# Local Human Capital Spillovers*

Nathaniel Baum-Snow†
*University of Toronto*

Nicolas Gendron-Carrier‡
*McGill University*

Ronni Pavan§
*University of Rochester*

March 26, 2026

Abstract: Using 2001-2018 administrative tax records for nearly all adult residents in the Toronto metropolitan region, this paper estimates the profiles of earnings and labour market participation spillovers that operate between neighbours in their residential locations. Using a high dimensional fixed effects framework, estimates relate annual earnings or employment status to various aggregations of fixed components of these same outcomes for neighbours within 75 meters. We find evidence of small average spillovers with considerable heterogeneity across individual and neighbourhood attributes. While the average elasticity of earnings with respect to its average fixed component among peers is 0.025, spillovers can be more than three times larger when imparted on low human capital individuals. In contrast, while above median human capital individuals impart positive spillovers, they do not benefit from being surrounded by more highly skilled peers. A model of labour supply and human capital accumulation shows how larger estimated peer effects for employment outcomes can be interpreted as reservation wage spillovers. Both earnings and reservation wage spillovers exhibit very rapid spatial decay and the former exhibits rapid temporal decay.

*Grant Benjamin provided superb research assistance. We thank audiences at CREI; Helsinki GSE; JKU Linz; the London School of Economics (Geography); RFBerlin; the Universitat de Barcelona; the University of British Columbia; the Canadian Summer Conference in Real Estate and Urban Economics (Rotman School of Management); and the Workshop on Structural Transformation, Urbanization, and Development (Collegio Carlo Alberto) for helpful comments. Financial support from the David Feldman Centre for Real Estate and Urban Economics and the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

†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

The nature of externalities that operate between neighbours in residential locations influences the observed spatial distribution of demographic and income groups across neighbourhoods, the optimal design of spatially and demographically targeted policies, and the rate of human capital accumulation in the economy. Conceptual work from Schelling (1971) to Cutler et al. (1999) demonstrates how direct preferences over the composition of neighbours can lead to high levels of residential segregation. More recent quasi-experimental work estimates the nature and magnitudes of local externalities ("neighbourhood effects") for various population sub-groups, including those incentivized to move out of public housing (Kling et al., 2007; Chetty et al., 2016), those forced to move (Chyn, 2018), and cross-metro area or census-tract migrant children (Chetty and Hendren, 2018; Chetty et al., 2026). These studies identify the composite treatment effects of different types of locations on their residents, rather than direct spillovers in earnings or human capital between nearby residents.

Conclusions in studies that quantify general equilibrium models using aggregate neighbourhood level data to evaluate the welfare consequences of housing and place-based policies depend crucially on the forms and magnitudes of spillovers that operate across individuals ("endogenous amenities"). As they take forms that are necessarily tractable and model-specific, these studies must estimate or calibrate simple specifications of neighbourhood or labour-market level endogenous amenities that operate across demographic groups (Diamond, 2016; Almagro and Dominguez-Iino, 2025; Almagro et al., 2026; Weiwu, 2024; Chyn and Daruich, 2025; Fogli et al., 2026a,b). The specification of spillovers determines how spatially targeted policy or infrastructure changes influence group-specific residential sorting and welfare.

While they rely less on model structure for identification, flexible neighbourhood choice demand estimation approaches using microdata also face identification challenges (Bayer et al., 2007; Cook et al., 2023; Davis et al., 2023). As the demographic and income composition of neighbours may be correlated with local unobservables, identification of spillovers as a function of neighbourhood demographic composition can require strong assumptions. With most of the related research effort focused on estimating demographic preferences as revealed through individuals' location choices, no prior study identifies earnings or human capital spillovers in residential locations as distinct from job referral effects (Bayer et al., 2008).

Using panel data on the universe of tax-filing adults living in the Toronto metropolitan region between 2001 and 2018, this paper estimates the magnitudes and profiles of earnings and labour market participation spillovers that operate between neighbours in their

residential locations. Building on prior work that estimates productivity spillovers across nearby firms Baum-Snow et al. (2024), we develop a dynamic framework that allows for the recovery of causal effects of neighbourhood demographic environments on residents' labour market outcomes. Estimation equations relate an individual's labour market outcome, either annual earnings or employment status, to an individual fixed effect, spillovers from the fixed components of this same outcome among current and past non-family members living on the same residential block, and small neighbourhood-by-year fixed effects that absorb contextual effects of neighbourhoods and changes in fundamentals. Residential blocks are defined as microgeographic hexagonal areas with 75-meter sides (approximately $0.015\text{km}^2$) and neighbourhoods are spatial aggregates of blocks forming larger hexagons with sides of approximately 525 meters and an area of $0.72\text{km}^2$. In the context of a model of human capital accumulation and endogenous labour supply with exogenous peer effects, estimated spillover parameters can be interpreted as a combination of labor supply responses, rates of human capital transmission, and reservation wage spillovers between residential neighbours. "Reservation wage spillovers" are peer effects that influence labour supply through channels other than human capital accumulation, including social norms surrounding work or preferences for non-market time. Spillovers are allowed to vary with individual characteristics and neighbourhood attributes. Among other things, our empirical setup allows us to characterize the magnitudes by which lower-income neighbours impart costs or benefits on higher-income counterparts, the extent to which more disadvantaged neighbours benefit from alternative neighbourhood human capital compositions, and the dynamics of these externalities.

In a baseline specification, we estimate the elasticity of annual earnings with respect to the average fixed component of peers' annual earnings to be 0.025. Put differently, peers' exogenous attributes are approximately 2.5 percent as important as an individual's own exogenous attributes in determining earnings. A one standard deviation increase in average peer quality raises own earnings by 0.8 percent, corresponding to 1.1 percent of a standard deviation of earnings. Spillovers in labour force participation are larger in magnitude than earnings spillovers. Within the structure of our model, this difference identifies static reservation wage spillovers, which we estimate to average 0.011. Finally, through inclusion of firm fixed effects, we find no evidence that earnings spillovers operate through neighbour referrals to higher-paying firms.

Building on the baseline estimates, we examine heterogeneity along several dimensions. Temporally, human capital spillovers exhibit limited persistence: we estimate a small one-year lag effect and no evidence of effects at two- or three-year lags. In contrast, reservation wage spillovers display modest persistence, with contemporaneous and one- to three-year lag elasticities of approximately 0.005. Spatially, spillovers decay rapidly with distance. We find

negligible human capital and reservation wage spillovers operating beyond 75-meter hexagonal areas. This rapid spatial decay is consistent with evidence in Oreopoulos (2003), who finds using data from Toronto public housing placements that variation in broader neighbourhood demographic composition has little impact on low-income families' labour market outcomes conditional on more immediate residential peer quality. We also document substantial heterogeneity by observable attributes. Earnings spillovers are larger for men than for women, for immigrants relative to natives, for individuals without children, and for residential movers. They are particularly pronounced in dense and lower-income neighbourhoods.

Most strikingly, we document important non-linear effects across the human capital distribution. Earnings spillover effects vary both with an individual's own position in the distribution and with the composition of peers. Individuals with below median human capital enjoy all of the spillover benefits. For this group, doubling average peer quality increases own earnings by 6.1 percent, holding constant the share of peers above the median. Increasing the fraction above the median from zero to one increases the spillover effect by an additional 2.4 percentage points. These estimates indicate that the nature of earnings spillovers onto lower human capital individuals is convex in average peer group quality. These individuals benefit much more from being exposed to high human capital peers. By contrast, we find little evidence of reciprocal spillovers onto individuals above the median of the human capital distribution.

These results have important implications both for our understanding of residential location equilibria and the design of effective place-based policies. A common assumption in spatial models is log-supermodularity, whereby spillovers are both larger from higher human capital peers and that this effect is bigger when experienced by people with more human capital. While we find evidence for the first component, we do not find support for the second. The log-supermodular force for strong sorting on human capital thus only potentially exists at the lower end of the human capital distribution. If earnings spillovers were the only residential externality, that would tend to push for neighbourhood stratification at the lower end but more income mixing at the higher end of the neighbourhood average human capital distribution. In addition, the fact that lower human capital individuals benefit disproportionately from exposure to higher human capital peers indicates that there are potential aggregate gains from residential integration. For example, the siting of subsidized housing in higher human capital blocks is estimated to have large positive effects on its residents' earnings and employment rates, with no deleterious impact on those of surrounding incumbent workers. Of course, the full welfare implications are still ambiguous given the potential negative local amenity effects (Diamond and McQuade, 2019) and potential difficulty recruiting lower human capital residents into subsidized housing in high income neighbourhoods (Cook

et al., 2023).

This analysis of residential spillovers at a microgeographic scale for the near full adult population of a large labour market relies on population-wide administrative data. A central advantage of these data is the availability of six-character postal codes for place of residence, which provide an unusually fine-grained spatial representation of the population. In denser parts of the city, a postal code typically corresponds to a single building or apartment complex, while in lower-density areas it generally corresponds to a city block face. This spatial resolution is essential for measuring highly local human capital and labour force participation spillovers and for our research design, which exploits variation in neighbour composition within small geographic areas. Unlike many matched employer-employee datasets, the data includes most individuals who are unemployed or out of the labour force, as tax filing is required to claim benefits in Canada. This feature allows us to distinguish between intensive- and extensive-margin labour market responses to changes in neighbourhood demographic composition.

The longitudinal and administrative nature of the data is central to our approach. Following residents over time allows us to recover estimates of the pre-market component of individual human capital through fixed effects and to estimate the dynamic effects of exposure to different residential environments. As in empirical studies that exploit movers to study place effects (e.g., Chetty and Hendren, 2018), identification relies on within-individual variation over time. This structure accommodates long-run sorting of individuals into particular blocks or neighbourhoods, allowing residential location choices to depend arbitrarily on time-invariant individual characteristics. To address dynamic residential sorting in response to changes in location fundamentals or to time-varying shocks to earnings potential, we additionally include detailed neighbourhood-by-year fixed effects, in the spirit of Bayer et al. (2008). Identification therefore exploits within-neighbourhood, cross-block variation in residual peer composition within a given year. Our approach combines two well-established research designs, mover-based identification and within-neighbourhood comparisons, into a unified empirical framework. By integrating these designs, we relax the identifying assumptions required by either approach in isolation. Moreover, the use of population-wide administrative data allows granular insights into the composition of interpersonal spillovers between residential neighbours across the full distribution of households.

Our empirical method draws on Arcidiacono et al. (2012), with extensions as in Baum-Snow et al. (2024) to allow for more flexible forms of peer effects terms in estimation equations. We estimate a flexible dynamic version of the framework developed in Arcidiacono et al. (2012), which allows us to capture exposure effects of neighbourhood composition in an accumulation model. Methodologically, the closest related papers estimate peer effects in

schools or workplaces rather than at residential locations. Using a similar empirical setup to estimate contemporaneous peer effects among coworkers performing routine tasks in Germany, Cornelissen et al. (2017) reports wage elasticities of 0.05 with respect to average peer quality. Similar to our findings, they document smaller wage spillovers among higher human capital work groups. Cornelissen et al. (2017) interpret this pattern as reflecting an effort motive rather than human capital spillovers, consistent with evidence in Mas and Moretti's (2009) estimates of spillovers among grocery checkout workers. In a companion paper (Baum-Snow et al., 2026), we explore sources of learning in the workplace and find limited persistence of peer effects in wages among coworkers.

Considerable observational evidence of homophily and residential segregation by income and race (e.g. Massey and Denton (1988)) has led some to conclude that neighbourhoods have strong influences on lifetime outcomes. Indeed, Chetty et al.'s (2026) Opportunity Atlas data indicate wide variation across nearby neighbourhoods in the opportunity they provide to children. This study and Chetty and Hendren (2018) use variation in the timing of cross-location moves to demonstrate that children exposed longer to areas with higher rates of upward income mobility experience improved labour market outcomes in adulthood. These studies are only able to examine the composite causal role of neighbourhoods or metro areas in a nonparametric way, with no source of identifying variation that can be used to separate out mechanisms explaining why certain neighbourhoods engender greater rates of upward mobility. Instead, we focus on estimating two components of neighbourhood effects that we can observe in our data.

The most well-known (quasi)-experimental evidence on truly local neighbourhood effects comes from the Moving to Opportunity (MTO) experiment. Chetty et al. (2016) demonstrate that, in the long-run, girls benefited from (randomized) incentives to move to lower-poverty neighbourhoods while the evidence for boys is more mixed. Parents had some improvements in mental health, but no detectable labour market consequences (Kling et al., 2007). Chyn (2018) finds evidence from forced moves using public housing demolitions to be largely corroborative. While well-identified, these studies necessarily focus on very disadvantaged sub-populations who, in the case of MTO, choose to accept the offer of moving to a different neighbourhood. There is essentially no well-identified evidence of neighbourhood effects for a broad segment of the population. For this reason, the neighbourhood effects literature has been necessarily disconnected from the place-based literature, including that on inclusionary housing and the siting of new housing developments.

Evaluation of place-specific policies requires a modeling framework that accommodates general equilibrium responses. What is often described as "neighbourhood effects" in the reduced-form literature is represented as "endogenous amenities" in the quantitative spatial

literature. Beyond peer effects, Brueckner (2000) highlights additional sources of local externalities, including reduced access to open space, increased traffic congestion, and higher local service provision costs. Local externalities that depend on the demographic and skill composition of residents are accommodated in parsimonious ways in much of the urban and spatial economics literatures. This work typically relies on aggregate data at the metropolitan or neighbourhood levels and adopts simple specifications of endogenous local residential externalities (Diamond, 2016; Su, 2022; Couture et al., 2023; Almagro and Dominguez-Iino, 2025; Almagro et al., 2026). Using aggregate data, however, econometric identification is a major challenge. For example, Diamond and McQuade (2019) and Rossi-Hansberg et al. (2010) present evidence that people value some combination of higher income neighbours and new housing development but are not able to separate out these two forces. Nor are they able to separate out why people may value higher income neighbours. Here we quantify the magnitudes of two central components of such local endogenous amenities.

The remainder of the paper proceeds as follows. Section 2 develops the conceptual framework, which provides structural interpretations for the estimated parameters and clarifies the conditions under which the reduced-form earnings and labour force participation coefficients identify the human capital and reservation wage spillover parameters of interest. Section 3 describes the empirical specification, identification strategy, and estimation. Section 4 presents the data and sample construction. Section 5 reports the main results. Section 6 concludes.

