Local Productivity Spillovers*

Nathaniel Baum-Snow†
University of Toronto
Nicolas Gendron-Carrier‡
McGill University
Ronni Pavan§
University of Rochester

May 6, 2020

Abstract: This paper presents evidence of revenue and productivity spillovers across firms at fine spatial scales. For groups of firms within 75 meter radius areas, estimates indicate an average elasticity of firm revenue to the mean of peer firm revenue of 0.018. The elasticity with respect to aggregate peer revenue is at most an additional 0.003 percent. Impacts are very local, fully decaying within 250 meters, and are greater for higher quality firms. Tests for mediation through input-output and occupational similarity relationships yield stronger evidence for the occupational similarity channel. Theory indicates that TFP spillovers are if anything greater than these revenue spillover estimates.

*Rolando Campusano and Xiaoyang Li provided superb research assistance.
†Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, ON, Canada, M5S3E6. e-mail: Nate.Baum-Snow@rotman.utoronto.ca.
‡Department of Economics, McGill University, 855 Sherbrooke St. West, Montreal, QC, Canada, H3A2T7. e-mail: nicolas.gendron-carrier@mcgill.ca.
§Department of Economics, University of Rochester, 280 Hutchinson Road, Rochester, NY, USA, 14627. e-mail: rpavan@ur.rochester.edu.

1 Introduction

Considerable evidence quantifies the scale and nature of agglomeration economies at the regional and local labor market levels. Greenstone et al. (2010), Ellison et al. (2010), Bloom et al. (2013), Faggio et al. (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.

Using panel data derived from corporate tax records on the universe of single-location firms in Canada, this paper provides some of the first causal estimates of productivity spillovers at small spatial scales for a broad set of firms and quantifies the underlying mechanisms driving these spillovers. We find strong evidence of revenue and productivity externalities that operate between firms within areas of 75 meter to 250 meter radii. For groups of firms within 75 meter radius areas, we find that the elasticity of revenue with respect to mean revenue in a firm's peer group is 0.016-0.018 for the average firm. We estimate an elasticity of revenue with respect to aggregate peer group revenue of at most 0.003 percent. For a sense of magnitude, these estimates indicate that going from the 10th to 90th percentile of peer group spillovers across firms in our data increases revenue by 5 percent via linear-in-means spillovers and up to an additional 2 percent via aggregate spillovers. This latter impact is very skewed such that the majority of firms hardly benefit from local aggregate effects. These two impacts are additive. Tests for mediation through industry input-output and occupational similarity relationships across firms yield stronger evidence for the occupational similarity channel. We also find that firms benefit slightly more from exposure to higher quality firms in other two-digit industries than their own. Coupled with the dominance of linear-in-means type externalities, we interpret our evidence as showing that learning or knowledge transfer between nearby firms is the primary mechanism driving spillovers at microgeographic spatial scales.

Through the use of counterfactuals, we investigate the extent to which sorting across space and into peer groups on unobserved firm quality drives agglomeration economies. While we find strong evidence of sorting of firm quality on peer quality, this sorting is mostly accounted for by differences in location fundamentals. Absent such sorting, aggregate firm revenue through linear-in-means spillovers would be 0.6 percent lower, mostly because the highest quality firms receive smaller spillovers in this environment. The sorting of higher quality firms into denser employment subcenters is an important additional force, leading to up to an additional 0.8 percent in aggregate revenue. When the linear-in-means and agglomeration spillovers are both considered, our estimates indicate that aggregate revenue would be 0.7% lower absent sorting across fixed locations with firms and 1.4% lower absent sorting across all potential locations.

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 40,000 firms in more than 30,000 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. With reasonable assumptions about the data generating process for revenue, also employed by De Loecker (2011), our identification strategy allows revenue spillover estimates to be interpreted as TFP spillovers. Robustness checks using direct TFPR estimates or revenue adjusted for endogenous price responses to changes in peer group composition corroborate these assumptions.

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 beforehand 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) (henceforth, "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, to recover the degree of complementarity between a firm's own unobserved fixed attributes and those of its peers, to distinguish between the relative importance of different types of connectivity weights, and to measure the extent to which spillovers decay spatially. 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]

[1] Liu et al. (2014), which simultaneously estimates impacts of linear-in-means and aggregate type spillovers

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, 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.

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 area fixed effects of a 500 meter radius interacted with year. In robustness checks we additionally include 250 meter radius area fixed effects not interacted with year. In this most saturated specification, identifying variation comes from a combination of cross-sectional differences in firm composition in adjacent regions of 75 meter radius and differential changes in firm composition over time in these same peer groups when compared within larger regions with a 500 meter radius. The inclusion of neighborhood fixed effects coupled with changes in firm composition are important to identify peer effects separately from location fundamentals.

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 gives support to our identification strategy. In order to hedge against business cycle risk, landlords typically rent out space on a rolling basis with 10 year commercial leases, generating smoother variation in such turnover and making it more difficult for firms to coordinate on location. As a result, in any given year there are typically few options available for new commercial space within a 500 meter radius. Bayer et al. (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 firm location choices could be correlated with annual shocks to small area fixed effects.

in the context of looking at peer effects on effort in studying and participating in school sport activities, is an exception.

Our focus on high skilled services 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 some model structure to account for endogenous price responses corroborate our more reduced form estimates.

At first blush, it might appear that our evidence that linear-in-means (peer effects) type spillovers dominate simple aggregation (agglomeration) spillovers is at odds with observed productivity and wage premia that are associated with city size. Coupled with our evidence that higher quality firms experience larger spillovers from peer groups of the same quality than do lower quality firms, however, our baseline results indicate an important interaction between sorting and firm externalities that generates aggregate increasing returns at the city level. That is, evidence in this paper shows that 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. That is, the "Plant Size-Place Effect" of larger firms in larger cities (Manning, 2009) also means there will be larger firm-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 et al. (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 vs. between firms. Our very localized evidence is in line with Moretti (2004), Kantor and Whalley (2014), and Serafinelli (2013)'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 other firm's outcome (Manski, 1993; Angrist, 2014), we do not attempt to isolate this component of our spillover estimates. Instead, we follow Gibbons et al. (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 anyhow, 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 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 2, we develop a theoretical framework that justifies and interprets our use of revenue as the main outcome variables of interest. Section 3 describes our empirical model, identification, and estimation. Section 4 describes the data and sample. Section 5 discusses the main results. Section 6 presents counterfactuals oriented toward isolating the impacts of firm sorting. Section 7 concludes.

2 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. Starting with a standard firm 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 to the weighted aggregate of its peers' fixed effects. We show that under certain conditions this parameter measures the average total factor productivity (henceforth, "TFP") spillover between firms within each peer group.

Our main estimation equation accommodates both perfectly competitive and monopolistically 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 adjustment for the variable input share. If output prices are endogenous and specific to the firm, an increase in TFP reduces marginal cost, thereby reducing the 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. Given measures of this demand elasticity, we can adjust revenue to account for this endogenous price response, allowing us to recover measures of TFP spillovers under imperfect competition as well. We note that estimated unadjusted revenue responses to changes in peer composition if anything understate the true magnitude of spillovers, as they reflect in part the negative impact on the output price. We discuss details of the perfect competition case here and relegate a detailed discussion of the monopolistic competition case to Appendix 1.

[2] Some of the most credible evidence of endogenous productivity spillovers uses a supply chain network structure for identification, as in Bazzi et al. (2017).

We think of firms as operating on an amount of space that is fixed in the short run. The only way a firm can adjust the total amount of space it uses is to move to a different block b. In the empirical work we vary the size of the block by aggregating postal codes to areas of radius 75 to 250 meters. The 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 the TFP parameter $A_{i,b,k,t}$, 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 so output prices are likely to be determined at a broader level of spatial aggregation than the block. With market power, output prices differ across firms as developed Appendix 1. The fixed cost $F_{i,b,k,t}$ captures real estate and capital inputs, which are fixed in the short run, but whose price can vary over time and space.

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 (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 goal of the empirical work is to piece out productivity spillovers from the relationship between variation in peers' log revenue and firms' own log revenue. Doing so requires holding constant location-specific attributes of wages and output prices, which we control for with various fixed effects described below. Conditional on output prices and wages, (2) thus indicates the extent to which shocks to log revenue that spill over from nearby firms fully reflect log TFP spillovers between firms. In particular, conditional on the output price and variable input cost, an observed 10 percent shock to log revenue would reflect a 3.3 percent change in TFP given a variable input share of 70 percent.

2.1 TFP Spillovers

To complete the structural representation of our estimation equation, we specify the process through which we conceptualize TFP spills over 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:

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

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 the aggregation of firm-specific component of TFP that is fixed over time across other firms in firm i's peer group. In the bulk of our empirical work, we summarize this object as one unified "average treatment effect" parameter, as is standard in the peer effects and TFP spillovers literatures. We also investigate the degree of complementarity between α_i^A and peer group quality.

Local area-industry-year fixed effects $\phi_{B(b),k,t}^A$ capture a combination of location fundamentals and industry level TFP shocks. In our baseline specification, connectivity weights $\omega_{ij}(M_{b,t})$ are equal across peers. To study the nature of spillovers, we also impose weights measuring input-output relationships, occupational similarity, or industry similarity between firm i's industry and firm j's industry, with details in Section 4.3. Weights are normalized in "linear-in-means" specifications and are unscaled in "agglomeration" specifications. $M_{b,t}$ denotes the set of firms in block b at time t.

In order to distinguish between mechanisms driving agglomeration spillovers at a microgeographic scale, some of our empirical work looks at "horse races" between different weighting schemes. These horse races are either between linear-in-means and agglomeration type spillovers or between different types of spillovers given linear-in-means or agglomeration aggregation schemes.[3] In these cases, (3) becomes

$$\ln A_{i,b,k,t} = \alpha_i^A + \phi_{B(b),k,t}^A + \gamma_1^A \left[\sum_{j \in M_{b,t}, \neq i} \omega_{ij}^1(M_{b,t}) \alpha_j^A \right] + \gamma_2^A \left[\sum_{j \in M_{b,t}, \neq i} \omega_{ij}^2(M_{b,t}) \alpha_j^A \right] + \varepsilon_{i,b,k,t}^A.$$

[3] For computational reasons and because we only have independent identifying variation across peer group composition, we limit all horse races to be between only two different peer group compositions at a time.

