peer group have different variable input shares. In the following subsection, we develop this idea further to additionally allow for imperfect competition.

As they have the same structural relationships with TFP, we use log employment and log total payroll as alternative outcome variables to corroborate the log revenue results. Payroll can be viewed as a quality adjusted version of the labor input.

D. Accommodating Imperfect Competition

To accommodate imperfect competition, we conceptualize an environment in which each firm in industry k has the same markup over marginal cost because it faces the same demand elasticity for its product η_k, in addition to having the same variable input share θ_k. While various modeling frameworks can deliver common markups, in Appendix A.A1 we derive it from the setup in De Loecker (2011), in which firms are monopolistically competitive and consumers have constant elasticity of substitution preferences over firm-specific varieties in each industry. Pass-through from TFP to revenue depends on the output demand elasticities faced by firms. As demand becomes more elastic, markups decline and the pass-through from TFP shocks to revenue gets stronger. In particular, the structural revenue equation, analogous to equation (2), becomes

$$\ln R_{i,b,k,t} = \frac{1+\eta_k}{\eta_k(1-\theta_k)-\theta_k} \ln A_{i,b,k,t} - \frac{\theta_k(1+\eta_k)}{\eta_k(1-\theta_k)-\theta_k} \ln w_{B(b),k,t} + \xi_{k,t} + e_{i,b,k,t}, \quad (5)$$

where the structural interpretations of $\xi_{k,t}$ and $e_{i,b,k,t}$ are laid out in Appendix A.A1. As demand gets more elastic and firms in industry k lose market power, $\frac{1+\eta_k}{\eta_k(1-\theta_k)-\theta_k}$ increases, converging toward $\frac{1}{1-\theta_k}$ and the perfect competition case seen in equation (2). The structural equation for the variable input $\ln L_{i,b,k,t}$ has the same coefficient on $\ln A_{i,b,k,t}$.

Substituting for $\ln A_{i,b,k,t}$ in equation (5) using equation (3) and comparing the structural revenue equation with our reduced form estimation equation (4), one can see that the log revenue spillover parameter γ^R is equal to γ^A only if all firms within each peer group have the same variable input share and output demand elasticity. This observation reflects one advantage of focusing on high-skilled services firms only, as their variable input shares and market power are likely to be similar across firms.

Under heterogeneous output demand elasticity and variable input shares within peer groups, we recover estimates of TFP spillovers γ^A under the data generating process described by equation (5). In particular, we show in Appendix A.A2 that using log revenue divided by $\frac{1+\eta_k}{\eta_k(1-\theta_k)-\theta_k}$ as the dependent variable makes the spillover parameter equal to γ^A. We explain in Appendix A.A3 how we measure θ_k and η_k in the data.

Most of our empirical analysis uses unadjusted log revenue as an outcome. As resulting peer effect estimates can incorporate price responses, they capture something closer to "revenue TFP" rather than "quantity TFP" spillovers. While log revenue based spillover estimates are reduced form in nature, we see a number of advantages to using this as our primary outcome. As it is a required reporting line for corporations, revenue is measured accurately and consistently across firms. Moreover, revenue TFP spillovers are of interest in their own right. As input demand responds to both TFP and output price shocks, log revenue spillovers estimates are informative about the economic geography of cities. They help explain the spatial concentration of employment and economic activity observed in the data.

Accurate recovery of quantity TFP spillovers depends crucially on a combination of strong modeling assumptions and accurate measurement of variable input shares and output demand elasticities. As it is impossible to know the true form of TFP spillovers, one key modeling assumption is that TFP spillovers follow the data generating process described in equation (3). Moreover, TFP must be backed out from strong assumptions about the demand system.[6] Finally, even with firm level

[6]An alternative approach would be to estimate firm level TFP using procedures proposed in Ackerberg, Caves and Frazer

balance sheet information, calibration of model parameters is subject to potentially serious measurement error difficulties. Nevertheless, we show below that results using adjusted log revenue (TFP) and unadjusted log revenue yield similar spillover estimates.

II. Empirical Implementation

Commensurate with the structural equations developed in the prior section, our baseline estimation equation relates outcome $y_{i,b,k,t}$ of firm i in peer group (and location) b operating in industry k at time t to peer outcomes using the following specification:

$$y_{i,b,k,t} = a_i + \phi_{B(b),k,t} + \gamma \sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t}) a_j + \varepsilon_{i,b,k,t}. \tag{6}$$

We use log firm sales revenue as our primary outcome of interest. Robustness checks use adjusted log revenue, log employment, and log total payroll as alternative outcomes.

In equation (6), a_i is a firm fixed effect and $\phi_{B(b),k,t}$ is a combination of local area fixed effects, year fixed effects, and industry fixed effects that captures access to local productive amenities, local labor supply conditions, and secular trends in industry-specific productivity, wages and/or output prices. While we explore various combinations of these fixed effects in the empirical work, in order to maintain sample size our primary specification has separate location-year and industry-year fixed effects.

The key predictor variable, $\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t}) a_j$, is an aggregation of the fixed component of the outcome variable in peer firms at time t, in which the weights are $\omega_{ij}(M_{b,t}) = \frac{1}{|M_{b,t}|-1}$ in the linear-in-means specification and $\omega_{ij}(M_{b,t}) = 1$ in the agglomeration specification. γ is the main parameter of interest and captures the average total spillover effect of peers' fixed attributes on the outcome for firm i. Subject to normalization discussed below, firm fixed effects a_i are economically informative measures of firm quality. We use estimates of components of a_i that are identified to investigate the importance of sorting across peer groups on firm quality and to quantify the extent to which such sorting has consequences for aggregate revenue. Section II.A discusses which components of a_i are identified under various scenarios.

The key spillover parameter γ can be interpreted in two useful ways. Most obviously, it is the elasticity of y with respect to an aggregation of the fixed component of peers' y. Perhaps more informatively, γ can also be viewed as the ratio of the importance of fixed peer attributes to fixed own attributes for generating variation in y. To see this, we imagine that each firm has a vector

(2015) or Gandhi, Navarro and Rivers (2020). However, because they use lagged input quantities as instruments and incorporate price taking assumptions, these approaches are not well suited to isolating annual variation in firm TFP or market power, especially for new arrivals to a location.

of fixed unobserved exogenous attributes X_i that contribute to y_i. These attributes are aggregated by index weights β into the scalar $\widetilde{X}_i$. That is, $a_i = \delta_o X_i \beta = \delta_o \widetilde{X}_i$, where δ_o is a common scalar parameter describing the importance of a firm's attribute index in contributing to its overall quality. Much of the peer effects literature conceptualizes "exogenous effects" as the causal impacts of exogenous peer attributes on outcomes (e.g., Gibbons, Overman and Patacchini, 2015). Rewriting the peer effects term in equation (6) with the exogenous effects spillover parameter δ_p, we have

$$\gamma \sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t}) a_j = \delta_p \sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t}) \widetilde{X}_j.$$

Substituting for $\widetilde{X}_j$ from above, we have $\gamma = \frac{\delta_p}{\delta_o}$, which is equivalent to our second interpretation. Absent endogenous effects, fixed peer attributes are 100γ percent as important as a firm's own fixed attributes in determining the outcome y.[7]

A. Measuring Firm Quality

In this sub-section, we discuss identification of firm fixed effects under various specifications and implications for the measurement of firm quality. It is informative to partition the firm fixed effects into common and idiosyncratic components:

$$a_i = \bar{\alpha} + \alpha_i.$$

Under linear-in-means aggregation schemes in which $\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t}) = 1$, the common component of the firm fixed effect factors out of the peer effects term as the constant $\gamma\bar{\alpha}$. $\bar{\alpha}$ is thus not separately identified from contextual effects $\phi_{B(b),k,t}$ under linear-in-means spillovers. As a normalization, in linear-in-means models we allocate the full constant term $\bar{\alpha}(\gamma + 1)$ to location-industry-time fixed effects $\phi_{B(b),k,t}$.

Empirical implementation of specifications in which $\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})$ is not constant across locations does allow for separate identification of $\bar{\alpha}$ by using variation in peer group size if $\gamma \neq 0$. In these cases, we can separately identify $\bar{\alpha}$ by including the sum of the weights as a separate control

[7]The addition of endogenous effects, in which $y_{i,b,k,t}$ depends structurally on $y_{j,b,t,k}$, would make the analysis more complicated. Several example models are discussed in the appendix of Arcidiacono et al. (2012). One relevant result is that interpretation of γ changes to be close to the sum of exogenous and endogenous spillovers if firms react strategically to expectations about (rather than actual) peer outcomes. In our empirical setting with heterogeneous firms operating in high-skilled services, we think it is unlikely that firms set revenue, factor quantities, or unobserved time-varying contributors to these outcomes strategically with their peers. Therefore, we interpret γ as capturing exogenous spillovers only.

variable. That is, we can rewrite equation (6) as

$$y_{i,b,k,t} = \alpha_i + \tilde{\phi}_{B(b),k,t} + \gamma \sum_{j \in M_{b,t}, \neq i} [\omega_{ij}(M_{b,t})\alpha_j] + \sigma[\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})] + \varepsilon_{i,b,k,t}, \quad (7)$$

where $\tilde{\phi}_{B(b),k,t} = \phi_{B(b),k,t} + \bar{\alpha}$ and $\sigma = \gamma\bar{\alpha}$. From this equation, as long as there is variation in the sum of the weights $\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})$ across peer groups, α_i, $\tilde{\phi}_{B(b),k,t}$, γ, and σ can all be separately identified. Therefore, $\bar{\alpha}$ can also be separately identified as $\frac{\sigma}{\gamma}$ as long as $\gamma \neq 0$. In the agglomeration specification in which $\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t}) = |M_{b,t}| - 1$, we thus include the number of other firms in the peer group as a separate independent variable. Intuitively, the impact of an additional low quality firm to a peer group raises aggregate peer group quality but reduces mean peer group quality, with the spillover parameters scaled appropriately in estimation to match this normalization. If the common portion of a_i were not identified, it would be more difficult to distinguish between these two types of spillovers, as additional firms could even reduce aggregate peer quality in the agglomeration specification. This issue does not arise in the linear-in-means specification, as peer group quality depends only on relative rather than absolute firm quality and does not depend on peer group size.

B. *Spillover Comparisons*

We explore a number of specifications to make comparisons across different types of spillovers. For our main analysis, we compare linear-in-means type spillovers with agglomeration type spillovers, meaning we primarily consider estimation equations of the form

$$y_{i,b,k,t} = \alpha_i + \ddot{\phi}_{B(b),k,t} + \frac{\gamma_{\text{LIM}}}{|M_{b,t}| - 1} \sum_{j \in M_{b,t}, \neq i} \alpha_j + \gamma_{\text{Agg}} \sum_{j \in M_{b,t}, \neq i} \alpha_j + \ddot{\sigma}(|M_{b,t}| - 1) + \varepsilon_{i,b,k,t}, \quad (8)$$

where $\ddot{\phi}_{B(b),k,t} = \phi_{B(b),k,t} + \bar{\alpha}(1 + \gamma_{\text{LIM}})$ and $\ddot{\sigma} = \bar{\alpha}\gamma_{\text{Agg}}$. Such estimates allow us to determine the extent to which linear-in-means versus agglomeration type spillovers dominate.[8]

To determine which types of firm-to-firm connections best facilitate spillovers, in subsequent analyses we add the additional term $\beta_{\text{W}} \sum_{j \in M_{b,t}, \neq i} \omega^{\text{W}}_{ij}(M_{b,t})$ to equation (8). Here, $\sum_{j \in M_{b,t}, \neq i} \omega^{\text{W}}_{ij}(M_{b,t})$ is specified as the fraction of peers in the top tercile of some connectivity type W with firm i. Therefore, β_{W} is interpreted as the additional spillover a typical firm would receive by going from a peer group composition without any close peer connections to one with the same mean and aggregate

[8]Appendix B.B2 discusses estimation of the more general specification in which $\frac{\gamma_{\text{LIM}}}{|M_{b,t}|-1} \sum_{j \in M_{b,t}, \neq i} \alpha_j$ is replaced by $\gamma_A \sum_{j \in M_{b,t}, \neq i} \omega^A_{ij}\alpha_j + \sigma_A \sum_{j \in M_{b,t}, \neq i} \omega^A_{ij}$ and $\gamma_{\text{Agg}} \sum_{j \in M_{b,t}, \neq i} \alpha_j + \ddot{\sigma}(|M_{b,t}| - 1)$ is replaced by $\gamma_B \sum_{j \in M_{b,t}, \neq i} \omega^B_{ij}\alpha_j + \sigma_B \sum_{j \in M_{b,t}, \neq i} \omega^B_{ij}$.

quality but in which all peers are in the top tercile of connections of type W to firm i. As is discussed further in Section III.C below, we consider bilateral input-output relationships, occupational similarity, prevalence of labor flows between industries, and a simple indicator for being in the same two-digit industry.

We similarly estimate heterogeneous spillover effects by firm quality. As with the industry connections analysis, we focus on estimating the impact of having a higher fraction of peers in the top tercile of the local 500 meter radius area's α distribution. These results speak to the log supermodularity assumption often used in theoretical modeling of cities with heterogeneous agents (e.g. Davis and Dingel, 2019). As the inclusion of fixed effects $\ddot{\phi}_{B(b),k,t}$ precludes us from estimating the full distribution of α across all locations, looking within 500 meter radius areas is the furthest we can go in evaluating spillover heterogeneity across firm quality while still controlling for changes in location fundamentals.[9]

C. Estimation

Arcidiacono et al. (2012) proves that γ in the linear-in-means specification of equation (6) can be identified by nonlinear least squares (NLLS) provided at least one peer group experiences variation in group composition. If each peer group has at least one firm that has a non-missing outcome for at least two periods, all firm fixed effects are identified jointly with γ. Moreover, this setup accommodates missing data on outcomes as long as each firm is observed with non-missing data at least once. In Appendix C, we extend this proof to accommodate arbitrary exogenous connectivity weights linking firms. Evidence in the following section shows that there exists considerable variation in peer group composition in our data, meaning that we can identify estimates of α_i for the vast majority of firms. As we show in Appendix C, the identification proof can be extended to accommodate additional spillover terms as in equation (8) as long as there exists sufficient variation in changes in peer group composition. We estimate empirical models using the iterative algorithm proposed by Arcidiacono et al. (2012).[10]

If the weights do not sum to a constant, the nonlinear least square estimator for parameters in

[9]Rather than adding an additional term to the specification in equation (8), in unreported results we replace the agglomeration terms in equation (8) with $\gamma_{\mathrm{W}} \sum_{j\in M_{b,t},\neq i} \omega_{ij}^{\mathrm{W}}(M_{b,t})\alpha_j + \ddot{\sigma}_W \sum_{j\in M_{b,t},\neq i} \omega_{ij}^{\mathrm{W}}(M_{b,t})$. In this expression, $\omega_{ij}^{\mathrm{W}}(M_{b,t}) = \frac{w_{ij}^{\mathrm{W}}}{|M_{b,t}|-1}$, where w_{ij}^{W} is an indicator for whether the firm i-to-j connection is above the median, whether firm i is in the top tercile of the area's α distribution, or whether firm j is in the top tercile of the area's α distribution. We drop the agglomeration term in this case since our estimates of γ_{Agg} are not significant. The same qualitative messages as from the simpler specifications ensue, though parameter convergence is more fragile.

[10]As we demonstrate in Appendix C, consistency requires that one of two environments hold. Either (1) the number of peer groups N goes to infinity for a fixed T and there is no serial correlation in the errors or (2) both N and T go to infinity, in which case the errors can be serially correlated. Below we demonstrate with Monte Carlo simulations that the low serial correlation in errors married with sufficiently large T leads to negligible biases in estimates in our empirical setting.

our main estimation equation (7) solves

$$\min_{\alpha_i,\tilde{\phi}_{B(b),k,t},\sigma,\gamma} \sum_t \sum_i \left(y_{i,b,k,t} - \alpha_i - \tilde{\phi}_{B(b),k,t} - \sigma \sum_{j\in M_{b,t}\neq i} \omega_{ij}(M_{b,t}) - \gamma \sum_{j\in M_{b,t}\neq i} \omega_{ij}(M_{b,t})\alpha_j \right)^2 .$$

Taking first-order conditions with respect to α_i yields updating equations for each α_i. Arcidiacono et al. (2012) propose to solve for parameters using a two-step iterative algorithm. In the first step of model estimation, the firm fixed effects are taken as given and estimates of γ, σ, and $\tilde{\phi}_{B(b),k,t}$ are obtained by a standard fixed effect estimator. In the second step, γ, σ, and $\tilde{\phi}_{B(b),k,t}$ are taken as given and new estimates of the firm fixed effects are obtained using first order conditions. After a number of iterations, this procedure converges to the nonlinear least square solution. In our primary specification, we initialize α_i to be estimates from a regression of $y_{i,b,k,t}$ on firm, local area-year, and 2-digit industry-year fixed effects, assigning the constant to α_i. In the linear-in-means specification, σ is not separately identified and thus cannot be estimated. Estimation of specifications that include both linear-in-means type spillovers and agglomeration type spillovers follows the same procedure, though with a more complicated updating rule for α_i. Appendix B details updating equations for α_i for all of the specifications we estimate. We use a symmetric wild bootstrap (MacKinnon, 2006) clustered by 75 meter radius peer group areas b to calculate standard errors.[11]

D. *Identification*

Consistent identification of γ requires variation in the composition of firms within blocks that is unrelated to time-varying unobservables driving outcomes. By using changes in peer group composition for identification, this setup is not subject to the classic identification challenge faced in much of the empirical agglomeration literature, that firms (or workers) systematically sort across locations on their own fixed unobserved attributes. In our context, such sorting would occur if more productive or high paying firms located in higher quality locations. For example, if more productive firms are the high bidders for commercial real estate near train stations and highway interchanges, there could be a correlation between firm and peer outcomes that is not causal but is instead driven by this contextual natural advantage. By including firm fixed effects, this empirical setup controls for such sorting on levels.

Our main empirical specifications include 500 meter radius area fixed effects interacted with year as controls. The key identifying assumption is thus that variation in changes in peer group composition within 500 meter radius areas is not related to very local trends in unobservables that

[11] Unclustered standard errors are typically about 40% smaller.

drive firm outcomes. For example, one may be concerned with the possibility that certain types of locations receive shocks that both attract better firms and directly impact incumbent firm outcomes. That is, neighborhood trends in firm productivity, output demand, or labor supply conditions may predict both changes in firm composition (the mix of α_js) and changes in the productivity of incumbent firms ($\varepsilon_{i,k,b,t}$). Given that the key source of identifying variation in the empirical work comes from firm entry to and exit from blocks, we must clean out any such unobservables that predict both composition changes in peers' fixed effects because of firm turnover and changes in outcomes for incumbents.

One sufficient condition for clean identifying variation is sufficiently tight commercial real estate markets within 500 meter radius areas such that firms cannot choose exact locations within these small areas. As a result, spillover estimates that accrue from fixed attributes of neighboring firms are isolated from the impacts of potentially correlated neighborhood fundamentals. Thin commercial real estate markets put a constraint on the amount of information firms can act upon when deciding which building into which to move. This is similar to the identification strategy employed in Bayer, Ross and Topa (2008) for estimating the rate of job referrals across residential neighbors, though unlike Bayer, Ross and Topa (2008), our analysis is not subject to sorting bias on levels within small neighborhoods.

The use of panel data is central to our analysis. Without panel data, it would be impossible to isolate each firm's individual exogenous quality α_i that is fixed over time. Moreover, panel data is required to account for sorting across peer groups on unobserved firm characteristics. As much of the peer effects literature has not had access to panel data, it has had a difficult time separately identifying spillovers from unobserved agent attributes absent explicit randomization into peer groups. As such, much of the peer effects literature is only able to look at settings in which peer group assignment is conditionally random. Even in these cases, this literature has had a difficult time estimating the full magnitudes of exogenous peer effects.

Panel data also allows us to implement a number of post-estimation identification checks that validate our empirical strategy. These include demonstrating that errors are neither correlated with shocks to peer group quality in future locations before moves in nor in past locations after moves out. We also show more reduced form event study evidence that firms have better outcomes when their peer group changes with the addition of high-quality firms

III. Data and Descriptive Evidence

A. Data and Sample

The data set used for the analysis incorporates Canadian administrative tax records on firms and workers. The main source is T2 Corporation Income Tax Return files for all incorporated firms in Canada in each year 2001-2012. All corporations in Canada must file a T2 return every year, even if there is no tax payable. The T2 files contain information on firm revenues, expenses, and assets. Additional information on payroll and employment is derived from linked firm records on employment remuneration (Form T4). We also observe anonymized six-character postal code identifiers for the location of each firm's primary operations and a distance matrix for these anonymized postal codes out to one kilometer. Canadian postal codes in the central areas of cities typically cover blockfaces or individual buildings.

We keep all firm-years in the Montreal, Toronto and Vancouver census metropolitan areas with evidence that the firm is operating. We focus on using information about sales of goods and services (revenue), employment, and payroll as these are required reporting lines in the corporate tax filings. We drop firms that cycle back and forth between postal codes, with missing location information, or with no 4-digit industry information. We identify a firm's entry and exit years as the first and last years it has positive reported revenue, employment, or payroll. As the empirical setup admits missing values on outcomes, we keep firm-years with missing information on any of these measures in between entry and exit years. Because we only observe one postal code per firm, our primary estimation sample only includes single-location firms. As firms are defined as tax reporting units, many acquired firms and subsidiaries are kept in our data since they report as separate tax entities. We perform robustness checks assigning multi-location firms to their headquarters locations.

Table 1 presents summary statistics on the firms in our data. Columns (1) and (2) show statistics for firms in all industries and columns (3) and (4) show those for the 42% of firms that are in high-skilled services (NAICS 5), the largest 1-digit sector by firm count. The next biggest sector is recreation, accommodation and food services (NAICS 7). We elect not to include NAICS 7 firms because their demand conditions commonly vary at a microgeographic scale.[12] We observe approximately 181,500 single-location NAICS 5 firms operating in at least one year 2001-2012 in Montreal, Toronto, and Vancouver. The typical NAICS 5 single-location firm is smaller than the average single-location firm. It has lower revenue (CAD 300,000 per year) and fewer employees (4) but greater payroll per worker (CAD 48,000). These single-location firms are sufficiently small that

[12] Many studies of agglomeration focus on manufacturing, which accounts for only about 10% of firms in our study area.

their individual movement is unlikely to influence local factor prices.