## 2 Conceptual Framework

In this section, we develop a conceptual model that delivers the estimating equations we use to separately identify human capital and reservation wage spillovers between neighbours.[1]

### 2.1 Human Capital

Let $H_{i,b,t}$ denote the stock of human capital (or skill) possessed by individual $i$ living in residential block $b$ in year $t$. We assume human capital is homogeneous across individuals in the sense that it commands a common rental rate in the labour market, though individuals differ in their endowments and accumulation of human capital. Empirically, we define residential blocks $b$ as microgeographic hexagonal areas with 75-meter or 200-meter sides (see Section 4 for details).

[1]We emphasize that this is just one candidate data generating process that delivers these estimation equations. Intensive margin labour supply decisions ("effort") are yet to be incorporated.

We specify a data generating process for human capital that accommodates human capital spillovers operating between neighbours within residential locations. Each individual $i$ has an individual-specific, time-invariant component $\alpha_{i,0}$ that captures innate ability and skills accumulated prior to labour market entry. This fixed component of human capital is allowed to spill over directly through peer effects.

In addition to fixed ability, individual $i$'s human capital at time $t$ depends on three components: (i) contemporaneous and accumulated residential human capital spillovers, (ii) general labour market experience and other exogenous time-varying individual characteristics $X_{i,t}$, and (iii) local productive amenities and location accessibility defined at the neighbourhood rather than block level, captured by $\phi_{B(b),t}$, where $B(b)$ maps residential block $b$ to its surrounding neighbourhood. Put together, worker $i$'s stock of human capital at time $t$ is given by:

$$\ln H_{i,b,t} = \alpha_{i,0} + \sum_{\tau=0}^{K} \pi_\tau \bar{\alpha}_{i,t-\tau} + \beta' X_{i,t} + \phi_{B(b),t} \tag{1}$$

where $\bar{\alpha}_{i,t-\tau}$ denotes the average fixed human capital component of individual $i$'s residential peers at time $t-\tau$. We exclude other individuals in $i$'s household from an individual's peer group to focus on spillovers between neighbours only.

Formally, let $M_{i,b,t}$ denote the set of individual $i$'s peers at time $t$, which includes all non-family members residing in block $b$ in that year. While we generalize to allow them to take various forms, initially we think of spillovers as taking a linear-in-means form, meaning

$$\bar{\alpha}_{i,t} = \sum_{j \in M_{i,b,t}} \frac{1}{|M_{i,b,t}|} \alpha_{j,0}. \tag{2}$$

The vector $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_K)'$ is the key object of interest that we aim to estimate. The parameter $\pi_\tau$ denotes the elasticity of individual $i$'s human capital with respect to the average fixed component of human capital of residential neighbours $\tau$ periods earlier. That is, $\pi_\tau$ captures the effect of current or past peer composition on current human capital, allowing for potentially heterogeneous effects across different lags.

## 2.2 Skill Prices and Wages

Human capital is sold on a competitive labour market. For flexibility, we allow the rental price of skill to vary across time, space, and employer. Let $P_{B(b),t}$ denote the baseline rental price of human capital in neighbourhood $B(b)$ at time $t$. In addition to neighbourhood-level skill price variation, compensation includes a firm-specific component. There is a finite set

of firms, indexed by $j \in \mathcal{J} = \{1, 2, \ldots, J\}$. Each firm $j$ has its own skill price factor $p_j$ that scales the local baseline rental price. Let $j(i,t)$ denote the identity of the firm employing worker $i$ at time $t$. The wage of worker $i$ residing in block $b$ and employed at firm $j(i,t)$ at time $t$ is therefore given by:

$$w_{i,b,t} = p_{j(i,t)} P_{B(b),t} H_{i,b,t} \tag{3}$$

After taking logs and substituting for $H_{i,b,t}$ from equation (1), the following section demonstrates that this suggests an estimating equation that relates individual $i$'s log earnings to the average fixed component of her peers' log earnings.

## 2.3 Firm Choice and Employment

**Firm Choice.** Conditional on labour market participation, workers can choose to work at any of the $J$ firms in the economy. The utility they get from each option depends on their wage and a firm-specific term that captures unobserved preferences for working at firm $j$. Specifically, the utility worker $i$ receives from working at firm $j$ at time $t$ is:

$$U_{i,j,t} = z_{i,j,t} p_j P_{B(b),t} H_{i,b,t}$$

where $z_{i,j,t}$ captures idiosyncratic preferences for firm $j$. The preference shocks $z_{i,j,t}$ are independently and identically distributed across firms and workers according to a Fréchet (Type-II Extreme Value) distribution with shape parameter $\theta$. The scale parameter is normalized to one without loss of generality. Workers choose the firm that maximizes utility. The expected utility from employment is given by:

$$\mathbb{E}\left[\max_{j \in \{1,\ldots,J\}} U_{i,j,t}\right] = \gamma \Phi P_{B(b),t} H_{i,b,t}$$

where $\gamma = \Gamma(1 - 1/\theta)$ and $\Gamma(\cdot)$ is the gamma function and the term $\Phi \equiv \left[\sum_{j=1}^{J} p_j^{\theta}\right]^{1/\theta}$ is a wage index summarizing the distribution of firm-level skill prices.

**Employment Decisions.** We adopt a two-stage information structure for labour supply. In the first stage, individual $i$ decides whether to participate in the labour market at time $t$ by comparing the expected maximum utility from employment to a reservation value associated with non-employment. Conditional on participation, the worker then draws $\{z_{i,j}\}_{j=1}^{J}$ and chooses the utility-maximizing firm. This second stage is described by the firm-choice decision above and delivers the wage outcomes observed among workers.

Each worker has a time-varying reservation wage $w_{i,t}^R$. Employment occurs if

$$\ln \mathbb{E}\left[\max_{j\in\{1,\ldots,J\}} U_{i,j,t}\right] > \ln w_{i,t}^R.$$

We assume the reservation wage can be decomposed as:

$$\ln w_{i,t}^R = \sum_{\tau=0}^{K} \pi_\tau^R \bar{\alpha}_{i,t-\tau} + \beta'^R X_{i,t} + \phi_{B(b),t}^R + \nu_{i,b,t}^R \tag{4}$$

where $\beta^R$ captures the effect of exogenous time-varying individual characteristics, $\phi_{B(b),t}^R$ captures neighbourhood-by-year variation in reservation wages, and $\nu_{i,b,t}^R$ is an idiosyncratic shock. This specification allows current and past average residential peer quality to affect labour supply through non-human capital channels, such as social norms or preferences for non-market time. The coefficients $\pi_\tau^R$ capture such peer effects. For example, if the presence of lower human capital individuals nearby increases the value of non-market time, then $\pi_\tau^R$ would be negative.

For simplicity, we assume the idiosyncratic reservation wage shock follows a uniform distribution on the unit interval. The probability of employment is then:

$$\begin{aligned}
&\Pr\left[\ln \gamma\Phi P_{B(b),t} + \alpha_{i,0} + \sum_{\tau=0}^{K}(\pi_\tau - \pi_\tau^R)\bar{\alpha}_{i,t-\tau} + (\beta' - \beta'^R)X_{i,t} + (\phi_{B(b),t} - \phi_{B(b),t}^R) > \nu_{i,b,t}^R\right] \\
&= \ln \gamma\Phi P_{B(b),t} + \alpha_{i,0} + \sum_{\tau=0}^{K}(\pi_\tau - \pi_\tau^R)\bar{\alpha}_{i,t-\tau} + (\beta' - \beta'^R)X_{i,t} + (\phi_{B(b),t} - \phi_{B(b),t}^R)
\end{aligned} \tag{5}$$

where the second equality in equation (5) follows from the fact that for $\nu \sim U[0,1]$, we have $\Pr(\nu < x) = x$ for $x \in [0,1]$. The following section demonstrates that this suggests an estimating equation that relates an indicator for whether individual $i$ is employed to the average fixed component of her peers' employment status.

## 3 Empirical Implementation

In the empirical analysis, we seek to estimate the magnitudes and profiles of local spillovers that operate between residential neighbours. To do so, we employ a dynamic framework that allows for the recovery of causal effects of neighbourhood demographic environments on residents' labour market outcomes, building on econometric results in Arcidiacono et al. (2012), Baum-Snow et al. (2024), and Baum-Snow et al. (2026). The empirical models we estimate are directly motivated by the conceptual framework laid out in Section 2 above. Specifically,

we estimate variants of the intensive-margin earnings equation (3) and the extensive-margin employment equation (5), relating individual labour market outcomes to current and lagged measures of residential peer quality.[2] The framework shows how to assign structural interpretations to each empirical model parameter and clarifies the conditions under which the reduced-form coefficients we recover identify the human capital spillover parameters of interest.

As the conceptual framework laid out in Section 2 does not explicitly model residential location choice, the empirical analysis must confront the nonrandom allocation of individuals across space. Similar to Bayer et al. (2008), our identification strategy relies on the idea that housing markets are thin at small spatial scales, limiting the extent to which households can select into precise locations in any given year. Even if a household wishes to relocate to a specific neighbourhood in response to changes in local fundamentals, it may not be able to choose the exact block on which they reside, as suitable housing units that match household preferences may not be available at the time of the move. As explained below, our empirical strategy allows for both (i) residential sorting on fixed unobserved location characteristics, whereby higher-skill individuals select into particular blocks in the long-run, and (ii) residential sorting on contextual effects of slightly broader neighbourhood spatial definitions and changes in these neighbourhoods' fundamentals.

## 3.1 Baseline Estimation Equations

We define an individual's peer group as all non-family members living on the same residential block in period $t$. We vary the size of the area within which we define peer groups ("blocks") by constructing hexagonal areas with side lengths ranging from 75 meters to 200 meters. These blocks are nested within larger hexagons with 525-meter sides, which we refer to as neighbourhoods. Section 4 provides additional details on the spatial extent of peer group areas and neighbourhoods.

**Intensive Margin Responses.** Substituting the human capital equation (1) into the wage equation (3) yields the following data generating process for the log earnings of worker $i$ residing in block $b$ and employed at firm $j(i,t)$ at time $t$:

$$\ln w_{i,b,t} = \alpha^W_{i,0} + \sum_{\tau=0}^{K} \pi^W_\tau \bar{\alpha}^W_{i,t-\tau} + \beta^{W\prime} X_{i,t} + \phi^W_{B(b),t} + \rho^W_{j(i,t)} + \varepsilon^W_{i,b,t}. \tag{6}$$

[2] An important feature of our empirical models is that they are not subject to the reflection problem (Manski, 1993), as we focus on estimating spillovers from exogenous attributes of nearby individuals, as captured in their estimated fixed effects, rather than from their contemporaneous outcomes.

The superscript $W$ distinguishes the parameters and error terms in the reduced-form log earnings equation from the structural parameters in the conceptual framework.

This equation relates the log earnings of individual $i$ at time $t$ to an individual fixed effect capturing permanent earnings capacity, $\alpha^W_{i,0}$, and the average fixed effect of individual $i$'s current and past residential peers. The coefficients $\pi^W_\tau$ capture the elasticity of individual earnings with respect to average peer quality measured $\tau$ periods earlier. $\alpha^W_{i,0}$ identifies innate ability and skills accumulated prior to labour market entry, $\alpha_{i,0}$, and $\pi^W_\tau$ identifies the human capital spillover parameter of interest in equation (1), $\pi_\tau$. In the empirical analysis, we allow for $K = 3$ lags.[3]

Neighbourhood-by-year fixed effects $\phi^W_{B(b),t}$ control for contextual effects and neighbourhood trends at a broader spatial aggregation $B(b)$ than the block, such as 525-meter side hexagonal areas. Conceptually, $\phi^W_{B(b),t}$ captures a combination of skill price variation across space and time, $P_{B(b),t}$, and local productive amenities and location accessibility, $\phi_{B(b),t}$. Firm fixed effects $\rho^W_{j(i,t)}$ absorb persistent differences in compensation policies across firms and correspond to $\ln p_j$ in the conceptual framework. $\beta^{W\prime} X_{i,t}$ captures the effect of time-varying observable individual characteristics. $\varepsilon^W_{i,b,t}$ is an idiosyncratic shock which affects realized earnings in period $t$ only.

An empirical challenge is in separating human capital or effort spillovers from referral effects. Referral effects occur when high-ability individuals help residential peers obtain employment at higher-paying firms through social networks. If high-ability individuals are more likely to refer peers to higher-paying firms, then the correlation between average peer quality and wages would partially reflect sorting into firms rather than productivity effects. By including firm fixed effects, we eliminate this mechanism, as the identification of spillover effects comes from within-firm variation in wages only.[4]

**Extensive Margin Responses.** Let $E_{i,b,t}$ denote an indicator variable equal to one if individual $i$ living in block $b$ is employed at time $t$ and zero otherwise. We estimate the following extensive-margin labour supply equation:

$$E_{i,b,t} = \alpha^E_{i,0} + \sum_{\tau=0}^{K} \pi^E_\tau \bar{\alpha}^E_{i,t-\tau} + \beta^{E\prime} X_{i,t} + \phi^E_{B(b),t} + \varepsilon^E_{i,b,t} \tag{7}$$

[3]Identification and interpretation of $\pi$ parameters is most straightforward if the deepest lag of average peer quality has a coefficient of 0. Experimentation reveals no evidence of lag effects for wages beyond $\tau = 3$.

[4]This does not account for referral effects that operate within firms for instance, if neighbours help workers obtain better positions, projects, or promotions within the same firm. Such within-firm referral effects would remain confounded with human capital spillovers.

Here, $\alpha^E_{i,0}$ is interpreted as the permanent employability of individual $i$ relative to other individuals, holding constant current and past average peer group quality, time-varying characteristics $X_{i,t}$, and neighbourhood-by-year fixed effects $\phi^E_{B(b),t}$. $\pi^E_\tau$ captures the total effect of average residential peer quality measured $\tau$ periods ago on current employment probability.