2.2 Structural Interpretation of Revenue Spillovers

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

$$\ln R_{i,b,k,t} = \alpha_i^R + \phi_{B(b),k,t}^R + \gamma^R \left[\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t}) \alpha_j^R \right] + \varepsilon_{i,b,k,t}^R. \tag{4}$$

The model shows how to assign structural interpretations to each empirical model parameter in (4) and be clear about the conditions under which the reduced form parameter γ^R identifies the structural parameter γ^A. Inserting (3) into the total revenue equation (2) delivers the structural interpretation of each parameter in (4).

We first consider the interpretation of local area-year and industry-year fixed effects $\phi_{B(b),k,t}^R$. Once these are understood, it is more straightforward to see what firm-specific factors remain. Conceptually, $\phi_{B(b),k,t}^R$ are intended to control for variable input prices, location fundamentals, and industry-specific components of productivity and output demand. The empirical work explores two specifications of these fixed effects: 1) 500 meter radius area indicators interacted with year fixed effects and 2-digit industry indicators interacted with year fixed effects and 2) the same with the addition of 250 meter radius area fixed effects. Future robustness analysis will also consider 500 meter radius areas doubly interacted with year fixed effects and 2-digit industry fixed effects.

Under perfect competition, the structural interpretation of the fixed effects in (4) are

$$\phi_{B(b),k,t}^R = \frac{\theta_k}{1-\theta_k} \ln \theta_k + \frac{1}{1-\theta_k} \ln p_{B(b),k,t} - \frac{\theta_k}{1-\theta_k} \ln w_{B(b),k,t} + \frac{1}{1-\theta_k} \phi_{B(b),k,t}^A.$$

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

The structural interpretation of α_i^R is determined jointly by the firm-specific fixed effect term and the spillover term. If the firm-specific fixed effect in (4) is set to $\alpha_i^R = \frac{1}{1-\theta_{k(i)}} \alpha_i^A$, the remaining terms in (4) are

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

If all firms in firm i's peer group have the same variable input share, revenue spillovers γ^R directly measure TFP spillovers γ^A. Our empirical setting also allows for interpretation of unadjusted revenue spillover estimates as TFP spillovers under monopolistic competition and CES demand (constant markups) if all firms in firm i's peer group have the same variable input share and markup, as discussed in Appendix 1. This argument exhibits 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. If variable input shares and markups are not the same for all firms in a peer group, the structural interpretation of α_i^R is more complicated and the structural error $\varepsilon_{i,b,k,t}^R$ includes a peer group aggregation component. In this case, γ^R no longer strictly measures TFP spillovers. However, subject to validity of our primary identification strategy discussed in the following section, it still measures revenue spillovers holding input costs, a component of output prices and local area-industry-time drivers of TFP constant.

In a robustness analysis, we employ two strategies to accommodate differences within peer groups in variable input shares and markups in order to recover estimates of TFP spillovers. Our first strategy uses revenue adjusted for market power as an outcome. This adjustment allows us to isolate firm fixed effects as the permanent firm-specific component of TFP and the TFP spillover parameter γ^A is then directly estimated. As a second alternative strategy, we use a direct measure of firm-year TFPR as an outcome. This strategy has the disadvantage of not separating out impacts on prices from quantities. See Appendix 2 for details.

3 Empirical Implementation

We consider data generating processes in which there is productivity diffusion between nearby firms. Our baseline estimation equation relates outcome $y_{i,k,b,t}$ of firm i operating in industry k and peer group b at time t to peer outcomes using the following specification:

$$y_{i,k,b,t} = \alpha_i + \phi_{B(b),t} + \rho_{k,t} + \gamma \sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})\alpha_j + \varepsilon_{i,b,k,t}. \tag{5}$$

We use log firm sales revenue as our primary outcome of interest. Robustness checks instead use a measure of TFPR and revenue adjusted for market power as alternative outcomes.

In (5), α_i is a firm fixed effect, $\phi_{B(b),t}$ is a local area-year specific fixed effect that captures access to local productive amenities or local labor supply conditions, and $\rho_{k,t}$ is an industry and year specific fixed effect capturing secular trends in industry-specific productivity, wages and/or output prices. The key predictor variable, $\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})\alpha_j$, is an aggregate of the fixed component of this same outcome in nearby firms at time t, where the weights depend on some measure of proximity between firm i and firm j at time t. Weights are normalized to sum to $\frac{\sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})}{|M_{b,t}|-1}$ in linear-in-means specifications and are unscaled in agglomeration specifications. Most of our empirical work uses “basic weights” in which $\omega_{ij}(M_{b,t}) = \frac{1}{|M_{b,t}|-1}$ 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 a firm outcome. The firm fixed effects α_i are economically informative measures of firm quality. We use estimates of α_i to investigate the importance of sorting across space on estimated firm quality to quantify the extent to which such sorting accounts for aggregate firm spillovers.

We estimate a number of different variants of (5) to understand heterogeneity in treatment effects and to make comparisons across different types of spillovers. To investigate heterogeneous treatment effects, we estimate

$$y_{i,k,b,t} = \alpha_i + \phi_{B(b),t} + \rho_{k,t} + \gamma^0 \sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})\alpha_j + \gamma^1 \alpha_i \sum_{j \in M_{b,t}, \neq i} \omega_{ij}(M_{b,t})\alpha_j + \varepsilon_{i,b,k,t}.$$

We also investigate a number of “horse race” specifications. For these, we estimate

$$y_{i,k,b,t} = \alpha_i + \phi_{B(b),t} + \rho_{k,t} + \gamma^A \sum_{j \in M_{b,t}, \neq i} \omega^A_{ij}(M_{b,t})\alpha_j + \gamma^B \sum_{j \in M_{b,t}, \neq i} \omega^B_{ij}(M_{b,t})\alpha_j + \varepsilon_{i,b,k,t}.$$

This allows us to determine the extent to which linear-in-means versus agglomeration type spillovers dominate for a given weighting scheme, to determine which types of connectivity weights best accommodate each type of spillover, and to determine the spatial extent of spillovers.

Empirical implementation requires imposing a normalization that assigns the constant either to each firm's α_i or to one set of fixed effects included for identification, $\phi_{B(b),t}$ or $\rho_{k,t}$. In order to distinguish linear-in-means from agglomeration spillovers, we assign the constant to α_i. With this choice, the impact of an additional low quality firm to a peer group raises peer group quality in the agglomeration specification but reduces peer quality in the linear-in-means specification, with the spillover parameters scaled appropriately in estimation to match this normalization. If the constant were not included in α_i, it would be more difficult to distinguish these two types of spillovers as additional firms could reduce aggregate peer quality in the agglomeration specification.

3.1 Identification

Identification of our key parameters of interest requires variation in the composition of firms within blocks that is unrelated to unobservables driving outcomes. The empirical setup we adopt from Arcidiacono et al. (2012) uses such identifying variation in a way that accommodates the classic identification challenge of sorting on fixed unobserved attributes. In our context, such sorting would occur if particularly productive or high paying firms located in certain blocks because of block attributes that promote such productivity rather than because of productivity spillovers. For example, more productive firms may be the high bidders for commercial real estate near train stations and highway interchanges because of their workers' higher values of commuting time, thereby inducing a correlation between average outcomes of firms in a block that is not causal. If fixed components of neighbors' outcomes are uncorrelated with shocks to firm i's outcomes, then we can identify γ in (5) directly by nonlinear least squares even while excluding industry-time and location-time fixed effects. Firm fixed effects would be separately identified given variation over time in the composition of firms in blocks. The spillover parameter would be identified from relationships between firm outcomes $y_{i,k,b,t}$ residualized for firm fixed effects and the weighted aggregate of neighbors' fixed effects in the firm's block.

Our baseline conceptualization of the data generating process in (5) incorporates location fixed effects as controls for two reasons. First, we think about spillovers as accruing from fundamental fixed attributes of neighboring firms rather than location characteristics themselves as they operate through neighboring firms. Appropriate specification of this process thus requires partialing out the location fixed effects $\phi_{B(b),t}$. Second, local shocks may be a driving force for the location choice of firms, even conditional on their α_is. We are 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 and exit to and from blocks, we must clean out any such unobservables that predict both composition changes in neighbors' fixed effects because of firm turnover and changes in outcomes for incumbents.

In the main analysis, we use the full sample of firms and blocks in downtown Montreal, Toronto, and Vancouver and flexibly control for neighborhood-time fixed effects. The lack of correlation between the fixed effects of entering firms and shocks to incumbent firms thus hinges on locally thin markets for commercial real estate in these areas. Thin commercial real estate markets put a constraint on the amount of information firms can act upon when deciding which building to move into. In one specification, we define neighborhoods as sets of blocks in areas of 500 square meters, justifying the inclusion of $\phi_{B(b),t}$ in our empirical specification discussed above. This controls for secular trends in productivity, demand or labor supply conditions and is similar to the identification strategy employed by Bayer et al.

(2008). The benefit of this approach is that it allows us to recover more broadly based treatment effects and also to have enough variation in the data to explore mechanisms driving the results through comparisons across different connectivity weights. Moreover, because identification comes from changes in firm composition, it accommodates some types of unobserved changes that come from arrivals or departures of multi-location firms.

The key identifying assumption necessary to obtain unbiased estimates of the spillover parameter is that trends in local economic conditions driving changes in firm composition and changes in outcomes operate at a broader spatial scale than do the firm spillovers. Controls for neighborhood-specific nonparametric time trends allows identification to come from variation in the peer group area in which new firms and relocating firms locate. Once again, this assumes the existence of sufficiently tight commercial real estate markets at the neighborhood level such that firms cannot choose the exact block in which to locate within a small area. In robustness checks, we also consider specifications which add 250 meter radius fixed effects, accommodating fixed differences in fundamentals across even smaller areas.

3.2 Estimation

Estimation of (5) requires solving a nonlinear least squares problem, as there is a large number of firm fixed effects that need to be recovered jointly with the spillover parameter γ. We estimate the models outlined above using the iterative algorithm proposed by Arcidiacono et al. (2012). As long as 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.

