Nix, Emily

Working Paper

Learning spillovers in the firm

Working Paper, No. 2020:14

Provided in Cooperation with:
IFAU - Institute for Evaluation of Labour Market and Education Policy, Uppsala

Suggested Citation: Nix, Emily (2020) : Learning spillovers in the firm, Working Paper, No. 2020:14, Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala

This Version is available at:
https://hdl.handle.net/10419/246023

WORKING PAPER 2020:14

Learning spillovers in the firm

Emily Nix

IFAU
INSTITUTE FOR
EVALUATION OF
LABOUR MARKET AND
EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Learning spillovers in the firm

Emily Nix*

Abstract

To produce output for a firm, coworkers often interact. This paper examines the possibility that as a byproduct of these interactions, there are learning spillovers: coworkers learn general skills from each other that increase future productivity. In the first part of the paper I show that learning spillovers imply externalities in the return to human capital which firms may not internalize when there is asymmetric information. As a result, individuals may inefficiently invest in their own education. Next, I show that learning spillovers are empirically relevant. Using matched administrative data from Sweden and a combination of fixed effects and controls to address bias from worker sorting and firm heterogeneity, I find that increasing the average education of a given worker's coworkers by 10 percentage points increases that worker's wages in the following year by 0.3%, which is significant at the 1% level. The effect is persistent, decreases with age, and is higher for workers in occupations where they interact more regularly with their coworkers.

JEL Codes: E24, J24.

Keywords: Human Capital Accumulation, Diffusion of Knowledge, Learning.

*University of Southern California. Email: enix@usc.edu.

I am grateful to Joseph Altonji, Costas Meghir, Lisa Kahn, and Nancy Qian for their invaluable guidance and support. This research is based upon work completed at the Institute for Evaluation of Labour Market and Education Policy in Uppsala, Sweden, and I am particularly indebted to Olof Åsland, Anders Forslund, Lisa Laun, and Jörgen Moen. For helpful comments, I am also thankful to Noriko Amano, Vittorio Bassi, Thomas Cornelissen, Benjamin Friedrich, Eric Grönqvist, Ran Gu, Alex Mas, Rebecca McKibbon, Corina Mommaerts, Ana Reynoso, Martin Söderström, and seminar participants at Barcelona GSE, Census Bureau, Federal Trade Commission, IFAU, IIES Stockholm, Indiana University, LSE, McGill, Notre Dame, UCLA, University of Illinois Urbana-Champaign, University of Namur, USC Economics, USC Marshall, University of South Carolina, UVA, and Yale. I gratefully acknowledge financial support from a National Science Foundation Graduate Research Fellowship, the Yale Economics Department, and the Institution for Social and Policy Studies.

1 Introduction

Producing output in groups is a mainstay of modern economies. To produce output, coworkers often interact, possibly learning from and teaching one another. In this paper, I focus on these potential "learning spillovers" in firms. The idea that peers in one's firm might be an important source of human capital accumulation is not new. For example, Acemoglu (1996, p. 782) states that "excluding education and R&D, major human capital interactions happen among employees within a firm: for example, young workers learn from their more experienced coworkers. But these interactions should be internalized within the firm, and no economy wide human capital externalities should be observed".

This paper provides a theoretical and empirical analysis of learning spillovers in firms. I first define a simple model of learning spillovers in the firm. This yields the novel result that learning spillovers may not be fully internalized. Second, I construct a unique data set and use a combination of fixed effects and controls to show that learning spillovers are empirically relevant: increasing the average education[1] of a given worker's coworkers by 10 percentage points increases that worker's wages in the following year by 0.3%, which is significant at the 1% level. Third, I provide conditions under which social returns to education may exceed private returns to education and decompose the social returns to education into the part due to the direct effect of a college education versus the part due to learning spillovers in the firm. I find that the social returns of adding an additional college-educated worker range from 0.194 to 0.222, with 12.61% to 14.43% of the total increase attributable to learning spillovers.

In my theoretical model, workers increase their stock of general skills as a byproduct of working together to produce output for firms. The amount of general skills workers obtain (the size of learning spillovers) depends on the average education of the firm. I use a general equilibrium framework to solve for wages and find that in contrast to the consensus in the literature, learning spillovers are not straightforward for firms to internalize.[2] Three conditions make it particularly challenging for firms to internalize learning spillovers. First, learning spillovers increase future productivity, even after a worker leaves the current firm. Second, the size of the spillover depends on a worker's learning type, which may not be known to the firm. Third, coworkers are non-excludable and (partially) non-rival inputs in the production of learning spillovers. Under these condi-

[1] Average education is measured by the total number of college (or more) educated workers in the worker's plant divided by the total number of workers in the plant.

[2] Later in the introduction I review statements consistent with the prior view in the literature that these spillovers will be internalized, from Acemoglu (1996), Moretti (2004b), Barro (1996), and Lange and Topel (2006).

tions, coworkers impose externalities on each other that are particularly challenging to internalize. As a result, individuals may not take the full extent of education externalities into account when making education decisions and the number of educated workers in a competitive equilibrium may be inefficient.

To test for learning spillovers, I use Swedish administrative data to construct a unique data set covering the universe of workers, their peers, and firms in Sweden from 1985 to 2012. Motivated by the theoretical model, I test for learning spillovers by looking at the relationship between the education level of past coworkers and current wages. To control for unobserved time-invariant firm heterogeneity and worker sorting, I include firm and worker fixed effects. To address time-varying omitted variables, I include county$\times$time and industry$\times$time dummies.

I find that increasing the average education of a given worker's colleagues by 10 percentage points increases that worker's wages in the following year by approximately 0.3%, which is significant at the 1% level. This result stands up to a number of controls and robustness checks. The effect is also persistent. Average education of coworkers impacts wages at least seven years into the future. Compared to average wage growth in Sweden over this time period of roughly 1.7% to 2% per year, this estimated effect is non-negligible.

In addition, I document heterogeneity by age and occupation that is consistent with learning spillovers. The spillover is largest for younger workers, for whom human capital accumulation is most important, with no impact for workers who are older than 40. Using data from O*NET I construct a ranking of occupations by importance of interactions with coworkers. I find that on average workers in occupations that have higher interpersonal rankings according to O*NET also receive greater learning spillovers. For example, professionals and managers obtain the largest spillovers from their coworkers, while drivers, who interact little with coworkers, experience the smallest impact.

These findings have important implications for the returns to education. Using the theoretical results from the first part of the paper, I present conditions under which the social and private returns to college are not perfectly aligned. I combine these conditions with my empirical estimates of learning spillovers to provide bounds for the social returns to college. I then decompose the social returns to college into the fraction attributable to learning spillovers versus the fraction attributable to the direct increase in productivity that adding a college-educated worker versus high-school-educated worker to the workplace confers. My findings suggest that the social return of adding an additional college worker in terms of income ranges from 0.195 to 0.22, with 11.36% to 12.82% of the total increase attributable to learning spillovers.