Comparing equation (7) with equation (5) in the conceptual framework implies $\pi^E_\tau = \pi_\tau$ if peer effects operate solely through human capital accumulation. If peer effects also operate through other mechanisms beyond human capital, such as information about job opportunities, social norms regarding labour force participation, etc., then $\pi^E_\tau$ will differ from $\pi_\tau$. Specifically, in the conceptual framework, $\pi^E_\tau = \pi_\tau - \pi^R_\tau$ where $\pi^R_\tau$ captures the effect of peer composition on the reservation wage via these other non-human-capital channels. $\pi^E_\tau > \pi_\tau$ would suggest that higher quality peers decrease reservation wages (e.g., through positive social norms that encourage work), amplifying the positive effect through human capital. Estimating both intensive- and extensive-margin models allow us to separately identify human capital spillovers ($\pi_\tau$) from other non-human-capital channels such as social norms regarding labour force participation ($\pi^R_\tau$).

## 3.2 Heterogeneous Effects

We extend the baseline models along three dimensions: (i) heterogeneous responses by observable attributes such as gender and neighbourhood type; (ii) nonlinear effects by own position in the human capital distribution; and (iii) nonlinear effects by the composition of peers across the human capital distribution.

**Heterogeneous Responses by Observable Attributes.** We extend the baseline empirical models to capture heterogeneous responses to peers based on own observable attributes (such as gender, immigration status, tenure in the neighbourhood) and neighbourhood type (such as neighbourhood density and average income). Let $Z_{i,t}$ denote a vector of such attributes. We estimate the following equation for an outcome $y$:

$$y_{i,b,t} = \alpha^y_{i,0} + \sum_{\tau=0}^{K} \pi^y_\tau(Z_{i,t})\bar{\alpha}^y_{i,t-\tau} + \phi^y_{B(b),t} + \beta^{y'} X_{i,t} + \varepsilon^y_{i,b,t} \tag{8}$$

where the superscript $y$ can take values $W$ when using log earnings as the outcome or $E$ for employment. In the heterogeneous response to peers model, all variables in $Z_{i,t}$ are included in $X_{i,t}$ if they are time-varying.

**Non-Linear Effects Across the Human Capital Distribution.** We also evaluate multiple dimensions of asymmetry in spillovers across the human capital distribution. This allows us to characterize the magnitudes by which lower human capital neighbours impart costs or benefits on higher human capital counterparts, the extent to which more disadvantaged neighbours benefit from alternative neighbourhood demographic compositions, and the importance of human capital mixing in residential locations.

Let $H_i(x)$ denote an indicator variable equal to one if individual $i$'s fixed effect lies above the $x^{th}$ percentile of the overall $\alpha$ distribution and zero otherwise. We allow responses to peers to vary across the human capital distribution by estimating:

$$y_{i,b,t} = \alpha^y_{i,0} + \sum_{\tau=0}^{K} \pi^y_\tau(H_i(x))\bar{\alpha}^y_{i,t-\tau} + \phi^y_{B(b),t} + \beta^{y'} X_{i,t} + \varepsilon^y_{i,b,t} \tag{9}$$

These results speak to the log supermodularity assumption often used in theoretical modeling of cities with heterogeneous agents, which posits that high-skill individuals benefit more from interactions with other high-skill individuals than do low-skill individuals.

We also allow spillovers to depend on the fraction of peers in certain regions of the overall human capital distribution. Specifically, let $D_{i,t}(x_1, x_2)$ denote the fraction of individual $i$'s peers whose fixed effects lie between the $x_1^{th}$ and $x_2^{th}$ percentile of the overall $\alpha$ distribution. We estimate:

$$y_{i,b,t} = \alpha^y_{i,0} + \sum_{\tau=0}^{K} \pi^y_\tau \bar{\alpha}^y_{i,t-\tau} + \sum_{\tau=0}^{K} \kappa^y_\tau D_{i,t-\tau}(x_1, x_2) + \phi^y_{B(b),t} + \beta^{y'} X_{i,t} + \varepsilon^y_{i,b,t} \tag{10}$$

The parameter $\kappa^y_\tau$ captures the additional effect on individual $i$'s outcome of having a higher fraction of peers in the specified percentile range, measured $\tau$ periods earlier, holding constant average peer quality.

Finally, we consider fully interacted specifications where both $\pi^y_\tau$ and $\kappa^y_\tau$ are allowed to depend on $H_i(x)$, allowing the effects of average peer quality and the fraction of peers in specific ranges to vary across the individual's position in the human capital distribution.

## 3.3 Identification

Consistent identification of the spillover parameters $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_K)'$ requires variation in local residential peer composition that is conditionally orthogonal to unobserved determinants of individual labour market outcomes. In our setting, we confront two central identification challenges related to endogenous residential sorting.

First, individuals may sort across residential locations on the basis of fixed unobserved

characteristics. More able or more productive individuals may systematically select into higher-amenity blocks or neighbourhoods within the city. For example, workers with higher human capital may bid more aggressively for residences near employment centres, metro stations, or highway interchanges because they have a higher opportunity costs of commuting time. Such sorting on fixed unobserved characteristics would generate a spurious positive correlation between average peer quality and individual labour market outcomes.

Second, individuals may sort dynamically in response to changes in location fundamentals or to time-varying shocks to their earnings potential. Neighbourhoods experiencing improvements in local fundamentals may attract individuals with higher unobserved human capital or more favourable idiosyncratic productivity shocks. In addition, if the price of human capital contains an individual-specific stochastic component, workers experiencing positive skill-price draws may be more likely to relocate toward higher-amenity locations. Absent appropriate controls, such dynamic sorting would confound causal spillover effects with compositional changes driven by selective migration.

**Empirical Strategy.** Our empirical strategy addresses endogenous residential sorting by combining individual fixed effects with detailed neighbourhood-by-year fixed effects. We additionally control for time-varying observable individual characteristics to account for differential life-cycle profiles and observable changes in labour supply determinants. This approach draws on two well-established research designs in the labour and urban economics literatures and weakens the identifying assumptions required by either design in isolation.

First, as in the "AKM" literature concerned with identifying firm effects and in empirical work exploiting movers designs to study place effects, our baseline specifications include individual fixed effects.[5] Individual fixed effects absorb permanent differences in ability, preferences, and labour market attachment. Identification therefore relies on within-individual variation over time. This accommodates long-run sorting of individuals into particular blocks or neighbourhoods, allowing residential location choices to depend arbitrarily on time-invariant individual characteristics without biasing estimates of the spillover parameters. Our use of panel data is therefore central for identification in the presence of sorting on levels.

Second, we include detailed neighbourhood-by-year fixed effects $\phi_{B(b),t}$. These fixed effects absorb time-varying neighbourhood-level shocks and trends that may simultaneously affect local amenities, housing demand, and labour market outcomes. As a result, identification leverages within-neighbourhood, across-block variation in peer composition within a given year rather than comparisons across neighbourhoods or over time within neighbourhoods. This within-neighbourhood research design has been employed extensively in cross-sectional

[5]See, for example, Finkelstein et al. (2016) and Chetty and Hendren (2018).

contexts to identify spatial spillovers in outcomes such as employment referrals, foreclosures, school choice, and financial investments.[6] Intuitively, one source of identifying variation compares the residualized outcomes of incumbent individuals residing in blocks that experience the arrival of new neighbours with different levels of human capital to the residualized outcomes of other incumbent individuals living one or two blocks away within the same neighbourhood in the same year. Because these comparisons are made within neighbourhood-years, they are robust to neighbourhood-level shocks and to sorting driven by such shocks.

Identification thus relies on the randomness in the composition of new arrivals into blocks conditional on individual and neighborhood-year fixed effects. This requires that housing markets are sufficiently thin at fine spatial scales that, within a given year, migrating households can choose destination neighbourhoods $B(b)$ but face limited scope to select specific residential blocks $b$ within those neighbourhoods. Conditional on individual fixed effects and neighbourhood-by-year fixed effects $\phi_{B(b),t}$, variation in block-level peer composition is therefore driven primarily by the timing and availability of vacant housing units rather than by endogenous responses to block-specific shocks. Under this assumption, residual within-neighbourhood, within-year changes in local residential peer composition are independent of time-varying unobserved determinants of individual labour market outcomes. Importantly, this assumption allows for long-run sorting on permanent block-level attributes: households may relocate to preferred housing blocks when suitable vacancies arise idiosyncratically. However, because housing availability at the block level is limited in any given year, households generally cannot adjust their residential location contemporaneously in response to short-run, block-specific shocks.

The plausibility of this assumption is supported by the microgeographic scale of our analysis. Blocks are defined as hexagonal areas of approximately 0.015 $km^2$, while neighbourhoods correspond to hexagonal areas of roughly 0.7 $km^2$. At these scales, the number of housing units available on a given block in a given year is limited. Moreover, households have heterogeneous preferences over housing attributes, further constraining the availability of suitable units at the block level at any given point in time. While households may observe changes in neighbourhood-level demographic composition and amenities, fine-grained variation across blocks within neighbourhoods is likely more difficult to observe ex ante due to information frictions and search costs. As a result, conditional on choosing a neighbourhood, the exact block of residence is largely determined by housing availability at the time of the move rather than deliberate selection based on block-level shocks.

The key identifying assumption is partially testable. In particular, if changes in block-

[6] See, for example, Bayer et al. (2008), Linden and Rockoff (2008), Campbell et al. (2011), Currie et al. (2011), Anenberg and Kung (2014), Currie et al. (2015), and Bayer et al. (2021).

level peer composition were driven by unobserved shocks to individual outcomes, current residual outcomes would be correlated with the characteristics of neighbours who enter the block in subsequent periods or who exited in prior periods. We explore such diagnostics in the results section. Relatedly, controlling for lagged peer group composition helps with identification.

A maintained assumption of our empirical framework is that a substantial share of social interactions occurs at very short spatial distances, among individuals residing on the same block. To the extent that interactions also occur across nearby blocks, our estimates capture the difference between very local spillovers and broader neighbourhood-level spillover effects. However, as we show very rapid spatial decay across 75m side hexagons within 525m side hexagons, we think it unlikely that we are missing broader spatial scale human capital spillovers that operate through interpersonal interactions. Most conservatively, however, our estimates can be interpreted as lower bounds on the overall magnitudes of neighbourhood human capital spillovers. Moreover, the fact that spatially lagged peer group composition has near 0 spillover effects is evidence that spatially correlated unobservables are unlikely to be biasing our estimated spillover parameters of interest.

## 3.4 Estimation

Estimation of the spillover parameters $\boldsymbol{\pi}$ proceeds jointly with individual fixed effects and the remaining model parameters. The regressors of interest, average peer quality $\bar{\alpha}_{i,t-\tau}$, are not directly observed, as they are functions of the fixed effects we estimate. We therefore solve the nonlinear least squares problem implied by the empirical model using an iterative algorithm that alternates between updating the common parameters and updating the individual fixed effects.

Arcidiacono et al. (2012) show that the first-order conditions with respect to each $\alpha_{i,0}$ yield updating equations that can be solved in closed form given the other parameters. We adopt their two-step procedure. In the first step, we take the current fixed effects as given, construct the residential peer quality measures $\bar{\alpha}_{i,t-\tau}$ for all $i$ and $\tau$, and estimate the spillover coefficients $\boldsymbol{\pi}$, the coefficients on time-varying controls $\boldsymbol{\beta}$, the neighbourhood-by-year fixed effects $\phi_{B(b),t}$, and, in some intensive-margin log earnings specifications, the firm fixed effects $\rho_j$, by ordinary least squares. In the second step, we take those estimates as given and update each $\alpha_{i,0}$ using the first-order conditions. We repeat these two steps until the spillover parameter estimates converge. Under weak regularity conditions, this procedure converges to the nonlinear least squares solution (Arcidiacono et al., 2012).[7] We initialize the individual

[7]We note that $\hat{\boldsymbol{\pi}}$ is root-$T$ consistent if errors are serially correlated.

fixed effects from a regression of the outcome on individual fixed effects, neighbourhood-by-year fixed effects with no peer quality regressors, with the intercept absorbed in the individual effects.

The exact form of the updating equations for $\alpha_{i,0}$ depends on the specification (the number of lags $K$, whether firm fixed effects are included, and how peer quality or heterogeneous effects are specified). Our setup accommodates missing outcomes (e.g., due to the training period). The Online Appendix provides the updating equations for $\alpha_{i,0}$ for each specification we estimate, including the case with missing data. Baum-Snow et al. (2026) discusses additional properties of the dynamic model we employ and performs Monte Carlo simulations showing the estimator performs well in finite samples.

Inference is conducted using a symmetric wild bootstrap (MacKinnon, 2006), clustered at the peer group location (75-meter hexagon) $b$ level. The wild bootstrap is advantageous in this setting since it allows for heteroskedasticity of an unknown form and does not require any resampling. Clustering at the peer group location level accounts for correlation in residuals within residential blocks over time and across individuals who share the same block in a given year.

## 4 Data

Our investigation uses population-wide panel data covering the universe of tax-filing individuals residing in the Toronto metropolitan region between 2001 and 2018. Figure 1 illustrates the spatial extent of the Toronto metropolitan region, which includes the city of Toronto and its surrounding municipalities. The dataset is constructed by linking multiple administrative tax files. Information on individuals comes from personal income tax returns (Form T1), which we merge to firm records of employment remuneration (Form T4) to construct a matched employer-employee dataset.

A central advantage of these administrative data is the availability of six-character postal codes for place of residence, which provide an unusually fine-grained spatial representation of the population. In denser parts of the city, a postal code typically corresponds to a single building or apartment complex, while in lower-density areas it generally corresponds to a city block face. This level of spatial resolution is essential for measuring local human capital and economic spillovers between residential neighbours and for our research design, which relies on variation in neighbour composition within small geographic areas.

In addition to place of residence, we observe individual characteristics such as date of birth, gender, immigration status, and family composition, along with detailed information on employment, firm identifiers and characteristics, and earnings. Unlike many matched

employer-employee datasets, we also observe most individuals who are unemployed or out of the labour force, as tax filing is required to claim benefits in Canada. This feature allows us to distinguish between intensive- and extensive-margin labour market responses to changes in neighbourhood demographic composition.

We focus on two primary outcomes: (i) an indicator for employment status and (ii) log earnings. We consider individuals to be employed if they have any employment income in a given year. We assign workers who hold multiple jobs in a given year to the firm from which they derive most employment income during the year.