The nonlinear least square estimator for parameters in our main estimation equation (5) minimizes the following objective function:[4]

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

Taking the first-order condition with respect to α_i yields an updating equation for each α_i. Arcidiacono et al. (2012) propose to solve for parameters using a two-step algorithm. In the first step, the firm fixed effects are taken as given and estimates of γ, $\phi_{B(b),t}$ and $\rho_{k(i),t}$ are obtained by a standard fixed effect estimator. In the second step, γ, $\phi_{B(b),t}$ and $\rho_{k,t}$ are taken

[4] In some variants of our analysis, we replace γ with $\gamma^0 + \gamma^1\alpha_i$. In other variants of our analysis, we replace $\gamma \sum_j \omega_{ij}(M_{b,t})\alpha_j$ with $\gamma^A \sum_j \omega_{ij}(M_{b,t})^A\alpha_j + \gamma^B \sum_j \omega_{ij}(M_{b,t})^B\alpha_j$, where $\omega_{ij}(M_{b,t})^A$ and $\omega_{ij}(M_{b,t})^B$ are defined to capture linear-in-means vs. agglomeration effects, different connectivity weights within linear-in-means, or different spatial extents of spillovers within linear-in-means or agglomeration type spillovers.

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. We initialize α_i to be estimates from a regression of $y_{i,k,b,t}$ on firm, local area-year and 2-digit industry-year fixed effects, assigning the constant to α_i. We use a symmetric wild bootstrap (MacKinnon, 2006) to calculate standard errors.

4 Data

4.1 Data Sources and Sample

Our primary data source is administrative data on all incorporated firms in Canada between 2001 and 2012. Our data set is derived from T2 Corporation Income Tax Return files. All corporations have to 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). This data set is particularly well suited for our analysis because it includes six-character postal code identifiers and distances between postal code centroids in 25 meter bands up to one kilometer. Canadian postal codes in the central areas of cities usually cover blockfaces or individual buildings.

We process the firm information to keep only firm-years in the Montreal, Toronto and Vancouver CMAs with some 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 exclude firm-years with none of these three items reported. We also drop firms that cycle back and forth between postal codes, firm-years with missing location information, and firms with no 4-digit industry information. If any of these selections generates a hole in a firm's individual panel, we drop the entire firm. We identify a firm's entry and exit years as the first and last years it has positive reported revenue. We do keep some firms that have missing revenue in the middle of their panel as positive employment and/or payroll provides evidence that they were operating. Because we only observe one postal code per firm, we restrict our attention to firms that have a single location in each year we observe them. 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.

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, the largest 1-digit sector by firm count. The next biggest sector is recreation, accommodation and food services (NAICS 7). We elected not to include NAICS

7 firms because their demand conditions commonly varies at a microgeographic scale.[5] The typical NAICS 5 firm in Montreal, Toronto or Vancouver is smaller than the average firm with lower revenue and fewer employees but higher payroll per worker. In particular, it has an annual revenue of about CAD 850,000 and 4 employees, with a typical worker earning about CAD 48,000. Single-location firms are small and their indivdual movement is unlikely to influence local factor prices.

We construct two different estimation samples using the information on single location firms operating in NAICS 5 industries. We endeavor to isolate peer groups of firms whose centroids are contained within 75 a meter radius. We first group postal codes by this criterion, fully segmenting each of the three CMAs in our data. We exclude all such groups 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 2 firms in any year 2001-2012. We iterate to additionally exclude peer group areas which include firms for which 500 meter by year fixed effects or 2-digit industry by year fixed effects are not identified at any point in their history. This is our robustness sample. Our main estimation sample additionally excludes firms that have missing revenue information within their panel, and their associated peer groups if this puts them below 2 firms.

The resulting samples have 269,144 and 216,704 firm-year observations respectively. The smaller of the two has about 6,000 postal codes and 3,000 peer groups in each year with an average peer group size of 6.5 firms. We cover about 25% of single location NAICS 5 firms in the three CMAs, with the exclusions primarily due to firms being in peer groups that have fewer than 2 firms and 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 of 2.1 firms. We cover firms that are in the denser areas of the three cities. Firms in our estimation sample also are somewhat larger than all single location NAICS 5 firms at about CAD 1.3 million per year with 5 employees. Figure 1 Panel A shows the distributions of firm log revenue in Sample 1 as compared to all single-location NAICS 5 firms. Figure 1 Panel B shows the distributions of peer group sizes in our estimation sample and for all such firms. Importantly, the distribution of peer group size is highly skewed to the right, with the largest peer groups having about 125 members. This dispersion in peer group size means that we have sufficient independent variation in aggregate and mean peer quality to separately identify linear-in-means from agglomeration spillovers.

Figure 1 Panels C and D show the distributions of average and aggregate peer log revenue respectively. Average peer revenue for our estimation sample has close to a lognormal distri-

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

bution, which is smoother than that for the full population of single location NAICS 5 firms. Aggregate peer log revenue is highly skewed, as should be expected given the distribution of peer group size.

Evidence in Figure 2 shows the extent to which firms sort into peer groups on revenue. Panel A shows that above the median, there is positive sorting on the mean log revenue of firm peers. Panel B shows that the same is true for the aggregate log revenue of peers, with a huge bump in the right tail of the distribution, meaning that the very high revenue firms tend to be located in highly agglomerated areas. In Section 6 below, we revisit relationships like this after accounting for the component of revenue due to spillovers to see that firms do sort on fundamental peer quality as well in both dimensions (α_i is correlated with $\sum_{j\in M_{b,t}\neq i}\omega_{ij}(M_{b,t})\alpha_j$).

4.2 Connectivity Weights

Our framework allows for cross-firm productivity spillovers that are mediated through industry input-output relationships and occupational similarity. We develop connectivity weights $\omega_{ij}(M_{b,t})$ with two key attributes. The first characterizes the type of spillover we consider, either linear-in-means or unscaled aggregation (agglomeration). The second attribute describes how we capture linkages between industries. To evaluate the relative importance of different mechanisms driving spillovers, we run "horse races" between aggregations of peer α_js under various different weighting schemes.

In the first regard, we consider connectivity weights that are of the form

$$\omega_{ij}(M_{b,t}) = \begin{cases} \frac{\omega_{ij}}{|M_{b,t}|-1} \text{ in the linear-in-means model} \\ \omega_{ij} \text{ in the agglomeration model} \end{cases}$$

That is, weights are normalized to sum to some fraction of the number of firms in the peer group in the linear-in-means model and are unconstrained in the agglomeration model.

In the second regard, we consider the following options for industry pair weights. As a baseline, we use "basic weights" in which $\omega_{ij}^{BASIC} = 1$ as a benchmark against which we evaluate other types of connections.

Similar to Greenstone et al. (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, $\omega_{ij}^{\text{SAME}} = 1$ if $k(i) = k(j)$ at the 2-digit NAICS level and 0 otherwise and $\omega_{ij}^{\text{OTHER}} = 1 - \omega_{ij}^{\text{SAME}}$.

Following Ellison et al. (2010), we build input-output weights using the Basic Price version of the 4-digit NAICS 2015 Statistics Canada input-output table. Consistent with other papers in the literature, these weights are defined as the maximum of upstream and downstream input and output shares

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

Finally, we build occupational similarity measures using the 2002 National Industry Occupation Employment Matrix, which is built from Occupational Employment Statistics survey data. This survey is conducted by the US Bureau of Labor Statistics. For each industry, it gives the share of employees in each four-digit occupation. We define our occupational similarity weights as follows:

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

Weights are intended to capture both relative and absolute connectivity between firms. It is for this reason that we scale the linear-in-means weights to sum to $\frac{\sum_{j\neq i}\omega_{ij}}{|M_{b,t}|-1}$ rather than 1 for firm i. In the following section, we discuss results of "horse races" in which IO or OCCSIM weights are run against basic weights in a linear-in-means context. If weights and firm quality are uncorrelated, that is $E\left(w_{ij}\alpha_j\right) \approx E\left(w_{ij}\right)E\left(\alpha_j\right)$, we can interpret spillover parameters in these horse races as follows:

$$\frac{(\gamma^{\text{BASIC}} + \gamma^{\text{W}}\bar{\omega}_{ij}^{\text{W}})}{|M_{b,t}|-1}\sum_{j\neq i}\alpha_j$$

That is, the sign of γ^{W} indicates whether this weight adds to or reduces the mean peer quality spillover. For the case in which we run a horse race between $\omega_{ij}^{\text{SAME}}$ and $\omega_{ij}^{\text{OTHER}}$, the resulting spillover is the weighted average of the two estimated spillover parameters.

5 Results

5.1 Main Estimates

Table 2 presents our main results. The first two rows present estimates of γ^{LIM} and γ^{Agg} for separate and horse race specifications of the model, respectively. The central result is that γ^{LIM} is a statistically significant 0.016 and γ^{Agg} is close to 0 (with a point estimate that is slightly negative). That is, we find robust evidence that linear-in-means type spillovers quantitatively dominate agglomeration spillovers at small spatial scales. As the estimates are identical whether the coefficients on peer group aggregates are estimated separately or simultaneously, we conclude that we have independent identifying variation for the two types of aggregators. This allows us to dig further into the mechanisms driving the linear-in-means results using horse races across weights below.

We can interpret the linear-in-means results in two ways. First, an approximate doubling of average peer quality leads to a 1.6 percent increase in firm revenue. If firms are not price-takers this implies an even greater increase in firm TFP. As the standard deviation in average peer quality is about 1 (at 1.10, Column 3), this is also approximately the impact of increasing peer quality by one standard deviation. Equivalently, this estimate can be interpreted as saying that absent endogenous effects, peers' attributes are 1.6% as important as a firm's own attributes for determining revenue – with a greater fraction for TFP.[6]

Table 2 Column 4 shows the implied average fraction of revenue explained by spillovers. It is calculated as Column 1 multiplied by Column 2. We emphasize that the resulting 19% is well outside of the support of the data. Our assignment of the constant to α_i rather than one of the other fixed effects means that average peer quality is not below 7 for any observation in our data. We report this number as a benchmark for comparison across linear-in-means specifications rather than as an indicator of the fraction of revenue we infer comes from spillovers. Column 5 reports the implied difference in the fraction of revenue accounted for by spillovers in the 90th percentile firm relative to the 10th percentile firm. This 90-10 gap of 4.6 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. Moreover, we find treatment effects that are increasing in firm quality (unreported).

The near 0 agglomeration estimates reported in Columns 6-10 should be viewed in the context of the inclusion of 500 meter radius region-year fixed effects.[7] 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

[6]To see this result, consider the following data generating process from the peer effects literature, as in Gibbons et al. (2015), expressed for one cross-section:

$$y = X\beta_x + \mathbf{W}X\delta_x + u \tag{6}$$

This is the linear-in-means equal weights model with M members of the peer group. $\mathbf{W}$ is an MXM matrix in which each off-diagonal element is $\frac{1}{M-1}$ and the diagonal elements are 0. Connecting this equation to our conceptual framework, relabel $X_i\beta_x$ as α_i so $\gamma \equiv \frac{\delta_x}{\beta_x}$. Because γ is fully identified through variation in peer group composition, any potential endogenous effects would be transitory and be part of the error term. Therefore, if peer group composition is uncorrelated with the error term, γ does not include endogenous effects.