This paper contributes to the literature estimating peer effects in the workplace. There is little doubt that peers shape individual outcomes, as demonstrated by the massive body of evidence on peer effects in schools and within neighborhoods. There is comparatively less evidence on peer effects at work, although this literature is rapidly growing. Bandiera, Barankay, and Rasul (2010) find that when working with friends, workers are more productive when paired with their more able friends. Mas and Moretti (2009) use high-frequency data from a supermarket chain and find strong evidence of productivity spillovers, primarily driven by internalization of free-riding externalities.[3] Azoulay, Graff Zivin, and Wang (2010) estimate sizable impacts from the loss of an academic superstar on his or her collaborators; in contrast, using variation induced by the Nazi government's expulsion of scientists from Germany, Waldinger (2012) finds no peer effects in academic departments. Amodio and Martinez-Carrasco (2018) show that the quality of other inputs has important interactions with the amount of learning spillovers from colleagues. Jarosch, Oberfield, and Rossi-Hansberg (2019) show that having higher paid coworkers is associated with future wage growth both in reduced form analysis and when they add more structure when estimating a model of firm production with learning spillovers. Martins and Jin (2010) estimate contemporaneous social returns to education in firms in Portugal and find large social returns, between 14% and 23%.[4]

Even more closely related, Brune, Chyn, and Kerwin (2020) randomly vary peers in tea production in Malawi and find that a 10% increase in average education of one's peers increases one's own productivity by 0.3%, although they find this is primarily driven by a motivation effect. Jackson and Bruegmann (2009) find that students in classrooms led by teachers exposed to better colleagues experience larger test gains.[5] Cornelissen, Dustmann, and Schönberg (2017) estimate contemporaneous social returns to peer fixed effects in firms in Germany and find small evidence of peer effects on wages. Herkenhoff, Lise, Menzio, and Phillips (2018) estimate a frictional labor market featuring a nonlinear production function in which workers learn from more knowledgeable coworkers but are not inhibited by less knowledgeable ones, and find in equilibrium that there is inefficient sorting across firms.

Relative to these papers, I add the result that in some cases learning spillovers from coworkers may not be fully internalized, and may cause inefficient prior investments in

[3]Herbst and Mas (2015) summarize 35 different studies and find consistent evidence that an increase in coworker productivity consistently increases a worker's own productivity.

[4]There have also been a few excellent papers leveraging the sports environment to identify peer effects. See, for example, Arcidiacono, Kinsler, and Price (2017) and Guryan, Kroft, and Notowidigdo (2009).

[5]In a later paper which conducts an experiment pairing low- and high-skilled teachers together, Papay, Taylor, Tyler, and Laski (2020) find a similar result, that test scores rise for students of low-skilled teachers who were paired with high-skilled teachers.

human capital. This result stands in contrast to numerous prior statements to the opposite effect in the literature. For example, Moretti (2004b, p. 661) states that "potential spillovers that occur within a plant...are likely to be internalized". Barro (1996, p. 147) says that "the spillover cannot represent just the ill effect of incompetent oldsters on aspiring youngsters within a firm (an interaction that would be internalized by the firm's wage policy), but must involve more wide-ranging effects that require government intervention". Lange and Topel (2006, p. 462) summarize the literature by stating that "when productive interactions occur within firms they are merely complementarities that will be internalized and priced". I also provide additional empirical proof that these learning spillovers within the firm are important sources of human capital accumulation into the future, while a number of the existing empirical papers discussed in the preceding paragraphs focus primarily on the concurrent impact of peers.

The idea that there may be on-the-job learning from coworkers that makes workers more productive in the future is also related to the much larger learning by doing literature (e.g., Thornton and Thompson (2001); Levitt, List, and Syverson (2013); Haggag, McManus, and Paci (2017); and many others). Learning by doing is thought to be a key source of human capital accumulation and growth (Lucas Jr, 1988), and in this paper I show that "learning from others while doing" may also play an important role.

The paper is organized as follows: In Section 2 I define a theoretical model with learning spillovers and show that under realistic conditions this leads to under investment in education. In Section 3 I use the theoretical results to motivate my empirical model, describe the threats to identification, and outline an estimation strategy. In Section 4 I describe the data construction and present descriptive evidence. In Section 5 I present the main results. In Section 6 I use the estimates to bound the private returns to education and in Section 7 I conclude.

2 A Model of Learning Spillovers in the Firm

In this section I write and solve a simple model that demonstrates how and why learning spillovers in the firm may not be internalized, causing inefficient investments in human capital. The model is overly simplified in order to make this theoretical point and will not be estimated directly, although I will use aspects of the model to inform the empirical analysis. The main aspect of the model is that workers may "learn from others while doing," in that workers learn general skills from their coworkers that make them more productive in the future. For example, suppose that I work with a co-author to write a paper. By working together on the paper, I also learn some tactics and new skills from my

co-author that will help me write a better paper on my own. These are general skills that will make me more productive in the future even when I am not working directly with this co-author.

In the model, the economy consists of a continuum of workers denoted by $i \in I$ and a continuum of identical firms denoted by $f \in J$. There are three periods. In the first period workers choose to get high education or low education at an individual specific cost θ^i which for all individuals is uniformly distributed over $[0, 1]$ and is uncorrelated with the amount each individual will learn on the job.

In the second period workers are hired by firms, receive a wage, consume, and as a byproduct, learn from their coworkers. The amount workers learn from their peers depends on their learning type, α, and the average education of the peers in their firm, S_f. Half of the population are A types who learn more from a given average education in the firm than the other half of the population, who we label B types. Firms hire college and high school workers of both types, $H_f = H_f^A + H_f^B$ and $L_f = L_f^A + L_f^B$, respectively. Letting

$$\bar{S_f} = \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B}$$

denote the average education at the firm, A types receive learning spillovers

$$s_f^A = \alpha^A \bar{S_f} \tag{1}$$

and B types receive learning spillovers

$$s_f^B = \alpha^B \bar{S_f} \tag{2}$$

where α is the learning parameter and $\alpha^A > \alpha^B$.[6] These skills are general and increase the worker's human capital, which makes her more productive in the third and final period, when workers simply work for a wage and consume (for simplicity there are no learning spillovers in the third period).

Turning to the firms, in the second period firms hire workers in order to produce consumption goods. The J firms are all identical. This assumption combined with assumptions on total production (outlined in Appendix B.1) allows me to rule out sorting driven by learning spillovers. I do so in order to show why learning spillovers may not be internalized in the simplest setting possible, but I discuss some possible implications

[6] I discuss the theoretical and empirical reasons for the particular functional form I chose for learning spillovers in Appendix B.2.

of relaxing this assumption in the conclusion.

Consumption goods are produced by firms using college-educated labor hired by a firm f, $H_f = H_f^A + H_f^B$, and high-school-educated labor hired by a firm f, $L_f = L_f^A + L_f^B$. The amount produced is given by $F\left(H_f, L_f\right)$, which is constant returns to scale. As a byproduct of being hired to produce consumption goods, these same college and high school graduates also gain on-the-job learning spillovers from each other, as given in equations 1 and 2. These learning spillovers enter the problem in two ways. First, they impact total production of consumption goods in the second period.[7] I assume that each worker's marginal productivity increases by exactly the amount of his learning spillover. Thus, with spillovers, total second period production of consumption goods at a firm f is

$$F\left(H_f, L_f\right) + \underbrace{\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\bar{S_f}}_{second\ period\ spillovers} \tag{3}$$

which is also constant returns to scale, given that F is constant returns to scale.[8]

Second, the spillovers increase production in the third period, but subject to depreciation, denoted δ. Thus, the total increase in consumption goods produced in the second and third periods due to learning spillovers is given by:

$$\underbrace{\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\bar{S_f}}_{second\ period\ spillovers} + \underbrace{\delta\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\bar{S_f}}_{third\ period\ spillovers}$$

For simplicity, in my theoretical model I assume that individuals simply consume their learning spillovers in the third period. This captures the fact that learning spillovers increase future wages, without having to explicitly model wages in future periods.