The longitudinal and administrative nature of the data is central to our approach. Following residents over time allows us to recover estimates of the pre-market component of each individual's human capital through individual fixed effects and to estimate the dynamic effects of exposure to different residential environments. As discussed in Section 3.3 above, our use of individual fixed effects and small neighbourhood-by-year fixed effects allows credible identification of causal effects of peer composition on outcomes. The use of population-wide administrative data allows granular insights into the composition of residential spillovers between residential neighbours across the full distribution of households.

## 4.1 Spatial extent of peer groups and neighbourhoods

We define the spatial extent of peer groups and neighbourhoods using a hexagonal grid with 75-meter sides that spans the Toronto metropolitan region. Relative to administrative boundaries, hexagons are uniform in shape and area and can naturally be aggregated to construct broader spatial units. We assign individuals to hexagons using the geographic centroid of their residential postal code.

Our primary definition of a peer group is a small hexagon with 75-meter sides, which has an area of approximately 0.015 $km^2$. These units capture highly local residential environments and serve as the building blocks for the broader geographic units used in the analysis, as illustrated in Figure 2 Panel A. To characterize spatial decay in residential externalities, we use intermediate-sized hexagons as an alternative peer group definition. These units are spatial aggregates of seven small hexagons and have sides of approximately 200 meters. Our primary definition of a neighbourhood is a spatial aggregate of 49 small hexagons. These groups of 49 small hexagons form a larger hexagonal area with sides of approximately 525 meters and an area of about 0.7 $km^2$. Figure 2 Panel B illustrates the size of these neighbourhoods for a subset of the municipality of Markham, a suburban municipality within the Toronto metropolitan region. In our primary empirical specification, we control for such small neighbourhood-by-year fixed effects to capture contextual effects of neighbourhoods and neighbourhood trends. Figure 2 Panel C outlines postal codes in a low density neighbourhood.

Postal code centroids are indicated by red dots. Figure 2 Panel D illustrates the size of our peer group locations within this neighbourhood and the mapping from postal codes to peer group locations.

## 4.2 Sample Restrictions

We now describe the main sample restrictions. Complete details on the construction of the estimation samples are provided in the Online Appendix.

We focus on individuals aged 22-55 living in the Toronto metropolitan region. Our analysis requires tracking individuals across residential locations over time. We therefore exclude observations with missing information on place of residence, as well as observations for individuals residing outside of the Toronto Census Metropolitan Area (CMA). When an individual leaves the Toronto CMA (or the dataset) for more than three years and subsequently returns, we treat the post-return observations as belonging to a new individual and assign them a new individual identifier. If an individual has more than one such prolonged absence, we remove them from the sample.

Our empirical specification includes lagged measures of peer group composition at lags $\tau = 1, 2, 3$. To address the resulting "cold start" problem, we impose a training period for each individual spell during which lagged peer characteristics cannot be constructed (i.e., the first 3 years after an individual appears or reappears in the dataset). During this training period, the outcome variable is set to missing so that any spillovers received in these periods do not contribute to the estimation. However, individuals remain in the data in these observations and their fixed effect continue to contribute to the peer environments of others.

We construct two estimation samples. The first sample includes both employed and non-employed individuals and is used to study extensive-margin labour supply responses to changes in neighbourhood environments. The second sample focuses on workers by setting the outcome variable to missing in years in which the individual earns less than \$10,000.[8] In both cases, we restrict the estimation sample to the largest connected set of observations in a high-dimensional fixed-effects model that includes individual fixed effects, neighbourhood-by-year fixed effects, and, for the worker sample only, firm fixed effects. This connected set consists of the largest subset of individuals, neighbourhood-years, and firms that are linked, directly or indirectly, through mobility, and for which relative fixed effects are identified.

For the broad estimation sample that includes non-employed individuals, the data comprise 40.7 million person-year observations (28.7 million with non-missing outcomes after imposing the training period), covering 3.47 million individuals, 879,400 peer groups, and

[8] We use a minimum income threshold to minimize the effect of part time work. \$10,000 is roughly equal to 26 weeks of work full-time at the prevailing minimum wage during the sample period.

65,900 peer group locations. Individuals are observed on average for 11.75 years. Over the sample period, the Toronto metropolitan region had an average population of 2.7 million residents aged 22 to 55.[9] Our broad estimation sample therefore captures approximately 85% of the prime working-age population.[10] For the worker-only estimation sample, the data comprise 31.8 million person-year observations (19.6 million with non-missing outcomes after imposing the training period), covering 2.59 million individuals, 855,200 peer groups, and 64,700 peer group locations. Individuals in the worker-only sample are observed on average for 12.27 years.

## 4.3 Descriptive Statistics

Table 1 reports summary statistics for the two estimation samples. Column (1) presents statistics for the broad sample that includes non-employed individuals, while Column (2) reports statistics for the worker-only sample.

The demographic composition of the two samples is similar. The average age is approximately 41 years. Among observations with non-missing outcomes, 13 percent correspond to individuals aged 25-29, 31 percent to those aged 30-39, 36 percent to those aged 40-49, and 20 percent to those aged 50-55. Women account for slightly more than half of the sample (53 percent in the broad sample and 51 percent among workers). A majority of individuals have children (71 percent), and roughly half are foreign-born (54 percent in the broad sample and 50 percent among workers), consistent with the demographic composition of the Toronto metropolitan region.

Average annual earnings, including zeros, are $46,000 in the broad sample and $65,600 in the worker-only sample, with corresponding mean log earnings of 10.53 and 10.81. In the broad sample, 79 percent of individuals report positive earnings in a given year, and 71 percent report earnings above $10,000. Figure 3 plots the distribution of log annual earnings (solid line) and the distribution of contemporaneous peer-average log annual earnings (dashed line) in the worker-only estimation sample. Both distributions are approximately log-normal. The distribution of peer-average earnings is more concentrated, reflecting the mechanical smoothing induced by aggregation across peers.

Peer groups contain, on average, 46 individuals when averaging across peer group-years (37 in the worker-only sample). When averaging instead across person-year observations, the mean peer group size increases to 155 in the broad sample and 115 in the worker-only sample.

[9] Authors' calculations using Canadian Census data.

[10] This coverage rate is consistent with estimates of tax non-filing rates based on survey data. For example, Robson and Schwartz (2020) estimate a non-filing rate of 12.41% for individuals aged 25-64 in Canada in 2016.

This difference reflects the skewed distribution of peer group sizes, as larger groups receive greater weight when averaging at the individual level. As shown in Figure 4, the distribution is heavily skewed to the right, with some peer groups exceeding 1,500 members, primarily in the urban core. The standard deviation of peer group size is 385 in the broad sample and 290 in the worker-only sample.

Figure 5 documents substantial cross-sectional sorting: higher-income individuals systematically live near higher-income neighbours, particularly above the median of the earnings distribution. This pattern highlights the importance of accounting for endogenous residential sorting when estimating peer effects and motivates our empirical strategy based on within-neighbourhood, across-block variation in changes in peer composition over time.

To identify local peer effects, we require sufficient residential mobility to separately estimate individual and neighbourhood-year fixed effects and to generate residual variation in local peer composition. Table 2 provides additional information on the structure of our sample, focusing on mobility and variation in outcomes and peer quality observed in the data. Among observations with non-missing outcomes, 11 percent correspond to individuals in their first year in a new residential block, and 79 percent have been at their current residential block for more than two years.

In our baseline specifications based on the intensive-margin log earnings equation (6) and the extensive-margin employment equation (7), the standard deviation of the estimated individual fixed effects ($\alpha_{i,0}^{W}$ and $\alpha_{i,0}^{E}$) is 0.60 and 0.34, respectively. The corresponding standard deviation of the outcome is 0.69 and 0.41. The standard deviation of average peer quality is 0.30 and 0.10, or about 40 percent of the overall standard deviation of the corresponding labour market outcome.

As explained in Section 3.3, one source of identifying variation compares the residualized outcomes of incumbent individuals residing in blocks that experience the arrival of new neighbours with different levels of human capital, as captured by their fixed effects, to the residualized outcomes of other incumbent individuals living one or two blocks away within the same neighbourhood in the same year. The standard deviation of the change in average peer quality from one year to the next is 0.12 for log earnings and 0.05 for employment status. Unsurprisingly, this variation is more than four times larger for movers than for stayers. Nevertheless, even among stayers there is substantial within-block variation in peer composition, amounting to about 25 percent of the cross-sectional variance in average peer quality in the worker-only sample. Taken together, these patterns indicate that there is meaningful within-neighbourhood, over-time variation in peer composition to support the identification of local spillover effects.

# 5 Results

In this section, we present and discuss earnings (intensive-margin) and labour force participation (extensive-margin) spillover parameter estimates for various baseline empirical model specifications, alternative peer group definitions, and spillover specifications capturing heterogeneous treatment effects. Standard errors reported in all tables are analytical, clustered at the peer-group location (75-meter hexagon) level.[11]

## 5.1 Intensive Margin Peer Effect Estimates

We begin by presenting estimates of average earnings spillovers within hexagonal peer group areas of approximately 0.015 $km^2$ (75-meter sides). These headline estimates correspond to equation (6) and come from nonlinear least squares regressions of log annual earnings on individual fixed effects, age and age-squared, small neighbourhood-by-year fixed effects, and current and lagged measures of average residential peer quality. We emphasize that there exists important heterogeneity underlying these linear-in-means spillover estimates, which we explore below.

Table 3 presents estimates from a static specification with no lags ($K = 0$). Specification (2) is our preferred specification. Our headline estimate of the elasticity of earnings with respect to the average of residential peers' fixed components of earnings is 0.025. As discussed in Section 3, this coefficient identifies the human capital spillover parameter of interest in equation (1) and we interpret it as the elasticity of human capital with respect to the average fixed component of human capital of residential neighbours. Another way of interpreting this estimate is that peers' exogenous characteristics are 2.5 percent as important as an individual's own exogenous characteristics in contributing to that person's stock of human capital. A one-standard deviation increase in average peer fixed effects (of 0.30) is estimated to increase an individual's human capital by 0.8 percent, or 1.1 percent of a standard deviation of log earnings.

As discussed, Figure 5 illustrates substantial cross-sectional sorting, with higher-income individuals systematically residing near higher-income neighbours. The dynamic migration of higher $\alpha$ individuals into certain neighbourhoods over time is also a central feature of the data. Estimates in specifications (1) and (4), which exclude neighbourhood-by-year fixed effects, are roughly 40 percent larger and indicate that dynamic sorting is an important potential threat to identification. Controlling for the number of peers in specification (3) has

[11] As described in Section 3, our preferred inference procedure uses a symmetric wild bootstrap clustered at the same level. Bootstrap-based inference is in progress and will be reported in a subsequent version. To the extent that bootstrapped standard errors are larger than analytical ones, some of the smaller estimates discussed below may not be statistically significant at conventional levels.

no effect on the estimate and the coefficient on the number of peers is zero, indicating that neighbourhood density does not confound our estimates.

An empirical challenge is in separating human capital spillovers from referral effects. Including firm fixed effects in specification (5) increases the estimate slightly from 0.025 to 0.028, indicating that the headline spillover coefficient does not operate through the referral of neighbours to higher-paying firms. If referrals were driving the results, controlling for the identity of the employing firm would attenuate the estimate; instead, within-firm variation in earnings reveals a spillover that is, if anything, slightly larger.

When estimated in dollar terms, results in the final column of Table 3 present a peer effects estimate of 0.042. This amounts to approximately CAD $1,800 per year in earnings per standard deviation increase in average peer quality, or 1.6 percent of a standard deviation.

This static formulation has two interrelated limitations. First is that human capital accumulation is a dynamic process. Spillovers from past peer groups can matter for workers' current human capital stocks. Second is the identification concern that peer group compositions may be serially correlated. Including past peer group compositions in the regression addresses both of these issues.

Table 4 presents results analogous to those in Table 3, but with the addition of three lags of average peer group quality ($K = 3$ in the empirical model equation (6)). Reflecting a small amount of serial correlation in peer group composition, the contemporaneous spillover estimate in our primary specification (2) drops slightly from 0.025 to 0.022. The one-year lag effect has a point estimate of 0.008, which is statistically significant based on analytical standard errors. Two- and three-year lag effects are small and economically negligible at $-0.002$ each. In dollar terms (specification 6), we find no evidence of lagged effects.

The central threat to identification is that higher $\alpha$ individuals may dynamically sort into blocks that are trending differentially from their surrounding 525-meter-side hexagonal neighbourhood. We plan to test for this by examining correlations of the estimated residuals from equation (6) with future peer group composition, future peer group entrants, and past peer group leavers. We will carry out analogous exercises separately for workers who move residential locations between year $t$ and $t+1$, for whom the scope for endogenous block selection is largest. These diagnostics using errors from Table 4 will be reported in a subsequent version of the paper.

## 5.2 Extensive Margin Estimates

Using the broader sample of all individuals 22–55, Table 5 presents our static extensive margin results. Our headline estimate in column (2) is 0.035. This coefficient captures the total effect of average residential peer quality on current employment probability. A 10 percentage point

increase in the average permanent employability of residential peers increases the probability of being employed by 0.35 percentage points. Interpreted in the context of our model in Section 2, the extensive-margin coefficient identifies $\pi_\tau^E = \pi_\tau - \pi_\tau^R$, so that the 0.010 gap between the extensive-margin estimate of 0.035 and the intensive-margin estimate of 0.025 identifies $-\pi_0^R = 0.010$. Higher-quality peers reduce reservation wages, encouraging labour force participation beyond the human capital channel alone. The final column of Table 5 uses annual earnings in dollars of everybody in the data, including 0s, as the outcome variable, thereby combining intensive and extensive margin spillover effects. Given that the extensive margin effect is larger than the intensive margin effect, it is consistent that this combined estimate of 0.046 exceeds the intensive-margin-only estimate of 0.042.