[7]Preliminary bootstrapped standard error estimates indicate that both estimation and robustness sample agglomeration results are marginally significant at the 5% level.

economies, knowledge transfers may be more likely to occur as a function of average than aggregate peer group quality.

Results in the third and fourth rows of Table 2 are estimated off of a broader sample that includes firms with missing revenue information in the middle of their individual panels. The inclusion of these firms has the advantage of capturing more of the economic activity at the cost of including firms that might have temporarily ceased operation for a few years. Firm-years with no revenue can impart spillovers on other firms (via fixed effects estimated off of nonmissing observations from other years) but not receive spillovers. These robustness sample linear-in-means results are a little bit larger than the main estimation sample results, perhaps because including more potential peers reduces measurement error in peer composition. The associated agglomeration estimate turns slightly positive but remains small at 0.003%. As including all firms may be more important for the agglomeration aggregator, we have more confidence in this slightly positive estimate.

Results in the fifth and sixth rows of Table 2 give a sense of the extent of sorting on location fundamentals. This "simple" specification excludes 500 meter radius-year fixed effects from the estimation equation. The fact that estimates of γ^{LIM} and γ^{Agg}, 0.025 and 0.00005 respectively, are both larger reflects positive sorting of higher quality firms into more productive and denser locations. This echoes the descriptive evidence in Figure 2. 500 meter area fixed effects are key controls to account for such sorting across space.

Results in the final two rows of Table 2 are for a more saturated specification that additionally includes 250 meter radius fixed effects. These controls for fixed attributes of smaller regions do not affect LIM estimates but do push the agglomeration estimates even closer to 0. Based on this and evidence from an alternative robustness specification in which 500 meter-year fixed effects are interacted with two-digit industry, we conclude that we have imposed sufficiently detailed fixed effects to successfully achieve identification.

5.2 Spatial Decay

Table 3 presents results that speak to the extent to which estimates decay spatially. To look at this, we estimate specifications identical to those in Table 2 except with peer group radii extended to 125 or 250 meters. Resulting sample sizes are 26 and 75 percent larger respectively, as larger radii have the potential to include more firms in each group, leading to more groups meeting the 2 high skilled service sector firm minimum threshold. Average peer group size grows from 6.5 for 75 meters to 6.9 at 125 meters and 8.3 at 250 meters radii. We stop at 250 meters in order to maintain enough groups within 500 meter radius fixed effects.

Linear-in-means results in Table 3 show a lot of stability out to 250 meters while agglomeration results are even closer to 0 than for the 75 meter radius. While the estimated spillover parameter is informative, to fully understand the meaning of these results, we also need to account for changes in estimated firm quality. As we expand the peer group size, estimates of α_i tend to fall.[8] This likely comes through a selection effect, that the number of firms in the sample rises as we expand the peer group radius, thereby including firms located in less dense areas. With positive sorting on density, this reduces the average quality of firms in the sample. Putting the spillover parameters together with the average peer group quality, we see in Table 3 Column 4 that the fraction of revenue due to spillovers falls for the average firm with increases in peer group radius. Another way to quantify spatial decay is to evaluate the impact of increasing one peer's α_j by one standard deviation. This calculation results in growth of 0.49 percent within a 75 meter radius and 0.35 percent within the two larger radii. If anything, the agglomeration results show full decay within 125 meters.

Finally, we estimate horse races between peer groups of 250 and 75 meter radii. We operationalize this by running a horse race between aggregators with two different weights. The first assigns a "basic" weight of 1 to all firms. The second assigns a weight of 1 only to those firms in the same 75 meter radius peer group, nested inside the 250 meter radius peer group. These results are in progress.

5.3 Weights

The domination of linear-in-means over agglomeration spillovers leads us to focus only on the former in evaluating causal mechanisms through looking at different weights. Table 4 presents results of horse races between two weights at a time. The first two rows show basic weights against occupational similarity and input-output weights, respectively. The third row shows these two weights against each other. The fourth row looks at own 2-digit industry versus other 2-digit industry. With basic weights set as the first weight, the interpretation of the second spillover parameter is in how much additional impact the associated mechanism has.

Results in the first three rows of Table 4 show that if anything occupational similarity is more likely to be driving the linear-in-means results than are input-output relationships. Relative to a benchmark with equal weights, however, both types of connections have negative additional impacts. Because the two sets of weights are scaled differently, the average treatment impact in Column 9 is more informative than γ^B for quantifying the importance of these different types of connections. Occupational similarity adds -0.02 and input-output relationships add -0.04 to the baseline with equal weights. The horse race between occupational similarity and input-output weights in the third row shows that occupational similarity

[8] $\bar{\alpha}_i$ falls from 11.84 for the 75 meter radius to 11.79 for the 125 meter radius and 11.61 for the 250 meter radius.

wins.

Results in the final row of Table 4 speak to the relative importance of localization and urbanization economies in high skilled services. Here, γ^A and γ^B can be compared, as the two weights are both scaled as dummy variables. These estimates show that firms in other 2-digit industries within high skilled services impart greater spillovers than those in the same 2-digit industry. With about two-thirds of peers being in different 2-digit industries on average, the total spillover is 14 percent from firms in other industries and only 5 percent from firms in the same industry. This evidence is consistent with that in Henderson et al. (1995) that young innovative industries benefit more from cross-industry spillovers and contrasts with evidence for manufacturing in Greenstone et al. (2010).

The picture painted by our evidence on mechanisms is as follows. Some but not all of the benefit firms accrue from proximity is via interactions between their workers. This is likely knowledge transfer but could be simple incentives to work hard, as in Cornelissen et al. (2017)'s evidence about within firm work groups. The potential for new knowledge acquisition is greater between different industries with similar worker task requirements than within the same industry.

6 Firm Sorting and Agglomeration Economies

In this section, we document the extent of sorting on firm quality across space and evaluate the extent to which sorting of firms across space matters for both the distribution of spillovers and aggregate spillovers. We carry out this analysis by applying our spillover and firm quality (α_i) estimates from Table 2 to two counterfactual spatial distributions of firms. In the first counterfactual, we assess the importance of sorting on firm quality while holding density (peer group size) constant. Second, we additionally assess the role of firm density for generating the distribution of spillovers. All exercises use estimates of the spillover parameter γ and firm-specific quality α_i reported in the third row of Table 2. We use estimates from the "robustness" sample because this broader sample is more inclusive and produces better agglomeration results.

Results indicate that while there is positive sorting on firm quality and density, this sorting does not account for a large fraction of the additional aggregate revenue that can be attributed to spillovers. Our calculations indicate that absent any sorting across space, aggregate firm revenue would be 1.4 percent lower, with 0.8 of this coming from agglomeration effects and the remainder coming from linear-in-means effects. Half of this total effect is from variation in firm density across space and half is from sorting conditional on peer group size. The aggregate importance of sorting is similar for linear-in-means and agglomeration type spillovers. That is, as a fraction of the treatment effect of spillovers, sorting plays a much larger role for agglomeration than linear-in-means spillovers. But the aggregate impact of sorting is small. As a result, we conclude that sorting mostly occurs on location fundamentals: higher quality locations (those with larger $\phi_{B(b),t}$) attract higher quality firms. But the higher quality firms impart only slightly larger spillovers on their neighbors.

Figure 3 documents the relationship between estimated firm quality $\widehat{\alpha_i}$ and the size of the treatment imparted through spillovers enjoyed by firm i. Panel A shows the demeaned linear-in-means treatment $\frac{\widehat{\gamma^{\text{LIM}}}}{1-|M_{b,t}|}\sum_{j\subseteq M_{b,t},\neq i}\widehat{\alpha_j^{LIM}}$ and Panel B shows the demeaned agglomeration treatment $\widehat{\gamma^{\text{Agg}}}\sum_{j\subseteq M_{b,t},\neq i}\widehat{\alpha_j^{Agg}}$. Results in Panel A show a monotonic and near linear relationship between average peer quality and firm quality. The lowest quality firm receives a treatment that is about 0.02 less than the highest quality firm on average. With a standard deviation in peer quality of 1.1 (from Table 2), this is less dispersion than can be explained from all the variation in the data in peer quality. Panel B shows much less dispersion in agglomeration treatment impacts across firms of different quality. This profile is much more bimodal, with below average quality firms typically getting about 0.002 less in treatment than above average quality firms. This bimodalism reflects differences in the spatial distribution of firms. The broader message is that higher quality firms do benefit from greater spillovers by sorting into areas with other higher quality firms. But the associated magnitudes are small.

Figure 4 presents counterfactual treatment distributions relative to actual treatment distributions under our two sorting scenarios. For Counterfactual 1, we randomly allocate firms to fixed locations across space. 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. For Counterfactual 2, we additionally impose that all peer groups are of the same size. This allows us to see the impact of variations in density on aggregate spillovers. Results in panel A show two almost perfectly symmetric distributions around 0. This means that about half of firms experience increased linear-in-means spillovers and half of firms experience reduced such spillovers absent sorting across space. Getting rid of dense concentrations increases the variance in these impacts a bit but otherwise has little impact. Figure 4 Panel B shows different counterfactual effects under aggregate spillovers. While given fixed locations, there is a symmetric impact of randomly shuffling firms, the impact of getting rid of dense concentrations (Counterfactual 2) is mostly negative.

Results in Figure 5 display how these impacts of sorting differ by firm quality. The linear-in-means results in Panel A show that the lowest quality firms benefit the most, by up to 0.01, by imposing random peers. This impact declines approximately linearly with firm quality down to a negative impact of -0.01 for the highest quality firms. These magnitudes are very similar to the relative relationships between firm quality and average peer quality in Figure 3. Under Counterfactual 1, Figure 5 Panel B shows very small impacts with expected signs under agglomeration models. Once again, low quality firms benefit and high quality firms are hurt by randomly shuffling all firms across fixed locations. However, flattening the density gradient is bad for all firms, though it is worse for the highest quality firms. Under this second counterfactual, average firm revenue declines by about 0.4%-0.6% across the full distribution of firm quality.