Our estimation sample consists of small peer group areas within which we observe the population of single-location NAICS 5 firms. To build these peer group areas, we first group postal codes into regions in which the distance between the centroid of a nodal postal code and all other postal code centroids is less than 75 meters. These peer group areas fully segment each of the three cities in our data. We exclude all such areas that either have at least one member postal code with an area that is greater than $\pi 75^2$ sq meters (0.018 sq km) or have fewer than two high-skilled services firms in any year 2001-2012. We iterate to additionally exclude peer groups that include firms for which at least one contextual fixed effect required for estimation is not separately identified from the firm fixed effect. The primary estimation sample is thus constructed jointly with the primary empirical model specification, which has 500 meter radius area by year and 2-digit industry code by year fixed effects. The estimation sample tends to include denser areas and grows in robustness specifications with fewer fixed effects.

Figure 1 presents a map of postal codes and major streets in downtown Toronto. Rings of various radii around the centroid of the focal postal code for one example peer group area are indicated. This peer group area is centered immediately southwest of the corner of King and Yonge streets, which is in the financial district. Five other postal codes have centroids that are within the indicated 75 meter radius circle, putting them inside the same peer group area. Inclusion of full postal codes based on centroid location only means that most peer group areas have radii that are somewhat greater than 75 meters. In particular, the average firm in our sample is in a peer group of radius 117 meters and 0.043 sq km.

The primary sample has approximately 56,000 firms and 282,000 firm-year observations. Of these observations, 13,000 have missing revenue. The average firm in our sample has CAD 430,000 per year in revenue and 4.8 employees, who earn an average of CAD 55,000 per year. These firms are spread across 42,100 peer groups for an average peer group size of 6.7 firms. We cover about 30% of single-location NAICS 5 firms in the three cities, with the exclusions due to firms being alone in peer group areas and/or in postal codes that are too large. Indeed, the average single-location NAICS 5 firm is in a postal code with a radius of 169 meters and is in a peer group area of 2.1 firms. The firms in our sample generate about 30% of aggregate NAICS 5 firm revenue in the three cities.

B. Peer Group Composition

Identification of peer effects using our empirical strategy requires both a panel data structure and temporal variation in peer group composition. Firms appear in our primary sample for an average of 6.2 years out of 12 years of data, with a standard deviation of 3.9. We observe half of the firms in our primary sample for at least 6 years. Firms may operate in some years but not contribute to the estimation sample due to the sample restrictions described above. Estimation sample firms experience 1.4 75 meter radius peer group areas on average, with a standard deviation of 0.7. However, the typical firm is not very mobile. Only 34% of firms in our sample experience more than one peer group area in our data. When firms move, they move between 500 meter radius areas 94 percent of the time.

Higher revenue firms sort into peer groups of both higher average and aggregate revenue. Figure 2 shows non-parametric relationships between average peer log revenue (Panel A) or aggregate peer log revenue (Panel B) and firm log revenue. Above the median, there is strong positive sorting on the mean log revenue and aggregate log revenue of peers. As can be inferred from comparing Panels A and B, there is also a strong positive correlation between mean and aggregate peer log revenue. Without accounting for both simultaneously in the empirical work, it is thus easy to mistake linear-in-means type spillovers for aggregate type spillovers.[13]

Figure E1 provides a sense of the variation in log revenue and peer group composition in our data. Importantly, it shows that the peer group size distribution is highly skewed to the right, with the largest peer groups having about 150 members and the average firm exposed to 16 peers. As a result, there is much greater dispersion in aggregate peer log revenue than in average peer log revenue. The associated independent variation is needed to empirically distinguish between these two types of spillovers.

C. Connectivity Weights

Ellison, Glaeser and Kerr (2010) and Faggio, Silva and Strange (2017) describe the extent to which firms in manufacturing industries connected through input-output linkages, occupational similarity, and/or patent citations coagglomerate. Part of our analysis evaluates the extent to which cross-firm productivity spillovers within peer groups of firms in high-skilled services are mediated through these same types of inter-industry connections. As in the coagglomeration studies, we explore the relative

[13] In Section V below, we revisit relationships like this after accounting for the component of revenue due to spillovers. We will see, again, that firms positively sort on both average and aggregate peer quality. That is, $\widehat{\alpha}_i$ is more highly correlated with both $\frac{1}{|M_{b,t}|-1}\sum_{j\in M_{b,t}\neq i}\widehat{\alpha}_j$ and $\sum_{j\in M_{b,t}\neq i}\widehat{\alpha}_j$ than would be the case if firms were allocated randomly into peer groups. Relatedly, average and aggregate peer quality are positively correlated. Just as with log revenue, firms tend to assortatively match into peer groups on α_i when observed in the cross-section.

importance of input-output linkages and occupational similarity for the magnitudes of spillovers. In addition, we look at industry connections as defined by the prevalence of worker flows between industries, as in Serafinelli (2019). Finally, as in Greenstone, Hornbeck and Moretti (2010), we examine the extent to which being in the same 2-digit industry matters. We do not look at the prevalence of patenting or patent citations because patenting is rare in high-skilled services. Our connectivity weights analysis estimates the extent to which increasing the fraction of peers in the top tercile of each bilateral weights distribution, calculated for our primary estimation sample, affects firm outcomes. Additional details about connectivity weights can be found in Appendix D.

IV. Results

In this section, we present and discuss parameter estimates under various spillover specifications, aggregation weights, and peer group definitions. Equation (6) with $\omega_{ij} = \frac{1}{|M_{b,t}|-1}$ is the estimation equation for all linear-in-means estimates, delivering $\widehat{\gamma}_{\text{LIM}}$. In this case, mean firm quality $\bar{\alpha}$ is not identified. Equation (7) with $\omega_{ij} = 1$ is the estimation equation for agglomeration estimates, delivering $\widehat{\gamma}_{\text{Agg}}$. Horse races between linear-in-means and agglomeration aggregation schemes are estimated using equation (8), delivering both $\widehat{\gamma}_{\text{LIM}}$ and $\widehat{\gamma}_{\text{Agg}}$ simultaneously. The agglomeration and horse race models also deliver estimates of mean firm quality.

A. Main Estimates

Table 2 presents the main results of the paper. The first two columns show separate estimates of γ_{LIM} and γ_{Agg} with log revenue as the outcome. We find a statistically significant estimate of 0.018 for γ_{LIM} but an insignificant estimate for γ_{Agg} that is close to zero. The third column presents these parameters estimated jointly. The result is a slightly larger γ_{LIM} estimate of 0.024 and an estimate for γ_{Agg} that remains close to zero, turning slightly negative. This pattern reflects both a positive correlation between changes in mean and aggregate peer quality and the fact that agglomeration spillovers at a 75 meter radius area spatial scale are very close to zero. Standard errors for $\widehat{\gamma}_{\text{LIM}}$, clustered by 75 meter radius peer group area, are near 0.009 in both columns (1) and (3).

We can interpret the linear-in-means results in two ways. First, an approximate doubling of average peer quality leads to a 1.8 to 2.4 percent increase in firm revenue. As the standard deviation in average peer quality is 1.1, this is also approximately the impact of increasing peer quality by one standard deviation. Equivalently, this estimate can be interpreted as indicating that absent endogenous effects, peers' attributes are 1.8 to 2.4 percent as important as a firm's own attributes for determining revenue. The final row of Table 2 reports the implied difference in the fraction of revenue accounted for by spillovers in the 90^{th} percentile firm relative to the 10^{th} percentile firm. This 90-10 gap of 5-7 percent shows a wide range of spillovers across firms depending on the environment. Recall evidence in Figure 2 Panel A showing that high quality firms tend to colocate, which is part of what generates this dispersion.

The near zero agglomeration spillover estimates should be viewed in the context of the inclusion of 500 meter radius local area-year fixed effects. Our estimates cannot rule out the existence of aggregate increasing returns at higher levels of spatial aggregation. Sharing of inputs provided at high minimum efficient scales, sharing of output markets, and labor market pooling are all likely to operate at spatial scales at or above 500 meter radius regions. As such, we interpret our microgeographic scale results as primarily reflecting knowledge flows rather than these other forces. Of the forces driving agglomeration economies, knowledge transfer may also be more likely to occur as a function of average rather than aggregate peer group quality.

Results in columns (4)-(6) of Table 2 show analogous estimates using the more parsimonious specification that excludes 500 meter radius-year fixed effects. Comparison of these estimates with those in columns (1)-(3) indicate relationships between location fundamentals and peer group composition. This specification delivers a linear-in-means estimate of 0.029. The larger estimate in column (4) than column (1) indicates that through composition shifts, peer groups tend to improve in average quality in areas experiencing positive productivity trends and/or peer groups tend to decline in average quality in areas experiencing negative productivity trends. The agglomeration model in column (5) yields a statistically significant estimate of 0.0019, more than six times larger than its counterpart in column (2). This indicates a positive correlation between trends in aggregate peer group quality and location fundamentals. The horse race model in column (6) generates a γ_{LIM} estimate of 0.021, which is statistically indistinguishable from our primary specification estimate of 0.024. The estimate of γ_{Agg} falls some to 0.0011 but is still well above the corresponding estimate in column (3). Therefore, the composition bias primarily comes from higher α_i firms crowding into locations experiencing productivity growth or departing locations with productivity declines. That is, natural advantage and aggregate peer quality are positively correlated at small spatial scales in a way that is likely not causal. Recovery of credible estimates of γ_{LIM} thus requires controlling either for neighborhood-year fixed effects or aggregate peer quality. Because peer groups tend to be larger in places with better location fundamentals, even conditional on average peer quality, recovery of credible estimates of γ_{Agg} requires controls for both location fundamentals and average peer quality.

Results in columns (7)-(9) of Table 2 are estimated with controls for 500 meter radius area-year fixed effects but not industry-year fixed effects. These results are similar to the results from our main specification in columns (1)-(3). The conceptual interpretation based on the model in Section I is that either there is not much heterogeneity across NAICS 5 industries in variable input shares or market power or that there is not systematic sorting by industry across peer groups in a way that is correlated with local productivity shocks.

The fully saturated specification with a triple interaction between 500 meter radius area, 2-digit industry, and year fixed effects reduces the sample size by 38 percent and yields estimates of γ_{LIM} and γ_{Agg} that are indistinguishable from our main estimates reported in column (3) (not reported). We conclude that a specification with 500 meter radius area-year and industry-year fixed effects strikes a good balance between maintaining sample size and facilitating strong identification. As such, we maintain this specification throughout the remainder of our analysis.

Various statistics about estimated firm and peer quality distributions are listed near the bottom of Table 2. Statistics about $\hat{\alpha}_i$ distributions and estimated peer group compositions are quite stable across specifications. As such, we are confident in using this information to help evaluate the prevalence of sorting on firm quality across peer group and location quality in counterfactual experiments explored in Section V. For specifications that include agglomeration terms, mean firm quality $\bar{\alpha}$ can be calculated but is imprecisely estimated because $\hat{\gamma}_{\text{Agg}}$ is near 0, commensurate with our discussion of equation (8).

B. Identification Checks

Central to our identification strategy is that 500 meter radius area-year fixed effects must successfully control for the component of trends in location fundamentals that is correlated with peer group composition. Moreover, identification requires no selective migration of firms with positive revenue shocks or trends to higher average quality peer groups conditional on fixed effects. To evaluate the strengths of these conditions, we undertake exercises with the post-estimation errors from our primary specification in Table 2 column (3). These exercises verify that the error term is not correlated with various attributes of future or past peer quality that could reflect trends in location fundamentals that operate at spatial scales smaller than 500 meter radius areas or selective migration. Table 3 presents these results. All standard errors are calculated using a symmetric wild bootstrap of post-estimation errors with 100 replications and peer group area level clustering.

The first two columns of Panel A report regressions of errors on estimated average peer quality (means of peer estimated α_j) one and two years in the future (column 1) or one and two years in the past (column 2). These estimates are all less than one-fifth the magnitude of our main estimate of γ_{LIM} and are not statistically significant. For the firms that do not move between periods t

and $t+1$, these small magnitudes indicate that any potential bias from local neighborhood trends in fundamentals that are correlated with changes in peer group composition are small. For firms that do move, these estimates indicate that firms are not responding to positive (negative) shocks or quality trends by moving to higher (lower) quality peer groups in a statistically significant way. Evidence in columns (3) and (4) corroborate these observations by showing that the average qualities of future peer group entrants and past peer group leavers are also not significantly related to the error term.

Table 3 Panel B reports results from a similar set of exercises using only estimated errors for firms that move across peer group locations between periods t and $t+1$. The first column shows no statistically significant relationship between errors in period t (before the move) and the contemporaneous average peer group quality in the future location. The negative estimate indicates that if anything movers receiving positive shocks before moving actually tend to move to locations with *lower* average peer quality, obviating the potential identification concern that firms experiencing positive shocks or trends tend to sort into better quality peer groups. Columns (2) and (3) show smaller relationships between the error term in period t and the average qualities of firms entering or departing to and from the future location in period t (before the firm arrives). This is evidence that firms receiving positive shocks are not sorting into improving locations and that firms receiving negative shocks are not sorting into declining locations. In the same spirit, columns (4) and (5) show no correlation between errors in period $t+1$ (after the move) and the average quality of new arrivals or departures in the previous location at time $t+1$. Firms that experience positive shocks or trends do not selectively migrate to higher quality or improving peer groups conditional on fixed effects, nor do firms that experience negative shocks selectively migrate to lower quality or deteriorating peer groups.[14]

Appendix C demonstrates that consistent estimation of peer effect parameters using our empirical model requires either that the number of observations associated with each location tends to infinity or that the error term exhibits no serial correlation or heteroskedasticity. However, the result in Table 3 Panel A column (5) shows a moderate serial correlation in the error term of 0.27, which may reflect trends in firm quality.

Monte Carlo simulations confirm that heteroskedasticity and serial correlation of the errors of this magnitude hardly influence estimates of interest given the length of the panel. To carry out this analysis, we parameterize heteroskedasticity as normally distributed across firms. Using post-

[14] While we find no evidence of sorting on changes, below we show strong evidence of sorting in the cross-section: higher quality firms do tend to co-locate in peer groups. Such cross-sectional sorting is controlled for with firm fixed effects in the empirical work.

estimation data, we calculate the variance of the error term for each firm and fit it to a normal distribution. Monte Carlo simulations sample from error distributions with this firm-level distribution of variances and a within-firm serial correlation of 0.27. Across 100 simulations, the resulting average estimates of γ_{LIM} are 0.016 using the specification in Table 2 column (1) and 0.022 using the specification in Table 2 column (3). Monte Carlo simulations that assume other forms of error heteroskedasticity all yield implied estimates that are even closer to those reported in Table 2.

A series of exercises using data from years surrounding large shocks to average peer quality further corroborate robustness of our estimates. We isolate events in which average peer quality declines by more than the 10^{th} or 25^{th} percentiles or increases by more than the 75^{th} or 90^{th} percentiles of the distribution of change in average peer quality across all firm-year observations in our primary estimation sample. These are mean changes in average peer quality of -0.73, -0.43, 0.49, and 0.82, respectively. Using the same post-estimation data set used to generate Table 3, we form a firm-event time panel keeping only incumbent firms exposed to these large changes in average peer quality in event time t. We run separate regressions of firm log revenue (residualized from the estimated fixed effects) on average peer quality, aggregate peer quality, and the number of peers for each event time $t-2$ to $t+2$ and event-type sub-sample. Table E1 reports the resulting estimates of γ_{LIM}, which remain remarkably stable across these sub-samples and close to our main parameter estimate at about 0.024.

We verify that the same qualitative result holds in the few clean events induced by the arrival of new firms that exist in our data. We isolate the 93 events in which all incumbent firms in a peer group experience a decline in average peer quality below the 10^{th} percentile and the 25 events in which all incumbent firms in a peer group experience an increase in average peer quality above the 90^{th} percentile, along with no other changes in firm composition within two years prior and one year after these shocks. While the restriction to such clean events results in small samples, event study results in Figure E2 show statistical significance, post-event persistence in treatment effects of expected signs, and a lack of differential pre-trends.

C. Alternative Outcomes

Inspired by model predictions in Section I, Table 4 presents results using three alternative outcomes: adjusted log revenue, log employment, and log payroll. Moreover, we explore the extent to which excluding multi-location firms influences the analysis.

The first column shows parameter estimates using log revenue adjusted for cross-industry heterogeneity in variable input shares and market power as the outcome. Specifically, the outcome is log revenue divided by $\frac{1+\eta_k}{\eta_k(1-\theta_k)-\theta_k}$, where θ_k is the variable input share and η_k is the output demand elasticity faced by firms in industry k calculated as described in Appendix A.A3. Spillover parameter estimates are quite similar to those reported in Table 2 column (3). Our estimate of γ_{LIM} is only slightly smaller at 0.021 and the estimate of γ_{Agg} becomes more negative at -0.0012. This is evidence that heterogeneous treatment effects for log revenue because of industry heterogeneity are not seriously biasing our main coefficients of interest.

The model predicts that variable inputs should exhibit peer effects that are identical to those for revenue. Employment is by far the largest component of variable cost and can be measured consistently across firms in our data.[15] We view payroll as a quality-adjusted measure of employment (Fox and Smeets, 2011). Here we see linear-in-means estimates of about 0.016 for employment and 0.013 for payroll with large standard errors. Though smaller than the revenue results, these estimates are not statistically different. These slightly smaller estimates may reflect hiring and firing frictions and the fact that these measures may not capture the full variation in hours worked, as most of the employment we see is salaried. Moreover, employment and payroll may be understated for smaller firms due to the unobserved labor provided by firm owners.

The final column of Table 4 examines robustness of our main estimates in Table 2 to using a sample that includes multi-location firms. Here, we are constrained to assign all firm revenue to the location reported on firm tax filings. This is a source of non-classical measurement error in the dependent variable, in which we expect greater overstatements of average and aggregate peer revenue in locations with better fundamentals and better peers, thereby biasing estimates downwards. Nonetheless, we find only a slightly smaller γ_{LIM} spillover estimate of 0.019.

We note that some of the agglomeration literature examines relationships between firm level outcomes and city or region level aggregates that are of a somewhat different functional form from those examined in this paper. A common model specification might make firm log revenue or TFP an increasing function of aggregate population, employment, or GDP in the city or more local region. Our main agglomeration specification relates firm log revenue to something close to the sum of peer log revenue rather than the log of the sum of peer revenue. Unfortunately, our empirical setup limits us to linear aggregations of peer α_j, precluding us from directly examining peer group aggregates like $\ln[\sum_{j\in M_{b,t+1},\neq i}\exp\{\alpha_j\}]$. Attempts to estimate spillovers using firm revenue in dollars rather than its log as the outcome presents estimator convergence challenges with very large standard errors. The log revenue specification fits the data much better.

[15] While firms do report materials costs, this measure is small for NAICS 5 firms and exhibits wide heterogeneity across firms.

D. Spatial Decay

Results in Table 5 provide evidence of rapid spatial decay, with negligible linear-in-means spillovers operating beyond 75 meters. We demonstrate this result in two steps. First, we show that estimates are similar to those reported in Table 2 when peer groups are defined over broader areas. We explain how this stability is evidence of very local effects only. We then verify this claim by jointly estimating spillovers with respect to average peer quality in concentric expanding rings around a firm using post-estimation data, finding no evidence of impacts beyond 75 meters.

Table 5 Panel A reports estimates from specifications identical to those in Table 2 column (3), except with peer group area radii extended to 150, 200, or 250 meters. We cannot go beyond the 250 meter radius peer group area while maintaining separate identification of 500 meter radius area-year fixed effects. To maintain comparison with the primary estimation sample, we consolidate the peer groups used in Table 2 column (3) to create larger peer groups. Therefore, all samples exclude firms in 75 meter radius peer group areas with one or more member postal code with an area that is greater than $\pi 75^2$ sq meters and 75 meter radius peer group areas with fewer than two high-skilled services firms in any year 2001-2012. As for the primary estimation sample, we exclude peer groups that include firms for which at least one contextual fixed effect required for estimation is not separately identified from the firm fixed effect. As a result, slightly more observations are excluded when using the broader peer group area definitions compared to the primary estimation sample in column (1).

Estimates of γ_{LIM} show considerable stability across peer group sizes at 0.022-0.025. We interpret this parameter stability as reflecting rapid spatial decay in peer effects, as parameters can be scaled by the size of 75 meter radius peer group areas. To get a sense of magnitude, we note that the average firm-year in our primary estimation sample is exposed to a 75 meter radius peer group area of size of 0.04 sq km, a 150 meter radius peer group area 2.3 times as large, a 200 meter radius peer group area 3.6 times as large, and a 250 meter radius peer group area 5.0 times as large (Table E2). If firms are uniformly spatially distributed within peer group areas, estimates reported in Table 5 Panel A column (2) indicate that a 10 percent increase in average peer quality within a typical 75 meter radius peer group that is contained within a 150 meter radius peer group thus leads to about a $\frac{0.253}{2.3} = 0.11$ percent increase in revenue on average, rather than the 0.24 percent estimated within the smaller peer group areas. This suggests that most of the peer effect is generated from firms within 75 meters. Agglomeration estimates are close to zero for all peer group definitions, indicating that any agglomeration impacts that exist must operate at spatial scales at or above 500 meter radius areas.

Panel B column (1) takes the post-estimation data from Panel A column (1) and jointly estimates spillovers with respect to average peer quality in concentric expanding rings around a firm. Only the innermost 75 meter radius is significant, with a coefficient of 0.022. Columns (2)-(4) report analogous exercises using the same post-estimation data, though with only two peer group areas at a time in the regression. In each case, only average peer quality within 75 meters has significant coefficients, and each remains near our headline estimate of 0.024. Taken together, the evidence in Table 5 suggests rapid spatial decay in linear-in-means spillovers.[16] [17]

E. Industry Connections and Firm Quality

In this sub-section, we examine how spillovers depend on peer attributes. To do so, we begin with the specification in Table 2 column (3) and add the fraction of peers with some attribute as an additional regressor. We first examine impacts of having more peers in the same 2-digit industry and in top terciles of input-output, occupational similarity, and worker flow industry relationships. Details of how we construct these measures of firm connectedness are in Section III.C and Appendix D. Results are reported in Table 6 columns (1)-(5).