Individuals all have the same linear utility functions over the three periods:

$$U^i = c_1^i + c_2^i + c_3^i$$

There are perfect credit markets, the interest rate is 0, and there is no discounting.

[7] One can easily remove the second period spillovers and the results remain. Based on the prior literature which suggests important concurrent spillovers (see the discussion in the introduction) I include them in my model.

[8] Note that an alternative way of incorporating the spillovers would be to write: $F\left(H_f + \alpha^A H_f^A \bar{S_f} + \alpha^B H_f^B \bar{S_f}, L_f + \alpha^A L_f^A \bar{S_f} + \alpha^B L_f^B \bar{S_f}\right)$. This should not change the results, so for simplicity I use the current specification.

2.1 Competitive Equilibrium with Learning Spillovers and Implications for Efficiency

First, I present a worst-case (and implausible) scenario where the externalities generated by learning spillovers are ignored. As expected, when I solve for these conditions in Proposition 1, equilibrium wages do not include the marginal productivity of each worker in terms of producing future learning spillovers. As a result, the outcome is inefficient. Workers underinvest in education. Note that the inefficiency arises here because education, which produces the spillovers, is endogenously chosen, and produces the learning spillovers for coworkers.[9]

Proposition 1. *Suppose that firms are unaware of the learning spillovers workers are providing to each other and do not attempt to adjust wages accordingly. In that case, the competitive equilibrium exists and is unique, but is not Pareto efficient. Workers underinvest in education. Equilibrium wages by education and type are:*

$$\begin{aligned} w_f^{H^K} &= F_1 + \alpha^K \bar{S_f}^* \\ &\quad + \left(\alpha^A \left(H_f^{A*} + L_f^{A*}\right) + \alpha^B \left(H_f^{B*} + L_f^{B*}\right)\right) \left(\frac{1}{H_f^* + L_f^*} - \frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}\right) \\ w_f^{L^K} &= F_2 + \alpha^K \bar{S_f}^* \\ &\quad - \left(\alpha^A \left(H_f^{A*} + L_f^{A*}\right) + \alpha^B \left(H_f^{B*} + L_f^{B*}\right)\right) \left(\frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}\right) \\ K &= A, B \end{aligned}$$

In addition, workers receive their type-specific learning spillovers in the third period.
Proof: See Appendix *A.2*.

In order for a competitive equilibrium to get the right amount of college-educated workers, it must provide the right incentives to go to college, with the effects of learning spillovers on future productivity priced into wages. For college workers, this should imply an increase in wages. College workers increase average education in the firm, and thus impose positive externalities on coworkers. For high-school-educated workers, this

[9] More generally, a competitive equilibrium with learning spillovers is guaranteed to be efficient whenever the spillovers do not depend on prior investments. Under this condition, the total amount produced is correct whether or not learning spillovers are internalized in wages. How the surplus from learning spillovers is divided among workers merely moves the competitive equilibrium along the Pareto frontier.

should imply a decrease in wages. Such workers decrease average education in the firm, and thus impose negative externalities on coworkers. In the next proposition, I assume that firms perfectly observe workers' learning types and can pay personalized wages to account for the total amount learned by each worker. This is similar to the conditions for a Lindahl equilibrium for public goods (Lindahl, 1919).[10]

Firms maximize profits relative to each worker's participation constraint. The participation constraints are determined by the workers' problems. Workers work at a given firm f in the second period if the total compensation provided by that firm exceeds their reservation compensation level, w^{H^A}, w^{H^B}, w^{L^A}, and w^{L^B}, which they take as given. These reservation compensations are determined in equilibrium.

Total compensation provided by a given firm includes wages paid plus the learning spillovers workers receive and consume in the third period. Learning spillovers are subject to depreciation, given by δ. Thus, the participation constraints by education and learning type are:

$$\begin{aligned} w_f^{H^K} + \delta \alpha^K \bar{S}_f &\geq w^{H^K} \\ w_f^{L^K} + \delta \alpha^K \bar{S}_f &\geq w^{L^K} \\ K &= A, B \end{aligned}$$

These conditions make explicit the trade-off between wages and spillovers that in turn affects the firm's demand for each type of worker by education level. Under these conditions, I prove the following proposition:

Proposition 2. *Suppose that firms have perfect information on workers' learning types and can pay personalized wages by education and learning type. Then the competitive equilibrium exists, is unique, and is Pareto efficient. The equilibrium wages by education and type are:*

$$\begin{aligned} w_f^{H^K} &= F_1 + \alpha^K \bar{S}_f{}^* \\ &\quad + (1+\delta)\left(\alpha^A\left(H_f^{A*} + L_f^{A*}\right) + \alpha^B\left(H_f^{B*} + L_f^{B*}\right)\right)\left(\frac{1}{H_f^* + L_f^*} - \frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}\right) \\ w_f^{L^K} &= F_2 + \alpha^K \bar{S}_f{}^* \\ &\quad - (1+\delta)\left(\alpha^A\left(H_f^{A*} + L_f^{A*}\right) + \alpha^B\left(H_f^{B*} + L_f^{B*}\right)\right)\left(\frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}\right) \end{aligned}$$

[10] Note that this implicitly assumes that firms are able to charge workers for the externality (Coase, 1960).

$$K = A, B$$

In addition, workers receive their type-specific learning spillovers in the third period.
Proof: See Appendix A.3.

The intuition for the result is straightforward. Since firms are able to trade off paying workers in wages versus providing learning spillovers (as shown in the worker participation constraints), this drives up demand for college-educated workers relative to high-school-educated workers. In addition, by effectively restricting each type of worker to purchase only the spillover that will be gained by that learning type (through the type-specific deductions in wages), firms are able to deduct more from high-learning workers than from low-learning workers. This in turn drives up demand for high-learning workers relative to low-learning workers.

Combined, these mechanisms result in an equilibrium that appropriately internalizes learning spillovers into wages. The equilibrium is able to account for the total amount learned and the relative contribution of college- and high-school-educated workers by directly internalizing the amount learned by each type. As a result, the incentives for college education are correct and the outcome is efficient.

However, in reality it is unlikely that firms have perfect information on each worker's learning type, as was assumed for Proposition 2. In the third and last result, I assume that there is asymmetric information. Under these conditions, will workers voluntarily reveal what learning type they are? Formally, are different contracts incentive compatible, where contracts consist of type-specific wages and the amount of spillover a given type receives from exposure to the average education of the firm's workers:

$$\begin{aligned} w_f^{H^B} + \delta\alpha^A\bar{S}_f &\geq w_f^{H^A} + \delta\alpha^A\bar{S}_f \\ w_f^{H^A} + \delta\alpha^B\bar{S}_f &\geq w_f^{H^B} + \delta\alpha^B\bar{S}_f \\ w_f^{L^B} + \delta\alpha^A\bar{S}_f &\geq w_f^{L^A} + \delta\alpha^A\bar{S}_f \\ w_f^{L^A} + \delta\alpha^B\bar{S}_f &\geq w_f^{L^B} + \delta\alpha^B\bar{S}_f \end{aligned}$$

These incentive compatibility constraints imply that

$$\begin{aligned} w_f^{H^A} &= w_f^{H^B} \\ w_f^{L^A} &= w_f^{L^B} \end{aligned}$$

which means that firms cannot induce workers to reveal their learning type by offering different contracts. The reason a separating equilibrium is not possible is because all

workers within a firm are exposed to the same average education, regardless of their type. Given that, workers will always claim to be whatever type receives the highest wage.