The aggregate impact of sorting on revenue through spillovers is small. We compute this by aggregating revenue under the two counterfactual scenarios and comparing it to total observed firm revenue. We carry out this calculation as follows for each counterfactual c and peer group M_i:

$$\ln Y^c = \ln\left[\sum_i exp(y_i + \gamma[M_i^c - M_i])\right]$$

That is, we calculate aggregate revenue in the counterfactual environment in which peer group quality M_i is replaced by peer group quality M_i^c. This way of calculating impacts of sorting is not sensitive to the normalization of firm fixed effects, as any normalization differences out in $M_i^c - M_i$. The top row of Table 5 shows the linear-in-means version, where $M_i^{LIM} = \frac{1}{1-|M_{b,t}|}\sum_{j\subseteq M_{b,t},\neq i}\widehat{\alpha_j^{LIM}}$ and the second row shows agglomeration version, where $M_i^{LIM} = \sum_{j\subseteq M_{b,t},\neq i}\widehat{\alpha_j^{Agg}}$. 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. That is, this integrates plots in Figure 5 over the distribution of firm quality to determine aggregate impacts. Given our evidence that linear-in-means and agglomeration effects are additive, we add up the total impact in the bottom row.

Results show that under Counterfactual 1, aggregate revenue would decrease by 0.38% through linear-in-means channels and 0.33% through agglomeration channels. This reflects mild positive sorting of higher quality firms into peer groups. Harmonizing the size of each peer group to be the same reduces revenue by an additional 0.24% through linear-in-means channels and 0.47% through agglomeration channels. The linear-in-means result reflects the mean-preserving spread seen in Figure 4. The agglomeration result reflects the reduction in peer group size experienced by the average firm. The total impact is 1.4 percent of revenue.

7 Conclusions

Considerable evidence from metropolitan area level spatial scales exists on the magnitude of aggregate increasing returns to scale. Yet little empirical evidence exists at microgeographic spatial scales. Evidence in this paper shows that firms benefit from being near higher quality peers. In particular, the elasticity of firm revenue and TFP with respect to the average quality of other firms within 75 meters is 0.016-0.018. However, after conditioning on 500 meter radius areas, the average firm benefits at most only marginally from being surrounded by a greater amount of economic activity within 75 meters. That is, 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. However, because of the huge amount of dispersion in density, firms in the most dense locations benefit considerably from being in these locations.

Occupational similarity relationships are more important drivers of linear-in-means spillovers than input-output relationships at this small spatial scale. However, spillovers are slightly larger from other 2-digit industries. This is evidence that knowledge transfer may be more valuable within occupations but across industries.

With metro level elasticities of TFP with respect to population estimated to be in the 0.03-0.05 range (Combes and Gobillon, 2014), additional mechanisms are required to go from our micro evidence to the macro evidence. 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 operate at higher levels of aggregation. An important question for future research is how microgeographic estimates like those reported here aggregate up to the local labor market level.

References

Ackerberg, D., K. Caves, and G. Frazer (2016): "Identification Properties of Recent Production Function Estimators," *Econometrica*.

Angrist, J. D. (2014): "The perils of peer effects," *Labour Economics*, 30, 98–108.

Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012): "Estimating spillovers using panel data, with an application to the classroom," *Quantitative Economics*, 3, 421–470.

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

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

Bazzi, S., A. V. Chari, S. Nataraj, and A. D. Rothenberg (2017): "Identifying Productivity Spillovers Using the Structure of Production Networks," Working Papers WR-1182, RAND Corporation.

Bloom, N., M. Schankerman, and J. V. Reenen (2013): "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81, 1347–1393.

Combes, P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection," *Econometrica*, 80, 2543–2594.

Combes, P.-P. and L. Gobillon (2014): "The Empirics of Agglomeration Economies," IZA Discussion Papers 8508, Institute for the Study of Labor (IZA).

Cornelissen, T., C. Dustmann, and U. Schonberg (2017): "Peer Effects in the Workplace," *American Economic Review*, 107, 425–56.

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

De Loecker, J. (2011): "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity," *Econometrica*, 79, 1407–1451.

De Loecker, J. and J. Eeckhout (2018): "Global Market Power," .

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

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

Faggio, G., O. Silva, and W. C. Strange (2017): "Heterogeneous Agglomeration," *The Review of Economics and Statistics*, 99, 80–94.

Gandi, A., S. Navarro, and D. Rivers (2020): "On the Identification of Gross Output Production Functions," *Journal of Political Economy.*

Gibbons, S., H. G. Overman, and E. Patacchini (2015): *Spatial Methods*, Elsevier, vol. 5 of *Handbook of Regional and Urban Economics*, chap. 0, 115–168.

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

Hanlon, W. W. and A. Miscio (2017): "Agglomeration: A long-run panel data approach," *Journal of Urban Economics*, 99, 1–14.

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

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

Liu, X., E. Patacchini, and Y. Zenou (2014): "Endogenous Peer Effects: Local Aggregate or Local Average," *Journal of Economic Behavior and Organization*, 103, 39–59.

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

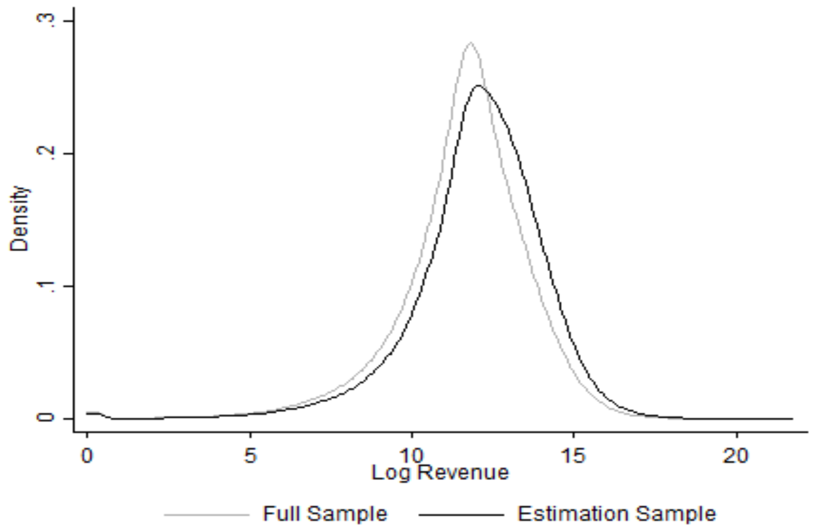
Manning, A. (2009): "The Plant Size-Place Effect: Agglomeration and Monopsony in Labour Markets," *Journal of Economic Geography*, 10, 717–744.

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

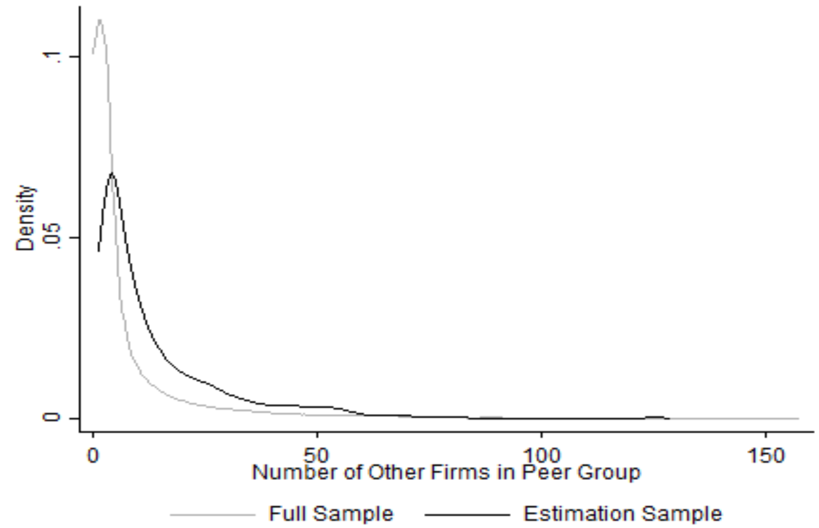
Moretti, E. (2004): "Human capital externalities in cities," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson and J. F. Thisse, Elsevier, vol. 4 of *Handbook of Regional and Urban Economics*, chap. 51, 2243–2291.

Serafinelli, M. (2013): "Good Firms, Worker Flows and Productivity," Mpra paper, University Library of Munich, Germany.

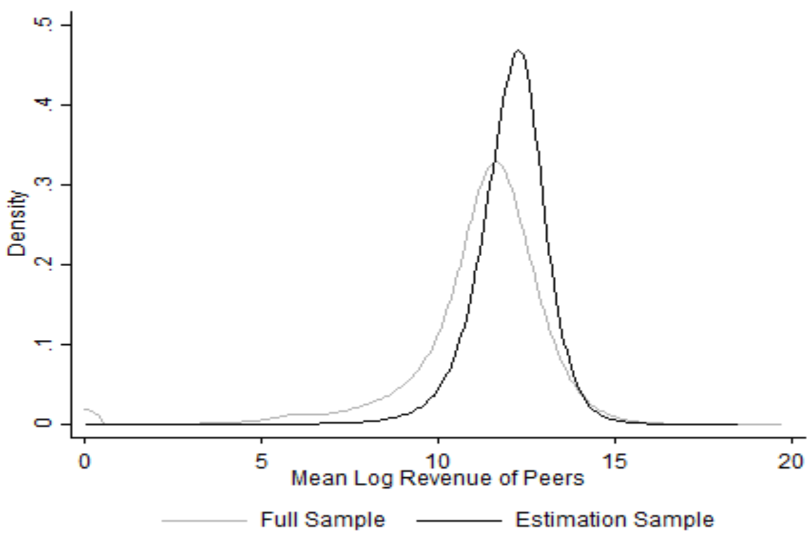
Figure 1: Descriptive Graphs



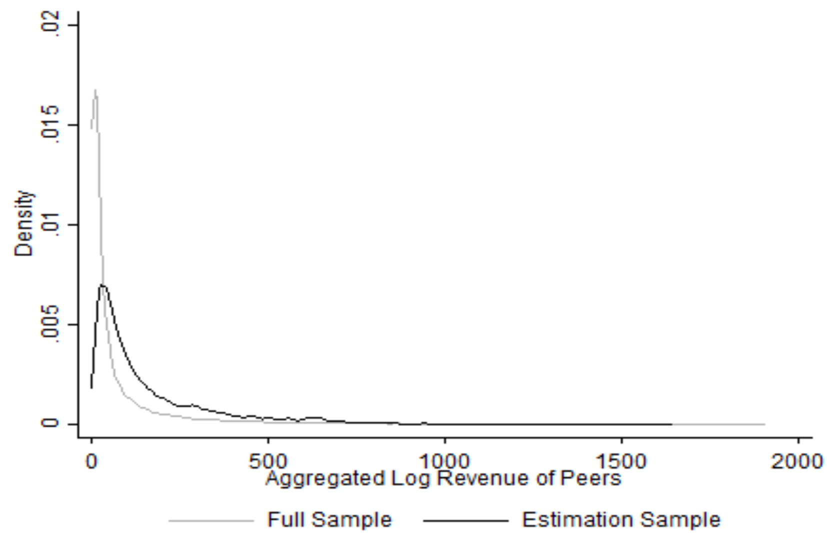
Panel A: Distribution of Firm log Revenue