Results reveal that industry relationships matter in addition to average peer quality. Results in column (1) and column (2) show that having a greater fraction of peers in the same 2-digit industry or in industries that are more closely connected upstream or downstream may result in lower revenues, though these estimates are not significant. In contrast, results in columns (3) and (4) show that having a greater fraction of peers with closer labor market connections or job task compositions, as measured by occupational similarity or worker flows, results in higher revenues, though again estimates are not significant. Examining all of these forces together in column (5), we see that each individual estimate strengthens. Based on the estimates in column (5) and recognizing that about one-third of the typical firm's peers are in the top tercile of each distribution, the typical firm loses an estimated 0.9 percent of revenue from having peers in the same industry (significant at the 11 percent level) and 1.3 percent of revenue through close input-output relationships (significant at the 16 percent level).[18] However, it gains 1.5 percent of revenue from having peers in industries with closer worker flow connections (significant at the 8 percent level). Addition of these regressors does not influence the linear-in-means spillover estimate of 0.024 nor the conclusion that aggregate type spillovers are negligible at this spatial scale. The associated lack of correlation between peer

[16]We note that the near zero estimates for peer group area definitions beyond 75 meters in 5 Panel B obviate the possibility that spatial correlation in peer group quality within 500 meter radius areas could be driving results.

[17]Exercises similar to those in Table 5 Panel A using an expanded sample that includes all firms in the broader peer group areas yields similar results except a smaller coefficient for the 250 meter radius peer group area. These alternative peer group definitions include more low density neighborhoods and many small spatially isolated firms.

[18]About 20 percent of peers are in the same 2-digit industry for the average firm-year in our data.

industry composition and average or aggregate peer quality is further evidence in support of well-identified estimates in Table 2 column (3).

Results in column (6) of Table 6 show compelling evidence that about two-thirds of our estimated linear-in-means spillover of 0.024 from Table 2 is driven by peers in the top tercile of the firm quality distribution. When controlling for the fraction of peers in the top tercile of the local 500 meter radius area's firm quality distribution, the main linear-in-means estimate declines to 0.007.[19] The coefficient on the fraction of peers in the top tercile of the local area's firm quality distribution is 0.086. The final column presents an analogous regression that instead controls for the fraction of peers that are above median quality. The fact that the coefficient of 0.027 on this control is much smaller than the corresponding coefficient in column (6) is evidence that linear-in-means spillovers are convex in average peer firm quality. These results help rationalize the observation discussed in the following section that higher quality firms exhibit stronger assortative matching into peer groups than do lower quality firms.

F. Discussion

Given the small spatial scales involved and the focus on high-skilled service industries, we interpret the evidence presented in this section primarily as reflecting knowledge transfer between workers across firms. We come to this conclusion in part from process of elimination. The negative estimated impact of a greater fraction of peers in industries with stronger input-output connections argues against input sharing as being a central driver of results, as does the lack of local external economies of scale. The lack of scale effects also argues against matching as a primary driver of estimates, though we cannot rule out the possibility that the better access to information about good potential hires, or reduced search frictions for workers and jobs because of proximity, may be driving some of what we find. The fact that our data is dominated by firms producing services that trade over long distances argues against price competition effects as a major driver of results.

Knowledge flows are also fully consistent with the patterns of estimates. The potential for knowledge acquisition is greater from workers in different industries with similar worker requirements. The positive estimated impacts of having more peers in industries with a greater prevalence of labor flows and in industries with similar occupational structure point to knowledge flows as an important mechanism driving the results. While knowledge acquisition may happen through input-output connections, the associated relevant spatial scale is much broader than 75 meter radius areas. Our

[19] As most firms are not mobile across 500 meter radius areas, we cannot reliably compare firm quality estimates across these areas without sacrificing strength of identification. We can use the same estimator as for our baseline horse race specification because the first order condition in the updating rule for α_i is not affected.

results thus do not conflict with evidence in Bernard, Moxnes and Saito (2019) and Bazzi et al. (2017) that firm productivity propagates through closer buyer-supplier relationships. Our finding that peer quality in industries with which a firm's workers are likely to have more interaction, including those with stronger input-output relationships, contribute negatively to spillovers is evidence that the most useful knowledge flows from nearby firms are likely to come in an undirected way. The positive impacts of peer diversity for these types of firms is consistent with evidence in Henderson, Kuncoro and Turner (1995) that firms in young innovative industries benefit from cross-industry spillovers and contrasts with evidence for manufacturing in Greenstone, Hornbeck and Moretti (2010). The fact that spillovers are greater from higher quality peers is also consistent with such knowledge flows driving our results.

One key implication of our results is that while firms have some heterogeneity in incentives to seek out peer groups of particular types, higher quality firms have greater incentives to sort into locations with higher average quality peers. The following section demonstrates the existence of such sorting and considers implications for aggregates.

V. Firm Sorting and Agglomeration Economies

In this section, we provide evidence that higher quality firms are more likely to have peers of both higher average and aggregate quality. Moreover, the peer groups populated by higher quality firms tend to be in more productive locations, with stronger sorting into these locations among above median quality firms. We show that allocating firms randomly to peer groups generates weaker relationships between own firm quality, average peer quality, and location fundamentals than exist in equilibrium. The direct evidence documented here using estimated firm fixed effects reprises the more indirect evidence of such sorting from Figure 2 and Table 2.

We then turn to an analysis of whether this sorting matters for aggregates. Because spillovers primarily take a linear-in-means form, positive sorting into peer groups manifests itself in only small aggregate impacts on firm outcomes. The aggregate revenue reduction from eliminating the positive assortative matching into peer groups is 0.27 to 0.74 percent, with most coming from firms in the top quintile of the firm quality distribution.

Most exercises carried out in this section use estimates of firm quality α_i and spillovers γ_{LIM}, γ_{Agg}, and $\ddot{\sigma}$ from our primary specification in Table 2 column (3). Because in this specification, firm fixed effects α_i are primarily identified within local areas due to limited firm mobility, most analysis is carried out within these 500 meter radius areas. We use estimates from the specification without local area-year fixed effects in Table 2 column (6) as a point of comparison to gain a sense of the magnitude of firm sorting across locations. To distinguish them from primary specification estimates, we denote estimates from this alternative specification as $\widehat{\alpha}_i^6$, $\widehat{\gamma}_{\text{LIM}}^6$, $\widehat{\gamma}_{\text{Agg}}^6$, and $\widehat{\ddot{\sigma}}^6$. α_i^6 embodies an unknown combination of firm i's quality and fundamentals of the location in which firm i has spent the most time. Because only 32 percent of firms in our sample operate in more than one location, we see below that the distribution of $\widehat{\alpha}_i^6$ demeaned within 500 meter radius areas is almost identical to the distribution of $\widehat{\alpha}_i$. We take peer group compositions from the 2006 cross-section, as it is in the middle of the sample. For notational convenience, we drop t subscripts for the purposes of our discussion in this section.

A. *Relationships Between Average and Aggregate Peer Quality*

One important observation in our data is that because of firm heterogeneity, it is possible to confuse linear-in-means type spillovers for agglomeration type spillovers, as mean and aggregate peer quality are positively correlated. Figure 3 shows evidence to this effect. It shows relationships between average estimated peer quality $\frac{1}{|M_b|-1}\sum_{j\in M_b\neq i}\widehat{\alpha}_j$ and either aggregate peer quality $\sum_{j\in M_b\neq i}\widehat{\alpha}_j$ (left axis, solid line) or log aggregate peer revenue $\ln\sum_{j\in M_b\neq i}R_j$ (right axis, dashed line). Both plots show mostly monotonic positive relationships, indicating that higher average quality peers tend to be in peer groups of greater aggregate quality as well. That is, in 2006 there was positive sorting on levels of higher quality peers into larger and higher aggregate revenue peer groups.

Empirical relationships seen in Figure 3 reprise evidence from comparing estimates of γ_{Agg} in columns (2) and (3) of Table 2. After controlling for average peer quality, the estimated aggregate peer quality elasticity changes in a statistically insignificant way from slightly positive to slightly negative. A similar magnitude decline in $\widehat{\gamma}_{\text{Agg}}$ appears going from column (5) to column (6) in Table 2, though in this case both estimates are positive. Both of these comparisons reflect positive sorting of higher α_j firms into higher aggregate quality peer groups when evaluated in changes. The stronger sorting of higher quality firms into better aggregate quality peer groups in the 2006 cross-section relative to our empirical model estimates reported in Table 2 columns (2) and (3) reflects the fact that our empirical setup controls for such sorting on levels. Our evidence is thus that there is only a small amount of such sorting on changes remaining within 500 meter radius peer group areas, to the point of statistical insignificance. The strong sorting on levels seen in Figure 3 highlights an important drawback of cross-sectional studies of agglomeration using firm level data.

B. *Assortative Matching into Peer Groups and Locations*

Here we show evidence of a stronger relationship between firm and peer group quality than would be expected by chance. Because of "exclusion bias" (Caeyers and Fafchamps, 2020), the relationship between $\widehat{\alpha}_i$ and average peer quality would be negative if firms were randomly assigned to peer groups. To make informative comparisons that account for this bias, we carry out simulations in which we randomly assign firms to peer groups while holding each firm's estimated quality, $\widehat{\alpha}_i$, constant. This exercise is akin to that in Duranton and Overman (2005), who examine how much less localized firms in particular industries would be if allocated randomly to fixed locations across UK postal codes. Comparing observed peer group composition to average simulated peer group composition, we show that the equilibrium assignment of firms to peer groups involves a stronger relationship between firm quality and average peer quality than would exist under random assignment.

Comparisons of the relationships between $\widehat{\alpha}_i$ and the difference between observed average peer quality and counterfactual average peer quality under various scenarios allow us to characterize the magnitude of sorting across peer groups and locations. Figure 4 depicts these relationships. The solid black line shows the local linear polynomial relationship between $\widehat{\alpha}_i$ and $\frac{1}{|M_b|-1}\sum_{j\in M_b \neq i}\widehat{\alpha}_j - \overline{\alpha_i}$, in which $\overline{\alpha_i}$ is the average of average peer quality across 100 simulations of randomly allocating firms to peer groups within each area $B(b)$. Both $\widehat{\alpha}_i$ and $\frac{1}{|M_b|-1}\sum_{j\in M_b \neq i}\widehat{\alpha}_j - \overline{\alpha_i}$ have been demeaned within areas $B(b)$. That is, Figure 4 shows the relationship between firm quality and average peer quality after accounting for exclusion bias. The fact that this line is upward-sloping means that there is more sorting of higher quality firms into higher quality peer groups within local areas $B(b)$ than would exist through random allocation to peer groups. While this line is slightly upward-sloping up to about $\widehat{\alpha}_i = -1$, it turns more steeply upward for higher firm quality. This strengthening of positive sorting into peer groups with firm quality is consistent with our evidence on the heterogeneity in linear-in-means spillovers as functions of $\widehat{\alpha}_j$. Magnitudes are also informative. Because of sorting, the highest quality firms are in peer groups that are of 10-15 percent higher average quality than are firms with $\widehat{\alpha}_i = -1$. Lower quality firms exhibit approximately random sorting across peer groups. Because all analysis is performed within areas $B(b)$, the slope of the solid line represents a lower bound on the full magnitude of sorting across peer groups.

We next show that use of $\widehat{\alpha}_i^6$ rather than $\widehat{\alpha}_i$ to characterize sorting yields the same conclusions. In particular, the long dashed line in Figure 4 shows the same relationship as the solid line but is built using estimates of firm quality from Table 2 column (6) demeaned within 500 meter radius areas $B(b)$. As the solid and long dashed lines coincide, the two estimates of α_i are very similar once area fixed effects are taken out. We take this as evidence that it is reasonable to use $\widehat{\alpha}_i^6$ not demeaned within $B(b)$ to form comparisons that can be used to characterize sorting between areas.

The slope of the short dashed line in Figure 4 reflects a composite of firm sorting across locations and peer groups. It is built analogously to the solid line but using $\widehat{\alpha}_i^6$ (demeaned universally) rather than $\widehat{\alpha}_i$ (demeaned within local areas) as a basis. Because they are estimated without area fixed effects, $\widehat{\alpha}_i^6$ embody a combination of firm quality and location fundamentals. The fact that the short dashed line is more steeply upward sloping than the other two lines is thus evidence that beyond positive sorting into peer groups within areas $B(b)$, firms additionally positively sort either between areas on location fundamentals and/or across peer groups that are located in different areas $B(b)$. Such additional sorting is very strong, such that the average quality of peers and location for the typical high quality firm at $\widehat{\alpha}_i^6 = 3.5$ is over 50 percent greater than that at $\widehat{\alpha}_i^6 = -1$. This comes despite the fact that there is less dispersion in $\widehat{\alpha}_i^6$ than $\widehat{\alpha}_i$, as seen in Table 2. Our estimates exhibit a combination of large differences in location fundamentals that interact with sorting on peer quality and strong sorting on peer quality across locations. Strong polarization in the sum of location fundamentals and average firm quality across locations is apparent from our estimates.

The fact that average and aggregate peer group quality are positively correlated is one central finding of this paper and merits some speculation about potential mechanisms that could generate this pattern in equilibrium. With positive linear-in-means peer effects only, all firms have an incentive to chase higher quality peers. This force would push peer groups with high quality firms to have higher aggregate revenue and employment. Convexity in spillovers as a function of peer quality, as seen in Table 6 column (6), only strengthens this incentive. Local rents would potentially be bid up in these locations with the higher cost of doing business sustained with the larger spillovers. Because higher quality firms also benefit more from high quality peers in dollar terms, there is positive assortative matching of firms into peer groups. Finally, such agglomerations of larger high quality peer groups will tend to locate in high local productivity "prime locations" (Ahlfeldt, Albers and Behrens, 2020). There are some parallels to the conceptual observation in the local public finance literature that locations with strong tax bases and high quality public goods are likely to be crowded by those that benefit the most from such spillovers, as in the model in Calabrese, Epple and Romano (2012).

C. Aggregate Impacts of Sorting

We now quantify the consequences of sorting for aggregate firm revenue. We use estimates of α_i, γ_{LIM}, and γ_{Agg} from Table 2 columns (3) and (6) to construct aggregate firm revenue under two different simulated random allocations of firms to peer groups. The first randomization procedure holds the number and size of all peer groups constant whereas the second procedure only holds the number of peer groups constant. Both procedures are consistent with the randomization carried out for the analysis in the prior sub-section. While the consequences of linear-in-means type spillovers are identical under the two randomization procedures, we find it instructive to consider the implications of adding estimated aggregate type spillovers. We carry out randomization within 500 meter radius areas only for the column (3) estimates and implement both local and universal randomization for the column (6) estimates.

Table 7 reports impacts of firm sorting across peer groups by aggregating revenue under the two counterfactual scenarios discussed above and comparing it to total observed firm revenue.[20] Entries in Table 7 show means and standard deviations of revenue impacts in percentage terms from carrying out 100 simulations of counterfactual revenue given random allocation of firms to peer groups. Results in the two columns under the header "Fixed Group Size" are generated holding peer group size fixed and those under the header "Equal Group Size" are generated given full randomization of firms across peer groups. In each column headed by LIM, aggregate firm revenue under counterfactual allocation C is given by

$$\ln Y^C_{LIM} = \ln\left[\sum_i exp\left(y_i + \frac{\widehat{\gamma}_{\text{LIM}}}{|M^C_{b(i)}| - 1}\sum_{j\in M^C_{b(i)},\neq i}\widehat{\alpha}_j - \frac{\widehat{\gamma}_{\text{LIM}}}{|M_{b(i)}| - 1}\sum_{j\in M_{b(i)},\neq i}\widehat{\alpha}_j\right)\right].$$

That is, we calculate aggregate revenue in the counterfactual environment in which actual peer group quality is replaced by peer group quality determined under counterfactual allocation C. This way of calculating impacts of sorting is not sensitive to the normalization of firm fixed effects, as any normalization differences out. Comparison against aggregate revenue $\ln Y = \ln\left[\sum_i e^{y_i}\right]$ shows how much aggregate revenue would be impacted if there were no sorting across locations. In each column headed by LIM+AGG, counterfactual revenue is constructed with the addition of $\widehat{\gamma}_{\text{Agg}}[\sum_{j\in M^C_{b(i)},\neq i}\widehat{\alpha}_j - \sum_{j\in M_{b(i)},\neq i}\widehat{\alpha}_j] + \widehat{\widehat{\sigma}}[|M^C_{b(i)}| - |M_{b(i)}|]$ within the exponential.

Results in the first row of Table 7 show that the sorting of higher quality firms into higher average quality peer groups within local areas increases aggregate firm revenue by 0.27 percent through linear-in-means effects. Randomly allocating firms across peer groups tends to make average peer group size smaller for larger high α_i firms and larger for smaller low α_i firms. The result is larger reductions in revenue due to spillovers for larger firms than corresponding increases for smaller

[20]This is a partial equilibrium analysis in the sense that it assumes reshuffling firms across peer groups only affects log revenue through the peer effects mechanisms studied here.

firms in dollar terms, netting out to a small aggregate effect. We emphasize that this 0.27 percent result understates the true aggregate impact of sorting because it does not include sorting impacts between different 500 meter radius areas, which we consider further in the context of our discussion of results in the third row of Table 7 below. Because our estimate of γ_{Agg} is slightly negative, the sum of the linear-in-means and agglomeration forces is greater and near 0, as seen in column (2), though we discount this evidence given that our estimate of γ_{Agg} is not significant and is of opposite sign than expected. As seen in Figure 4, firms with lower than average α_i would tend to benefit from imposing random sorting whereas the reverse is true for firms with greater than average α_i. In particular, aggregate revenue of firms in the bottom quintile of the quality distribution would be 0.18 percent higher under randomized peer groups through linear-in-means forces whereas that of firms in the top quintile would be 0.35 percent lower (not reported). If peer group size is also randomized (columns 3 and 4), the linear-in-means effect is by construction essentially unchanged but the overall agglomeration influence switches sign to reinforce the small negative linear-in-means impact. This comes from the positive but insignificant estimated coefficient on the number of peers $\widehat{\ddot{\sigma}}$.

The second row of Table 7 shows analogous objects but using parameter estimates from Table 2 column (6) while maintaining random allocation of firms to peer groups within 500 meter radius areas. As $\widehat{\gamma}^6_{\text{LIM}}$ is very close to $\widehat{\gamma}_{\text{LIM}}$, entries in the first and third columns of Table 7 are very similar in the first two rows. However, adding the impact of a positive $\widehat{\gamma}^6_{\text{Agg}}$ now results in aggregate revenue impacts of much greater magnitudes than in the first row. Incorporating the small aggregate elasticity of $\widehat{\gamma}^6_{\text{Agg}} = 0.0011$ means that randomization across peer groups of fixed size reduces aggregate revenue by 0.68 percent and that across peer groups of variable size reduces aggregate revenue by 1.35 percent. These results give a sense of the upper bound of aggregate implications of sorting within local areas.

The third row of Table 7 is constructed analogously to the second row, except that randomization is carried out universally rather than within 500 meter radius areas only. Results thus show the consequences of a combination of eliminating sorting between local areas and attributing location fundamentals to firm fixed effects. Results in the third row are thus likely an upper bound on the true aggregate impacts of sorting. Results in the first and third columns show that absent sorting aggregate revenue would be 0.74 percent lower through reduction in linear-in-means type spillovers. Including the estimated size of aggregate spillovers as well, this impact rises by about 1.5-1.8 additional percentage points.

As the aggregate impacts in Table 7 use our main specification, heterogeneous linear-in-means treatment effects are not accommodated. While we do not have the statistical power to precisely estimate the relative magnitudes of spillovers imparted by higher quality firms, we have shown evidence that firms in the top tercile of the quality distribution impart larger spillovers than do lower quality firms. To provide a sense of how important this can be for aggregates, we carry out the same counterfactual exercises using estimates from Table 6 column (6) instead, combining the heterogeneous ("HET") impact of the fraction of peers in the top tercile with the linear-in-means component of the counterfactual. Reported in Table E3, we see the aggregate effects of the LIM and HET effects in column (1) are 0.34 percent of aggregate firm revenue rather than 0.27 percent absent the accommodation of heterogeneous treatment effects. Other impacts under the counterfactual scenarios we consider, also reported in Table E3, follow the same patterns as in Table 7, though with slightly larger magnitudes.

VI. Conclusions

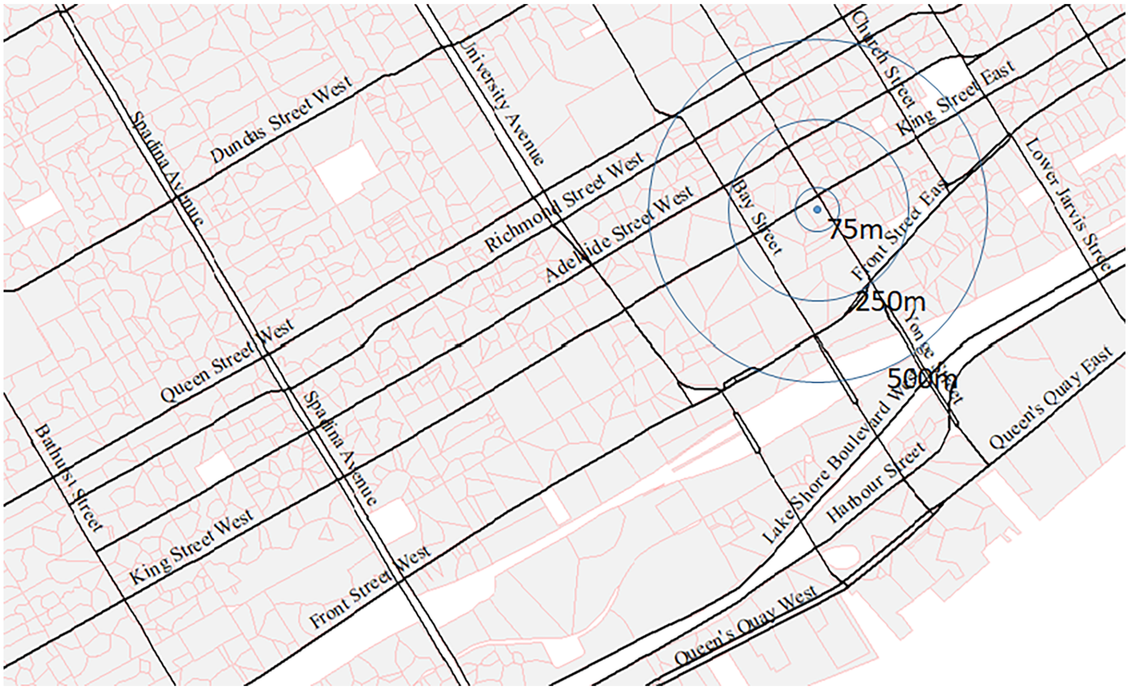
Considerable evidence exists on the magnitude of aggregate increasing returns to scale at the local labor market level. Yet little empirical evidence exists at microgeographic spatial scales. Using estimates from a nonlinear fixed effects empirical model of peer effects, evidence in this paper shows that firms benefit from being near higher quality peers, but that the nature of spillovers is entirely through average rather than aggregate peer quality. In particular, the elasticity of firm revenue and TFP with respect to the average quality of other firms within 75 meters is about 0.024. This elasticity decays quickly with distance such that the average spillover beyond 75 meters is not distinguishable from zero. When making comparisons within 500 meter radius regions, we find no evidence that the average firm benefits from being surrounded by a greater amount of economic activity within 75 meters. To the extent that scale matters, it is the amount of activity in regions of 500 meter radius or larger that is mostly important, not the very local scale.