This results in the following updated worker participation constraints, when there is asymmetric information:

$$w_f^H \geq w^{H^K} - \delta\alpha^A\bar{S}_f \tag{4}$$
$$w_f^H \geq w^{H^B} - \delta\alpha^B\bar{S}_f \tag{5}$$
$$w_f^L \geq w^{L^A} - \delta\alpha^A\bar{S}_f \tag{6}$$
$$w_f^L \geq w^{L^B} - \delta\alpha^B\bar{S}_f \tag{7}$$

In summary, the personalized wages that allow for an efficient outcome in Proposition 2 are similar to the personalized prices required for a Lindahl equilibrium for public goods. Naturally, the concerns are also similar (asymmetric information and thin markets). Insofar as a Lindahl equilibrium is not a realistic solution to the public goods problem, firms will not be able to fully internalize learning spillovers by paying personalized wages. More formally, I obtain the following solution under asymmetric information:

Proposition 3. *Suppose there is asymmetric information such that workers know their learning parameters and firms do not. Then efficiency is no longer a guaranteed outcome in the competitive equilibrium. The set of possible equilibrium wages by education are:*

$$w_f^H = F_1 + E[\alpha] + \delta\alpha^B\frac{L_f^*}{H_f^* + L_f^*} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L_f^*}{\left(H_f^* + L_f^*\right)^2} \tag{8}$$
$$w_f^L = F_2 + E[\alpha] - \delta\alpha^B\frac{H_f^*}{H_f^* + L_f^*} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H_f^*}{\left(H_f^* + L_f^*\right)^2} \tag{9}$$

with

$$\begin{aligned}\lambda_1 &\in \left[0, H_f^*\right]\\ \lambda_2 &= H_f^* - \lambda_1\\ \lambda_3 &\in \left[0, I - H_f^*\right]\\ \lambda_4 &= I - H_f^* - \lambda_3\end{aligned}$$

where λ_1 is the Lagrange multiplier on the college-educated, high-learning-type participation constraint, λ_2 is the Lagrange multiplier on the college-educated, low-learning-type participation

constraint, λ_3 is the Lagrange multiplier on the high-school-educated, high-learning-type participation constraint, and λ_4 is the Lagrange multiplier on the high-school-educated, low-learning-type participation constraint. In addition, workers receive their type-specific learning spillovers in the third period.

Proof. See Appendix A.4.

In summary, the conclusions from these three propositions are that learning spillovers may be challenging for firms to internalize. In a plausible setting with asymmetric information, a competitive equilibrium is unlikely to be efficient (see Proposition 3). A few additional points are relevant. First, the challenges in markets with learning spillovers do not occur with traditional training inputs, as I confirm in Appendix B.3. The intuition is that since firms can choose different amounts of traditional inputs for different workers, firms can simply announce a menu of input amounts and corresponding prices. Asymmetric information is not an issue since it is incentive compatible for workers to choose different packages according to their learning types. Second, the inefficiency only occurs when learning spillovers impact future productivity and are produced by human capital which is endogenously chosen. Otherwise the result is always efficient. Third, if the competitive equilibrium does not fully internalize learning spillovers, social returns may exceed private returns. I return to this point and provide some estimates based on both the results in this section and my empirical results in Section 6.

3 Empirical Framework

The goal of the empirical section is to test for the existence and importance of learning spillovers. All else equal, does a given worker exposed to more-educated coworkers learn more general skills relative to an identical worker exposed to less-educated coworkers? The theoretical model suggests looking at the effect of average education of coworkers in a worker's firm in the previous year on his current wage. This follows from the model, which shows that the effect of current coworkers on current wages is ambiguous.[11] It depends on the complementarity of inputs in producing consumption goods, the degree to which spillovers are internalized, the size of learning spillovers, and a worker's own education. Given these opposing forces, both positive and negative coefficients on current coworkers could be consistent with a model of learning spillovers.

In contrast, the theoretical model predicts that if there are learning spillovers such that past coworkers increase one's human capital and make one more productive, then this

[11] See equations 4–7 and 8 and 9.

will cause an increase in wages in future periods provided wages capture the productivity of workers. All else equal, the theoretical model therefore predicts that a worker exposed to a firm with higher average education in the previous year will experience higher wages in the current year. This implies the following regression, where wages in a given year t, for a given worker i can be written as

$$w_{it} = \pi_1 \bar{H}_{it-1} + \pi_0 h_i + d_t + \varepsilon_{it} \tag{10}$$

where d_t represents year dummies, $\bar{H}_{it-1}$ denotes the average education the worker was exposed to in his firm(s) last year, and h_i denotes the individual worker's own education.

An estimate of $\hat{\pi}_0 > 0$ implies that college graduates are more productive than high school graduates ($F_1 > F_2$), but may also capture the positive externality highly educated workers impose on coworkers relative to the negative externality less educated workers impose.[12] In contrast, the coefficient on average education the worker was exposed to in his firm(s) last year only captures learning spillovers. If π_1 is unbiased, $\hat{\pi}_1 = 0$ implies there are no learning spillovers and $\hat{\pi}_1 > 0$ implies there are positive learning spillovers.

There are two reasons why OLS estimates of 10 will be biased. First, there could be time-invariant omitted variables, such as worker sorting and unobserved firm heterogeneity. I use either firm$\times$worker or firm and worker fixed effects to deal with any time-invariant omitted variables. For example, individual fixed effects deal with upward bias from workers with higher ability sorting into more educated firms. Firm fixed effects deals with the possibility that firms employing more-educated workers also provide more formal training opportunities.

Second, there could be time-varying omitted variables. For example, suppose that increases in average education within firms are driven by influxes of college graduates into certain counties. This could drive up both the average education of workers in firms in treated counties and increase demand for local goods, which may also drive up future wages. Alternatively, suppose there is skill-biased technological change. Skill-biased technological change would affect both the number of college graduates (through an increase in demand for college graduates) and the returns to skill. While year dummies will capture general trends in skill-biased technological change, if its intensity varies by industry my estimates may be biased upward.

To control for these additional sources of bias, I include county$\times$time fixed effects, d_{ct},

[12] $\hat{\pi}_0$ will only capture learning spillovers insofar as these spillovers are internalized. I return to this point in Section 6.

and industry×time fixed effects, d_{kt}. This leads to the following regression:

$$w_{it} = \pi_1 \bar{H}_{it-1} + \pi_x X_{it} + d_t + d_{i\mathbf{f}(i,t-1)} + d_{ct} + d_{kt} + \varepsilon_{cfikt} \quad (11)$$

X_{it} is a vector of time-varying individual controls consisting of number of children and marital status. d_t represents year dummies. d_{ct} is the county×time dummy, and d_{kt} is the industry×time dummy. In my most robust specification I estimate a county×industry×time dummy, d_{ckt}. In equation 11, $d_{i\mathbf{f}(i,t-1)}$ is a firm (where the worker was employed last year) by worker fixed effect.