Adding three lags of peer group composition attenuates the contemporaneous extensive margin estimate from 0.035 to 0.026, as seen in column (2) of Table 6. The contemporaneous gap with the intensive margin estimate from Table 4 narrows from 0.010 in the static specification to just 0.004, indicating that much of the reservation wage spillover operates through a dynamic process. Estimated one-, two-, and three-year lags of 0.014, 0.006, and 0.006 each exceed their intensive-margin counterparts, implying extensive-intensive gaps of approximately 0.006 to 0.008 at each lag. This pattern is consistent with the gradual adoption of social norms regarding labour force participation: repeated exposure to higher-quality peers progressively reduces reservation wages. The lag-3 coefficient subsumes the effects of deeper lags as well, given serial correlation in peer composition, so the total dynamic reservation wage spillover is likely larger than the sum of the four estimated gaps.[12]

## 5.3 Spatial Decay

Many studies of neighbourhood effects use fairly large areas like census tracts to define neighbourhoods. Our setting provides a unique opportunity to assess the spatial extent of spillovers, and to show that interpersonal interactions at residential locations operate at a far more local spatial scale. A key finding is that human capital spillovers decay rapidly with distance: they are almost entirely concentrated within 75 meters and are negligible beyond that range. This finding is relevant both for understanding the spatial scale at which peer effects operate at residential locations and as an identification check. Much like adding time lags accounts for serially correlated peer group composition, adding spatial lags accounts for potential spatially correlated peer group composition. We assess the magnitudes of spatial decay by either redefining peer group areas to be larger or by using post-estimation data to

[12]We will experiment with additional lags to assess the persistence of reservation wage spillover effects. However, adding lags changes the estimation sample, potentially making results more difficult to compare to those reported here.

explicitly estimate spatial decay.

Our larger peer group areas are seven times the size of the 75-meter-side hexagons used thus far, including a focal hexagon and all six surrounding it. Within each broader neighbourhood area, there are now 7 intermediate-sized blocks rather than the 49 we were working with before. The results of estimating the same specification as in Tables 3 and 5 specification (2) but with these broader peer groups are reported in Table 7 Columns (2) and (5). The resulting estimates of 0.037 (intensive-margin) and 0.057 (extensive-margin) are somewhat larger than their baseline counterparts of 0.025 and 0.035. However, as they apply over regions that are 7 times larger, average per-hexagon treatment effects are much smaller at 0.005 and 0.008, respectively.

A more direct approach jointly estimates separate spillover parameters for own 75-meter sides hexagon and intermediate-sized 200 meter sides hexagon peer group quality. We take estimates of $\alpha_{i,0}$ from the regressions in Tables 3 or 5 specification (2) and use them to construct average peer group quality for these two peer group definitions. Results of regressing log earnings or employment status on both peer composition terms and the same set of fixed effects are reported in Table 7 Columns (3) and (6). To interpret the magnitudes, consider the thought experiment of doubling average peer quality in the focal hexagon (holding quality constant in all other hexagons) versus in one of the surrounding hexagons. In the focal hexagon, the resulting treatment effect is the 75-meter hexagon coefficient plus one-seventh of the 200-meter hexagon coefficient. If in a surrounding hexagon, it is one-seventh of the 200m hex coefficient only.

Results show that peer effects drop off precipitously within very small distances. Intensive margin estimates are 0.022 for an individual's own hexagon and 0.007 for the intermediate-sized group of seven hexagons. Doubling average peer quality in a single surrounding hexagon would increase log earnings by approximately $0.007/7 \approx 0.001$, compared to a total effect in the focal hexagon of $0.022+0.007/7 \approx 0.023$. That is, human capital spillovers almost entirely decay to 0 beyond 75 meters.[13] Extensive-margin spillover effects decay somewhat more slowly, with own-hexagon and intermediate-sized area spillover estimates of 0.031 and 0.012, respectively. An improvement in peer quality in an adjacent hexagon has approximately 5% of the employment impact of that in the focal hexagon. Finally, there is evidence of only a very small amount of spatial correlation in average peer group quality within broader neighbourhood-year, with the treatment effects of own-hexagon improvements in peer group quality falling by only 8 percent once controlling for the average quality of neighbours slightly further away. This is supportive evidence for our identification assumption that, conditional

[13]We tried jointly estimating individual fixed effects and peer effects in a "horse race" model with average peer quality in individuals' own 75 meter side hexagon and the larger group of 7 as separate regressors. Obtaining parameter convergence is difficult in this one-step estimator.

on individual fixed effects and neighbourhood-by-year fixed effects, variation in block-level peer composition is driven primarily by the timing and availability of vacant housing units rather than by endogenous responses to block-specific shocks.

## 5.4 Heterogeneous Responses by Observable Attributes

Table 8 reports intensive-margin earnings spillover estimates allowing peer effects to vary by observable individual characteristics and neighbourhood type. We control for these attributes if they are time-varying. Spillovers vary substantially across demographic groups. Estimates are larger for men (0.049) than for women (0.001), for immigrants (0.030) than for natives (0.021), for individuals without children under 18 at home (0.0545) than for parents (0.0167), and for recent movers who arrived in the peer group location within the past three years (0.0323) relative to stayers (0.021). Across neighbourhood types, spillovers are significantly larger in lower-income areas and in denser peer groups. The magnitude of these differences across neighbourhood types is large: the estimate in the lowest-income quartile of neighbourhoods (0.045) is roughly six times that in the highest-income quartile (0.008), and the estimate in the densest peer groups (0.0647) is more than four times that in the lowest-density groups (0.0141).

Table 9 presents corresponding extensive-margin labour force participation spillover estimates. These display similar patterns across neighbourhood income and density and by immigrant status. However, the gender pattern reverses: women exhibit larger extensive-margin spillovers (0.038) than men (0.032), and individuals with children show larger participation spillovers (0.041) than those without children (0.022). Movers and stayers have similar extensive-margin estimates (0.034).

Implied reservation wage spillovers, measured as the difference between extensive- and intensive-margin estimates, vary substantially across groups. The finding for women is particularly striking. The near-zero intensive-margin estimate combined with a large extensive-margin estimate implies that most of women's response to peer quality operates through the reservation wage channel rather than through human capital accumulation.

## 5.5 Non-Linear Effects Across the Human Capital Distribution

Results in Table 10 examine how human capital spillovers vary as a function of both the composition of peers and the individual's own position in the human capital distribution. Columns (2)–(4) augment the baseline specification by adding the fraction of peers above the 33rd, 50th, and 67th percentiles, respectively, of the overall $\alpha$ distribution. Holding constant average peer quality, reducing the share of peers in the lower tail of the distribution

is associated with positive treatment effects. As we move the threshold up the distribution, the coefficient on the fraction of peers above the threshold declines and that on average peer quality increases. By the 67th percentile (column 4), the composition term is negligible and the linear-in-means coefficient returns to 0.025.

Columns (5) and (6) interact spillover terms with the individual's own position relative to the median of the overall $\alpha$ distribution. Column (5) shows that individuals in the top half of the distribution do not benefit from increases in average peer quality: their estimated spillover elasticity is approximately zero ($0.075 - 0.079$). Our headline human capital spillover elasticity of 0.025 is therefore entirely driven by those in the bottom half of the distribution, who have a spillover elasticity of 0.075, roughly three times the average. Column (6) confirms this asymmetry by additionally including the fraction of peers above the median and its interaction with an individual's own position relative to the median of the $\alpha$ distribution. For below-median individuals, the coefficients on average peer quality (0.061) and the fraction of peers above the median (0.024) are both positive. The corresponding interaction terms are of similar magnitude but opposite sign, yielding near-zero net effects for above-median individuals. For individuals below the median, a unit increase in average peer group quality such that every peer goes from below to above the median implies an 8.5 percent increase in human capital. If the fraction of peers above the median does not change, a unit increase in average peer group quality implies a 6.1 percent increase.

These findings have direct implications for policy. The estimates suggest that relocating below-median individuals into higher-quality peer environments, for example through the siting of subsidized housing in higher-human-capital blocks, would generate substantial gains for these individuals while imposing little earnings costs to incumbent above-median residents.

## 6 Conclusion

Evaluating the consequences of locally targeted policies depends crucially on the nature of spillovers that operate between neighbours. This paper is the first to estimate profiles of earnings or labour force participation spillovers operating at small spatial scales across the full population of heterogeneous adults in a large metropolitan area. While the average elasticity of earnings with respect to average peer quality is only 0.025, there is considerable heterogeneity as a function of own and peer human capital. Individuals in the bottom half of the human capital distribution benefit from higher human capital neighbours, with a convex relationship seen in the data. However, those in the top half are not affected by the composition of neighbours' human capital. The fact that contemporaneous earnings spillovers

are much larger than their lags suggests an important role for effort or intensive margin labour supply effects, with only a small remaining role for dynamic human capital externalities. The greater persistence of extensive margin labour force participation spillovers of about 0.005 for up to 3 lags suggests neighbourhoods can fall into endogenous non-participation traps that may be difficult to reverse.

Beyond their implications for the efficacy of policies that encourage demographic mixing, these results also have implications for the way in which we model equilibrium neighbourhood choice. The lack of log-supermodularity in the amenity valuation of own interacted with peer human capital suggests the importance of accommodating more sophisticated endogenous amenity specifications in neighbourhood choice models. With this, our evidence suggests that it may be possible to support mixed income neighbourhoods in equilibrium, especially when the fraction of low human capital residents is small. Moreover, our evidence demonstrates how certain locations can end up in low-employment equilibria driven by endogenous amenities. The design of effective policies to unwind such poverty traps will require the sort of nuanced understanding of endogenous neighbourhood effects that we recover in this paper.

## References


Almagro, M., E. Chyn, and B. A. Stuart (2026): "The Welfare Impacts of Urban Renewal and Forced Relocation: Evidence from Chicago's HOPE VI Demolitions," Working paper.

Almagro, M. and T. Dominguez-Iino (2025): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *Econometrica*, 93, 1031–1071.

Anenberg, E. and E. Kung (2014): "Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures," *American Economic Review*, 104, 2527–2551.

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., N. Gendron-Carrier, J. Kinsler, and R. Pavan (2026): "Dynamic Spillovers in the Workplace," Working paper.

Baum-Snow, N., N. Gendron-Carrier, and R. Pavan (2024): "Local Productivity Spillovers," *American Economic Review*, 114, 1030–1069.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A unified framework for measuring preferences for schools and neighborhoods," *Journal of political economy*, 115, 588–638.

Bayer, P., K. Mangum, and J. W. Roberts (2021): "Speculative Fever: Investor Contagion in the Housing Bubble," *American Economic Review*, 111, 609–651.

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.

Brueckner, J. K. (2000): "A Tiebout/tax-competition model," *Journal of public economics*, 77, 285–306.

Campbell, J., S. Giglio, and P. Pathak (2011): "Forced Sales and House Prices," *American Economic Review*, 101, 2108–2131.

Chetty, R., J. N. Friedman, N. Hendren, M. R. Jones, and S. R. Porter (2026): "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," *American Economic Review*, 116, 1–51.

Chetty, R. and N. Hendren (2018): "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *The Quarterly Journal of Economics*, 133, 1107–1162.

Chetty, R., N. Hendren, and L. F. Katz (2016): "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment," *American Economic Review*, 106, 855–902.

Chyn, E. (2018): "Moved to opportunity: The long-run effects of public housing demolition on children," *American Economic Review*, 108, 3028–3056.

Chyn, E. and D. Daruich (2025): "An equilibrium analysis of the effects of neighborhood-based interventions on children," *American Economic Review*, 115, 4476–4522.

Cook, C., P. Z. Li, and A. J. Binder (2023): *Where to build affordable housing?: Evaluating the tradeoffs of location*, US Census Bureau, Center for Economic Studies Rochester, NY.

Cornelissen, T., C. Dustmann, and U. Schonberg (2017): "Peer Effects in the Workplace," *American Economic Review*, 107, 425–456.

Couture, V., C. Gaubert, J. Handbury, and E. Hurst (2023): "Income Growth and the Distributional Effects of Urban Spatial Sorting," *The Review of Economic Studies*, 91, 858–898.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015): "Environmental Health Risks and Housing Values: Evidence from 1600 Toxic Plant Openings and Closings," *American Economic Review*, 105, 678–709.

Currie, J., M. Greenstone, and E. Moretti (2011): "Superfund Cleanups and Infant Health," *American Economic Review*, 101, 435–441.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999): "The rise and decline of the American ghetto," *Journal of political economy*, 107, 455–506.

Davis, M. A., J. Gregory, and D. A. Hartley (2023): "Preferences over the racial composition of neighborhoods: Estimates and implications," Working paper.

Diamond, R. (2016): "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *American economic review*, 106, 479–524.

Diamond, R. and T. McQuade (2019): "Who wants affordable housing in their backyard? An equilibrium analysis of low-income property development," *Journal of Political Economy*, 127, 1063–1117.

DMTI Spatial Inc. (2014): "Platinum Postal Code Suite v2014.2," URL: https://mdl.library.utoronto.ca/collections/geospatial-data/platinum-postal-code-suite-4. Accessed: April 15, 2016.

Finkelstein, A., M. Gentzkow, and H. Williams (2016): "Sources of Geographic Variation in Health Care: Evidence from Patient Migration," *The Quarterly Journal of Economics*, 131, 1681–1726.

Fogli, A., V. Guerrieri, M. Ponder, and M. Prato (2026a): "The End of the American Dream? Inequality and Segregation in US Cities," *Journal of Political Economy*, 0, 000–000.

——— (2026b): "The Macroeconomic Effects of Neighborhood Policies: a Dynamic Analysis," Tech. rep., National Bureau of Economic Research.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75, 83–119.

Linden, L. and J. Rockoff (2008): "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," *American Economic Review*, 98, 1103–1127.

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

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

Mas, A. and E. Moretti (2009): "Peers at work," *American Economic Review*, 99, 112–145.

Massey, D. S. and N. A. Denton (1988): "The dimensions of residential segregation," *Social forces*, 67, 281–315.

Oreopoulos, P. (2003): "The Long-Run Consequences of Living in a Poor Neighborhood*," *The Quarterly Journal of Economics*, 118, 1533–1575.