Using estimates of firm quality, we show that there is assortative matching of higher quality firms into peer groups of greater average and aggregate quality. As externalities imparted by higher quality firms are greater, there is an incentive for firms to locate in peer groups with higher quality peers. This force may lead to the positive observed association between average and aggregate quality of peer groups. Because spillovers are linear-in-means, there are mostly distributional consequences associated with harmonizing firm composition across peer groups, with an associated reduction in aggregate firm revenue of less than 1 percent.

Additional mechanisms beyond those documented in this paper are required to justify the magnitudes of metro level elasticities of TFP with respect to population, which are estimated to be in the

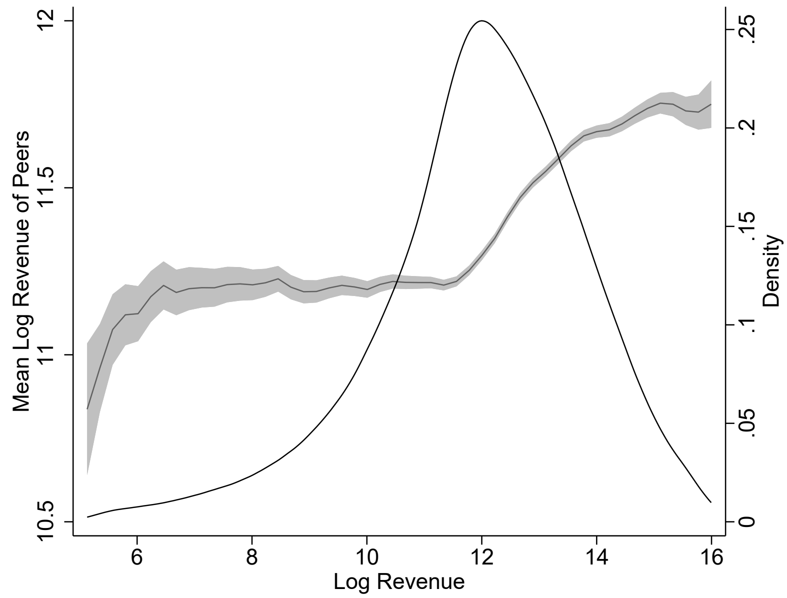
0.03-0.05 range (Combes and Gobillon, 2015). One important aspect held constant in this study is location fundamentals within 500 meter radius areas. As such, we provide evidence that a large fraction of aggregate increasing returns to scale must operate at higher levels of aggregation. An important question for future research is thus how microgeographic estimates like those reported here aggregate up to the local labor market level. Our evidence highlights the importance of considering essential firm heterogeneity for rationalizing observations about increasing returns to scale both at microgeographic and metro area level spatial scales.

FIGURE 1. – Map of Downtown Toronto

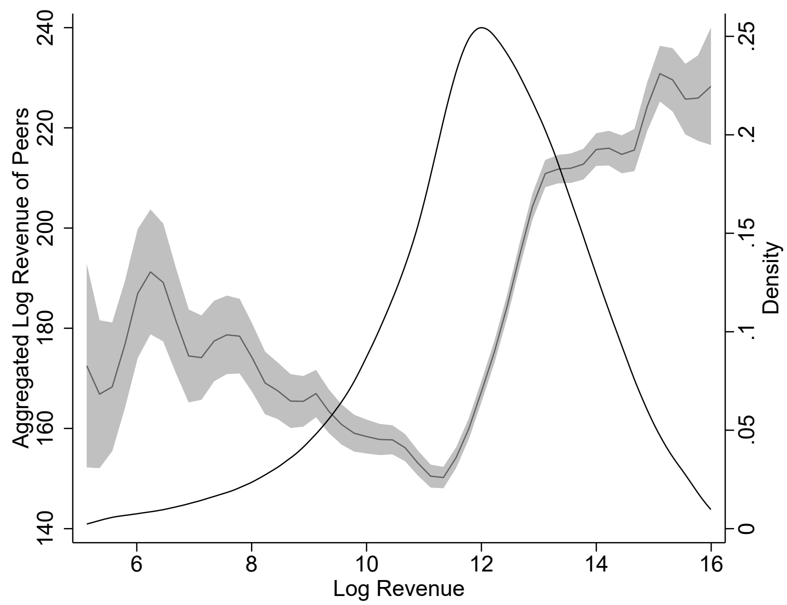


Notes: Postal codes are outlined by thin red lines. Major streets are in black. All postal codes with centroids within the indicated central 75 meter radius circle are included in the indicated example peer group area.

FIGURE 2. – Sorting on Peer Group Quality



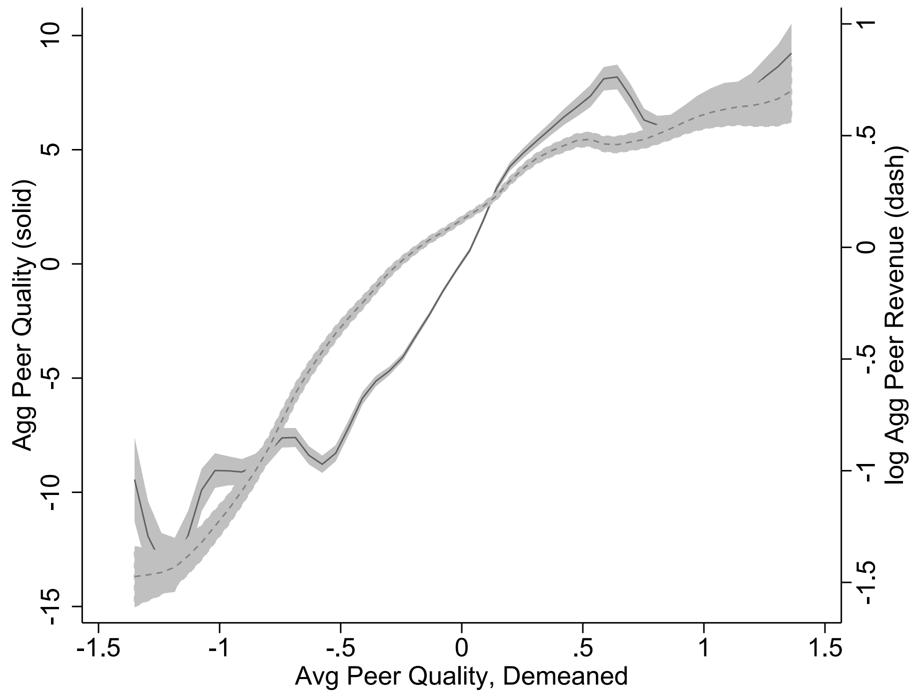
(a) Mean Peer Log Revenue by Firm Log Revenue



(b) Aggregate Peer Log Revenue by Firm Log revenue

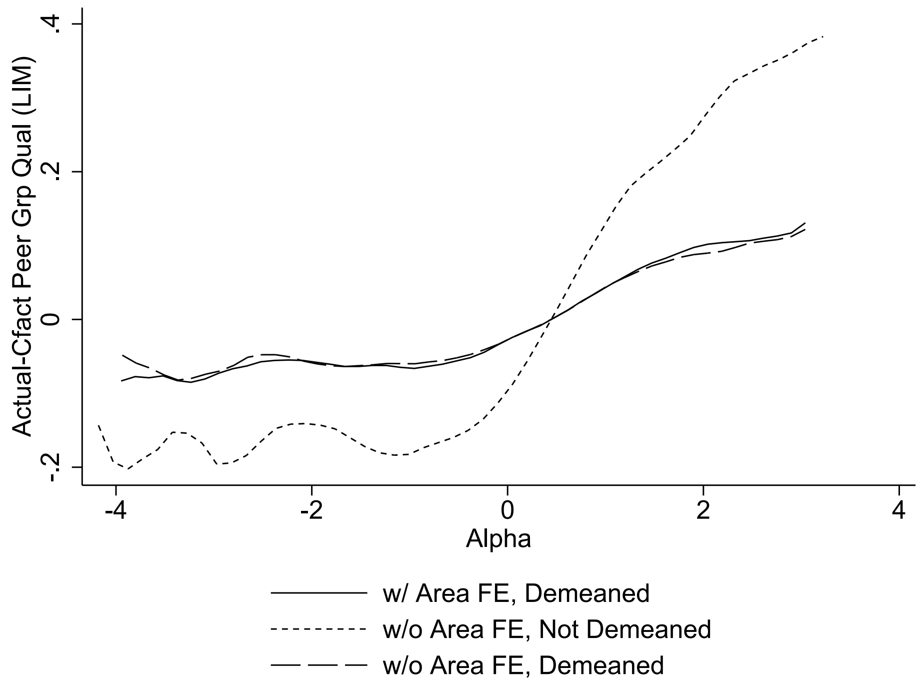
Notes: Plots show local polynomial relationships between firm log revenue and mean peer log revenue in Panel A and aggregate peer log revenue in Panel B. Shaded regions indicate 95% confidence intervals. Only firms in the primary estimation sample are included. The sample excludes multi-location firms and those in peer group areas with one or more member postal code with an area that is greater than $\pi 75^2$ sq meters (0.018 sq km) and peer group areas with fewer than two high-skilled services firms in any year 2001-2012. The sample only includes firms in the Montreal, Toronto, or Vancouver census metropolitan areas. To make the graph easier to read, the distribution of firm log revenue on the horizontal axis is trimmed at the 1^{st} and 99^{th} percentiles.

Figure 3. – Relationships Between Treatment Size and Peer Group Composition



Notes: This figure is constructed using data from the 2006 cross-section of the primary estimation sample. Results are based on estimates in Table 2 column (3). Plots show local polynomial relationships between estimated average peer group quality $\frac{1}{|M_b|-1}\sum_{j\in M_b,\neq i}\widehat{\alpha}_j$ and estimated aggregate peer group quality $\sum_{j\in M_b,\neq i}\widehat{\alpha}_j$ (solid line, left axis) or log aggregate peer revenue $\ln\sum_{j\in M_b,\neq i}R_j$ (dashed line, right axis). All objects are demeaned within 500 meter radius peer group areas. The distribution of demeaned estimated average peer group quality on the horizontal axis is trimmed at the 2.5^{th} and 97.5^{th} percentiles.

FIGURE 4. – Sorting into Peer Groups



Notes: This figure is constructed using data from the 2006 cross-section of the primary estimation sample. Results are based on estimates reported in Table 2 column (3) ($\widehat{\alpha}_i$) and column (6) ($\widehat{\alpha}_i^6$). The solid line is the nonparametric relationship between $\widehat{\alpha}_i$ and $\frac{1}{|M_b|-1}[\sum_{j\in M_b,\neq i}\widehat{\alpha}_j - \sum_{j\in M_b^C,\neq i}\widehat{\alpha}_j]$ in which counterfactual peer groups M_b^C are determined by random assignment of firms to peer groups within 500 meter radius areas. The long dashed line is the nonparametric relationship between $\widehat{\alpha}_i^6$ and $\frac{1}{|M_b|-1}[\sum_{j\in M_b,\neq i}\widehat{\alpha}_j^6 - \sum_{j\in M_b^C,\neq i}\widehat{\alpha}_j^6]$ under the same random assignment scheme and after demeaning $\widehat{\alpha}_i^6$ within local 500 meter radius areas. The short dashed line is the nonparameteric relationship between $\widehat{\alpha}_i^6$ and $\frac{1}{|M_b|-1}[\sum_{j\in M_b,\neq i}\widehat{\alpha}_j^6 - \sum_{j\in M_b^C,\neq i}\widehat{\alpha}_j^6]$ in which counterfactual peer groups M_b^C are determined by random assignment of firms to peer groups across all locations, without demeaning. The distribution of demeaned α on the horizontal axis is trimmed at the 2.5^{th} and 97.5^{th} percentiles.

Table 1. – Descriptive Statistics

	All Industries		High-Skilled Services (NAICS 5)		
	Multi	Single	Multi	Single	Primary
	(1)	(2)	(3)	(4)	(5)
Panel A: Statistics					
ln Revenue	15.06	12.05	14.50	11.60	11.93
	(2.21)	(1.98)	(2.42)	(2.03)	(2.09)
ln Payroll per Worker	10.65	10.13	10.80	10.29	10.42
	(0.75)	(0.92)	(0.87)	(0.99)	(0.98)
ln Employment	3.11	1.18	2.73	0.92	1.06
	(1.56)	(1.01)	(1.77)	(0.94)	(1.00)
Area of Postal Code	0.17	0.11	0.05	0.09	0.006
(sq km)	(12.90)	(11.31)	(0.87)	(10.28)	(0.005)
Panel B: Sample Sizes					
Observations	245,500	2,645,300	78,500	1,075,700	282,000
Obs., Non-Missing Rev.	233,200	2,520,300	74,300	1,023,200	269,100
# Firms	30,500	428,400	10,600	181,500	56,000
# Peer Group-Years	128,300	843,300	47,600	501,500	42,100

Notes: Statistics are for all firms in the Montreal, Toronto, and Vancouver census metropolitan areas for the 2001-2012 period. Panel A shows means with standard deviations in parentheses. The estimation sample in the final column excludes firms in peer group areas with one or more member postal code with an area that is greater than $\pi 75^2$ sq meters (0.018 sq km) and peer group areas with fewer than two high-skilled services firms in any year 2001-2012.

TABLE 2. – Results for 75 meter Radius Peer Groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Avg. Peer Firm F.E.	0.0179		0.0240	0.0285		0.0210	0.0212		0.0272
	(0.0088)		(0.0090)	(0.0063)		(0.0070)	(0.0083)		(0.0089)
Agg. Peer Firm F.E.		0.0003	-0.0006		0.0019	0.0011		0.0004	-0.0006
		(0.0007)	(0.0007)		(0.0006)	(0.0006)		(0.0007)	(0.0007)
Number of Peers		0.0003	0.0005		0.0003	0.0005		0.0003	0.0005
		(0.0006)	(0.0006)		(0.0004)	(0.0005)		(0.0006)	(0.0006)
Firm F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind.×Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Area×Year F.E.	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	269,100	269,100	269,100	270,700	270,700	270,700	269,100	269,100	269,100
# of Peer Group-Years	42,100	42,100	42,100	42,900	42,900	42,900	42,100	42,100	42,100
# of Firm	56,000	56,000	56,000	56,100	56,100	56,100	56,000	56,000	56,000
SD Firm F.E.	1.95	1.95	1.96	1.87	1.87	1.87	1.97	1.96	1.97
SD Avg. Peer Firm F.E.	1.11		1.12	0.98		0.97	1.11		1.12
SD Agg. Peer Firm F.E.		14.17	14.27		13.23	13.35		13.94	14.08
SD Number of Peers		19.52	19.52		19.49	19.49		19.52	19.52
Implied 90-10 Gap (LIM)	0.05		0.07	0.07		0.05	0.06		0.08

Notes: Estimates of γ in equation (7) under linear-in-means aggregation weights are shown in columns (1), (4), and (7) and agglomeration weights are shown in columns (2), (5), and (8). Estimates of γ_{LIM} and γ_{Agg} in equation (8) are estimated jointly in columns (3), (6), and (9). Log firm revenue is the dependent variable in all columns. "Ind." fixed effects are for 2-digit NAICS industry and "Area" fixed effects are for 500 meter radius regions. Symmetric wild bootstrapped standard errors clustered at the peer group area level are in parentheses.

Table 3. – Identification Checks

Panel A: Using all available observations

Outcome	Residual t				
	(1)	(2)	(3)	(4)	(5)
Avg. peer quality, $t+1$	0.0045 (0.0033)				
Avg. peer quality, $t+2$	0.0032 (0.0034)				
Avg. peer quality, $t-1$		-0.0037 (0.0032)			
Avg. peer quality, $t-2$		-0.0020 (0.0033)			
Avg. quality of entrants, $t+1$			0.0007 (0.0012)		
Avg. quality of departures, $t-1$				0.0012 (0.0010)	
Residual $t-1$					0.2663 (0.0017)
Observations	171,600	170,700	144,200	148,400	210,000

Panel B: Using only firms that change location between period t and $t+1$

Outcome	Residual t			Residual $t+1$	
	(1)	(2)	(3)	(4)	(5)
Avg. peer quality, next loc. t	-0.0150 (0.0102)				
Avg. quality of entrants, next loc. t		-0.0087 (0.0063)			
Avg. quality of departures, next loc. t			0.0032 (0.0078)		
Avg. quality of entrants, previous loc. $t+1$				0.0016 (0.0078)	
Avg. quality of departures, previous loc. $t+1$					0.0009 (0.0066)
Observations	8,500	7,200	7,000	6,600	7,400

Notes: Each column in each panel shows coefficients and standard errors from a separate regression. The dependent variable is the residual term associated with estimates in Table 2 column (3). Panel A uses all available observations whereas Panel B focuses on movers and only uses observations either immediately prior to a move across peer group areas (columns 1-3) or immediately afterwards (columns 4-5). All predictor variables (except Panel A, column 5) are calculated as the mean of estimated α_j in the indicated period and peer group location. The predictor variable in Panel A, column 5 is the lagged residual. Symmetric wild bootstrapped standard errors clustered at the peer group area level are in parentheses.

TABLE 4. – Alternative Outcomes

	Adj. Log Revenue	Log Emp.	Log Payroll	Incl. Multi
	(1)	(2)	(3)	(4)
Avg. Peer Firm F.E.	0.0209	0.0156	0.0125	0.0187
	(0.0114)	(0.0107)	(0.0144)	(0.0073)
Agg. Peer Firm F.E.	-0.0012	0.0001	0.0005	-0.0007
	(0.0010)	(0.0009)	(0.0012)	(0.0005)
Number of Peers	0.0004	-0.0002	0.0004	0.0007
	(0.0007)	(0.0003)	(0.0004)	(0.0005)
Observations	269,100	170,900	174,700	309,900
# of Peer Group-Years	42,100	36,100	36,400	45,300
# of Firms	56,000	35,400	36,000	62,700
Mean y	11.42	1.06	11.45	12.23
SD y	4.23	1.00	1.47	2.29
SD Firm F.E.	2.05	0.96	1.39	2.14
SD Avg. Peer Firm F.E.	1.15	0.47	0.71	1.18
SD Agg. Peer Firm F.E.	16.41	6.53	7.71	21.09
SD Number of Peers	19.52	19.95	19.94	23.76
Implied 90-10 Gap (LIM)	0.06	0.02	0.02	0.06

Notes: Estimates are analogous to those in Table 2 column (3) but using alternative outcome variables. Adjusted Log Revenue used in column (1) is calculated as log revenue divided by $\frac{1+\eta_k}{\eta_k(1-\theta_k)-\theta_k}$. Details are in Appendix A.A1. "Incl. Multi" in the final column is the same regression as in Table 2 column (3) except multi-location firms are included in the sample. For each of these firms, all revenue is necessarily assigned to the one location reported on the firm tax filing. Symmetric wild bootstrapped standard errors clustered at the peer group area level are in parentheses.

Table 5. – Spatial Decay

Panel A: Other Peer Group Definitions

Peer Group Area	75m Radius	150m Radius	200m Radius	250m Radius
	(1)	(2)	(3)	(4)
Avg. Peer Firm F.E.	0.0240	0.0253	0.0224	0.0233
	(0.0090)	(0.0106)	(0.0112)	(0.0109)
Agg. Peer Firm F.E.	-0.0006	-0.0001	-0.0007	0.0002
	(0.0007)	(0.0005)	(0.0005)	(0.0004)
Number of Peers	0.0005	-0.0003	0.0000	-0.0001
	(0.0006)	(0.0003)	(0.0002)	(0.0002)
Observations	269,100	269,000	268,900	268,800
# of Peer Group-Years	42,100	35,900	32,700	30,200
# of Firms	56,000	55,900	55,900	55,900
SD Avg. Peer Firm F.E.	1.12	1.05	1.02	1.00
SD Agg. Peer Firm F.E.	14.27	20.20	24.12	30.21
SD Number of Peers	19.52	42.08	53.25	64.50
Implied 90-10 Gap (LIM)	0.07	0.07	0.06	0.07

Panel B: Competing Peer Groups, Using Post-Estimation Data from Main Specification

	(1)	(2)	(3)	(4)
Avg. Peer Firm F.E. (75m)	0.0222	0.0216	0.0251	0.0227
	(0.0043)	(0.0040)	(0.0035)	(0.0031)
Avg. Peer Firm F.E. (150m)	0.0051	0.0027		
	(0.0047)	(0.0040)		
Avg. Peer Firm F.E. (200m)	-0.0074		-0.0014	
	(0.0048)		(0.0035)	
Avg. Peer Firm F.E. (250m)	0.0043			0.0016
	(0.0041)			(0.0031)

Notes: Estimates in Panel A are analogous to those in Table 2 column (3) except for the indicated peer group area definitions. All samples exclude peer groups that include firms for which at least one contextual fixed effect required for estimation is not separately identified from the firm fixed effect. As a result, slightly more observations are excluded when using the broader peer group area definitions compared to the primary estimation sample in column (1). Symmetric wild bootstrapped standard errors clustered at the peer group area level are in parentheses. Estimates in Panel B use post-estimation data from Panel A column (1) to run regressions of log revenue (residualized from the estimated fixed effects) on average peer quality within 75, 150, 200, and/or 250 meters. Aggregate peer quality and the number of peers within 75 meters are included as controls in all columns. Standard errors clustered at the peer group area level are in parentheses.