Panel B: Distribution of Peer Group Size

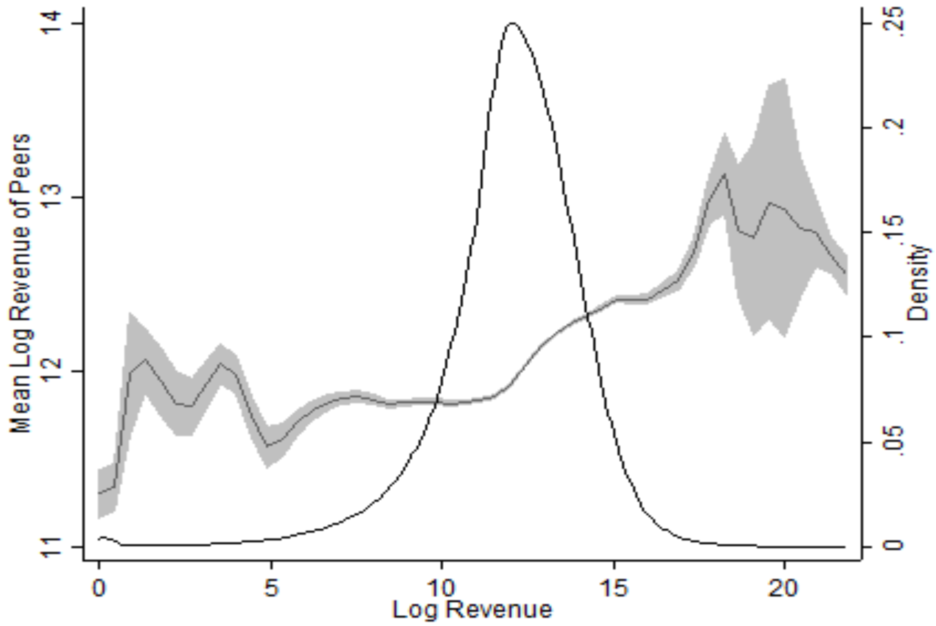


Panel C: Distribution of Mean Log Revenue in Peer Groups

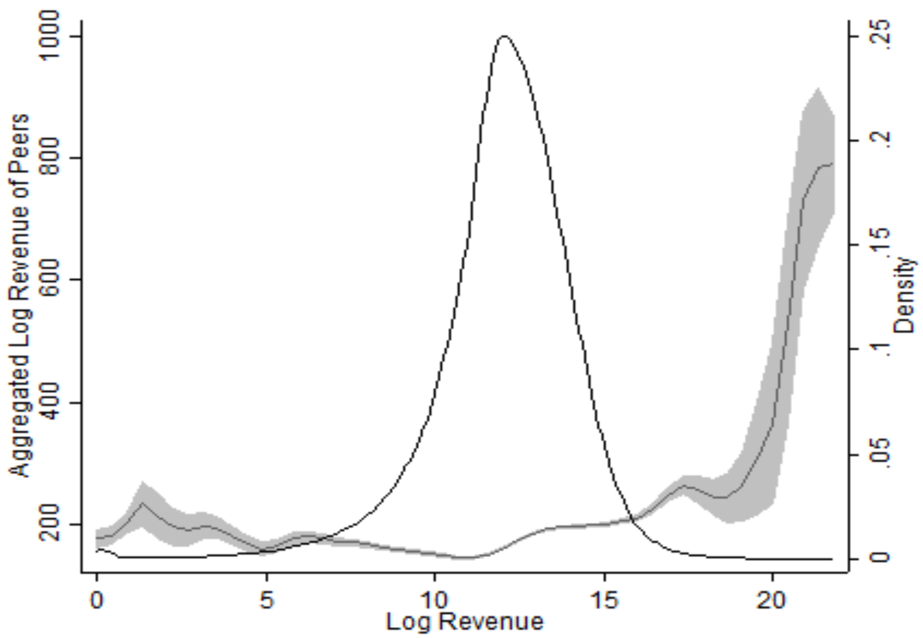


Panel D: Distribution of Aggregate Log Revenue in Peer Groups

Figure 2: Sorting on Peer Group Quality



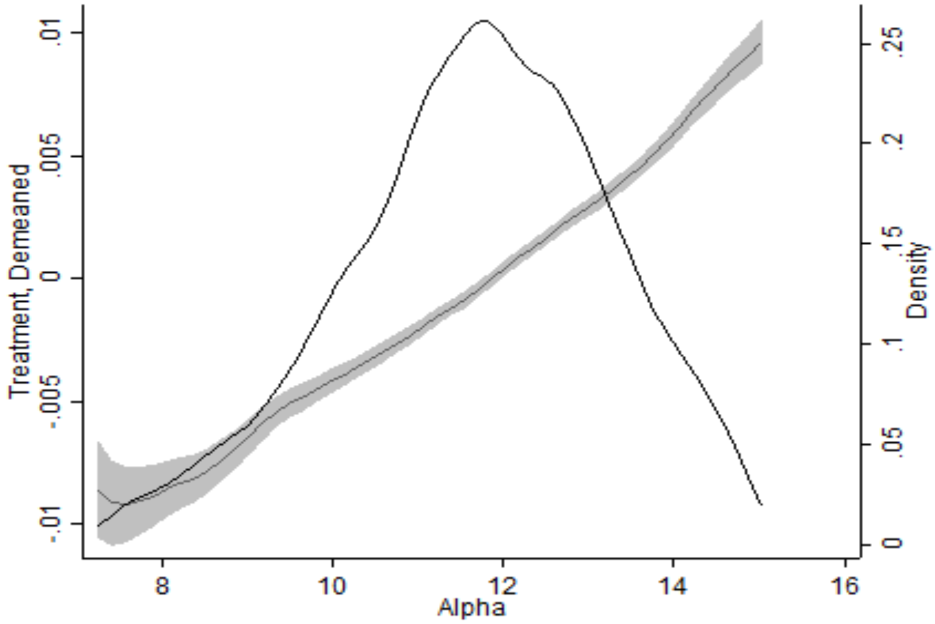
Panel A: Mean Peer log Revenue by Firm log Revenue



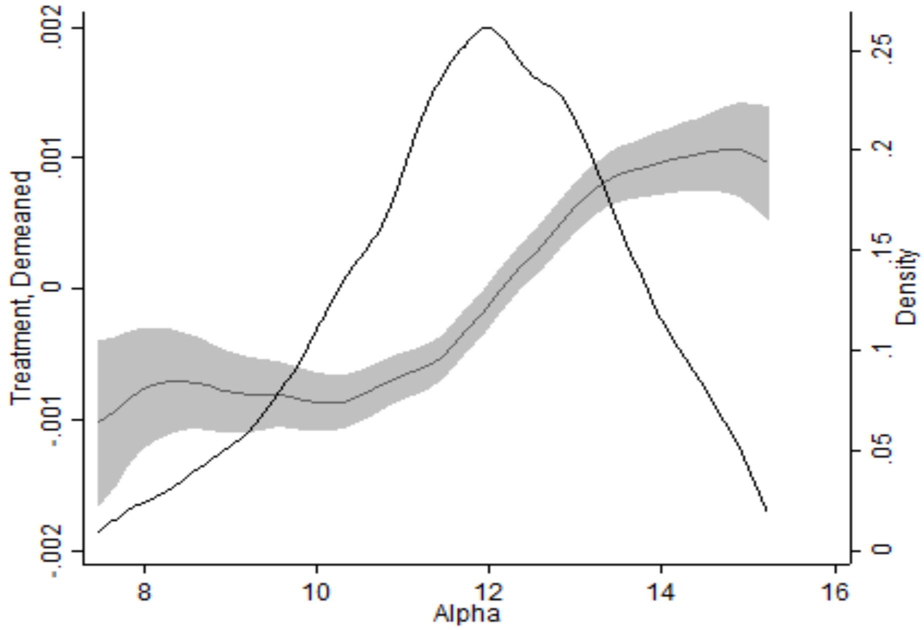
Panel B: Aggregate Peer log Revenue by Firm log Revenue

These graphs use the estimation sample.