In an alternative specification, I estimate separate firm and worker fixed effects:

$$w_{it} = \pi_1 \bar{H}_{it-1} + \pi_x X_{it} + d_t + d_i + d_{\mathbf{f}(i,t-1)} + d_{ct} + d_{kt} + \varepsilon_{cfikt} \quad (12)$$

where d_i are individual fixed effects and $d_{\mathbf{f}(i,t-1)}$ are firm fixed effects for the firm in which the worker was employed in the previous year, with individual and firm fixed effects estimated separately. The identification and estimation of firm, worker, and time fixed effects was pioneered by Abowd, Kramarz, and Margolis (1999) and I estimate these AKM fixed effects in equation 12 similarly to other papers in this literature. More details can be found in Appendix C.2.

I report estimates using either firm×worker fixed effects or firm and worker fixed effects because the two approaches identify the coefficient using different variations in the data.[13] Controlling for firm×worker matches restricts the identifying variation to movement of coworkers in and/or out of a given worker's firm that is not captured by industry×time and county×time trends. Estimating separate firm and worker fixed effects identifies the coefficient using changes in the average education workers are exposed to both from workers who moved firms as well as changes in coworkers among workers who stay in their current firm. While each approach may have its own distinct problems, if I find consistent results using both approaches, this provides reassurance that I am picking up a true underlying effect.[14]

[13] In contrast, most papers estimate AKM fixed effects instead of firm×worker fixed effects because they are interested in either the firm and worker fixed effects (d_i and $d_{\mathbf{f}(i,t-1)}$) themselves or estimates of time-invariant variables. Since I am not directly interested in those estimates, firm×worker fixed effects are technically sufficient to control for worker and firm fixed effects.

[14] Different results could either be cause for concern or simply indicate heterogeneous treatment effects. For example, learning spillovers may be larger for workers who experience an increase in average coworker education because they change firms compared to a worker who experiences a similar increase from a change in a few coworkers at his existing firm. A reason this could be true is if college workers all have more skills, but also have different types of skills, so that switching firms provides exposure to new coworkers with different skills.

The inclusion of firm, worker, time, county×time, and industry×time fixed effects naturally limits the scope for omitted variable bias.[15] In order to bias my estimates, an omitted variable must meet all of the following conditions at the same time:

1. Time varying;
2. Correlated with changes in future wages;
3. Correlated with changes in current average education in workers' firms; and
4. Not captured by the industry×time and county×time, or county×industry×time fixed effects.

While an omitted variable that fits all four of these conditions at the same time is unlikely, it is not impossible. For example, suppose that a given firm experiences a positive demand shock for its product. To meet the demand, the firm hires more workers. For some reason, the firm chooses to hire more college than high school graduates relative to its existing ratio of such workers, increasing the average education within the firm. However, due to labor market frictions the firm can't hire as many college graduates as it would like. This in turn causes the firm to increase its training of existing workers to increase their productivity. To address this possibility and others like it, in Sections 5.2 and 5.3 I document heterogeneity in the effect by occupation and age that is consistent with learning spillovers but is not consistent with alternative explanations.

Finally, one might worry about measurement error in a given worker's own education.[16] As shown in Griliches (1977) and extended to the peer effects framework in Acemoglu and Angrist (2001), estimates of social returns functions are biased upward if there is measurement error in a given worker's own education. This is unlikely to be a concern in my setting because I use administrative data. Furthermore, I show that the fixed effects remove any upward bias from measurement error.[17] Given the possible broader applicability of this solution to the upward bias in social returns estimates, in Appendix C.1 I derive the result formally, outline the conditions when it can be used successfully, and also demonstrate its usefulness though a simple simulation exercise.

[15] It also eliminates some of the true variation in average education.

[16] Measurement error in average education of past coworkers could also introduce bias. In Section 4, I discuss how I construct this variable in more detail, why it may be subject to measurement error, and why the expected bias is downward.

[17] Individual fixed effects control perfectly for time-invariant characteristics. Thus, to the extent that the worker's own education is time invariant, individual fixed effects control for it perfectly. As a result, measurement error in his own education only biases estimates of π_1 if it also introduces measurement error in average education. If that occurs, estimates of π_1 are biased downward.

4 Data and Descriptive Statistics

To estimate learning spillovers I construct a data set linking ten separate administrative and survey data sources from Sweden.[18] The raw data is compiled by Statistics Sweden. I link the data for the entire population from 1985 to 2012.

The data on employers comes from two sources. First, there is registry data which covers all companies. This is what I use for the majority of the analysis. Second, there is the Structural Business Statistics (SBS) database which consists of accounting and balance sheet data, which I use for some robustness checks. From 1997 onward, data is provided for all non-financial firms. From 1985 to 1996, companies with over 50 employees are included, as well as companies with 20 people or more in the industrial sector.

To obtain sufficiently rich data on employees, I pull from eight separate data sets. First is the Longitudinal Database on Education, Income and Employment (LOUISE). LOUISE contains variables on all working-age individuals in Sweden. From LOUISE I use educational attainment, age, county, municipality, gender, marital status, immigrant status, and number and ages of children.

Second is the Register-Based Labour Market Statistics (RAMS). This data set contains information on all employment spells each year for all employed individuals in Sweden. An employment spell is a set of contiguous months worked at a given firm. From RAMS I use the start and end month for each employment spell in a given year, annual income from each employment spell in a given year, and firm and plant identifiers. The third data set is SOKATPER, which provides information on unemployment spells for the working-age population in Sweden, with similar variables to RAMS.

For robustness exercises, I supplement the income data from RAMS with wage data. The wage data is provided in five separate data files, one each for private sector employees, private sector managers, and public employees at the local, county, and national level. However, this wage data is only available for all non-financial firms from 1997 onward. Prior to 1997, private employee wages are only available for workers employed at firms with over 50 employees. For this reason I rely primarily on the income data (which has full coverage of all workers), but provide robustness checks using the wage data in Table E.4 in Appendix E.

The two main variables of interest for my analysis are monthly wages and average education exposure each worker experiences at work. For the main analysis, I construct monthly wages by simply adding annual income across different employment spells and

[18] I build on the data set from Friedrich, Laun, Meghir, and Pistaferri (2019).

dividing total annual income by total months worked in the year.[19] Constructing average education exposure for a given worker is slightly more challenging. First, number of workers employed by education level is not reported in the firm data. Fortunately, this is not an issue since I have the universe of workers and their firm and plant identifiers. This allows me to construct average education within a given firm using worker data matched to firm data.

A second concern is that some workers have overlapping employments with associated income levels that indicate part-time work. Ignoring this issue could bias my measure of average education. To deal with this, I restrict each worker to 1 unit of total time each month to be allocated across employers, with time in overlapping employments weighted by monthly income.[20] Specifically, I use the RAMS data to add up the number of workers of each education type working at a given plant for each month, with workers employed at multiple firms within a given month weighted accordingly. Next, I take each worker and add up the education types she was exposed to in each month, based on the plants she worked at in a given month. I then add these up over all months and divide by 12. This gives me monthly exposure to workers who are graduates of college (who I denote as my high-education type) or high school (who I denote as my low-education type). Last, I divide monthly exposure to high-education workers over monthly exposure to all workers. This is the measure of average education in the firm in the previous year that I use for the analysis.