TABLE 6. – Peer Group Composition

	Industry Connections					Peer Quality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Avg. Peer Firm F.E.	0.0241	0.0239	0.0240	0.0241	0.0243	0.0073	0.0157
	(0.0090)	(0.0090)	(0.0090)	(0.0090)	(0.0090)	(0.0100)	(0.0103)
Agg. Peer Firm F.E.	-0.0006	-0.0006	-0.0006	-0.0006	-0.0006	-0.0005	-0.0005
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Number of Peers	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Frac. Same 2-Digit	-0.0256				-0.0426		
	(0.0264)				(0.0267)		
Frac. High Input-Output		-0.0359			-0.0381		
		(0.0280)			(0.0270)		
Frac. High Occ. Sim.			0.0076		0.0042		
			(0.0253)		(0.0285)		
Frac. High Worker Flows				0.0243	0.0453		
				(0.0226)	(0.0262)		
Frac. High Alpha (> 66th pctile)						0.0859	
						(0.0250)	
Frac. High Alpha (> 50th pctile)							0.0274
							(0.0223)
Observations	269,100	269,100	269,100	269,100	269,100	269,100	269,100
# of Peer Group-Years	42,100	42,100	42,100	42,100	42,100	42,100	42,100
# of Firm	56,000	56,000	56,000	56,000	56,000	56,000	56,000
SD Firm F.E.	1.96	1.96	1.96	1.96	1.96	1.95	1.96
SD Avg. Peer Firm F.E.	1.12	1.12	1.12	1.12	1.12	1.11	1.11
SD Agg. Peer Firm F.E.	14.27	14.28	14.27	14.26	14.26	14.26	14.27
SD Number of Peers	19.52	19.52	19.52	19.52	19.52	19.52	19.52

Notes: Table presents estimates of equation (8) with the addition of regressors indicated at left in the table. The first column includes the fraction of peers in the same 2-digit industry as an additional regressor. Columns (2)-(4) includes the fraction of peers in the top tercile of each indicated type of industry connection as additional regressors. Column (5) includes all additional controls used in columns (1)-(4) together. See Appendix D for more details about connectivity weights. The final two columns include the fraction of peers in the top tercile (column 6) or above the median (column 7) of the local 500 meter radius area's firm quality distribution as an additional regressor. Symmetric wild bootstrapped standard errors clustered at the peer group area level are in parentheses.

Table 7. – Aggregate Impacts of Counterfactual Firm Allocation Across Peer Groups

Randomization Type	Fixed Group Size		Equal Group Size	
Nature of Spillovers Considered	LIM	LIM + AGG	LIM	LIM + AGG
	(1)	(2)	(3)	(4)
Estimates w/ Area × Year F.E., Randomized Within Areas	-0.0027	-0.0019	-0.0029	-0.0060
	(0.0006)	(0.0007)	(0.0006)	(0.0003)
Estimates w/o Area × Year F.E., Randomized Within Areas	-0.0023	-0.0068	-0.0027	-0.0135
	(0.0005)	(0.0017)	(0.0005)	(0.0010)
Estimates w/o Area × Year F.E., Randomized Across All Locations	-0.0074	-0.0218	-0.0076	-0.0264
	(0.0008)	(0.0012)	(0.0010)	(0.0012)

Notes: Table presents the means and standard deviations of changes in aggregate revenue that would ensue under 100 simulations of various scenarios in which sorting of firms across peer groups is eliminated. Results in the two columns under the header "Fixed Group Size" are generated holding peer group size fixed and those under the header "Equal Group Size" are generated given full randomization of firms across peer groups. In each column headed by LIM, counterfactual firm revenue absent sorting is calculated adjusting for the linear-in-means component of the spillover and in each column headed by LIM+AGG, both linear-in-means and agglomeration terms are included in the calculation. The first row uses estimates from Table 2 column (3) and imposes demeaning and randomization across peer groups within 500 meter radius areas. The second row uses estimates from Table 2 column (6) instead with the same demeaning and randomization procedures. The third row uses estimates from Table 2 column (6) but demeans and randomizes across all peer groups.

REFERENCES

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification properties of recent production function estimators." *Econometrica*, 83(6): 2411–2451.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence from the Berlin Wall." *Econometrica*, 83(6): 2127–2189.

Ahlfeldt, Gabriel M, Thilo Albers, and Kristian Behrens. 2020. "Prime locations." CEPR Discussion Paper 15470.

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics*, 30(C): 98–108.

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. 2012. "Estimating Spillovers Using Panel Data, with an Application to the Classroom." *Quantitative Economics*, 3(3): 421–470.

Arzaghi, Mohammad, and J Vernon Henderson. 2008. "Networking Off Madison Avenue." *Review of Economic Studies*, 75(4): 1011–1038.

Atkin, David, M. Keith Chen, and Anton Popov. 2022. "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley." National Bureau of Economic Research Working Paper 30147.

Baum-Snow, Nathaniel, and Ronni Pavan. 2012. "Understanding the City Size Wage Gap." *Review of Economic Studies*, 79(1): 88–127.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy*, 116(6): 1150–1196.

Bazzi, Samuel, Amalavoyal V. Chari, Shanthi Nataraj, and Alexander D. Rothenberg. 2017. "Identifying Productivity Spillovers Using the Structure of Production Networks." Working Paper.

Bernard, Andrew B, Andreas Moxnes, and Yukiko U Saito. 2019. "Production networks, geography, and firm performance." *Journal of Political Economy*, 127(2): 639–688.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica*, 81(4): 1347–1393.

Caeyers, Bet, and Marcel Fafchamps. 2020. "Exclusion Bias and the Estimation of Peer Effects." Working Paper.

Calabrese, Stephen M, Dennis N Epple, and Richard E Romano. 2012. "Inefficiencies from Metropolitan Political and Fiscal Decentralization: Failures of Tiebout Competition." *Review of Economic Studies*, 79(3): 1081–1111.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics*. Vol. 5 of *Handbook of Regional and Urban Economics*, , ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, Chapter 5, 247–348. Elsevier.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection." *Econometrica*, 80(6): 2543–2594.

Conley, Timothy, Nirav Mehta, Ralph Stinebrickner, and Todd Stinebrickner. 2015. "Social Interactions, Mechanisms, and Equilibrium: Evidence from a Model of Study Time and Academic Achievement." National Bureau of Economic Research Working Paper 21418.

Cornelissen, Thomas, Christian Dustmann, and Uta Schonberg. 2017. "Peer Effects in the Workplace." *American Economic Review*, 107(2): 425–56.

Davis, Donald R, and Jonathan I Dingel. 2019. "A Spatial Knowledge Economy." *American Economic Review*, 109(1): 153–70.

De la Roca, Jorge, and Diego Puga. 2017. "Learning by Working in Big Cities." *Review of Economic Studies*, 84(1): 106–142.

De Loecker, Jan. 2011. "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity." *Econometrica*, 79(5): 1407–1451.

De Loecker, Jan, and Jan Eeckhout. 2018. "Global Market Power." CEPR Discussion Paper 13009.

Duranton, Gilles, and Henry Overman. 2005. "Testing for Localization Using Micro-Geographic Data." *Review of Economic Studies*, 72(4): 1077–1106.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review*, 100(3): 1195–1213.

Faggio, Giulia, Olmo Silva, and William C. Strange. 2017. "Heterogeneous Agglomeration." *Review of Economics and Statistics*, 99(1): 80–94.

Fox, Jeremy T, and Valérie Smeets. 2011. "Does Input Quality Drive Measured Differences in Firm Productivity?" *International Economic Review*, 52(4): 961–989.

Fujita, Masahisa, and Hideaki Ogawa. 1982. "Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations." *Regional Science and Urban Economics*, 12(2): 161–196.

Gandhi, Amit, Salvador Navarro, and David A Rivers. 2020. "On the identification of gross output production functions." *Journal of Political Economy*, 128(8): 2973–3016.

Gibbons, Steve, Henry G. Overman, and Eleonora Patacchini. 2015. "Spatial Methods." In *Handbook of Regional and Urban Economics*. Vol. 5 of *Handbook of Regional and Urban Economics*, , ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, 115–168. Elsevier.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy*, 118(3): 536–598.

Guryan, Jonathan, Kory Kroft, and Matthew J Notowidigdo. 2009. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." *American Economic Journal: Applied Economics*, 1(4): 34–68.

Hanlon, W. Walker, and Antonio Miscio. 2017. "Agglomeration: A Long-run Panel Data Approach." *Journal of Urban Economics*, 99(C): 1–14.

Henderson, Vernon, Ari Kuncoro, and Matthew Turner. 1995. "Industrial Development in Cities." *Journal of Political Economy*, 103(5): 1067–1090.

Kantor, Shawn, and Alexander Whalley. 2014. "Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks." *Review of Economics and Statistics*, 96(1): 171–188.

Kerr, William R, and Scott Duke Kominers. 2015. "Agglomerative Forces and Cluster Shapes." *Review of Economics and Statistics*, 97(4): 877–899.

Liu, Xiaodong, Eleonora Patacchini, and Yves Zenou. 2014. "Endogenous Peer Effects: Local Aggregate or Local Average." *Journal of Economic Behavior and Organization*, 103: 39–59.

MacKinnon, James G. 2006. "Bootstrap Methods in Econometrics." *The Economic Record*, 82: S2–S18.

Manning, Alan. 2009. "The Plant Size-Place Effect: Agglomeration and Monopsony in Labour Markets." *Journal of Economic Geography*, 10(5): 717–744.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–542.

Moretti, Enrico. 2004. "Human Capital Externalities in Cities." In *Handbook of Regional and Urban Economics*. Vol. 4 of *Handbook of Regional and Urban Economics*, , ed. J. V. Henderson and J. F. Thisse, Chapter 51, 2243–2291. Elsevier.

Rosenthal, Stuart S., and William C. Strange. 2003. "Geography, Industrial Organization, and Agglomeration." *Review of Economics and Statistics*, 85(2): 377–393.

Rosenthal, Stuart S, William C Strange, and Joaquin A Urrego. 2021. "JUE insight: Are city centers losing their appeal? Commercial real estate, urban spatial structure, and COVID-19." *Journal of Urban Economics*, 103381.

Serafinelli, Michel. 2019. "Good Firms, Worker Flows and Productivity." *Journal of Labor Economics*, 37(3): 747–792.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press.

Appendix A. The Imperfect Competition Case

This section develops structural equations that describe relationships between firm revenue or variable factor demand and peer group composition. Using these equations, we provide structural interpretations of empirical estimates. We study an environment in which the variable input share and output demand elasticity are industry-specific.

A1. Setup

With market power, each firm charges a markup over marginal cost that depends on the elasticity of demand it faces for its product. To model this phenomenon, we begin with an adapted version of the environment considered in De Loecker (2011). In this environment, consumers have CES preferences across firm-specific varieties within 2-digit industries. This yields industry-specific demand elasticities for each variety that are fixed over time. In particular, the demand faced by firm i can be written as

$$q_{i,b,k,t} = X_{k,t} p_{i,b,k,t}^{\eta_k} e^{\zeta_{i,b,k,t}}.$$

In this equation, one way of interpreting the industry-time effect $X_{k,t}$ is as capturing the following combination of industry-time specific demand shocks and an average price across varieties in industry k at time t:

$$X_{k,t} = \frac{Q_{k,t}}{P_{k,t}^{\eta_k}}$$

Alternatively, we can think of $X_{k,t}$ as representing a more reduced form demand shifter that is common to all varieties in industry k at time t. Either way, η_k is the demand elasticity faced by each firm in industry k for its product and $\zeta_{i,b,k,t}$ is an i.i.d demand shock that is uncorrelated with TFP shocks.

Profit maximization yields the following expression for the firm-year-industry specific price:

$$\text{(A1)} \quad \begin{aligned} \ln p_{i,b,k,t} = & -\frac{1}{D_k} \ln A_{i,b,k,t} + \frac{\theta_k}{D_k} \ln w_{B(b),k,t} - \frac{\theta_k}{D_k} \ln[\frac{1+\eta_k}{\eta_k}\theta_k] \\ & + \frac{1-\theta_k}{D_k}[\ln X_{k,t} + \zeta_{i,b,k,t}]. \end{aligned}$$

The denominator $D_k = -\eta_k(1-\theta_k) + \theta_k > 0$. As η_k approaches negative infinity, $\ln p_{i,b,k,t}$ goes to a constant by construction and firms have no market power. Otherwise, positive productivity shocks depress output prices. Associated negative shocks to marginal costs lead firms to increase output, moving further down marginal revenue and demand functions. That is, the more market power firms have, the greater the pass-through of positive productivity shocks to price discounts. Similarly, positive wage shocks and positive demand shocks get passed through to increased variety prices in this environment.

By definition, $\ln R_{i,b,k,t} = \ln p_{i,b,k,t} + \ln q_{i,b,k,t} = (1 + \eta_k) \ln p_{i,b,k,t} + \ln X_{k,t} + \zeta_{i,b,k,t}$. Insertion of equation (A1) into this condition delivers the following general expression for revenue, which matches equation (5) in the main text. This expression also holds under perfect competition, when $\eta_k = -\infty$.

$$\text{(A2)} \quad \begin{aligned} \ln R_{i,b,k,t} =& \frac{1+\eta_k}{\eta_k(1-\theta_k)-\theta_k} \ln A_{i,b,k,t} - \frac{\theta_k(1+\eta_k)}{\eta_k(1-\theta_k)-\theta_k} \ln w_{B(b),k,t} \\ & - \frac{\theta_k(1+\eta_k)}{D_k} \ln[\frac{1+\eta_k}{\eta_k}\theta_k] + \frac{1}{D_k}[\ln X_{k,t} + \zeta_{i,b,k,t}] \end{aligned}$$

If the firm is a price taker, this expression matches equation (2) with no change in price by l'Hôpital's Rule. As demand for the firm's product becomes less elastic, a given change in revenue must be driven by a larger TFP shock because the firm is more constrained in its optimal increase in quantity. For example, with $\theta_k = 0.7$ and $\eta_k = -2$, a 10 percent positive observed revenue change would reflect a 13 percent increase in TFP. However, with $\eta_k = -10$ instead, the associated TFP increase needed to achieve the same change in revenue is only 4 percent. Under perfect competition, this required TFP increase is further reduced to 3.3 percent.

A2. Derivation of an Estimating Equation

As seen in equation (A2), the pass-through of TFP shocks into revenue depends both on the strength of industry-specific market power and the importance of endogenous variable factor adjustments in response to TFP shocks. Within heterogeneous peer groups, there are thus variable revenue responses to the same TFP shock, making peer effects as described by a revenue based estimation equation heterogeneous within peer groups. This heterogeneous response mixes the TFP spillover parameter γ^A with market power and variable factor share parameters η_k and θ_k. In equation (4), the structural interpretation of the firm fixed effect is determined jointly by the firm-specific fixed effect term and the spillover term.

To see this mathematically, begin with equation (A2) and set the firm fixed effect α_i^R to equal

$-\frac{1+\eta_{k(i)}}{D_{k(i)}}\alpha_i^A$. Remaining firm-specific terms in equation (4) then have the structural interpretation

$$\gamma^R \sum_{j\in M_{b,t},\neq i}[\omega_{ij}(M_{b,t})\alpha_j^R] + \varepsilon_{i,b,k,t}^R = \frac{(1+\eta_{k(i)})}{D_{k(i)}}\gamma^A \sum_{j\in M_{b,t},\neq i}[\omega_{ij}(M_{b,t})\alpha_j^R \frac{D_{k(j)}}{1+\eta_{k(j)}}] - \frac{(1+\eta_{k(i)})}{D_{k(i)}}\varepsilon_{i,b,k,t}^A + \frac{\zeta_{i,b,k,t}}{D_{k(i)}}.$$

From this equation, it is clear that if firm i is in the same industry as all its peers, revenue spillovers γ^R directly measure TFP spillovers γ^A. However, if they are in different industries, the estimated spillover in the revenue equation γ^R mixes information about peer group composition and variable markups.

Our approach for recovering structural TFP spillovers is to adjust the dependent variable to homogenize treatment effects in estimation equations with the same form as equation (4). In particular, dividing both sides of equation (A2) by $-\frac{1+\eta_k}{D_k}$ yields the adjusted revenue measure

$$\text{(A3)} \qquad \ln \tilde{R}_{i,b,k,t} \equiv -\frac{D_k}{1+\eta_k}\ln R_{i,b,k,t}$$

for use as an outcome. Substituting equation (3) for $\ln A_{i,b,k,t}$, we have the following alternative structural equation for adjusted revenue, in which the spillover parameter equals the TFP spillover parameter γ^A:

$$\text{(A4)} \qquad \ln \tilde{R}_{i,b,k,t} = \alpha_i^A + \tilde{\phi}_{B(b),k,t} + \gamma^A \left[\sum_{j\in M_{b,t},\neq i}\omega_{ij}(M_{b,t})\alpha_j^A\right] + \tilde{\varepsilon}_{i,b,k,t}.$$

Because using adjusted revenue $\ln \tilde{R}_{i,b,k,t}$ as the dependent variable isolates firm fixed effects as the permanent firm-specific component of TFP α_i^A, the TFP spillover parameter γ^A can be directly estimated as the peer effect parameter.

The new structural interpretation of the fixed effects in equation (A4) is

$$\tilde{\phi}_{B(b),k,t} = \phi_{B(b),k,t}^A - \theta_k \ln w_{B(b),t} - \theta_k \ln \frac{\eta_k}{1+\eta_k} + \theta_k \ln \theta_k - \frac{1}{1+\eta_k}\ln X_{k,t}$$

and the error term in equation (A4) is

$$\tilde{\varepsilon}_{i,b,k,t}^R = \varepsilon_{i,b,k,t}^A - \frac{\zeta_{i,b,k,t}}{1+\eta_k}.$$

As in the perfect competition case, the fixed effects control for location fundamentals, input costs, and industry-time specific demand conditions.

A3. Measuring Factor Shares, Markups, and TFP

Our robustness analysis that explicitly accounts for firm-specific price endogeneity requires measures of variable factor shares θ_k and demand elasticities η_k for implementation, as described in equation (A3). We calculate these objects using revenue and payments to variable and fixed inputs as observed in the data.

Using the firm level cost minimization condition, De Loecker and Eeckhout (2018) show that the firm level markup can be calculated as $\theta_k \frac{R_{i,b,k,t}}{(wL)_{i,b,k,t}}$. This relationship can be verified as being identical for all firms in industry k in the context of the more restrictive model laid out above. In particular, we have an industry level markup which is equal to $\frac{\eta_k}{1+\eta_k}$ by profit maximization.

In the data, we observe firm level revenue $R_{i,b,k,t}$ and annual payments to labor and materials. We infer payments to capital as rental and repair costs plus the book value of capital (net of amortization) times a discount rate plus industry-specific depreciation rate. We set the discount rate to be the Bank of Canada prime rate plus 0.04 minus the inflation rate. We infer payments to real estate as building maintenance costs plus property taxes plus rent plus the value of buildings and land (net of amortization) times a mortgage rate plus depreciation rate minus a capital gains rate. The mortgage rate is the prime rate plus 0.02. The depreciation rate is non-zero for structures only and is reported by Statistics Canada for each 2-digit industry. The capital gains rate uses the CMA level Teranet residential home price index.

Using this information, we calculate the output elasticity with respect to variable factors $\theta_{k,t}$ and the markup $\frac{\eta_{k,t}}{1+\eta_{k,t}}$ at the 2-digit industry-year level. We calculate the output elasticity with respect to factor f, $\theta^f_{k,t}$, by aggregating payments to factors across all firms in each 2-digit industry-year bin, where the variable factor share $\theta_{k,t}$ is calculated as $\theta^{materials}_{k,t} + \theta^{labor}_{k,t}$. With $\theta_{k,t}$ in hand, we calculate the industry-year specific markup as

$$\frac{\eta_{k,t}}{1+\eta_{k,t}} = \theta_{k,t} \frac{\sum_i R_{i,k,t}}{\sum_i (wL)_{i,k,t}}.$$

Using this equation, we solve out for demand elasticities $\eta_{k,t}$ and average across years to recover calibrations of η_k. Our calibrations of θ_k are also averages of $\theta_{k,t}$ across years in our data.[21]

[21] We also experimented with using firm-specific markups but found them to be too noisy to be of use in estimation.

Appendix B. Estimation Details

This appendix derives the updating rules used for α_i in estimation.

B1. Case With One Peer Effect Term

We have the following generalized estimation equation which follows from equation (7):

$$y_{i,k,b,t} = \alpha_i + \bar{\alpha} + \gamma\bar{\alpha}W_{b,t}^{-i} + \phi_{B(b),k(i),t} + \gamma \sum_{j \in M_{b,t}\setminus\{i\}} \omega_{ij}(M_{b,t})\alpha_j + \varepsilon_{i,k,b,t},$$

where $W_{b,t}^{-i} = \sum_{j \in M_{b,t}\setminus\{i\}} \omega_{ij}(M_{b,t})$. If $W_{b,t}^{-i}$ is a constant (as in the linear-in-means specification), we get initial estimates of $\alpha_i, \gamma\bar{\alpha}W_{b,t}^{-i} + \bar{\alpha} + \phi_{B(b),k(i),t}$, and γ. If $W_{b,t}^{-i}$ is not a constant, we can separately identify $\sigma = \gamma\bar{\alpha}$ and $\bar{\alpha} + \phi_{B(b),k(i),t}$. α_i is then updated using the updating rule below, derived by minimizing the associated nonlinear least square objective function.

The nonlinear least square estimator minimizes the following objective function:

$$\sum_{i \in I}\sum_{t \in T_i}\Big(y_{i,k,b,t} - \alpha_i - \bar{\alpha} - \gamma\bar{\alpha}\sum_{j \in M_{b,t}\setminus\{i\}} \omega_{ij}(M_{b,t}) - \phi_{B(b),k(i),t} - \gamma \sum_{j \in M_{b,t}\setminus\{i\}} \omega_{ij}(M_{b,t})\alpha_j\Big)^2$$

For the linear-in-means specification, $\omega_{ij}(M_{b,t}) = \frac{1}{|M_{b,t}|-1}$ and for the agglomeration specification, $\omega_{ij}(M_{b,t}) = 1$.