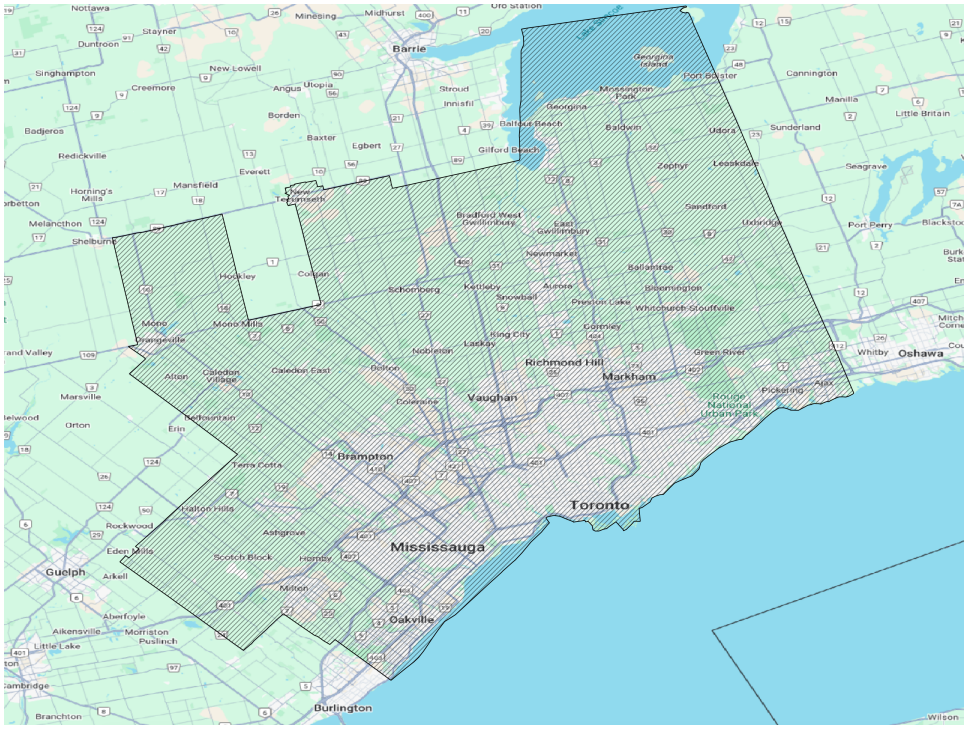
Rossi-Hansberg, E., P.-D. Sarte, and R. Owens III (2010): "Housing externalities," *Journal of political Economy*, 118, 485–535.

Schelling, T. C. (1971): "Dynamic models of segregation," *Journal of mathematical sociology*, 1, 143–186.

Su, Y. (2022): "The Rising Value of Time and the Origin of Urban Gentrification," *American Economic Journal: Economic Policy*, 14, 402–39.

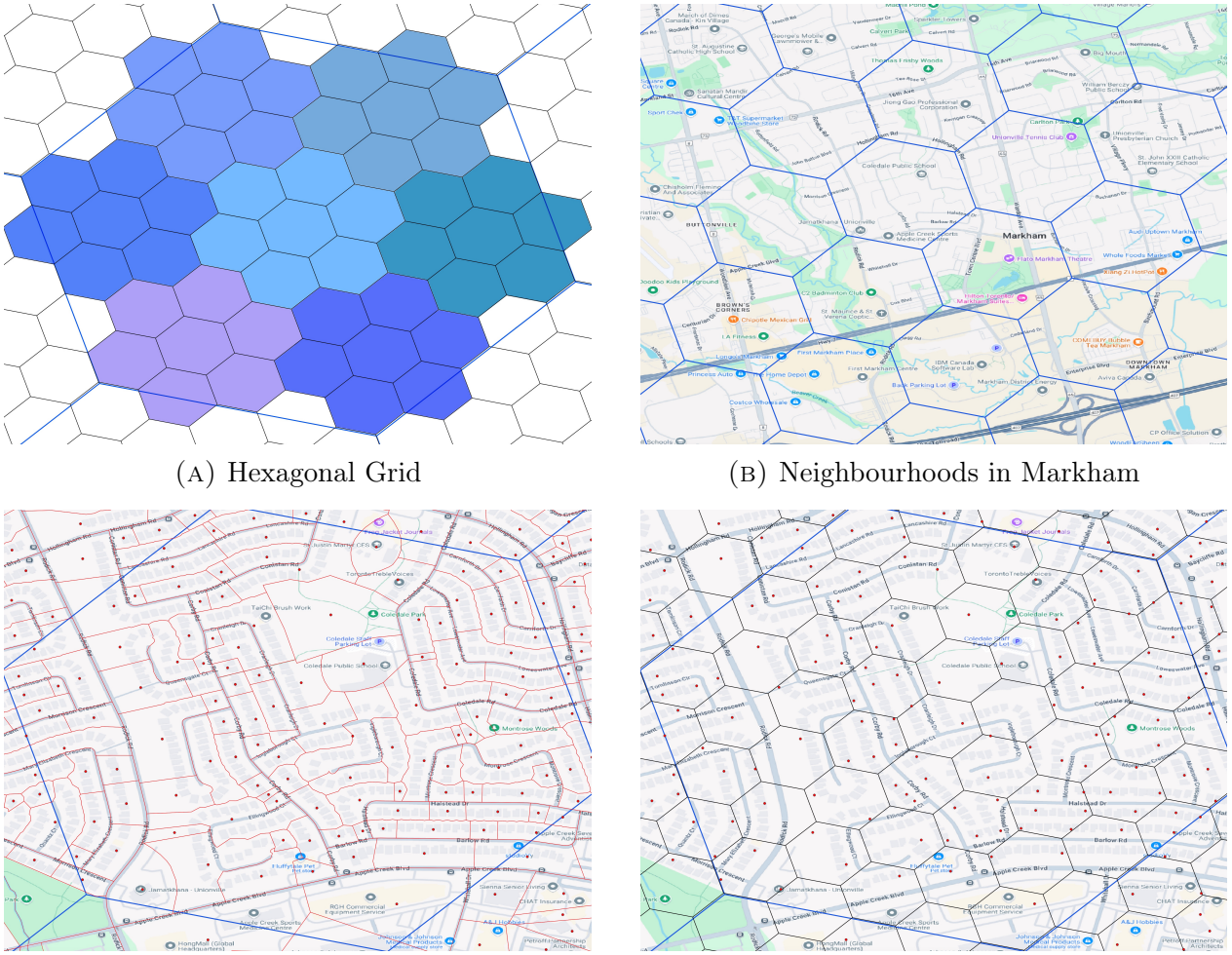
Weiwu, L. (2024): "Unequal access: Racial segregation and the distributional impacts of interstate highways in cities," Working paper.

FIGURE 1 – Geographic Coverage of the Sample




Notes: Toronto Census Metropolitan Area. Source data: Statistics Canada, 2021 Census Boundary File.

FIGURE 2 – Spatial Extent of Peer Groups and Neighbourhoods



(A) Hexagonal Grid

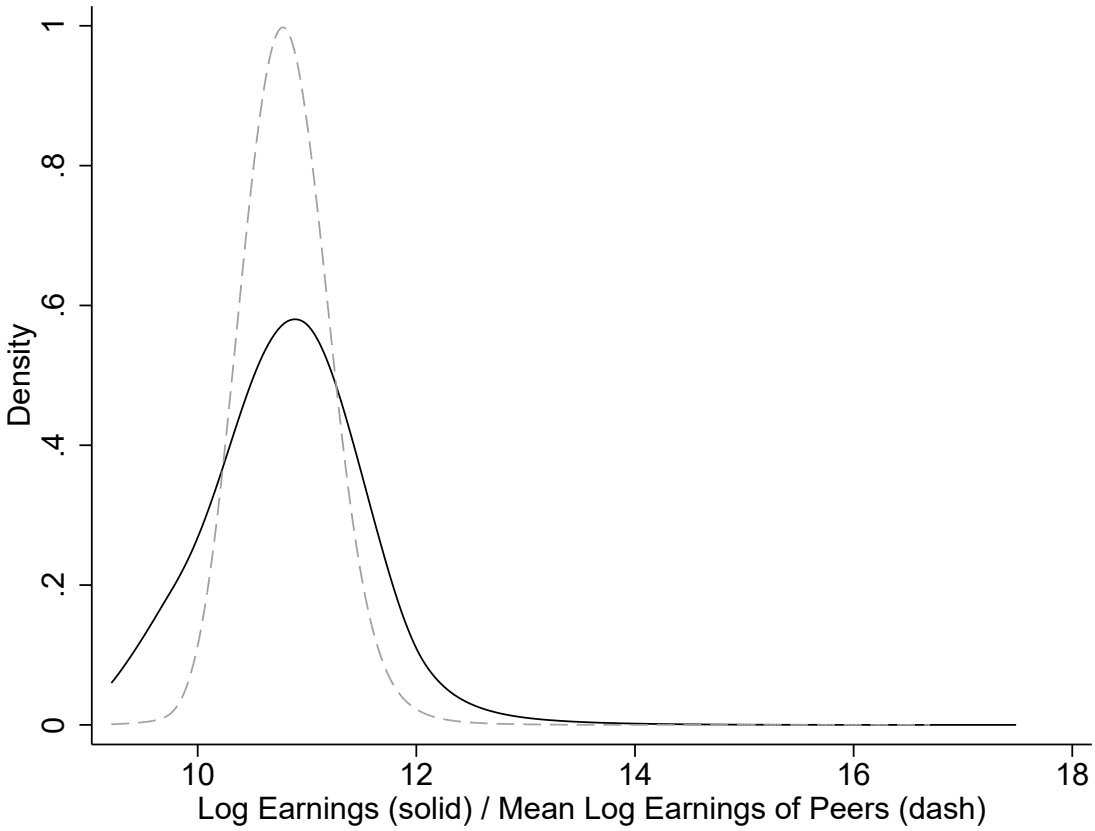
(B) Neighbourhoods in Markham

(C) Postal Codes within a Neighbourhood

(D) Mapping Postal Codes to Peer Groups

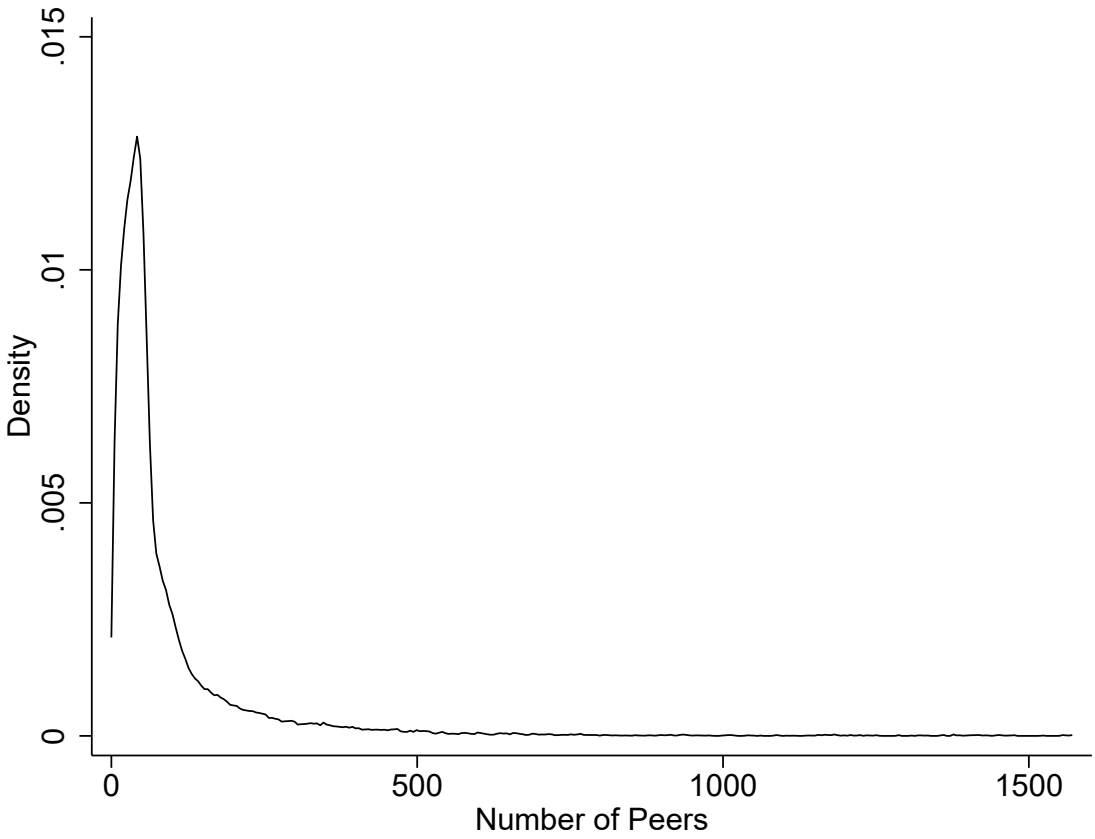
*Notes:* Panel (a) illustrates the hexagonal grid and the spatial aggregation of hexagons used in the analysis. Panel (b) shows neighbourhoods in Markham, where blue hexagons have 525-meter sides (approximately 0.7 $km^2$). Our baseline specifications include neighbourhood-by-year fixed effects at this spatial scale to absorb contextual effects of neighbourhoods and changes in location fundamentals. Panel (c) zooms in on one neighbourhood, with postal codes outlined by thin red lines and their centroids indicated by red dots. Panel (d) shows our primary definition of a peer group location: a hexagonal area with 75-meter sides (approximately 0.015 $km^2$), and the mapping from postal codes to peer group locations. Data source: DMTI Spatial Inc. (2014).