Figure 3: Relationships Between Estimated Firm and Peer Group Quality



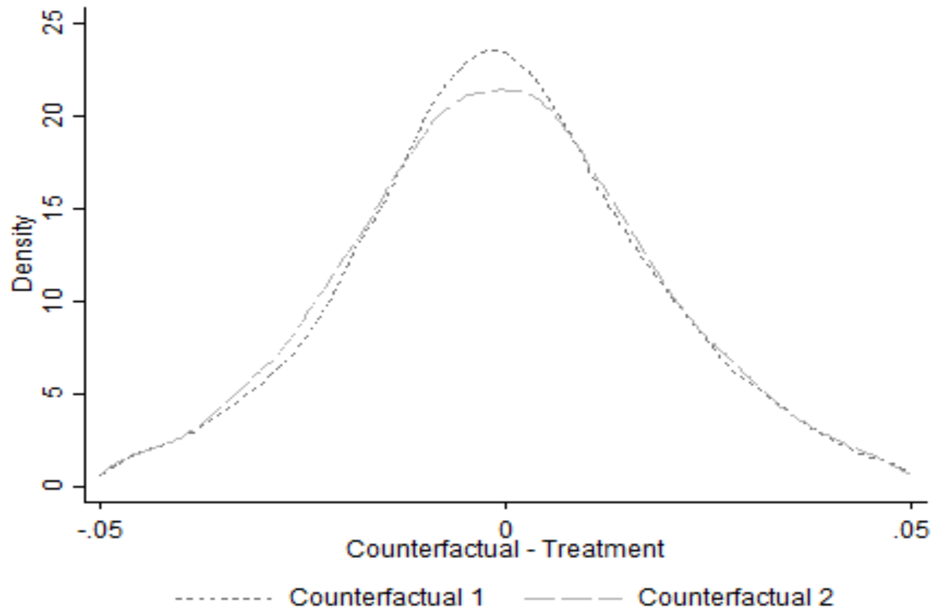
Panel A: Linear in Means Model



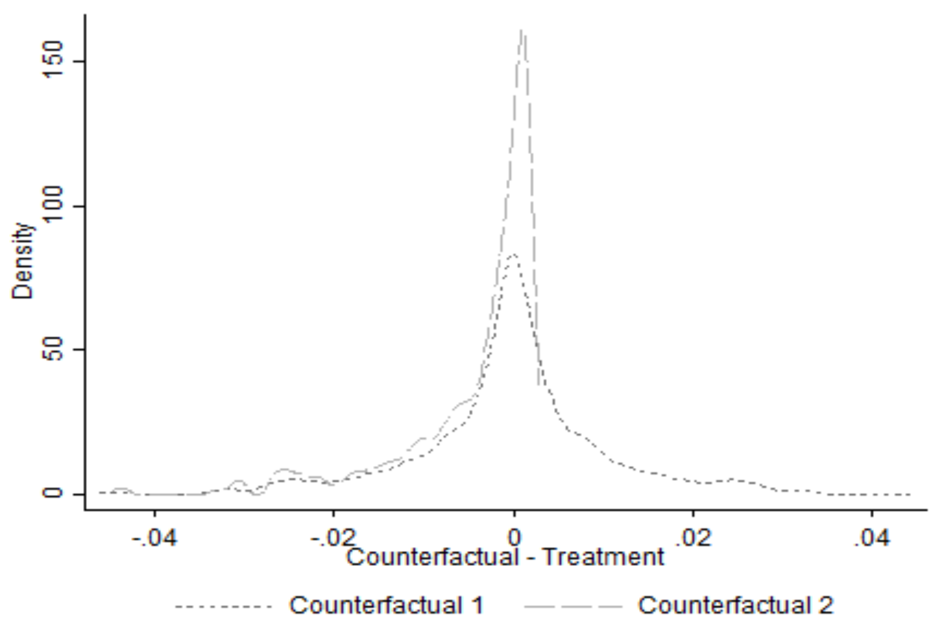
Panel B: Agglomeration Model

Graphs are based on the robustness sample results in Table 2.

Figure 4: Impacts of Sorting on Treatment Magnitudes



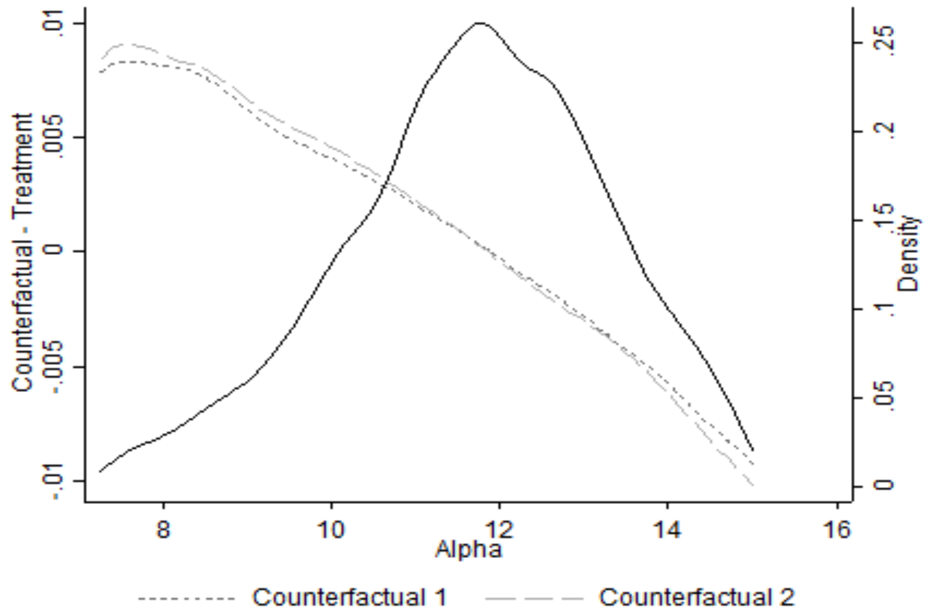
Panel A: Linear in Means Model, Robustness Sample



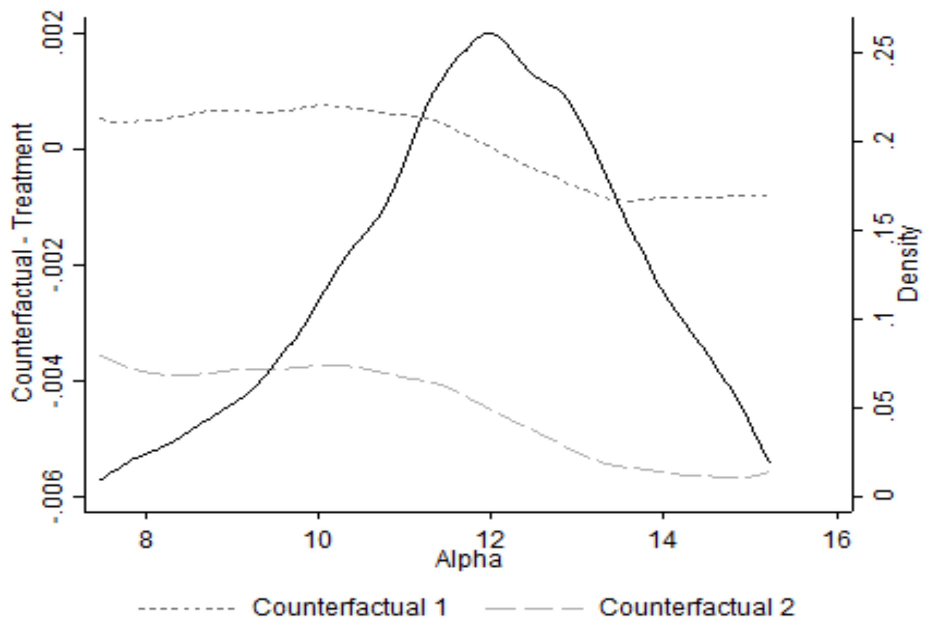
Panel B: Agglomeration Model, Robustness Sample

Graphs are based on the robustness sample results in Table 2.

Figure 5: Impacts of Sorting by Firm Quality



Panel A: Linear in Means Model, Robustness Sample



Panel B: Agglomeration Model, Robustness Sample

Graphs use the robustness sample results in Table 2.

Table 1: Descriptive Statistics

	All Industries		High-Skilled Services (NAICS 5)		
	Multi Loc	Single Loc	Multi Loc	Single Loc	
					Est. Sample
Panel A: Summary Statistics					
ln Revenue	15.06	12.05	14.50	11.60	12.03
	(2.21)	(1.98)	(2.42)	(2.03)	(2.03)
ln Payroll Per Worker	10.65	10.13	10.80	10.29	10.47
	(0.75)	(0.92)	(0.87)	(0.99)	(0.95)
ln Employment	3.11	1.18	2.73	0.92	1.09
	(1.56)	(1.01)	(1.77)	(0.94)	(1.00)
Area of Postal Code	0.166	0.111	0.050	0.090	0.006
(sq km)	(12.895)	(11.313)	(0.873)	(10.280)	(0.005)
Panel B: Sample Sizes					
# of Firm-Years (Obs)	245 517	2 645 291	78 484	1 075 672	216 704
# of Firms	30 464	428 377	10 643	181 496	44 830
# of Postal Code-Years	153 594	1 100 446	55 075	603 583	64 779
# of Peer Group-Years	128 284	843 305	47 562	501 458	33 580
Avg # in Peer Group	1.9	3.1	1.7	2.1	6.5

Statistics are for all firms in the Montreal, Toronto and Vancouver CMAs for the 2001-2012 period. Panel A shows means with standard deviations in parentheses. The estimation sample in the final column only includes firms in postal codes with areas less than 0.018 sq km ($\pi 75^2$ sq m) and in peer groups of at least 2 firms. This sample additionally excludes firms with missing revenue in some years surrounded by nonmissing revenue in other years. All samples drop firm-year observations in which revenue is missing at the beginning or end of the firm's panel.

Table 2: Results for 75 meter Radius Peer Groups

			Linear in Means					Agglomeration				
Sample	γ^{LIM} & γ^{Agg} Estimated …	Specification	γ^{LIM} (1)	Mn Avg Peer α (2)	SD Avg Peer α (3)	Avg. Treat (4)	90-10 Diff (5)	γ^{Agg} (6)	Mn Agg Peer α (7)	SD Agg Peer α (8)	Avg. Treat (9)	90-10 Diff (10)
Est.	Separately	Base	0.016	11.84	1.10	0.19	4.6%	-0.00003	170.8	202.1	0.00	-1.4%
Est.	Jointly	Base	0.016	11.84	1.10	0.19	4.6%	-0.00003	168.1	198.9	-0.01	-1.7%
Rob.	Separately	Base	0.018	11.68	1.11	0.21	5.1%	0.00003	190.7	236.7	0.01	1.9%
Rob.	Jointly	Base	0.018	11.68	1.11	0.21	5.0%	0.00003	187.5	232.7	0.01	1.7%
Est.	Separately	Simple	0.025	11.74	0.94	0.29	5.9%	0.00005	170.5	202.5	0.01	2.6%
Est.	Jointly	Simple	0.024	11.74	0.94	0.29	5.9%	0.00005	166.5	197.8	0.01	2.3%
Est.	Separately	Saturated	~0.017					0.00000	170.4	201.2	0.00	0.2%
Est.	Jointly	Saturated	~0.017					~0				

Sample sizes for the estimation (Est.) sample are reported in Table 1. The robustness (Rob.) sample (269,144 observations in 42,110 peer group-years) additionally includes firms with missing revenue observations in the middle of their panels. The base specification has 500 meter radius-year and 2-digit industry-year FE. The Simple specification (217,756 observations in 33,888 peer group-years) drops the 500 meter radius-year FE. The saturated specification (216,634 observations in 33,566 peer group-years) instead adds 250 meter radius FE.