The first-order condition with respect to α_i is:

$$\begin{aligned}
0 = &-2\sum_{t \in T_i}\Big(y_{i,k,b,t} - \alpha_i - \bar{\alpha} - \gamma\bar{\alpha}W_{b,t}^{-i} - \phi_{B(b),k(i),t} - \gamma \sum_{j \in M_{b,t}\setminus\{i\}} \omega_{ij}(M_{b,t})\alpha_j\Big) \\
&- 2\sum_{t \in T_i}\sum_{j \in M_{b,t}\setminus\{i\}}\Big(y_{j,k,b,t} - \alpha_j - \bar{\alpha} - \gamma\bar{\alpha}W_{b,t}^{-j} - \phi_{B(b),k(j),t} - \gamma \sum_{j' \in M_{b,t}\setminus\{j\}} \omega_{jj'}(M_{b,t})\alpha_{j'}\Big)\gamma\omega_{ji}(M_{b,t}).
\end{aligned}$$

Solving for α_i (Step 1/3):

$$\begin{aligned}
T_i\alpha_i = &\sum_{t \in T_i}\Big(y_{i,k,b,t} - \bar{\alpha} - \gamma\bar{\alpha}W_{b,t}^{-i} - \phi_{B(b),k(i),t} - \gamma \sum_{j \in M_{b,t}\setminus\{i\}} \omega_{ij}(M_{b,t})\alpha_j\Big) \\
&+ \sum_{t \in T_i}\sum_{j \in M_{b,t}\setminus\{i\}}\Big(y_{j,k,b,t} - \alpha_j - \bar{\alpha} - \gamma\bar{\alpha}W_{b,t}^{-j} - \phi_{B(b),k(j),t} - \gamma \sum_{j' \in M_{b,t}\setminus\{i,j\}} \omega_{jj'}(M_{b,t})\alpha_{j'}\Big)\gamma\omega_{ji}(M_{b,t}) \\
&- \sum_{t \in T_i}\sum_{j \in M_{b,t}\setminus\{i\}}\gamma^2\omega_{ji}(M_{b,t})^2\alpha_i
\end{aligned}$$

Solving for α_i (Step 2/3):

$$T_i\alpha_i + \sum_{t\in T_i}\sum_{j\in M_{b,t}\setminus\{i\}} \gamma^2\omega_{ji}(M_{b,t})^2\alpha_i =$$

$$\sum_{t\in T_i}\left(y_{i,k,b,t} - \bar{\alpha} - \gamma\bar{\alpha}W_{b,t}^{-i} - \phi_{B(b),k(i),t} - \gamma\sum_{j\in M_{b,t}\setminus\{i\}}\omega_{ij}(M_{b,t})\alpha_j\right)$$

$$+\sum_{t\in T_i}\sum_{j\in M_{b,t}\setminus\{i\}}\left(y_{j,k,b,t} - \alpha_j - \bar{\alpha} - \gamma\bar{\alpha}W_{b,t}^{-j} - \phi_{B(b),k(j),t} - \gamma\sum_{j'\in M_{b,t}\setminus\{i,j\}}\omega_{jj'}(M_{b,t})\alpha_{j'}\right)\gamma\omega_{ji}(M_{b,t})$$

Solving for α_i (Step 3/3):

$$\alpha_i =$$

$$\frac{1}{\left(T_i + \gamma^2\sum_{t\in T_i}\sum_{j\in M_{b,t}\setminus\{i\}}\omega_{ji}(M_{b,t})^2\right)}\times$$

$$\sum_{t\in T_i}\Bigg[\left(y_{i,k,b,t} - \bar{\alpha} - \gamma\bar{\alpha}W_{b,t}^{-i} - \phi_{B(b),k(i),t} - \gamma\sum_{j\in M_{b,t}\setminus\{i\}}\omega_{ij}(M_{b,t})\alpha_j\right)$$

$$+\gamma\sum_{j\in M_{b,t}\setminus\{i\}}\left(y_{j,k,b,t} - \alpha_j - \bar{\alpha} - \gamma\bar{\alpha}W_{b,t}^{-j} - \phi_{B(b),k(j),t} - \gamma\sum_{j'\in M_{b,t}\setminus\{i,j\}}\omega_{jj'}(M_{b,t})\alpha_{j'}\right)\omega_{ji}(M_{b,t})\Bigg]$$

In the linear-in-means specification with basic weights $\omega_{ij}(M_{b,t}) = \frac{1}{|M_{b,t}|-1}$, this expression is:

$$\alpha_i =$$

$$\frac{1}{\left(T_i + \gamma^2\sum_{t\in T_i}\frac{1}{|M_{b,t}|-1}\right)}\times$$

$$\sum_{t\in T_i}\Bigg[\left(y_{i,k,b,t} - \bar{\alpha}(1-\gamma) - \phi_{B(b),k(i),t} - \frac{\gamma}{|M_{b,t}|-1}\sum_{j\in M_{b,t}\setminus\{i\}}\alpha_j\right)$$

$$+\frac{\gamma}{|M_{b,t}|-1}\sum_{j\in M_{b,t}\setminus\{i\}}\left(y_{j,k,b,t} - \alpha_j - \bar{\alpha}(1-\gamma) - \phi_{B(b),k(j),t} - \frac{\gamma}{|M_{b,t}|-1}\sum_{j'\in M_{b,t}\setminus\{i,j\}}\alpha_{j'}\right)\Bigg]$$

In the agglomeration model with basic weights $\omega_{ji}(M_{b,t}) = 1$, this expression is:

$$\alpha_i = \frac{1}{\left(T_i + \gamma^2 \sum_{t\in T_i}(|M_{b,t}|-1)\right)} \times \sum_{t\in T_i}\left[\left(y_{i,k,b,t} - \bar{\alpha}(1-\gamma) - \gamma\bar{\alpha}M_{b,t} - \phi_{B(b),k(i),t} - \gamma \sum_{j\in M_{b,t}\setminus\{i\}} \alpha_j\right) + \gamma \sum_{j\in M_{b,t}\setminus\{i\}} \left(y_{j,k,b,t} - \alpha_j - \bar{\alpha}(1-\gamma) - \gamma\bar{\alpha}M_{b,t} - \phi_{B(b),k(j),t} - \gamma \sum_{j'\in M_{b,t}\setminus\{i,j\}} \alpha_{j'}\right)\right]$$

B2. Horse race

We carry out the analogous process for the horse race. For estimation, we replace $\gamma_{\text{Agg}} \sum_{j\in M_{b,t},\neq i} \alpha_j + \ddot{\sigma}(|M_{b,t}|-1)$ in the baseline horse race estimation equation (8) with its generalized counterpart $\gamma_{\text{W}} \sum_{j\in M_{b,t},\neq i} \alpha_j\omega_{ij}(M_{b,t})$ $+\ddot{\sigma}_{\text{W}} \sum_{j\in M_{b,t},\neq i} \omega_{ij}(M_{b,t})$. Define two weights, one for each element of the horse race

$$W^{-i}_{q,b,t} = \sum_{j\in M_{b,t}\setminus\{i\}} \omega_s(k(i),k(j),M_{b,t})$$

where $q \in \{m,s\}$. The nonlinear least square estimator minimizes the following objective function:

$$\sum_{i\in I}\sum_{t\in T_i}\left(y_{i,k,b,t} - \alpha_i - \bar{\alpha} - \gamma_s\bar{\alpha}W^{-i}_{s,b,t} - \gamma_m\bar{\alpha}W^{-i}_{m,b,t} - \phi_{k(i),B(b),t} - \gamma_s \sum_{j\in M_{b,t}\setminus\{i\}} \omega_s(k(i),k(j),M_{b,t})\alpha_j - \gamma_m \sum_{j\in M_{b,t}\setminus\{i\}} \omega_m(k(i),k(j),M_{b,t})\alpha_j\right)^2$$

The first-order condition with respect to α_i:

$$0 = -2\sum_{t\in T_i}\left(y_{i,k,b,t} - \alpha_i - \bar{\alpha} - \gamma_s\bar{\alpha}W^{-i}_{s,b,t} - \gamma_m\bar{\alpha}W^{-i}_{m,b,t} - \phi_{k(i),B(b),t} - \gamma_s \sum_{j\in M_{b,t}\setminus\{i\}} \omega_s\left(k\left(i\right),k\left(j\right),M_{b,t}\right)\alpha_j - \gamma_m \sum_{j\in M_{b,t}\setminus\{i\}} \omega_m\left(k\left(i\right),k\left(j\right),M_{b,t}\right)\alpha_j\right)$$

$$= -2\sum_{t\in T_i}\sum_{j\in M_{b,t}\setminus\{i\}}\Bigg[\Bigg(y_{j,k,b,t}-\alpha_j-\bar{\alpha}-\gamma_s\bar{\alpha}W^{-j}_{s,b,t}-\gamma_m\bar{\alpha}W^{-j}_{m,b,t}-\phi_{k(j),B(b),t}$$

$$-\gamma_s\sum_{j'\in M_{b,t}\setminus\{j\}}\omega_s(k(j),k(j'),M_{b,t})\alpha_{j'}-\gamma_m\sum_{j'\in M_{b,t}\setminus\{j\}}\omega_m(k(j),k(j'),M_{b,t}),M_{b,t})\alpha_{j'}\Bigg)$$

$$\times\Big(\gamma_s\omega_s(k(j),k(i),M_{b,t})+\gamma_m\omega_m(k(j),k(i),M_{b,t})\Big)\Bigg]$$

Solving for α_i (Step 1/2):

$$T_i\alpha_i+\alpha_i\sum_{t\in T_i}\sum_{j\in M_{b,t}\setminus\{i\}}\Big(\gamma_s\omega_s(k(j),k(i),M_{b,t})+\gamma_m\omega_m(k(j),k(i),M_{b,t})\Big)^2$$

$$=\sum_{t\in T_i}\Bigg(y_{i,k,b,t}-\bar{\alpha}-\gamma_s\bar{\alpha}W^{-i}_{s,b,t}-\gamma_m\bar{\alpha}W^{-i}_{m,b,t}-\phi_{k(i),B(b)}-\gamma_s\sum_{j\in M_{b,t}\setminus\{i\}}\omega_s(k(i),k(j),M_{b,t})\alpha_j$$

$$-\gamma_m\sum_{j\in M_{b,t}\setminus\{i\}}\omega_m(k(i),k(j),M_{b,t})\alpha_j\Bigg)$$

$$+\sum_{t\in T_i}\sum_{j\in M_{b,t}\setminus\{i\}}\Bigg[\Bigg(y_{j,k,b,t}-\alpha_j-\bar{\alpha}-\gamma_s\bar{\alpha}W^{-j}_{s,b,t}-\gamma_m\bar{\alpha}W^{-j}_{m,b,t}-\phi_{k(j),B(b)}$$

$$-\gamma_s\sum_{j'\in M_{b,t}\setminus\{i,j\}}\omega_s(k(j),k(j'),M_{b,t})\alpha_{j'}-\gamma_m\sum_{j'\in M_{b,t}\setminus\{i,j\}}\omega_m(k(j),k(j'),M_{b,t})\alpha_{j'}\Bigg)$$

$$\times\Big(\gamma_s\omega_s(k(j),k(i),M_{b,t})+\gamma_m\omega_m(k(j),k(i),M_{b,t})\Big)\Bigg]$$

Solving for α_i (Step 2/2):

$$\alpha_i =$$

$$\frac{1}{T_i + \sum_{t \in T_i} \sum_{j \in M_{b,t} \setminus \{i\}} \Big(\gamma_s \omega_s(k(j), k(i), M_{b,t}) + \gamma_m \omega_m(k(j), k(i), M_{b,t})\Big)^2} \times$$

$$\sum_{t \in T_i} \Bigg[\Bigg(y_{i,k,b,t} - \bar{\alpha} - \gamma_s \bar{\alpha} W^{-i}_{s,b,t} - \gamma_m \bar{\alpha} W^{-i}_{m,b,t} - \phi_{k(i),B(b)}$$

$$- \gamma_s \sum_{j \in M_{b,t} \setminus \{i\}} \omega_s(k(i), k(j), M_{b,t}) \alpha_j - \gamma_m \sum_{j \in M_{b,t} \setminus \{i\}} \omega_m(k(i), k(j), M_{b,t}) \alpha_j \Bigg)$$

$$+ \sum_{j \in M_{b,t} \setminus \{i\}} \Bigg[\Bigg(y_{j,k,b,t} - \alpha_j - \bar{\alpha} - \gamma_s \bar{\alpha} W^{-j}_{s,b,t} - \gamma_m \bar{\alpha} W^{-j}_{m,b,t} - \phi_{k(j),B(b)}$$

$$- \gamma_s \sum_{j' \in M_{b,t} \setminus \{i,j\}} \omega_s(k(j), k(j'), M_{b,t}) \alpha_{j'} - \gamma_m \sum_{j' \in M_{b,t} \setminus \{i,j\}} \omega_m(k(j), k(j'), M_{b,t}) \alpha_{j'} \Bigg) \times$$

$$\Big(\gamma_s \omega_s(k(j), k(i), M_{b,t}) + \gamma_m \omega_m(k(j), k(i), M_{b,t})\Big) \Bigg] \Bigg]$$

Appendix C. Proofs of Consistency

This appendix analyzes the consistency of an estimator of spillovers between firms based on the minimization of the squared prediction errors. The proofs shown here mimic the proof of Theorem 1 in Arcidiacono et al. (2012) (AFGK), in particular the first four lemmas where they show the consistency of their estimator. Throughout this section, firm i in peer group n at time t is characterized by a fixed effect α_i and a shock $\epsilon_{i,t,n}$. We analyze:

1) The consistency of an estimator where spillovers of firm j to firm i in peer group n are weighted by a known weight $\omega_{i,j,n}$;

2) The consistency of a horse race estimator where both aggregate spillovers ($\omega_{i,j,n} = 1$) and linear-in-mean spillovers ($\omega_{i,j,n} = \frac{1}{|M_{n,t}|-1}$, where $M_{n,t}$ denotes the set of firms in peer group n at time t) operate simultaneously;

3) The lack of consistency and bias of the estimator when the number of groups N goes to infinity, but the peer group has a fixed time dimension, in particular $T = 2$, and the shocks $\epsilon_{i,t,n}$ are autocorrelated;

4) The consistency of the estimator when both N and T go to infinity and the shocks are autocorrelated.

Throughout this section, we maintain most of AFGK's assumptions:

(i) $E\left(\epsilon_{i,t,n}\epsilon_{j,t,k}\right) = 0$ for all $j \neq i$ and $n \neq k$.

(ii) $E\left(\epsilon_{i,t,n}\alpha_j\right) = 0$ for all i, j, t, n.

(iii) $E\left(\alpha_{in}^4\right) < \infty$ for all i, n.

(iv) $E\left(\epsilon_{i,t,n}\right) = 0$ and $E\left(\epsilon_{i,t,n}^4\right) < \infty$ for all i, t, n.

(v) $E\left(\epsilon_{i,t,n}^2 | n, t\right) = E\left(\epsilon_{j,t,n}^2 | n, t\right)$ for all i, j, t, n.

(vi) The parameter $\gamma \in \Gamma$ where Γ is compact.

In the first two cases, as in AFGK, we also assume that $E\left(\epsilon_{i,t,n}\epsilon_{j,s,k}\right) = 0$ for $t \neq s$. This assumption is relaxed in the other two cases. Furthermore, we assume that the fixed effects $\{\alpha_{in}\}$ are not linear combinations of each other in i to guarantee uniqueness of the solutions, which was an implicit assumption in AFGK.

As in AFGK, we analyze consistency using a simplified structure of peer groups with a limited number of firms and periods (expect for Case 4). We do not expect this simplification to affect the general results. In the same spirit, we do not allow firm i's outcome to be affected by other covariates, except for peer effects. That is, we do not include industry-year or local area-year fixed effects as in our main analysis. Again, we do not believe that the general message of this section is affected by this choice of exposition.

Case 1 - General Weights

Outcome $y_{i,t,n}$ of firm i in peer group n at time t is:

$$y_{i,t,n} = \alpha_i + \gamma \sum_{j \in M_{n,t} \neq i} \omega_{i,j,n,t} \alpha_j + \epsilon_{i,t,n}$$

In addition to the six assumptions listed above, we assume that $E\left(\epsilon_{i,t,n}\epsilon_{j,s,k}\right) = 0$ for $t \neq s$. We consider the following limiting case:

(a) Firms are observed for at most two periods.

(b) Each peer group has two firms in each period.

(c) Within each peer group, one firm is observed for two periods and the other firm is observed for one period only.

The optimization problem is

$$\begin{aligned}\min_{\alpha,\gamma} \frac{1}{N} \sum_{n=1}^{N} ((y_{11n} - \alpha_{1n} - &\gamma\omega_{12n}\alpha_{2n})^2 + (y_{12n} - \alpha_{1n} - \gamma\omega_{13n}\alpha_{3n})^2 \\ &+ (y_{21n} - \alpha_{2n} - \gamma\omega_{21n}\alpha_{1n})^2 + (y_{32n} - \alpha_{3n} - \gamma\omega_{31n}\alpha_{1n})^2)\end{aligned}$$

where we omit the time period index in the weight given that two firms meet for only one time period.

Following AFGK, we first concentrate out the αs. Taking the first-order conditions and solving for the firm fixed effects (omitting the index n), we get:

$$\alpha_1 = \frac{\left(\left(1 + (\gamma\omega_{13})^2\right)(1 - \gamma\omega_{21}\gamma\omega_{12})(y_{11} - \gamma\omega_{12}y_{21}) + \left(1 + (\gamma\omega_{12})^2\right)(1 - \gamma\omega_{13}\gamma\omega_{31})(y_{12} - \gamma\omega_{13}y_{32})\right)}{\left((1 - \gamma\omega_{31}\gamma\omega_{13})^2\left(1 + (\gamma\omega_{12})^2\right) + (1 - \gamma\omega_{21}\gamma\omega_{12})^2\left(1 + (\gamma\omega_{13})^2\right)\right)}$$

$$\alpha_2 = \frac{\begin{pmatrix} \left(\left(1-\gamma\omega_{31}\gamma\omega_{13}\right)^2\gamma\omega_{12}-\gamma\omega_{21}\left(1+\left(\gamma\omega_{13}\right)^2\right)\left(1-\gamma\omega_{21}\gamma\omega_{12}\right)\right)y_{11} \\ +\left(\left(1-\gamma\omega_{31}\gamma\omega_{13}\right)^2+\left(1-\gamma\omega_{21}\gamma\omega_{12}\right)\left(1+\left(\gamma\omega_{13}\right)^2\right)\right)y_{21} \\ -\left(\gamma\omega_{21}+\gamma\omega_{12}\right)\left(1-\gamma\omega_{13}\gamma\omega_{31}\right)y_{12}+\left(\gamma\omega_{21}+\gamma\omega_{12}\right)\left(1-\gamma\omega_{13}\gamma\omega_{31}\right)\gamma\omega_{13}y_{32} \end{pmatrix}}{\left(\left(1-\gamma\omega_{31}\gamma\omega_{13}\right)^2\left(1+\left(\gamma\omega_{12}\right)^2\right)+\left(1-\gamma\omega_{21}\gamma\omega_{12}\right)^2\left(1+\left(\gamma\omega_{13}\right)^2\right)\right)}$$

$$\alpha_3 = \frac{\begin{pmatrix} \left(\left(1-\gamma\omega_{21}\gamma\omega_{12}\right)^2\gamma\omega_{13}-\gamma\omega_{31}\left(1+\left(\gamma\omega_{12}\right)^2\right)\left(1-\gamma\omega_{13}\gamma\omega_{31}\right)\right)y_{12} \\ +\left(\left(1-\gamma\omega_{21}\gamma\omega_{12}\right)^2+\left(1-\gamma\omega_{13}\gamma\omega_{31}\right)\left(1+\left(\gamma\omega_{12}\right)^2\right)\right)y_{32} \\ -\left(\gamma\omega_{31}+\gamma\omega_{13}\right)\left(1-\gamma\omega_{21}\gamma\omega_{12}\right)y_{11}+\left(\gamma\omega_{31}+\gamma\omega_{13}\right)\left(1-\gamma\omega_{21}\gamma\omega_{12}\right)\gamma\omega_{12}y_{21} \end{pmatrix}}{\left(\left(1-\gamma\omega_{31}\gamma\omega_{13}\right)^2\left(1+\left(\gamma\omega_{12}\right)^2\right)+\left(1-\gamma\omega_{21}\gamma\omega_{12}\right)^2\left(1+\left(\gamma\omega_{13}\right)^2\right)\right)}$$

Note that the above expressions simplify to the same ones as in AFGK when all weights are equal to 1:

$$\begin{aligned} \alpha_1 &= \frac{\left(y_{11}-\gamma y_{21}\right)+\left(y_{12}-\gamma y_{32}\right)}{2\left(1-\gamma^2\right)} \\ \alpha_2 &= \frac{y_{21}-\gamma^3 y_{11}-\gamma y_{12}+\gamma^2 y_{32}}{\left(1-\gamma^4\right)} \\ \alpha_3 &= \frac{y_{32}-\gamma y_{11}+\gamma^2 y_{21}-\gamma^3 y_{12}}{\left(1-\gamma^4\right)} \end{aligned}$$

After several substitutions, the original minimization problems becomes:

$$\min_{\alpha,\gamma} N \sum_{n=1}^{N} \frac{\left(\left(1-\gamma\omega_{31n}\gamma\omega_{13n}\right)\left(y_{11n}-\gamma\omega_{12n}y_{21n}\right)-\left(1-\gamma\omega_{21n}\gamma\omega_{12n}\right)\left(y_{12n}-\gamma\omega_{13n}y_{32n}\right)\right)^2}{\left(1-\gamma\omega_{31n}\gamma\omega_{13n}\right)^2\left(1+\left(\gamma\omega_{12n}\right)^2\right)+\left(1-\gamma\omega_{21n}\gamma\omega_{12n}\right)^2\left(1+\left(\gamma\omega_{13n}\right)^2\right)}$$

which is the same as in AFGK when the weights are equal to 1, i.e.

$$\min_{\alpha,\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{\left(y_{11n}-\gamma y_{21n}-y_{12n}+\gamma y_{32n}\right)^2}{2\left(1+\gamma^2\right)}.$$

Substituting for the true data generating process:

$$y_{i,t,n} = \alpha_i^o + \gamma_0 \sum_{j\in M_{n,t}\neq i} \omega_{i,j,n,t}\alpha_j^o + \epsilon_{i,t,n}^o$$

we obtain

$$\min_{\alpha,\gamma} \frac{1}{N}\sum_{n=1}^{N} q\left(y_n,\gamma\right)$$

where (omitting the index n):

$$q\left(y_n,\gamma\right) = \frac{\begin{pmatrix} (1-\gamma\omega_{31}\gamma\omega_{13})\left(\alpha_1^o+\gamma_0\omega_{12}a_2^o+\epsilon_{11}^\circ-\gamma\omega_{12}\left(\alpha_2^o+\gamma_0\omega_{21}a_1^o+\epsilon_{21}^o\right)\right) \\ -(1-\gamma\omega_{21}\gamma\omega_{12})\left(a_1^o+\gamma_0\omega_{13}a_3^o+\epsilon_{12}^\circ-\gamma\omega_{13}\left(\alpha_3^o+\gamma_0\omega_{31}a_1^o+\epsilon_{32}^o\right)\right)\end{pmatrix}^2}{(1-\gamma\omega_{31}\gamma\omega_{13})^2\left(1+(\gamma\omega_{12})^2\right)+(1-\gamma\omega_{21}\gamma\omega_{12})^2\left(1+(\gamma\omega_{13})^2\right)}.$$

Consider the expected value of the function $q(y_n,\gamma)$. Using assumptions (i), (ii), and (iv), the expression simplifies to:

$$\begin{aligned}&\mathbb{E}(q(y,\gamma)) = \sigma_\epsilon^2 \\ &+\mathbb{E}\left(\frac{\begin{pmatrix}((1-\gamma\omega_{31}\gamma\omega_{13})(1-\gamma\omega_{12}\gamma_0\omega_{21})-(1-\gamma\omega_{21}\gamma\omega_{12})(1-\gamma\omega_{13}\gamma_0\omega_{31}))\,\alpha_1^o \\ +(1-\gamma\omega_{31}\gamma\omega_{13})(\gamma_0-\gamma)\,\omega_{12}a_2^o-(1-\gamma\omega_{21}\gamma\omega_{12})(\gamma_0-\gamma)\,\omega_{13}a_3^o\end{pmatrix}^2}{(1-\gamma\omega_{31}\gamma\omega_{13})^2\left(1+(\gamma\omega_{12})^2\right)+(1-\gamma\omega_{21}\gamma\omega_{12})^2\left(1+(\gamma\omega_{13})^2\right)}\right).\end{aligned}$$

The first term does not depend on γ. Because the denominator of the second term is positive and the numerator is squared, it is equal to zero when $\gamma=\gamma_0$ while strictly positive when $\gamma\neq\gamma_0$. Hence, $\mathbb{E}[q(y,\gamma_0)]<\mathbb{E}[q(y,\gamma)]$ for all $\gamma\in\Gamma$ such that $\gamma\neq\gamma_0$.