The biggest limitation of this measure of average education of coworkers is that the finest level of interaction is at the plant level.[21] Not all workers in the same plant may interact, and I have no way of identifying which workers do interact. While more detail is rarely available in conventional data, finer data on coworker interactions can be quite helpful, as shown in Mas and Moretti (2009).

Table 1 presents summary statistics of my main variables of interest. While I used the entire population to construct the data and all variables in the analysis, because of computational constraints, I restrict estimation of learning spillovers in the plant to a

[19] Robustness exercises with the wage data use monthly reported wages directly.

[20] For example, suppose that Tom is college educated and works at plant A from January through March and earns a total of $3,000 (so $1,000 per month), and works at plant B from January to December and earns a total of $36,000 (so $3,000 per month). From January through March, when the employment spells overlap, Tom counts as .75 units of a high-education worker in plant B and .25 units of a high-education worker in plant A. Then, from April through December, Tom counts as 1 unit of a high-educated worker in plant B.

[21] The plant is defined as "every address, property, or group of neighboring property units in which a company operates."

5% sample.[22] I also restrict the sample to men aged 21–65.[23] Table 1 reports summary statistics for this sample. For more detailed definitions and notes on all the variables used in the empirical analysis, see Appendix D.

Table 1: Summary Statistics

	All	≤High School	College
Average year-worker observations	21.57	22.11	20.39
Real monthly earnings, 2012 SEK	27,861	24,778	34,555
Age	43.43	43.77	42.68
Married	0.51	0.49	0.56
Number of children aged 0–3	0.17	0.15	0.21
Number of children aged 4–6	0.13	0.12	0.15
Employed, of which			
Job stayer	0.89	0.89	0.89
Job mover	0.11	0.11	0.11
Industry			
Construction	0.12	0.15	0.04
Manufacturing	0.30	0.34	0.20
Retail trade	0.12	0.15	0.08
Services	0.46	0.36	0.68
Lagged average college share	0.31	0.19	0.57
Observations	1,673,605	1,145,816	527,789

Notes: Based on the 5

Figure 1 provides suggestive evidence on learning spillovers. I present a binned scatter plot of current wages and average education of coworkers in the previous year and the overlaid regression line,[24] which shows a strongly positive relationship. Panel B depicts the relationship conditional on the following controls: the worker's own education, marital status, number of children, a quadratic in experience, year dummies, industry dummies, and municipality dummies. The relationship remains strongly positive, and also becomes almost perfectly linear.

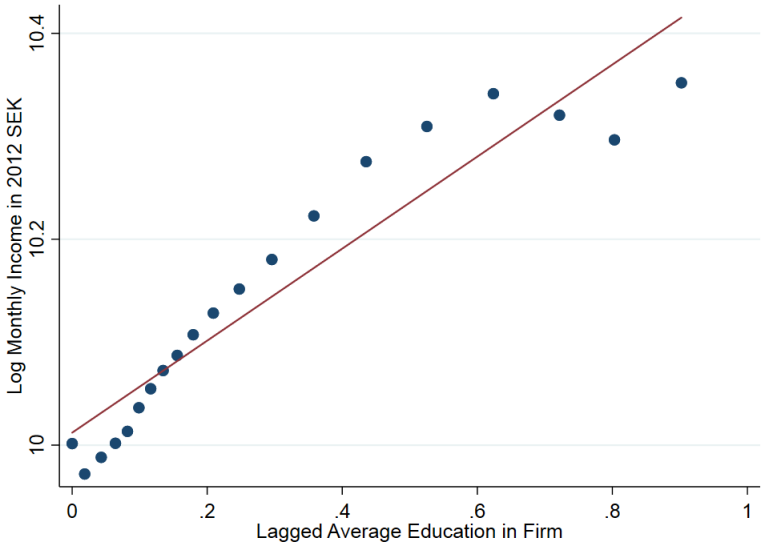
[22]The summary statistics for the full population are available upon request. As expected, they are virtually identical.

[23]I restrict the sample to men in the main estimates because of concerns over women transitioning in and out of full-time work more than men, but in Appendix Table E.1 I report estimates for women. The results are similar, although I find that women experience smaller learning spillovers compared with men.

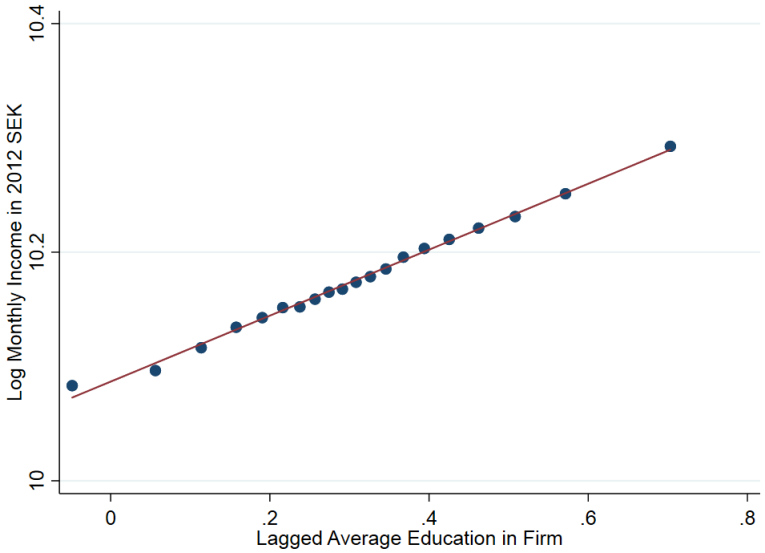
[24]These graphs were produced using binscatter, a user-written Stata command written by Michael Stepner, with input from Jessica Laird and Laszlo Sandor.

Figure 1: Binned Scatterplot of Current Log Monthly Income and Average Education of Coworkers Last Year

(a) Without Controls



(b) With Controls



Notes: Both figures show the relationship between income in the following year and average education of coworkers in one's firm in the current year. Panel A includes no controls. Panel B includes the following controls: the worker's own education, marital status, number of children, a quadratic in experience, year dummies, industry dummies, and municipality dummies.

5 Estimates of Learning Spillovers in the Firm

In column 1 of Table 2 I report estimates from equation 10. The coefficient on lagged average education (i.e., the learning spillover) is 0.195. It is identical to the estimated return to college, which is also 0.195. Naturally, this raw correlation suffers from many sources of bias. Column 2 adds individual fixed effects. The estimate of learning spillovers drops substantially, to 0.060. Including worker×plant fixed effects in column 3 reduces the coefficient a bit further, and controlling for county×time and industry×time dummies yields the smallest estimate of learning spillovers, at 0.025. These results imply that a 10 percentage point increase in average education of employees in a worker's plant increases wages the following year by approximately 0.3%. Compared to average wage growth in Sweden over this time period of roughly 1.7%–2% per year, an increase in wages of 0.3% is non-negligible. The results are also robust to numerous alternative specifications and additional controls.