Table 3: Results for Other Peer Group Radii, Base Specification

Peer Group Radius	Sample	γ^{LIM} & γ^{Agg} Estimated ...	Linear in Means γ^{LIM} (1)	Avg Peers (2)	SD Peers (3)	Avg. Treat (4)	90-10 Diff (5)	Agglomeration γ^{Agg} (6)	Avg Peers (7)	SD Peers (8)	Avg. Treat (9)	90-10 Diff (10)
75 m	Est.	Jointly	0.016	11.84	1.10	0.19	4.6%	-0.00003	168.1	198.9	-0.01	-1.7%
125 m	Est.	Jointly	0.013	11.79	1.03	0.15	3.4%	-0.00001	208.6	280.5	0.00	-0.9%
250 m	Est.	Jointly	0.016	11.61	0.92	0.18	3.7%	0.00001	352.6	626.0	0.00	1.2%
125 m	Est.	Separately	0.013	11.79	1.03	0.15	3.3%	-0.00001	211.2	284.0	0.00	-0.7%
125 m	Rob.	Separately	~0.014					0.00001	235.8	332.4	0.00	1.1%
250 m	Est.	Separately	0.016	11.61	0.92	0.19	3.8%	0.00001	358.1	635.6	0.00	1.5%

Estimates are analogous to those in Table 2, with the only difference being the peer group definitions. Radii listed at left indicate the peer group definition for each set of results. Sample sizes are 5-10 percent larger than those in Table 2.

Table 4: Comparison of Different Types of Firm Connections for LIM Results

Weight A	γ^A (1)	Avg Peers (2)	SD Peers (3)	Avg. Treat (4)	90-10 Diff (5)	Weight B	γ^B (6)	Avg Peers (7)	SD Peers (8)	Avg. Treat (9)	90-10 Diff (10)
Equal	0.018	11.84	1.10	0.21	5.0%	Occ Sim	-0.004	3.82	2.26	-0.02	-2.6%
Equal	0.020	11.83	1.10	0.23	5.5%	IO	-0.113	0.31	0.32	-0.04	-9.3%
Occ Sim	0.007	3.89	2.30	0.03	4.1%	IO	-0.129	0.32	0.33	-0.04	-10.8%
Same 2-Digit	0.013	3.99	3.51	0.05	11.9%	Other 2-Dig	0.018	7.85	3.58	0.14	16.4%

Note: Results in each row are estimated jointly. All results use the main estimation sample and 75 meter radius peer groups. Summary statistics and sample sizes are in Table 1. Weighted results for aggregate spillovers are not shown as they are inconclusive.

Table 5: Aggregate Revenue Reductions
Absent Sorting

	Counterfactual	
Model	1	2
Linear in Means	0.38%	0.62%
Agglomeration	0.33%	0.80%
Total	0.71%	1.42%

Appendix 1 The Monopolistic Competition Case

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 by De Loecker (2011). In this environment, CES preferences across firm-specific varieties within 2-digit industry yield industry-specific demand elasticities 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, $\ln X_{k,t} = \ln Q_{k,t} - \frac{1}{\eta_k} \ln P_{k,t}$ represents a combination of industry-time specific demand shocks and the industry-time price index.[9] These objects will be controlled for with fixed effects. η_k is the demand elasticity faced by each firm in industry k for its product. We note that all derivations in this section apply even if η is firm-specific. $\zeta_{i,b,k,t}$ 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:

$$\begin{aligned} \ln p_{i,b,k,t} = &- \frac{1}{D_k} \ln A_{i,b,k,t} + \frac{\theta_k}{D_k} \ln w_{B(b),k,t} - \frac{\theta_k}{D_k} (\ln \frac{1+\eta_k}{\eta_k} + \ln \theta_k) \\ &+ \frac{1-\theta_k}{D_k} [\ln X_{k,t} + \zeta_{i,b,k,t}] \end{aligned} \tag{7}$$

where $D_k = \theta_k(1+\eta_k) - \eta_k$. As η_k approaches negative infinity, $\ln p_{i,b,k,t}$ goes to $\ln P_{k,t}$ by construction, meaning that firms have no market power. Because η_k is always less than -1 for monopolists and $\theta_k < 1$, the common denominator D_k is always positive. Therefore, 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 smaller the pass-through of positive productivity shocks to price discounts. Similarly, positive wage shocks and positive demand shocks get passed through to increased prices.

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 (7) into this condition delivers the following general expression for revenue, which is also

[9] If preferences are not CES over varieties within industry, we can instead think of $X_{k,t}$ as representing a reduced form demand shifter that is common to all products in industry k at time t.

applicable under perfect competition (when $\eta_k = -\infty$)

$$\begin{aligned}\ln R_{i,b,k,t} = &-\frac{1+\eta_k}{D_k}\ln A_{i,b,k,t} + \frac{\theta_k(1+\eta_k)}{D_k}\ln w_{B(b),k,t} \\ &- \frac{\theta_k(1+\eta_k)}{D_k}(\ln\frac{1+\eta_k}{\eta_k} + \ln\theta_k) + \frac{1}{D_k}[\ln X_{k,t} + \zeta_{i,b,k,t}].\end{aligned} \tag{8}$$

If the firm is a price taker, this expression matches (2) with no change in price by l'Hopital's Rule. As demand for the firm's product becomes less elastic, revenue becomes more depressed 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. With $\eta_k = -10$ instead, the associated TFP increase is 4 percent, close to the 3.3 percent response under perfect competition. That is, as firms gain market power, between firm TFP spillovers measured through revenue shocks are likely to increase.

A.1.1 Structural Interpretation of Revenue Spillovers

Recall that our baseline estimation equation (4) takes the following form

$$\ln R_{i,b,k,t} = \alpha_i^R + \phi_{B(b),k,t}^R + \gamma^R\left[\sum_{j\in M_{b,t},\neq i}\omega_{ij}(M_{b,t})\alpha_j^R\right] + \varepsilon_{i,b,k,t}^R.$$

Inserting equation (3), the data generating process for firm i's TFP at time t, into the generalized total revenue equation (8) delivers the structural interpretation of each parameter in the baseline estimation equation above.

As discussed in Section 2, under perfect competition, the structural interpretation of the local area-industry-year fixed effects are

$$\phi_{B(b),k,t}^{R,pc} = \frac{\theta_k}{1-\theta_k}\ln\theta_k + \frac{1}{1-\theta_k}\ln p_{B(b),k,t} - \frac{\theta_k}{1-\theta_k}\ln w_{B(b),k,t} + \frac{1}{1-\theta_k}\phi_{B(b),k,t}^A.$$

Under monopolistic competition, they are interpreted as

$$\begin{aligned}\phi_{B(b),k,t}^{R,mc} = &\frac{-(1+\eta_k)}{D_k}\phi_{B(b),k,t}^A + \frac{\theta_k(1+\eta_k)}{D_k}\ln w_{B(b),k,t} - \frac{\theta_k(1+\eta_k)}{D_{k,t}}(\ln\frac{1+\eta_k}{\eta_k} + \ln\theta_k) \\ &+ \frac{1}{D_k}\ln X_{k,t},\end{aligned}$$

where $D_k = \theta_k(1+\eta_k) - \eta_k > 0$ as above. These fixed effects capture location and industry fundamentals, spatial variation in variable input prices, industry-specific markups, the industry-specific production technology, and industry demand conditions, respectively.

The structural interpretation of α_i^R is determined jointly by the firm-specific fixed effect term and the spillover term. If the firm-specific fixed effect in (4) is set to $\alpha_i^R = \frac{-(1+\eta_{k(i)})}{D_{k(i)}}\alpha_i^A$, the remaining terms in (4) are

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

If firm i is in the same industry as all its peers, revenue spillovers γ^R directly measure TFP spillovers γ^A.

Appendix 2 Alternative Specifications

We employ two strategies to accommodate differences within peer groups in markups and input shares in order to recover estimates of TFP spillovers.

Our first strategy divides both sides of (8) by $\frac{1+\eta_k}{\eta_k-\theta_k(1+\eta_k)}$ to build the adjusted revenue measure $\ln \tilde{R}_{i,b,k,t}$ for use as an outcome. Inserting (3) into (8), we have the following alternative structural estimation equation:

$$\ln \tilde{R}_{i,b,k,t} = \alpha_i^A + \tilde{\phi}_{B(b),k,t} + \gamma^A \left[\sum_{j\in M_{b,t},\neq i} \omega_{ij}(M_{b,t})\alpha_j^A\right] + \tilde{\varepsilon}_{i,b,k,t}. \quad (9)$$

This adjustment allows us to isolate firm fixed effects as the permanent firm-specific component of TFP and the TFP spillover parameter γ^A is then directly estimated.

The new structural interpretation of the fixed effects in (9) is

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

and the error term in (9) is

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

Following De Loecker and Eeckhout (2018), we calculate the industry level markup in the data using $\frac{\eta_k}{1+\eta_k} = \theta_k \frac{R_k}{(wL)_k}$, where θ, R and wL are aggregated from firms to the industry level. As a second alternative strategy, we use a direct measure of firm-year TFPR as an outcome. This strategy has the disadvantage of not separating out impacts on prices from quantities, thereby delivering underestimates of TFP spillovers.

A.2.1 Measuring Factor Shares, Markups, and TFP

Our robustness analysis that explicitly accounts for firm-specific price endogeneity requires measures of variable factor shares and markups. We construct information on payments to labor, materials, capital, and real estate, where we treat labor and materials as variable factors.

Payments to labor and materials are observed directly in the data. We infer payments to capital as rental and repair costs plus the book value of capital (net of amortization) times a discount rate plus 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 nonzero for structures only and is reported by Statistics Canada for each 2-digit sector. The capital gains rate uses the CMA level Teranet residential home price index.

We calculate the 2-digit industry-specific markup as

$$\frac{\eta_k}{1+\eta_k} = \theta_k \frac{R_k}{(wL)_k}.$$

We calculate the output elasticity with respect to factor f, θ_k^f, by factor costs across all firms in each 2-digit industry bin, where the variable factor share θ_k is calculated as $\theta_k^{materials}+\theta_k^{labor}$. Aggregate revenue R_k and payments to labor and materials $(wL)_k$ are observed directly in the data. Here, η_k is the demand elasticity faced by firms in industry k.[10]

We calculate TFPR as the following residual

$$\text{TFPR}_{i,b,k,t} = \ln R_{i,b,k,t} - \sum_f \theta_k^f (w^f F^f)_{i,b,k,t}$$

where $w^f F^f$ is the payments to factor f. We choose this method for estimating TFP rather than more sophisticated ones, as in Ackerberg et al. (2016) or Gandi et al. (2020), because we want to allow for year to year changes in semi-flexible inputs as a result of changes in peer group composition. Critically, this measure includes prices and thus we expect it to respond less than revenue to increases in TFPQ.

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