FIGURE 3 – Distribution of Log Earnings and Mean Log Earnings of Peers



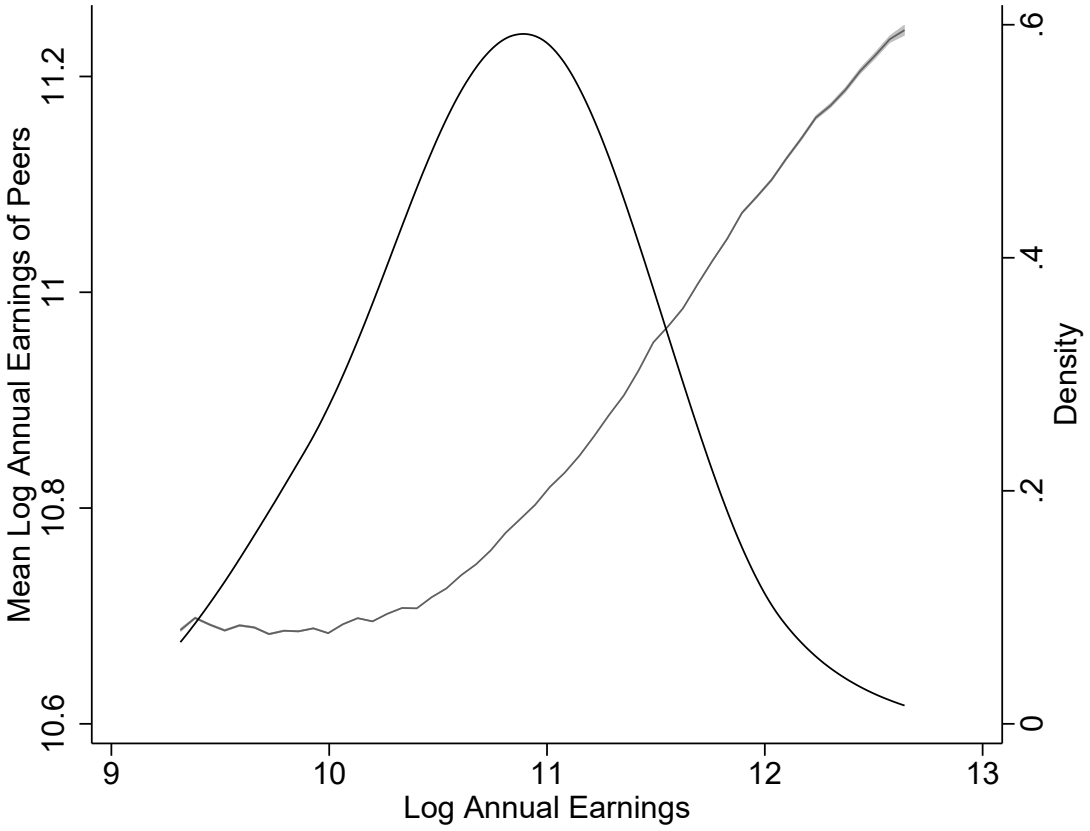

*Notes:* This figure plots the distribution of log annual earnings (solid line) and the distribution of the contemporaneous peer-average log annual earnings (dashed line) in our primary estimation sample. The sample includes all workers aged 22-55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides.

FIGURE 4 – Distribution of Peer Group Size




*Notes:* This figure shows the distribution of peer-group size in our primary estimation sample. The sample includes all workers aged 22-55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. To improve readability, the horizontal axis is trimmed at the $99^{\text{th}}$ percentile of the peer-group size distribution.

FIGURE 5 – Sorting on Peer Group Quality




*Notes:* This figure shows a local polynomial fit of mean peer log earnings on own log earnings. The shaded region indicates 95% confidence intervals. The figure includes workers in the primary estimation sample. The sample includes all workers aged 22-55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. To improve readability, the horizontal axis is trimmed at the $1^{st}$ and $99^{th}$ percentiles of the log earnings distribution.

TABLE 1 – Descriptive Statistics

| | Inc. non-emp. | Workers only |
|---|---|---|
| Observable characteristics | | |
| Average age | 41.00 | 40.83 |
| Below age 30 (%) | 0.13 | 0.13 |
| Age 30-39 (%) | 0.30 | 0.31 |
| Age 40-49 (%) | 0.36 | 0.36 |
| Age 50+ (%) | 0.21 | 0.20 |
| Women (%) | 0.53 | 0.51 |
| With kids (%) | 0.71 | 0.70 |
| Immigrant (%) | 0.54 | 0.50 |
| Labour market outcomes | | |
| Annual earnings | 46,000 | 65,600 |
| Log annual earnings | 10.53 | 10.81 |
| Any earnings (%) | 0.79 | 1.00 |
| Annual earnings above 10K (%) | 0.71 | 1.00 |
| Panel Structure | | |
| Total number of obs. | 40,706,000 | 31,823,000 |
| Number of obs., non-missing | 28,677,000 | 19,637,000 |
| Number of individuals | 3,466,900 | 2,592,000 |
| Number of peer groups | 879,400 | 855,200 |
| Number of peer group locations | 65,900 | 64,700 |
| Average number of peers | 155.00 | 115.00 |
| Standard deviation number of peers | 385.01 | 290.80 |

*Notes:* Statistics are reported for individuals aged 22–55 living in the Toronto census metropolitan area between 2001 and 2018. Column (1) reports the broad sample and includes both employed and non-employed individuals. Column (2) reports the worker sample, defined by setting the outcome variable to missing in years in which annual earnings are below $10,000. In both cases, we restrict the estimation sample to the largest connected set of observations in a high-dimensional fixed-effects model that includes individual fixed effects, neighbourhood-by-year fixed effects, and, for the worker sample only, firm fixed effects. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Neighbourhoods are aggregates of 49 small hexagons, forming larger hexagonal areas with sides of approximately 525 meters and area of about $0.7\,\mathrm{km}^2$.

TABLE 2 – Mobility and variation in outcomes and peer quality

Panel A: Tenure in neighbourhood

| | Inc. non-emp. | Workers only |
|---|---|---|
| More than 2 years (%) | 0.79 | 0.79 |
| First two years (%) | 0.22 | 0.21 |
| First year (%) | 0.11 | 0.11 |

Panel B: Variation in outcomes and peer quality

| | Inc. non-emp. | Workers only |
|---|---|---|
| | Any Earnings | Log Earnings |
| SD y | 0.41 | 0.69 |
| SD individual F.E. | 0.34 | 0.60 |
| SD average peer F.E. | 0.10 | 0.30 |
| SD $\Delta$ average peer F.E., t-1 to t | 0.05 | 0.12 |
| SD $\Delta$ average peer F.E., t-1 to t (stayers) | 0.04 | 0.07 |
| SD $\Delta$ average peer F.E., t-1 to t (movers) | 0.12 | 0.30 |

*Notes:* Statistics are reported for individuals aged 22-55 living in the Toronto census metropolitan area between 2001 and 2018. Column (1) reports the broad sample and includes both employed and non-employed individuals. Column (2) reports the worker sample, defined by setting the outcome variable to missing in years in which annual earnings are below $10,000. In both cases, we restrict the estimation sample to the largest connected set of observations in a high-dimensional fixed-effects model that includes individual fixed effects, neighbourhood-by-year fixed effects, and, for the worker sample only, firm fixed effects. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Individual fixed effects are obtained from the estimated models in Table 3, Column (2) (worker sample) and Table 5, Column (2) (broad sample).

TABLE 3 – Main results: Contemporaneous spillovers (intensive-margin)

| | Log Annual Earnings | | | | | Annual Earnings |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Avg. Peer F.E. | 0.0351 | 0.0251 | 0.0248 | 0.0367 | 0.0282 | 0.0419 |
| | (0.0003) | (0.0004) | (0.0004) | (0.0003) | (0.0005) | (0.0035) |
| Number of Peers | | | -0.0000 | | | |
| | | | (0.0000) | | | |
| Time-varying controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Year F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Area×Year F.E. | No | Yes | Yes | No | Yes | Yes |
| Firm F.E. | No | No | No | Yes | Yes | No |
| Observations | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 |
| # of Peer Group-Years | 855,200 | 855,200 | 855,200 | 855,200 | 855,200 | 855,200 |
| # of Peer Group Locations | 64,700 | 64,700 | 64,700 | 64,700 | 64,700 | 64,700 |
| # of Individuals | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 |
| Average $y$ | 10.81 | 10.81 | 10.81 | 10.81 | 10.81 | 65,600 |
| SD $y$ | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 | 111,043 |
| SD Individual F.E. | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 96,375 |
| SD Avg. Peer F.E. | 0.30 | 0.30 | 0.30 | 0.30 | 0.30 | 42,211 |

*Notes:* This table reports estimates of $\pi_0$ in equation (6) setting $K = 0$. The sample includes all workers aged 22–55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Neighbourhoods are aggregates of 49 small hexagons, forming larger hexagonal areas with sides of approximately 525 meters and area of about $0.7\,\text{km}^2$. Standard errors clustered at the block (peer-group) level are reported in parentheses.

TABLE 4 – Main results: Accumulation model (intensive-margin)

| | Log Annual Earnings | | | | | Annual Earnings |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Avg. Peer F.E. ($t$) | 0.0297 | 0.0217 | 0.0215 | 0.0299 | 0.0228 | 0.0396 |
| | (0.0007) | (0.0007) | (0.0007) | (0.0007) | (0.0008) | (0.0051) |
| Avg. Peer F.E. ($t-1$) | 0.0103 | 0.0078 | 0.0078 | 0.0112 | 0.0087 | 0.0002 |
| | (0.0007) | (0.0007) | (0.0007) | (0.0008) | (0.0008) | (0.0051) |
| Avg. Peer F.E. ($t-2$) | -0.0003 | -0.0019 | -0.0020 | 0.0002 | -0.0019 | 0.0081 |
| | (0.0007) | (0.0007) | (0.0007) | (0.0008) | (0.0008) | (0.0061) |
| Avg. Peer F.E. ($t-3$) | -0.0002 | -0.0017 | -0.0017 | 0.0017 | -0.0002 | 0.0033 |
| | (0.0006) | (0.0006) | (0.0006) | (0.0007) | (0.0007) | (0.0045) |
| Number of Peers | | | -0.0000 | | | |
| | | | (0.0000) | | | |
| Time-varying controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Year F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Area×Year F.E. | No | Yes | Yes | No | Yes | Yes |
| Firm F.E. | No | No | No | Yes | Yes | No |
| Observations | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 |
| # of Peer Group-Years | 855,200 | 855,200 | 855,200 | 855,200 | 855,200 | 855,200 |
| # of Peer Group Locations | 64,700 | 64,700 | 64,700 | 64,700 | 64,700 | 64,700 |
| # of Individuals | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 |
| Average $y$ | 10.81 | 10.81 | 10.81 | 10.81 | 10.81 | 65,600 |
| SD $y$ | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 | 111,043 |
| SD Individual F.E. | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 96,375 |
| SD Avg. Peer F.E. | 0.30 | 0.30 | 0.30 | 0.30 | 0.30 | 42,211 |

*Notes:* This table reports estimates of $\pi_0$ in equation (6) setting $K = 3$. The sample includes all workers aged 22–55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Neighbourhoods are aggregates of 49 small hexagons, forming larger hexagonal areas with sides of approximately 525 meters and area of about 0.7 $\text{km}^2$. Standard errors clustered at the block (peer-group) level are reported in parentheses.

TABLE 5 – Main results: Contemporaneous spillovers (extensive-margin)

| | Any Earnings | | Annual Earnings |
|---|---|---|---|
| | (1) | (2) | (3) |
| Avg. Peer F.E. | 0.0386 | 0.0349 | 0.0455 |
| | (0.0007) | (0.0007) | (0.0029) |
| Time-varying controls | Yes | Yes | Yes |
| Individual F.E. | Yes | Yes | Yes |
| Year F.E. | Yes | Yes | Yes |
| Area×Year F.E. | No | Yes | Yes |
| Observations | 28,677,000 | 28,677,000 | 28,677,000 |
| # of Peer Group-Years | 879,400 | 879,400 | 879,400 |
| # of Peer Group Locations | 65,900 | 65,900 | 65,900 |
| # of Individuals | 3,466,900 | 3,466,900 | 3,466,900 |
| Average $y$ | 0.79 | 0.79 | 46,000 |
| SD $y$ | 0.41 | 0.41 | 97,379 |
| SD Individual F.E. | 0.34 | 0.34 | 80,057 |
| SD Avg. Peer F.E. | 0.10 | 0.10 | 28,259 |

*Notes:* This table reports estimates of $\pi_0$ in equation (7) setting $K = 0$. The sample includes all individuals (both employed and non-employed) aged 22–55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Neighbourhoods are aggregates of 49 small hexagons, forming larger hexagonal areas with sides of approximately 525 meters and area of about $0.7\,\text{km}^2$. Standard errors clustered at the block (peer-group) level are reported in parentheses.

TABLE 6 – Main results: Accumulation model (extensive-margin)

| | Any Earnings | | Annual Earnings |
|---|---|---|---|
| | (1) | (2) | (3) |
| Avg. Peer F.E. ($t$) | 0.0268 | 0.0261 | 0.0416 |
| | (0.0010) | (0.0010) | (0.0045) |
| Avg. Peer F.E. ($t-1$) | 0.0178 | 0.0143 | 0.0034 |
| | (0.0010) | (0.0010) | (0.0047) |
| Avg. Peer F.E. ($t-2$) | 0.0083 | 0.0058 | 0.0085 |
| | (0.0010) | (0.0010) | (0.0048) |
| Avg. Peer F.E. ($t-3$) | 0.0077 | 0.0057 | 0.0076 |
| | (0.0009) | (.0010) | (0.0042) |
| Time-varying controls | Yes | Yes | Yes |
| Individual F.E. | Yes | Yes | Yes |
| Year F.E. | Yes | Yes | Yes |
| Area×Year F.E. | No | Yes | Yes |
| Observations | 28,677,000 | 28,677,000 | 28,677,000 |
| # of Peer Group-Years | 879,400 | 879,400 | 879,400 |
| # of Peer Group Locations | 65,900 | 65,900 | 65,900 |
| # of Individuals | 3,466,900 | 3,466,900 | 3,466,900 |
| Average $y$ | 0.79 | 0.79 | 46,000 |
| SD $y$ | 0.41 | 0.41 | 97,379 |
| SD Individual F.E. | 0.34 | 0.34 | 80,057 |
| SD Avg. Peer F.E. | 0.10 | 0.10 | 28,259 |

*Notes:* This table reports estimates of $\pi_0$ in equation (7) setting $K = 3$. The sample includes all individuals (both employed and non-employed) aged 22–55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Neighbourhoods are aggregates of 49 small hexagons, forming larger hexagonal areas with sides of approximately 525 meters and area of about $0.7\,\text{km}^2$. Standard errors clustered at the block (peer-group) level are reported in parentheses.