Table 2: Learning Spillovers

	(1)	(2)	(3)	(4)	(5)
Own education	0.195***				
	(0.0010)				
Lagged average education	0.195***	0.060***	0.028***	0.025***	0.028***
	(0.0016)	(0.0036)	(0.0044)	(0.0044)	(0.0044)
Individual effects		Yes			
Worker×Plant effects			Yes	Yes	Yes
County×Year				Yes	
Industry×Year				Yes	
County×Industry×Year					Yes

Notes: Dependent variable is current log wage. All models include year effects, as well as controls for number of children and marital status. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation within individual (column 2) and worker-plant spells (columns 3–5) are reported in parentheses. Column 1 also includes a quadratic control for experience, but this drops out in the other columns once individual fixed effects are added.

In Table 3, I report estimates from equation 12 with separate firm and plant fixed effects.[25] The estimates are similar to Table 2, although the coefficient on lagged average

[25] As described in Section 3, I estimate both plant×worker and separate plant and worker fixed effects in order to identify learning spillovers using different variations in the data. I present the estimates in

education is larger at 0.058. The interpretation is that a 10 percentage point increase in average education at a given worker's plant increases his wages in the following year by almost 0.6%.

Table 3: Separate Plant and Individual Fixed Effects

	(1)
Person and establishment parameters	
Number person effects	91,257
Number plant effects	65,670
Main effect of interest	
Lagged average education	0.058
	(0.004)
Summary of other parameter estimates	
Std. dev. of person effects (across person-year obs)	0.316
Std. dev. of plant effects (across person-year obs)	0.228
Correlation of person/plant effects (across person-year obs)	-0.404
Adjusted R-squared	0.792

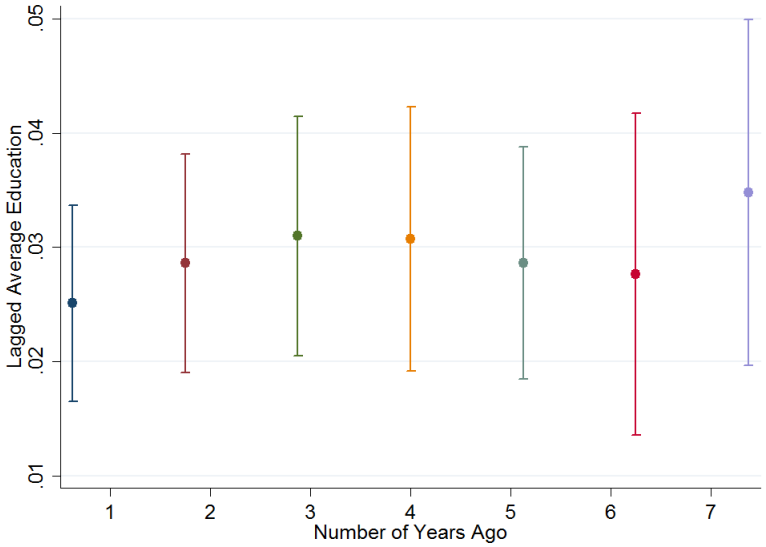
Notes: Dependent variable is current log wage. The model controls for year effects, number of children, marital status, industry-time dummies, and county-time dummies. Standard errors are based on 50 bootstrap replications and are reported in parentheses. See Appendix C.2 for more details on the estimation procedure.

Figure 2 documents the persistence of learning spillovers over time since exposure to coworkers.[26] The effect appears to be persistent, with similarly sized spillovers at least seven years into the future.

a separate table both to highlight the relevant results for firm and worker fixed effects and because the sample used for estimation is slightly different due to specific requirements for estimating separate firm and worker fixed effects. See Appendix C.2 for more details.

[26] See also Table E.7 in the Appendix.

Figure 2: Persistence of Spillovers over Time



Notes: This figure shows the effect of the average education of coworkers in the firm in the preceding year on workers' log wages one, two, three, and so on years later up to seven years later, with the lags estimated separately. Estimates can also be found in Table E.7 in the Appendix.

5.1 Robustness

In Appendix E, I report estimates using additional controls and alternative data samples to deal with remaining concerns. First, one might be concerned that with only 21 counties, industry×county×time dummies are too coarse to adequately capture local demand shocks. There was not sufficient variation to estimate the model with municipality×time dummies (there are 290 municipalities in Sweden), so instead I do two things to address this concern.

First, I construct Bartik shocks at the municipality level, and include them as controls. For more information on Bartik shocks and how I construct them for my setting, see Appendix E. The estimates of learning spillovers are slightly smaller with the inclusion of Bartik shocks (see Table E.2 in the Appendix). Second, I control for average education at the municipality level. The estimate does not change (see Table E.3 in the Appendix). In addition to addressing the concern that county×time dummies are too coarse, controlling for average education at the municipality level also ensures that I am picking up within-firm spillovers as opposed to across-firm spillovers.

Another concern is that in large plants individuals are not actually interacting together so I may not be capturing true spillovers. I address this by restricting the sample to plants

with less than 50 workers in Table E.6 in the Appendix and find similar results.[27]

To obtain the estimates reported in Table 2, I calculated monthly wages for each individual using total income earned from all firms and number of months worked in a given year.[28] I construct monthly wage because the data set that reports monthly wages directly does not cover all workers, raising selection concerns (see Section 4 for more details). However, one might be concerned that my monthly income measure is somehow biased. As a robustness exercise I reproduce Table 2 using the reported monthly wage (see Table E.4 in the Appendix) and find estimates using the wage data are slightly larger.

One might worry about collective bargaining and wage flexibility in Sweden. Historically, wages have been much more compressed in Sweden, at least in part due to collective bargaining. If wages are determined through collective bargaining, then it may not be possible for any given individual's wages to increase sufficiently to fully capture learning spillovers. This would bias my results downward. One way to address this concern is to compare the estimates in Table 2 to estimates restricting the sample to workers employed by private sector firms. Given Sweden's relatively stronger joint wage bargaining in public jobs versus private jobs, I would expect the spillover effect to be stronger if I restrict the sample to workers employed at private sector firms.[29] As predicted, the estimates of learning spillovers are larger, at 0.04 (see Table E.5, columns 1–5). Using the reported monthly wages for private sector workers increases the estimates even more: I find that a 10 percentage point increase in the average education of a worker's firm increases wages in the following year by 0.53% (see Table E.5, columns 6 and 7).[30]

All together, I interpret these results as evidence that learning spillovers exist, persist, and play an important role in determining wages. I discuss the results and their broader implications for welfare in more detail in Section 6. First, though, I document interesting

[27]One might also be concerned that there is complementarity in production and serial correlation in average education, so what is being captured by the coefficient on lagged average education of coworkers simply captures complementarity of current coworkers picked up due to serial correlation. As a check on this I ran a regression which included both lagged average education of coworkers and current average education of coworkers. The coefficients are positive and significant, and the coefficient on lagged average education of coworkers, while not quite as large in this specification as in others, is still substantive. These results are available on request.

[28]For more details on the construction of this variable, see Section 4.

[29]For example, according to Kjellberg (2011), in 2010 union density for public workers was 85%, while it was 65% for private sector workers. Also, the allowances for individual salary increases under union agreements differed across the sectors.

[30]This concern is particularly relevant for the external validity of my results. For example, in the U.S., collective wage setting is much weaker than in Sweden. According to the U.S. Bureau of Labor Statistics, 11.3% of workers were covered by collective bargaining in the U.S. in 2013 compared with 71% of workers in Sweden in 2010. Given this, along with the evidence presented here that suggests a downward bias from collective bargaining, I would expect the estimates of the impact of learning spillovers on wages to be larger for the U.S., reflecting more closely the full impact of learning spillovers on productivity.