As suggested in AFGK, most of the requirements to apply Theorem 12.1 in Wooldridge (2010) are satisfied with the exception of the following: For all $\gamma\in\Gamma$, $|q(\gamma,y)|\leq b(y)$ where b is a non-negative function such that $\mathbb{E}(b(y))<\infty$ Given that $q(\gamma,y)$ is always positive we can ignore the absolute value. Going back to the definition of $q(\gamma,y)$:

$$q(\gamma,y)=\frac{\left((1-\gamma\omega_{31}\gamma\omega_{13})(y_{11}-\gamma\omega_{12}y_{21})-(1-\gamma\omega_{21}\gamma\omega_{12})(y_{12}-\gamma\omega_{13}y_{32})\right)^2}{(1-\gamma\omega_{31}\gamma\omega_{13})^2\left(1+(\gamma\omega_{12})^2\right)+(1-\gamma\omega_{21}\gamma\omega_{12})^2\left(1+(\gamma\omega_{13})^2\right)}$$

and using the triangular inequality, we can show that $q(\gamma,y)<2\left(y_{11}^2+2y_{21}^2+2y_{12}^2+2y_{32}^2\right)$. The last step is to show that $\mathbb{E}\left(2y_{11n}^2+2y_{21n}^2+2y_{12n}^2+2y_{32n}^2\right)<\infty$. This follows exactly the proof in AFGK and we omit it here. Theorem 12.1 in Wooldridge (2010) can be applied to obtain $\hat{\gamma}\xrightarrow{p}\gamma_0$.

Case 2 - Horse Race

We now consider the consistency of a horse race estimator where both aggregate spillovers ($\omega_{i,j,n}^a=1$) and linear-in-mean spillovers ($\omega_{i,j,n}^b=\frac{1}{|M_{n,t}|-1}$, where $M_{n,t}$ denotes the set of firms in peer group n at time t) operate simultaneously. In this case, outcome $y_{i,t,n}$ of firm i in peer group n at time t

is:

$$y_{i,t,n} = \alpha_i + \gamma \sum_{j \in M_{n,t} \neq i} \alpha_{jn} + \rho \sum_{j \in M_{n,t} \neq i} \frac{\alpha_{jn}}{|M_{n,t}| - 1} + \epsilon_{i,t,n}.$$

We consider a situation where each peer group has the following composition:

(a) Firm $0n$ is observed for 3 periods.

(b) Firm $1n$ is observed only in the first period.

(c) Firm $2n$ is observed only in the second period.

(d) Firm $3n$ and $4n$ are observed only in the third period.

The optimization problem is

$$\min_{\alpha,\gamma,\rho} \frac{1}{N} \sum_{i=1}^{N} \left(\epsilon_{01n}^2 + \epsilon_{02n}^2 + \epsilon_{03n}^2 + \epsilon_{11n}^2 + \epsilon_{22n}^2 + \epsilon_{33n}^2 + \epsilon_{43n}^2 \right)$$

where

$$\begin{aligned}
\epsilon_{01n} &= y_{01n} - \alpha_{0n} - (\gamma + \rho)\alpha_{1n} \\
\epsilon_{02n} &= y_{02n} - \alpha_{0n} - (\gamma + \rho)\alpha_{2n} \\
\epsilon_{33n} &= y_{03n} - \alpha_{0n} - \left(\gamma + \frac{\rho}{2}\right)(\alpha_{3n} + \alpha_{4n}) \\
\epsilon_{11n} &= y_{11n} - \alpha_{1n} - (\gamma + \rho)\alpha_{0n} \\
\epsilon_{22n} &= y_{22n} - \alpha_{2n} - (\gamma + \rho)\alpha_{0n} \\
\epsilon_{33n} &= y_{33n} - \alpha_{3n} - \left(\gamma + \frac{\rho}{2}\right)(\alpha_{0n} + \alpha_{4n}) \\
\epsilon_{43n} &= y_{43n} - \alpha_{4n} - \left(\gamma + \frac{\rho}{2}\right)(\alpha_{0n} + \alpha_{3n}).
\end{aligned}$$

We first concentrate out α_1, α_2, α_3, and α_4 for all groups. From the first-order conditions, after several substitutions, we are able to rewrite the argument of the summation above as:

$$\begin{aligned}
&\epsilon_{01n}^2 + \epsilon_{02n}^2 + \epsilon_{03n}^2 + \epsilon_{11n}^2 + \epsilon_{12n}^2 + \epsilon_{33n}^2 + \epsilon_{43n}^2 \\
&= \frac{\left(y_{01n} - (\gamma+\rho)y_{11n} - \left(1 - (\gamma+\rho)^2\right)\alpha_{0n}\right)^2 + \left(y_{02n} - (\gamma+\rho)y_{22n} - \left(1 - (\gamma+\rho)^2\right)\alpha_{0n}\right)^2}{1 + (\gamma+\rho)^2} \\
&+ \frac{\left(\left(\gamma + \frac{\rho}{2}\right)y_{33n} + \left(\gamma + \frac{\rho}{2}\right)y_{43n} - \left(1 + \left(\gamma + \frac{\rho}{2}\right)\right)y_{03n} + \left(1 - \left(\gamma + \frac{\rho}{2}\right)\right)\left(1 + 2\left(\gamma + \frac{\rho}{2}\right)\right)\alpha_{0n}\right)^2}{\left(3\left(\gamma + \frac{\rho}{2}\right)^2 + 2\left(\gamma + \frac{\rho}{2}\right) + 1\right)}
\end{aligned}$$

Instead of concentrating out $\alpha_{0\mathrm{n}}$, we work directly with the minimization of the expected criterion function. To do so, we substitute for the true data generating process

$$
\begin{aligned}
y_{01n} &= \alpha^o_{0n} + (\gamma_0 + \rho_0)\,\alpha^o_{1n} + \epsilon^o_{01n} \\
y_{02n} &= \alpha^o_{0n} + (\gamma_0 + \rho_0)\,\alpha^o_{2n} + \epsilon^o_{02n} \\
y_{03n} &= \alpha^o_{0n} + \left(\gamma_0 + \frac{\rho_0}{2}\right)(\alpha^o_{3n} + \alpha^o_{4n}) + \epsilon^o_{03n} \\
y_{11n} &= \alpha^o_{1n} + (\gamma_0 + \rho_0)\,\alpha^o_{0n} + \epsilon^o_{11n} \\
y_{22n} &= \alpha^o_{2n} + (\gamma_0 + \rho_0)\,\alpha^o_{0n} + \epsilon^o_{22n} \\
y_{33n} &= \alpha^o_{3n} + \left(\gamma_0 + \frac{\rho_0}{2}\right)(\alpha^o_{0n} + \alpha^o_{4n}) + \epsilon^o_{33n} \\
y_{43n} &= \alpha^o_{4n} + \left(\gamma_0 + \frac{\rho_0}{2}\right)(\alpha^o_{0n} + \alpha^o_{3n}) + \epsilon^o_{43n}
\end{aligned}
$$

The expected value of that object can then be written as

$$
\begin{aligned}
&\mathbb{E}\left(\epsilon^2_{01n} + \epsilon^2_{02n} + \epsilon^2_{03n} + \epsilon^2_{11n} + \epsilon^2_{22n} + \epsilon^2_{33n} + \epsilon^2_{43n}\right) \\
&= \mathbb{E}\frac{\left(((\gamma_0 + \rho_0) - (\gamma + \rho))\,\alpha^o_{1n} + (1 - (\gamma + \rho)(\gamma_0 + \rho_0))\,\alpha^o_{0n} - \left(1 - (\gamma + \rho)^2\right)\alpha_{0n}\right)^2}{1 + (\gamma + \rho)^2} \\
&+ \mathbb{E}\frac{\left(((\gamma_0 + \rho_0) - (\gamma + \rho))\,\alpha^o_{2n} + (1 - (\gamma + \rho)(\gamma_0 + \rho_0))\,\alpha^o_{0n} - \left(1 - (\gamma + \rho)^2\right)\alpha_{0n}\right)^2}{1 + (\gamma + \rho)^2} \\
&+ \mathbb{E}\frac{\begin{pmatrix} \left(\left(\gamma + \frac{\rho}{2}\right)\left(1 + \left(\gamma_0 + \frac{\rho_0}{2}\right)\right) - \left(\gamma_0 + \frac{\rho_0}{2}\right)\left(1 + \left(\gamma + \frac{\rho}{2}\right)\right)\right)(\alpha^o_{3n} + \alpha^o_{4n}) \\ + \left(1 + \left(\gamma + \frac{\rho}{2}\right)\right)(\alpha_{0n} - \alpha^o_{0n}) + 2\left(\gamma + \frac{\rho}{2}\right)\left(\left(\gamma_0 + \frac{\rho_0}{2}\right)\alpha^o_{0n} - \left(\gamma + \frac{\rho}{2}\right)\alpha_{0n}\right) \end{pmatrix}^2}{2\left(\gamma + \frac{\rho}{2}\right)^2 + \left(1 + \left(\gamma + \frac{\rho}{2}\right)\right)^2} \\
&+ 3\sigma^2_{\epsilon n}
\end{aligned}
$$

using assumptions (i), (ii), and (v) and assuming $E\left(\epsilon_{i,t,n}\epsilon_{j,s,k}\right) = 0$ for $t \neq s$. Given that the first three terms on the right hand side are non-negative, the minimum is attained at $3\sigma^2_{en}$ when $\alpha_{0n} = \alpha^o_{0n}$, $\gamma = \gamma_0$, and $\rho = \rho_o$. To complete the proof of consistency, as in the previous case, one has to show that the original object can be bounded for all possible γs. We omit this step which can be achieved in a similar fashion to the previous case.

Case 3 - Autocorrelated Errors

As in the first case, we assume that peer groups have a simplified composition:

(a) We observe firms for at most two periods.

(b) Each peer group has two firms in each period.

(c) Within each peer group, one firm is observed for two periods and the other firm is observed for one period only.

Outcome $y_{i,t,n}$ of firm i in peer group n at time t is:

$$y_{i,t,n} = \alpha_i + \gamma \sum_{j \in M_{n,t} \neq i} \alpha_j + \epsilon_{itn}$$

where we assume $\epsilon_{i,t,n} = \rho \epsilon_{i,t-1,n} + u_{i,t,n}$. The optimization problem is

$$\min_{\alpha,\gamma} \frac{1}{N} \sum_{n=1}^{N} ((y_{11n} - \alpha_{1n} - \gamma\alpha_{2n})^2 + (y_{12n} - \alpha_{1n} - \gamma\alpha_{3n})^2 + (y_{21n} - \alpha_{2n} - \gamma\alpha_{1n})^2 + (y_{32n} - \alpha_{3n} - \gamma\alpha_{1n})^2).$$

As in the first case we concentrate out the αs. From the first-order conditions, we find that:

$$\alpha_{1n} = \frac{(y_{11n} - \gamma y_{21n}) + (y_{12n} - \gamma y_{32n})}{2(1-\gamma^2)}$$
$$\alpha_{2n} = \frac{y_{21n} - \gamma^3 y_{11n} - \gamma y_{12n} + \gamma^2 y_{32n}}{(1-\gamma^4)}$$
$$\alpha_{3n} = \frac{y_{32n} - \gamma y_{11n} + \gamma^2 y_{21n} - \gamma^3 y_{12n}}{(1-\gamma^4)}$$

Substituting, we get:

$$\min_{\alpha,\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{(y_{11n} - \gamma y_{21n} - y_{12n} + \gamma y_{32n})^2}{2(1+\gamma^2)}$$

Substituting y with the true data generating process yields

$$\min_{\alpha,\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{((\gamma_0 - \gamma)\alpha^o_{2n} - (\gamma_0 - \gamma)a^o_{3n} + \epsilon^o_{11} - \epsilon^o_{12n} + \gamma\epsilon^o_{32n} - \gamma\epsilon^o_{21n})^2}{2(1+\gamma^2)}.$$

Consider the expected value of the argument within the summation

$$\begin{aligned}
&\mathbb{E}\left(\frac{\left(\left(\gamma_0-\gamma\right)\alpha_2^o-\left(\gamma_0-\gamma\right)\alpha_3^o+\epsilon_{11}^o-\epsilon_{12}^o+\gamma\epsilon_{32}^o-\gamma\epsilon_{21}^o\right)^2}{2\left(1+\gamma^2\right)}\right)\\
&=\mathbb{E}\left(\frac{\left(\gamma_0-\gamma\right)^2\left(\alpha_{20n}-\alpha_{30n}\right)^2}{2\left(1+\gamma^2\right)}\right)\\
&+\mathbb{E}\left(\frac{\left(\epsilon_{11n}-\epsilon_{12n}+\gamma\epsilon_{32n}-\gamma\epsilon_{21n}\right)^2}{2\left(1+\gamma^2\right)}\right)\\
&+\mathbb{E}\left(\frac{2\left(\gamma_0-\gamma\right)\left(\alpha_{20n}-\alpha_{30n}\right)\left(\epsilon_{11n}-\epsilon_{12n}+\gamma\epsilon_{32n}-\gamma\epsilon_{21n}\right)}{2\left(1+\gamma^2\right)}\right).
\end{aligned}$$

All three terms on the right hand side are non-negative. The third term is equal to zero because of assumption (ii). Assuming that $\sigma_{\alpha\alpha}$ is the covariance between α_{20n} and α_{30n} and that the variance of α is the same across different αs, the whole expression becomes:

$$\frac{\left(\gamma_0-\gamma\right)^2}{\left(1+\gamma^2\right)}\left(\sigma_\alpha^2-\sigma_{\alpha\alpha}\right)+\sigma_\epsilon^2-\frac{\rho\sigma_\epsilon^2}{\left(1+\gamma^2\right)}$$

using assumptions (i), (iv), and (v). If $\rho=0$, this expression is minimized at $\gamma=\gamma_0$. If $\rho\neq 0$, minimizing the above leads to the following expression:

$$\left(\gamma_0-\gamma\right)\left(1+\gamma\gamma_0\right)=\frac{\gamma\rho\sigma_\epsilon^2}{\left(\sigma_\alpha^2-\sigma_{\alpha\alpha}\right)}$$

Assuming that γ_0 and ρ are positive, γ cannot be equal to γ_0 because the left hand side would be 0 while the right hand side would be strictly positive. The optimal solution to the limiting objective function is not γ_0, and hence the estimator of γ would be asymptotically biased.

Case 4 - $T\to\infty$

Here, we show that the bias in Case 3 disappears if we allow for T to diverge as well. As in Case 3, we assume (in addition to the standard six assumptions) that

$$\epsilon_{i,t,n}=\rho\epsilon_{i,t-1,n}+u_{i,t,n}.$$

We consider the following simplified structure of peer groups:

(a) Each peer group has two firms in each period.

(b) Within each group, one firm is observed for T periods and the other firm is observed for one period only.

Outcome $y_{i,t,n}$ of firm i in peer group n at time t is:

$$y_{i,t,n} = \alpha_i + \gamma \sum_{j \in M_{n,t} \neq i} \alpha_j + \epsilon_{i,t,n}$$

which is consistent with either linear-in-means or aggregate spillovers. To simplify notation, the staying firm is characterized by the indices $(0, t)$ and the one-period firms by (t, t) so the time t identifies those firms. Dropping the peer group index n, the optimization problem is:

$$\min_{a,\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{T} \left(\sum_{t=1}^{T} (y_{0t} - \alpha_0 - \gamma\alpha_t)^2 + (y_{tt} - \alpha_t - \gamma\alpha_0)^2 \right)$$

First we concentrate out the fixed effects for the one-period firms. From the first-order conditions, we have:

$$\alpha_t = \frac{\gamma\,(y_{0t} - \alpha_0) + (y_{tt} - \gamma\alpha_0)}{1 + \gamma^2}$$

Substituting back into the problem:

$$\min_{a,\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{T} \left(\sum_{t=1}^{T} \frac{\left((y_{0t} - \alpha_0) - \gamma\,(y_{tt} - \gamma\alpha_0)\right)^2}{1 + \gamma^2} \right).$$

The first-order condition for α_0 leads to

$$\alpha_0 = \frac{\sum_{t=1}^{T} (y_{0t} - \gamma y_{tt})}{T\,(1 - \gamma^2)}.$$

Substituting back in yields

$$\min_{\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{T} \left(\sum_{t=1}^{T} \frac{\left(\left(y_{0t} - \frac{\sum_{\tau=1}^{T}(y_{0\tau} - \gamma y_{\tau\tau})}{T(1-\gamma^2)}\right) - \gamma\left(y_{tt} - \gamma\frac{\sum_{\tau=1}^{T}(y_{0\tau} - \gamma y_{\tau\tau})}{T(1-\gamma^2)}\right)\right)^2}{1 + \gamma^2} \right).$$

Next, we substitute in the true data generating process:

$$y_{0t} = \alpha_0^o + \gamma_0\alpha_t^o + \epsilon_{0t}^o$$

$$y_{tt} = \alpha_t^o + \gamma_0\alpha_0^o + \epsilon_{tt}^o$$

to obtain

$$\min_{\gamma} q_{N,T}(\gamma, y) \equiv$$
$$\min_{\gamma} \frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\left(\sum_{t=1}^{T}\frac{\left((\gamma_0-\gamma)\left(\alpha_t^o-\sum_{\tau=1}^{T}\frac{\alpha_\tau^o}{T}\right)+\epsilon_{0t}^o-\gamma\epsilon_{tt}^o-\sum_{\tau=1}^{T}\frac{\epsilon_{0\tau}^o-\gamma\epsilon_{\tau\tau}^o}{T}\right)^2}{1+\gamma^2}\right).$$

We consider the performance of the estimator under a sequential asymptotic framework where we first let $N\to\infty$, and then let $T\to\infty$ (i.e. $(N,T)\overset{seq}{\to}\infty$).