TABLE 7 – Spatial Decay

| | Log Annual Earnings (workers only) | | | Any Earnings (full sample) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Avg. Peer F.E. (75m hex.) | 0.0251 | | 0.0220 | 0.0349 | | 0.0305 |
| | (0.0004) | | (0.0006) | (0.0007) | | (0.0008) |
| Avg. Peer F.E. (200m hex.) | | 0.0370 | 0.0068 | | 0.0569 | 0.0118 |
| | | (0.0006) | (0.0008) | | (0.0012) | (0.0014) |
| Time-varying controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Year F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Area×Year F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 19,637,000 | 19,637,000 | 19,637,000 | 28,677,000 | 28,677,000 | 28,677,000 |
| # of Peer Group-Years | 855,200 | 855,200 | 855,200 | 879,400 | 879,400 | 879,400 |
| # of Peer Group Locations | 64,700 | 64,700 | 64,700 | 65,900 | 65,900 | 65,900 |
| # of Individuals | 2,592,000 | 2,592,000 | 2,592,000 | 3,466,900 | 3,466,900 | 3,466,900 |
| Average $y$ | 10.81 | 10.81 | 10.81 | 0.79 | 0.79 | 0.79 |
| SD $y$ | 0.69 | 0.69 | 0.69 | 0.41 | 0.41 | 0.41 |

*Notes:* Estimates in Columns (1) and (4) are analogous to those reported in Table 3, Column (2), and Table 5, Column (2), respectively. Columns (2) and (5) use the same estimation samples as Columns (1) and (4), but redefine the peer-group area as a hexagonal area with 200-meter sides. Columns (3) and (6) use post-estimation data from Columns (1) and (4) to estimate regressions of residualized outcomes on two measures of average peer quality, constructed using hexagonal peer-group areas with 75-meter and 200-meter sides. Standard errors clustered at the block (peer-group) level are reported in parentheses.

TABLE 8 – Heterogeneous responses to peers (intensive-margin)

| Heterogeneity | Coeff. | Std. Err. | Mean $y$ | SD $y$ | SD $\bar{\alpha}$ | Heterogeneity | Coeff. | Std. Err. | Mean $y$ | SD $y$ | SD $\bar{\alpha}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A. Gender | | | | | | B. Family composition | | | | | |
| Men | 0.0490 | 0.0005 | 10.96 | 0.70 | 0.30 | No kid | 0.0545 | 0.0006 | 10.83 | 0.66 | 0.30 |
| Women | 0.0014 | 0.0005 | 10.67 | 0.64 | 0.29 | With kids | 0.0167 | 0.0005 | 10.81 | 0.70 | 0.30 |
| C. Immigration status | | | | | | D. Tenure in neighbourhood | | | | | |
| Native | 0.0214 | 0.0005 | 10.95 | 0.70 | 0.30 | Mover | 0.0323 | 0.0008 | 10.72 | 0.67 | 0.29 |
| Immigrant | 0.0300 | 0.0005 | 10.68 | 0.65 | 0.27 | Stayer | 0.0210 | 0.0004 | 10.84 | 0.69 | 0.30 |
| E. Neighbourhood income | | | | | | F. Peer group size | | | | | |
| Below 1st Q | 0.0454 | 0.0011 | 10.53 | 0.58 | 0.18 | Below 1st Q | 0.0141 | 0.0005 | 10.92 | 0.74 | 0.35 |
| 1st to 2nd Q | 0.0353 | 0.0012 | 10.73 | 0.60 | 0.14 | 1st to 2nd Q | 0.0270 | 0.0007 | 10.86 | 0.68 | 0.26 |
| 2nd to 3rd Q | 0.0198 | 0.0012 | 10.87 | 0.63 | 0.14 | 2nd to 3rd Q | 0.0396 | 0.0007 | 10.77 | 0.65 | 0.25 |
| Above 3rd Q | 0.0077 | 0.0009 | 11.12 | 0.77 | 0.25 | Above 3rd Q | 0.0647 | 0.0010 | 10.69 | 0.65 | 0.27 |

*Notes:* This table reports estimates of $\pi_0$ in equation (8) setting $K = 0$. The dependent variable is log annual earnings. As in Table 3, Column (2), all specifications include individual fixed effects and neighbourhood-by-year fixed effects. The sample includes all workers aged 22–55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Neighbourhoods are aggregates of 49 small hexagons, forming larger hexagonal areas with sides of approximately 525 meters and area of about 0.7 $\text{km}^2$. Standard errors clustered at the block (peer-group) level are reported in parentheses.

TABLE 9 – Heterogeneous responses to peers (extensive-margin)

| Heterogeneity | Coeff. | Std. Err. | Mean $y$ | SD $y$ | SD $\bar{\alpha}$ | Heterogeneity | Coeff. | Std. Err. | Mean $y$ | SD $y$ | SD $\bar{\alpha}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A. Gender | | | | | | B. Family composition | | | | | |
| Men | 0.0316 | 0.0009 | 0.81 | 0.39 | 0.10 | No kid | 0.0216 | 0.0013 | 0.81 | 0.39 | 0.11 |
| Women | 0.0378 | 0.0009 | 0.77 | 0.42 | 0.10 | With kids | 0.0405 | 0.0008 | 0.78 | 0.42 | 0.09 |
| C. Immigration status | | | | | | D. Tenure in neighbourhood | | | | | |
| Native | 0.0225 | 0.0009 | 0.83 | 0.38 | 0.10 | Movers | 0.0342 | 0.0013 | 0.79 | 0.41 | 0.09 |
| Immigrant | 0.0464 | 0.0010 | 0.75 | 0.43 | 0.09 | Stayers | 0.0344 | 0.0008 | 0.79 | 0.41 | 0.10 |
| E. Neighbourhood income | | | | | | F. Peer group size | | | | | |
| Below 1st Q | 0.0445 | 0.0013 | 0.72 | 0.45 | 0.12 | Below 1st Q | 0.0142 | 0.0009 | 0.79 | 0.41 | 0.13 |
| 1st to 2nd Q | 0.0197 | 0.0016 | 0.80 | 0.40 | 0.07 | 1st to 2nd Q | 0.0371 | 0.0014 | 0.80 | 0.40 | 0.08 |
| 2nd to 3rd Q | 0.0170 | 0.0015 | 0.82 | 0.39 | 0.07 | 2nd to 3rd Q | 0.0468 | 0.0014 | 0.80 | 0.40 | 0.07 |
| Above 3rd Q | 0.0181 | 0.0014 | 0.82 | 0.39 | 0.08 | Above 3rd Q | 0.0999 | 0.0016 | 0.76 | 0.43 | 0.10 |

*Notes:* This table reports estimates of $\pi_0$ in equation (8) setting $K = 0$. The dependent variable is an indicator for employment status. As in Table 3, Column (2), all specifications include individual fixed effects and neighbourhood-by-year fixed effects. The sample includes all individuals (both employed and non-employed) aged 22–55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Neighbourhoods are aggregates of 49 small hexagons, forming larger hexagonal areas with sides of approximately 525 meters and area of about $0.7\,\text{km}^2$. Standard errors clustered at the block (peer-group) level are reported in parentheses.

TABLE 10 – Non-Linear Effects Across the Human Capital Distribution

| | Log Annual Earnings | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Avg. Peer F.E. | 0.0251 | 0.0103 | 0.0124 | 0.0253 | 0.0751 | 0.0609 |
| | (0.0004) | (0.0008) | (0.0009) | (0.0008) | (0.0005) | (0.0009) |
| Frac. High Alpha ( > 33th pctile) | | 0.0275 | | | | |
| | | (0.0012) | | | | |
| Frac. High Alpha ( > 50th pctile) | | | 0.0123 | | | 0.0237 |
| | | | (0.0012) | | | (0.0012) |
| Frac. High Alpha ( > 67th pctile) | | | | -0.0006 | | |
| | | | | (0.0012) | | |
| Avg. Peer Firm F.E. X Own Alpha (> 50th pctile) | | | | | -0.0789 | -0.0649 |
| | | | | | (0.0006) | (0.0006) |
| Frac. High Alpha ( > 50th pctile) X Own Alpha (> 50th pctile) | | | | | | -0.0270 |
| | | | | | | (0.0002) |
| Time-varying controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Year F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Area×Year F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 | 19,637,000 |
| # of Peer Group-Years | 855,200 | 855,200 | 855,200 | 855,200 | 855,200 | 855,200 |
| # of Peer Group Locations | 64,700 | 64,700 | 64,700 | 64,700 | 64,700 | 64,700 |
| # of Individuals | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 | 2,592,000 |
| Average $y$ | 10.81 | 10.81 | 10.81 | 10.81 | 10.81 | 10.81 |
| SD $y$ | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 |

*Notes:* This table evaluates asymmetries in spillovers across the human capital distribution. Estimates in Column (1) replicate those reported in Table 3, Column (2), and are included for reference. Columns (2), (3), and (4) augment the baseline specification by including the fraction of peers whose fixed effects lie above the 33rd, 50th, and 67th percentiles of the overall $\alpha$ distribution, respectively. Column (5) reports estimates of $\pi_0$ in equation (9) setting $K = 0$ and allowing the effect of average peer quality to differ depending on whether an individual's own fixed effect lies above or below the median of the $\alpha$ distribution. Column (6) includes both average peer quality and the fraction of peers with $\alpha$ above the median, and allows both effects to differ by whether an individual's own fixed effect lies above or below the median. The sample includes all workers aged 22–55 living in the Toronto census metropolitan area between 2001 and 2018. We define an individual's peer group as all non-family members residing on the same residential block in a given year. Residential blocks $b$ are defined as microgeographic hexagonal areas with 75-meter sides. Neighbourhoods are aggregates of 49 small hexagons, forming larger hexagonal areas with sides of approximately 525 meters and area of about 0.7 $\text{km}^2$. Standard errors clustered at the block (peer-group) level are reported in parentheses.

## Online Appendix A Estimation

In this technical appendix we lay out the details of the updating rule used for $\alpha_{i,0}$ in the simple version of the model without covariates or variation in treatment effects as a function of exogenous characteristics. These are easy extensions to add.

Individual $i$ is characterized by a constant unobserved random variable $\alpha_{i,0}$. Conceptually, we think of it as innate ability and skills accumulated prior to labour market entry. This fixed component of human capital is allowed to spill over directly through peer effects.

Let $M_{i,t}$ denote the set of individual $i$'s peers at time $t$, which includes all non-family members residing in the same block $b(i,t)$ in that year. Spillovers are linear-in-means and average peer quality, $\bar{\alpha}_{i,t}$, is defined as:

$$\bar{\alpha}_{i,t} = \sum_{j \in M_{i,t}} \frac{1}{|M_{i,t}|} \alpha_{j,0}$$

We have the following data generating process for an outcome $y$ for individual $i$ residing in block $b$ at time $t$:

$$y_{i,b,t} = \alpha_{i,0} + \sum_{\tau=0}^{t-1} \pi_\tau \bar{\alpha}_{i,t-\tau} + \phi_{B(b),t} + \varepsilon_{i,b,t}$$

In the empirical work, we set $\pi_\tau = 0$ if $\tau > K$ for computational ease.

The nonlinear least square estimator minimizes the following objective function:

$$\sum_{i \in I} \sum_{t \in T_i} L_{i,t}^2$$

where

$$L_{i,t} = \left( y_{i,b,t} - \alpha_{i,0} - \phi_{B(b),t} - \sum_{\tau=0}^{t-1} \pi_\tau \bar{\alpha}_{i,t-\tau} \right)$$

The first-order condition with respect to $\alpha_{i,0}$ is

$$0 = -2 \sum_{t \in T_i} L_{i,t} - 2 \sum_{t \in T_i} \sum_{j \in M_{i,t}} \sum_{\tau=0}^{T_j - t} L_{j,t+\tau} \frac{\pi_\tau}{|M_{j,t}|}$$

Define $L_{i,t}^{\setminus\{\alpha_{j,0}\}}$ as $L_{i,t}$ minus the terms containing $\alpha_{j,0}$. The first-order condition can be

written as:

$$0 = \sum_{t \in T_i} L_{i,t}^{\setminus\{\alpha_{i,0}\}} - T_i \alpha_{i,0} + \sum_{t \in T_i} \sum_{j \in M_{i,t}} \sum_{\tau=0}^{T_j - t} \frac{\pi_\tau}{|M_{j,t}|} \left( L_{j,t+\tau}^{\setminus\{\alpha_{i,0}\}} - \left( \sum_{t'=1}^{t+\tau} \frac{\pi_{t+\tau-t'}}{|M_{j,t'}|} \mathbb{1}\left\{ b(j,t') = b(i,t') \right\} \right) \alpha_{i,0} \right)$$

Solving for $\alpha_{i,0}$ we get:

$$\alpha_{i,0} = \frac{\sum_{t \in T_i} L_{i,t}^{\setminus\{\alpha_{i,0}\}} + \sum_{t \in T_i} \sum_{j \in M_{i,t}} \sum_{\tau=0}^{T_j - t} \frac{\pi_\tau}{|M_{j,t}|} L_{j,t+\tau}^{\setminus\{\alpha_{i,0}\}}}{T_i + \sum_{t \in T_i} \sum_{j \in M_{i,t}} \sum_{\tau=0}^{T_j - t} \frac{\pi_\tau}{|M_{j,t}|} \sum_{t'=1}^{t+\tau} \frac{\pi_{t+\tau-t'}}{|M_{j,t'}|} \mathbb{1}\left\{ b(j,t') = b(i,t') \right\}}$$

For the case with missing out outcomes (but individuals are still in the peer group), this expression is:

$$\alpha_{i,0} = \frac{\sum_{t \in T_i} \mathbb{1}\left\{ y_{i,b,t} \neq \cdot \right\} L_{i,t}^{\setminus\{\alpha_{i,0}\}} + \sum_{t \in T_i} \sum_{j \in M_{i,t}} \sum_{\tau=0}^{T_j - t} \frac{\pi_\tau}{|M_{j,t}|} \mathbb{1}\left\{ y_{j,t+\tau} \neq \cdot \right\} L_{j,t+\tau}^{\setminus\{\alpha_{i,0}\}}}{\sum_{t \in T_i} \mathbb{1}\left\{ y_{i,b,t} \neq \cdot \right\} + \sum_{t \in T_i} \sum_{j \in M_{i,t}} \sum_{\tau=0}^{T_j - t} \frac{\pi_\tau}{|M_{j,t}|} \mathbb{1}\left\{ y_{j,t+\tau} \neq \cdot \right\} \sum_{t'=1}^{t+\tau} \frac{\pi_{t+\tau-t'}}{|M_{j,t'}|} \mathbb{1}\left\{ b(j,t') = b(i,t') \right\}}$$