Define the limiting objective function as

$$q(\gamma, y) \equiv \lim_{T\to\infty} \underset{N\to\infty}{plim}\, q_{N,T}(\gamma, y)$$

and we first consider $plim_{N\to\infty} q_{N,T}(\gamma, y)$:

$$\mathbb{E}\left[\frac{1}{T}\left(\sum_{t=1}^{T}\frac{\left((\gamma_0-\gamma)\left(\alpha_t^o-\sum_{\tau=1}^{T}\frac{\alpha_\tau^o}{T}\right)+\epsilon_{0t}^o-\gamma\epsilon_{tt}^o-\sum_{\tau=1}^{T}\frac{\epsilon_{0\tau}^o-\gamma\epsilon_{\tau\tau}^o}{T}\right)^2}{1+\gamma^2}\right)\right]$$
$$=\frac{1}{T}\frac{1}{1+\gamma^2}\sum_{t=1}^{T}\frac{(\gamma_0-\gamma)^2}{1+\gamma^2}\mathbb{E}\left(\alpha_t^o-\sum_{\tau=1}^{T}\frac{\alpha_\tau^o}{T}\right)^2+\frac{1}{T}\frac{1}{1+\gamma^2}\sum_{t=1}^{T}\mathbb{E}\left(\epsilon_{0t}^o-\gamma\epsilon_{tt}^o-\sum_{\tau=1}^{T}\frac{\epsilon_{0\tau}^o-\gamma\epsilon_{\tau\tau}^o}{T}\right)^2,$$

where the cross-products are 0 by assumption (ii). Notice that both terms are squared and non-negative. The first term is minimized at 0 when $\gamma=\gamma_0$. The expectation in the second term can be expressed as

$$
\begin{aligned}
&\mathbb{E}\left[\left(\epsilon^o_{0t} - \gamma\epsilon^o_{tt} - \sum_{\tau=1}^{T}\frac{\epsilon^o_{0\tau} - \gamma\epsilon^o_{\tau\tau}}{T}\right)\right]^2 \\
&= \mathbb{E}\left[\left(\left(1-\frac{1}{T}\right)\epsilon_{0t} - \gamma\left(1-\frac{1}{T}\right)\epsilon_{tt} - \sum_{\tau=1\neq t}^{T}\frac{\epsilon_{0\tau}-\gamma\epsilon_{\tau\tau}}{T}\right)^2\right] \\
&= \mathbb{E}\Bigg[\left(1-\frac{1}{T}\right)^2\epsilon^2_{0t} + \gamma^2\left(1-\frac{1}{T}\right)^2\epsilon^2_{tt} + \frac{1}{T^2}\Bigg(\sum_{\tau=1\neq t}^{T}\epsilon_{0\tau}-\gamma\epsilon_{\tau\tau}\Bigg)^2 - 2\gamma\left(1-\frac{1}{T}\right)^2\epsilon_{0t}\epsilon_{tt} \\
&+ 2\gamma\left(1-\frac{1}{T}\right)\epsilon_{tt}\sum_{\tau=1\neq t}^{T}\frac{\epsilon_{0\tau}-\gamma\epsilon_{\tau\tau}}{T} - 2\left(1-\frac{1}{T}\right)\epsilon_{0t}\sum_{\tau=1\neq t}^{T}\frac{\epsilon_{0\tau}-\gamma\epsilon_{\tau\tau}}{T}\Bigg] \\
&= (1+\gamma^2)\left(1-\frac{1}{T}\right)^2\sigma^2_\epsilon + \frac{2}{T^2}\Bigg(\sum_{\tau=1\neq t}^{T}\sigma^2_\epsilon + 2\sum_{\tau,\tau'=1\neq t,\tau\neq\tau'}^{T}\mathbb{E}(\epsilon_{0\tau}\epsilon_{0\tau'})\Bigg) \\
&+ \frac{2\gamma^2}{T^2}\Bigg(\sum_{\tau=1\neq t}^{T}\sigma^2_\epsilon + 2\sum_{\tau,\tau'=1\neq t,\tau\neq\tau'}^{T}\mathbb{E}(\epsilon_{\tau\tau}\epsilon_{\tau\tau'})\Bigg) - \frac{2}{T}\left(1-\frac{1}{T}\right)\sum_{\tau=1\neq t}^{T}\mathbb{E}(\epsilon_{0t}\epsilon_{0\tau})
\end{aligned}
$$

Note that

$$
\left|\sum_{\tau=1\neq t}^{T}\mathbb{E}(\epsilon_{0t}\epsilon_{0\tau})\right| \leq \left|\sum_{\tau,\tau'=1\neq t,\tau\neq\tau'}^{T}\mathbb{E}(\epsilon_{0\tau}\epsilon_{0\tau'})\right| \leq 2\sum_{k=1}^{T-1}|\rho^k\sigma^2_\epsilon| \leq 2\sigma^2_\epsilon\frac{1-|\rho|^T}{1-|\rho|} = O(1),
$$

even as $T\to\infty$ since $|\rho|<1$. The same can be said for the covariances of $\epsilon_{\tau\tau}$ since they are equivalent to that of $\epsilon_{0\tau}$.

We can thus show that the whole second term can be written as

$$
\left(\frac{T-1}{T}\right)^2\sigma^2_\epsilon + \frac{2}{T^2}\Bigg(\sum_{\tau=1\neq t}^{T}\sigma^2_\epsilon + 2\sum_{\tau,\tau'=1\neq t,\tau\neq\tau'}^{T}\mathbb{E}(\epsilon_{0\tau}\epsilon_{0\tau'})\Bigg) - \frac{2}{T^2(1+\gamma^2)}\left(1-\frac{1}{T}\right)\sum_{t=1}^{T}\sum_{\tau=1\neq t}^{T}\mathbb{E}(\epsilon_{0t}\epsilon_{0\tau}).
$$

Since the final term above is a function of γ, it is not guaranteed that setting $\gamma=\gamma_0$ minimizes the objective function and hence may be biased as shown in Case 3. However, when we take the limit as $T\to\infty$, this term is $O(1/T)$ and hence goes to 0. In addition, the second term is $O(1/T)$ and also approaches 0, and thus the whole term reduces to σ^2_ϵ in the limit.

This means that, in the limit as $T\to\infty$, we only have to consider the first term of $\text{plim}_{N\to\infty} q_{N,T}(\gamma,y)$ in optimizing γ, and hence the solution is $\gamma=\gamma_0$.

To have convergence for our estimator, we require uniform convergence of the objective function $q_{N,T}(\gamma,y)$ to $q(\gamma,y)$. As in the previous case (and as in AFGK), we only need to show that for all

$\gamma \in \Gamma$,

$$\left| \frac{1}{T} \sum_{t=1}^{T} \frac{\left(\left(y_{0t} - \frac{\sum_{\tau=1}^{T} (y_{0\tau} - \gamma y_{\tau\tau})}{T(1-\gamma^2)} \right) - \gamma \left(y_{tt} - \gamma \frac{\sum_{\tau=1}^{T} (y_{0\tau} - \gamma y_{\tau\tau})}{T(1-\gamma^2)} \right) \right)^2}{1+\gamma^2} \right| \leq b(\gamma)$$

where b is a non-negative function such that $E(b(\gamma)) < \infty$. This can be shown using repeatedly the triangular inequality.

Appendix D. Details about Connectivity Weights

This section provides details about the connectivity weights used in the empirical analysis.

Input-output weights allow for examination of the extent to which spillovers operate through the flow of goods. Stronger input-output linkages may facilitate knowledge transfer about production practices or demand conditions. We build input-output weights using the Basic Price version of the 4-digit NAICS 2015 Statistics Canada input-output table. As in Ellison, Glaeser and Kerr (2010), underlying continuous weights are the maximum of upstream and downstream input and output shares:

$$w_{ij}^{\text{IOC}} = \max[\text{Input}_{k(i),k(j)},\text{Input}_{k(j),k(i)},\text{Output}_{k(i),k(j)},\text{Output}_{k(j),k(i)}].$$

We also construct separate weights using each component of w_{ij}^{IOC}. These produce similar results.

Occupational similarity weights allow for examination the extent to which knowledge transfer that is specific to particular occupations is an important driver of firm spillovers.[22] We view results using these weights as informative about the extent to which industries with more similar occupational mixes have more productive knowledge flows. Closer occupational similarity with peers could mean that workers learn more about how to effectively perform their core occupational tasks, where such knowledge transfer may happen through chance encounters (Atkin, Chen and Popov, 2022). We build occupational similarity measures using the 2002 US National Industry Occupation Employment Matrix, which is built using data from the Occupational Employment Statistics survey conducted by the Bureau of Labor Statistics. For each industry, it gives the share of employees in each four-digit occupation. Similar to Ellison, Glaeser and Kerr (2010), we define occupational similarity weights as:

$$w_{ij}^{\text{OCCSIM}} = \max[\text{Corr}(\text{Occ. Share}_{k(i)}, \text{Occ. Share}_{k(j)}), 0].$$

Worker flows weights similarly capture the extent to which workers in firm i's industry are likely to have either direct experience working in peers' industries or to use a similar set of skills in their jobs. Seeing a high rate of worker flows from peers' industries is an indicator of closer connections in one or both of these dimensions. We build information on the prevalence of inter-industry worker flows by using the employer-employee match component of our data set. Using all employees in

[22] Ellison, Glaeser and Kerr (2010) interpret greater coagglomeration of firms in occupationally similar industries in local labor markets as reflecting labor market pooling. Their interpretation is likely to be less relevant at the small spatial scale of spillovers that we examine in this paper.

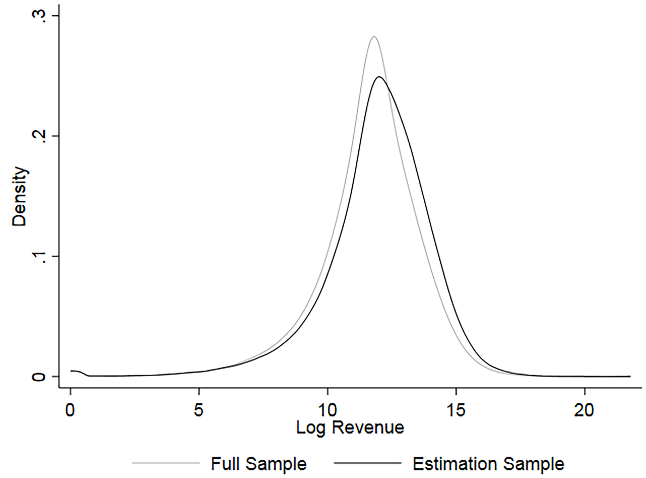
Canada earning at least CAD 5,000 that had different employers in 2001 and 2002, we calculate the share of worker flows from firms in each industry k' that go to each other industry k, adjusting for the share of industry k' in total employment. In particular,

$$w_{ij}^{\text{WFLOW}} = \frac{\text{fraction of industry job changers to industry } k(i) \text{ that are from } k'(j)}{\text{fraction of total job changers from industry } k'(j)}.$$

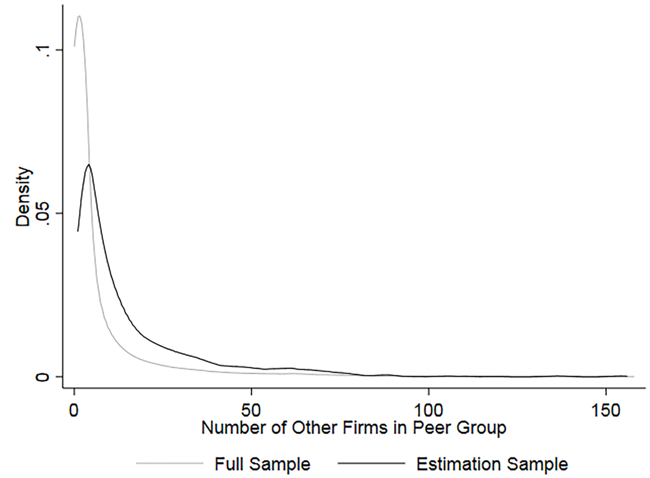
The denominator accounts for the fact that random choices out of industries with greater worker shares and/or mobility rates would mechanically generate greater flows to all other industries. Therefore, w_{ij}^{WFLOW} measures the extent to which worker flows from industry $k'(j)$ to industry $k(i)$ are greater or less than expected relative to random destination industry choices, taking transitions out of industry $k'(j)$ as given.

Finally, similar to Greenstone, Hornbeck and Moretti (2010), we also test whether firms in the same 2-digit industry generate differential spillovers to those in other 2-digit industries. In this case, $w_{ij}^{\text{SAME}} = 1$ if $k(i) = k(j)$ at the 2-digit NAICS level and 0 otherwise. Rather than using terciles, we implement this weight in the empirical work by examining impacts of having a higher fraction of peers in the same industry.

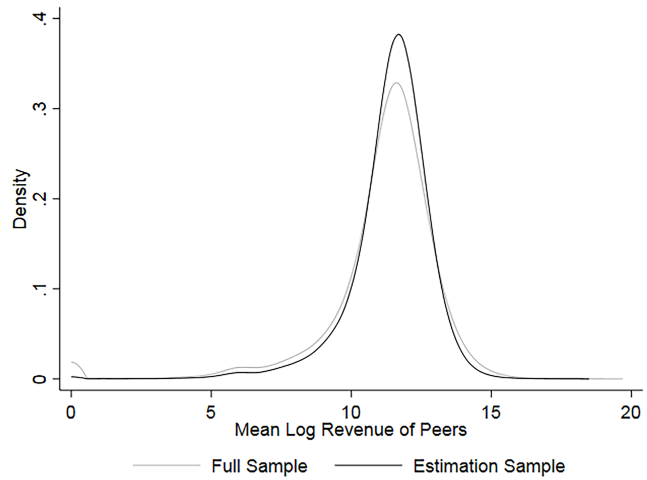
Figure E1. – Descriptive Graphs



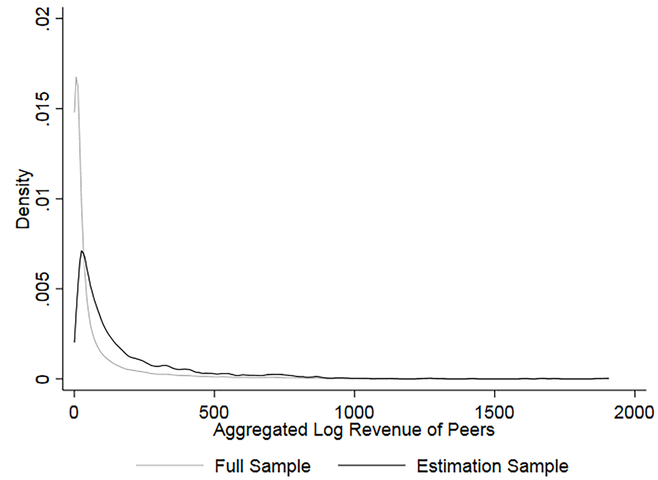
(a) Distribution of Firm Log Revenue



(b) Distribution of Peer Group Size



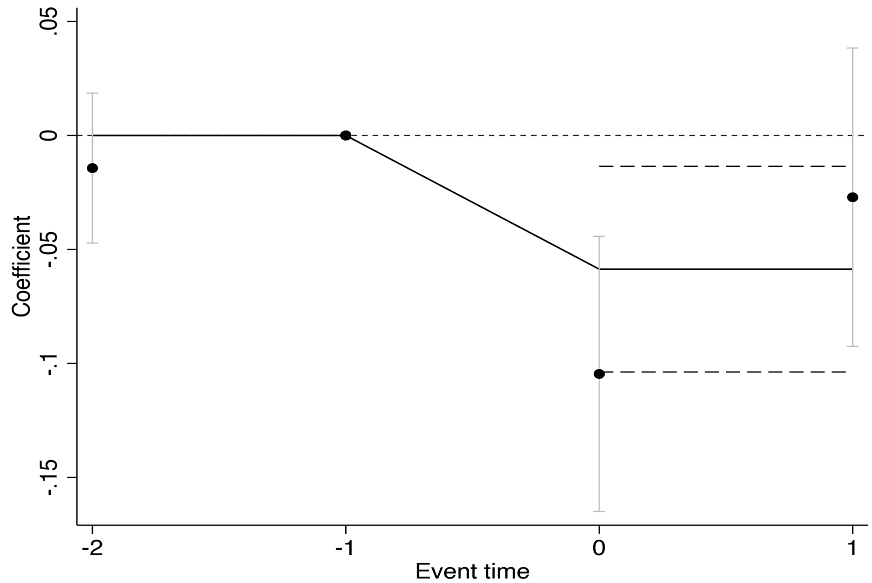
(c) Distribution of Mean Log Revenue in Peer Groups



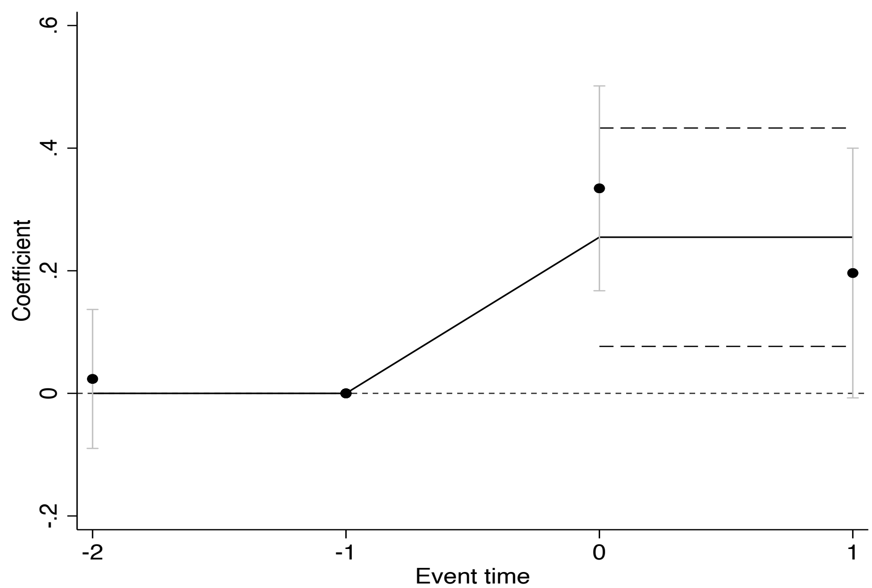
(d) Distribution of Aggregate Log Revenue in Peer Groups

Notes: Plots are for all single-location high-skilled services firms in Montreal, Toronto, and Vancouver census metropolitan areas. The estimation sample excludes firms in peer group areas with one or more member postal code with an area that is greater than $\pi 75^2$sq meters (0.018 sq km) and peer group areas with fewer than two high-skilled services firms in any year 2001-2012.

Figure E2. – Changes in Log Revenue Induced by Firm Entry



(a) Negative shock: firm entry causes a change in average peer quality below 10^{th} percentile



(b) Positive shock: firm entry causes a change in average peer quality above 90^{th} percentile

Notes: Figures show coefficients and confidence intervals from event-study regressions. The dependent variable is firm log revenue residualized for estimated fixed effects in Table 2 Column (3). Dots are coefficients on yearly event-time dummies, normalizing the coefficient on event-time -1 to 0. Horizontal solid lines are coefficients on biyearly event-time dummies, normalizing the coefficient on pooled event-time -1 and -2 to 0. Whiskers and dashed lines show associated 90% confidence intervals calculated using standard errors clustered at the event level. Estimation samples only include incumbent firms in peer groups in which all incumbents experience a change in average peer quality that is below the 10^{th} percentile (Panel A) or above the 90^{th} percentile (Panel B) of the change in average peer quality distribution across all firm-year observations in the primary estimation sample. Only such events that are induced by the arrival of new firms in a peer group location with no other changes in peer composition up to two years prior and one year after the event are included. Panel A has 93 events, 220 incumbents, and 2200 observations. Panel B has 25 events, 60 incumbents, and 560 observations.

Table E1. – Coefficient Stability Around Large Events

Event time	$t-2$	$t-1$	t	$t+1$	$t+2$
	(1)	(2)	(3)	(4)	(5)
Change in average peer quality $<$ 10th percentile					
Avg. Peer Firm F.E.	0.0214	0.0250	0.0215	0.0367	0.0107
	(0.0063)	(0.0057)	(0.0060)	(0.0068)	(0.0063)
$\bar{X}$	0.06	0.17	-0.56	-0.39	-0.30
Number of Obs.	16,100	19,600	19,500	16,300	12,700
Change in average peer quality $<$ 25th percentile					
Avg. Peer Firm F.E.	0.0262	0.0228	0.0167	0.0236	0.0158
	(0.0048)	(0.0042)	(0.0043)	(0.0055)	(0.0050)
$\bar{X}$	0.05	0.13	-0.30	-0.19	-0.14
Number of Obs.	36,700	44,800	44,500	36,900	28,200
Change in average peer quality $>$ 75th percentile					
Avg. Peer Firm F.E.	0.0243	0.0241	0.0162	0.0225	0.0196
	(0.0047)	(0.0042)	(0.0037)	(0.0044)	(0.0048)
$\bar{X}$	-0.30	-0.33	0.16	0.16	0.14
Number of Obs.	40,400	49,200	49,100	40,300	30,200
Change in average peer quality $>$ 90th percentile					
Avg. Peer Firm F.E.	0.0206	0.0272	0.0085	0.0170	0.0168
	(0.0075)	(0.0069)	(0.0075)	(0.0063)	(0.0071)
$\bar{X}$	-0.54	-0.63	0.19	0.18	0.13
Number of Obs.	16,900	20,600	20,500	16,900	12,700

Notes: Each entry is from a separate regression analogous to that in Table 2 Column (3) but using post-estimation data and different sub-samples. All firms exposed to changes in average peer quality of the amount indicated in each panel are assigned an event year. Regressions of firm log revenue residualized for estimated fixed effects in Table 2 Column (3) on average peer quality, aggregate peer quality, and number of peers are run separately by event time. $\bar{X}$ is the average of average peer quality in each estimation sample. Standard errors are clustered by peer group area.

Table E2. – More Information About Peer Groups

	Peer Group Area Radius			
	75m	150m	200m	250m
Panel A: Average and SD Across Firm-Years				
# of Peers	15.95	28.26	36.34	45.13
	(19.55)	(42.15)	(53.43)	(64.72)
Area (sq. km)	0.043	0.100	0.155	0.216
	(0.115)	(0.216)	(0.340)	(0.403)
Panel B: Average and SD Across Firms				
# of Peer Groups Experienced	1.45	1.44	1.44	1.43
	(0.72)	(0.71)	(0.71)	(0.70)

Notes: Averages and standard deviations (in parentheses) are for single-location high-skilled services firms in the primary estimation sample. The sample excludes firms in peer group areas with one or more member postal code with an area that is greater than $\pi 75^2$sq meters (0.018 sq km) and peer group areas with fewer than two high-skilled services firms in any year 2001-2012. The sample only includes firms in the Montreal, Toronto, or Vancouver census metropolitan areas. Statistics in Panel A are calculated using all firm-year observations. Statistics in Panel B are calculated using one observation per firm.

TABLE E3. – Aggregate Impacts of Counterfactual Firm Allocation Across Peer Groups, Het. Treatment

Randomization Type	Fixed Group Size		Equal Group Size	
Nature of Spillovers Considered	LIM + HET	+ AGG	LIM + HET	+ AGG
	(1)	(2)	(3)	(4)
Estimates w/ Area × Year F.E., Randomized Within Areas	-0.0034	-0.0029	-0.0037	-0.0071
	(0.0006)	(0.0008)	(0.0006)	(0.0004)
Estimates w/o Area × Year F.E., Randomized Within Areas	-0.0129	-0.0170	-0.0134	-0.0238
	(0.0007)	(0.0016)	(0.0006)	(0.0011)
Estimates w/o Area × Year F.E., Randomized Across All Locations	-0.0094	-0.0231	-0.0093	-0.0271
	(0.0013)	(0.0015)	(0.0013)	(0.0015)

Notes: Table presents the means and standard deviations of changes in aggregate revenue that would ensue under 100 simulations of various scenarios in which sorting of firms across peer groups is eliminated. Results in the two columns under the header "Fixed Group Size" are generated holding peer group size fixed and those under the header "Equal Group Size" are generated given full randomization of firms across peer groups. In each column headed by LIM + HET, counterfactual firm revenue absent sorting is calculated adjusting for the linear-in-means component of the spillover as well as the fraction of peers in the top tercile of the local 500-meter radius area's firm quality distribution, using coefficients from Table 6, column 6. In each column headed by +AGG, the same two terms plus the agglomeration term are included in the calculation, again using coefficients from Table 6, column 6. The first row uses fixed effects estimates from Table 2 column (3) and imposes demeaning and randomization across peer groups within 500-meter radius areas. The second row uses fixed effects estimates from Table 2 column (6) instead with the same demeaning and randomization procedures. The third row uses fixed effects estimates from Table 2 column (6) but demeans and randomizes across all peer groups.