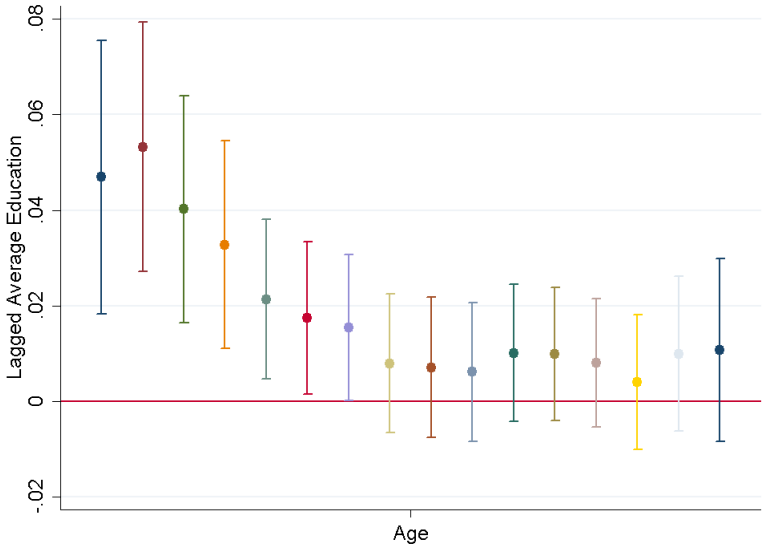
heterogeneity in the effects.

5.2 Heterogeneous Effects by Age

Consider learning spillovers by age. A reasonable prediction is that workers learn the most early in their careers and that the amount learned decreases as workers age. A decrease in learning spillovers as workers age is likely for two reasons. First, there may simply be a limit to the amount of relevant skills a given worker can obtain from his or her coworkers. Second, learning spillovers are most valuable to younger workers who have more time remaining in their career to reap the benefits from the additional skills obtained from coworkers.

In Figure 3, I graph the effects by age group. Estimates include worker×plant fixed effects, year dummies, industry×year dummies, and county×year dummies.[31] Consistent with a story of learning spillovers, the effect is largest at the youngest ages. More precisely, the effect increases at the earliest ages, and then decreases steadily until it is no longer statistically significantly different from zero from ages 38–48 onward. With most alternative explanations, I would expect the impact to be similar across ages.

Figure 3: Age Profile of Learning Spillovers: Overlapping 10-Year Increments



Notes: This figure shows the effect on workers' log wages of the average education of coworkers in the firm in the preceding year according to the age of the workers, where the ages in the x-axis represent overlapping 10-year age intervals, starting with the first blue line which represents ages 24–34 (the next line represents 26–36, and so on). Each point represents a separate regression.

[31] Figure 3 estimates the effect of learning spillovers on workers at overlapping 10-year intervals, starting with ages 24–34, then 26–36, 28–38, and so on. An alternative approach is to estimate the effects for non-overlapping 10-year age bins. I do so in Table E.8 in the Appendix. The pattern is the same.

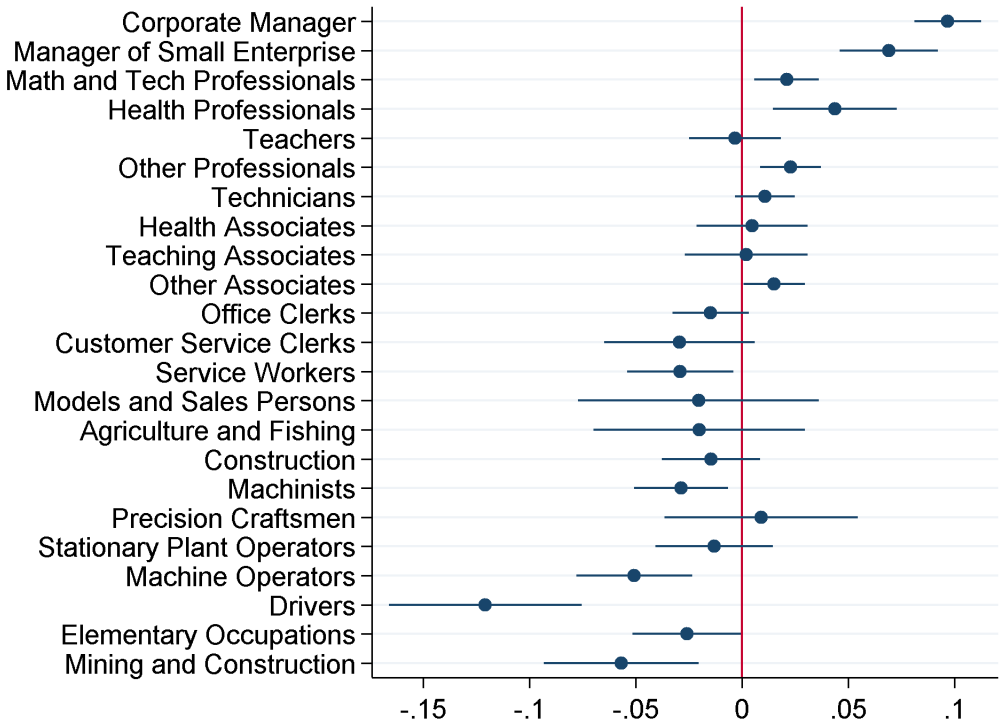
5.3 Heterogeneous Effects by Occupation

Learning spillovers should be larger when workers interact with each other more. To test this prediction, I estimate the amount of learning spillovers by occupation.[32] Certain occupations, such as driving, presumably offer fewer opportunities for interactions with coworkers than other occupations. The occupation groups in my data are defined by the Swedish Standard Classification of Occupations (SSYK), which is based on the International Standard Classification of Occupations.

Figure 4 graphs the effect by occupation. All estimates are relative to the omitted occupational category of legislators and senior officials. Figure 4 shows that workers in more isolated occupations (e.g., drivers, farmers, fisherman, machine operators, and those in elementary occupations, which includes janitors, garbage collectors, deliverers, and street vendors) experience the smallest effects. In contrast, those in occupations that likely have more opportunities for learning spillovers (e.g., managers and professionals) experience the largest effects.

[32]Note that the occupation data is only available in the wage data. This means that the sample is restricted (see Section 4). Additionally, Figure 4 only includes data from 2000 to 2010. Last, I omit occupation categories which had fewer than 100 individuals in the category. This includes the following categories: agricultural, fishery and related labourers, and other craft and related trades workers.

Figure 4: Occupational Profile of Learning Spillovers



Notes: This figure shows the effect of the average education of coworkers in the firm in the preceding year on workers according to the occupation of the workers. All estimates are relative to the omitted category of legislators and senior officials.

To further explore the heterogeneity of learning spillovers by occupation, I construct a ranking of occupations by interaction with peers using data from O*NET.[33] I convert the SSYK occupation categories to correspond to those of O*NET and rank them according to their average O*NET ranking of the importance of establishing and maintaining interpersonal relationships with coworkers.[34][35] Using this ranking, I compare the amount

[33]O*NET provides detailed information on activities, skills, and knowledge used in different occupations and was developed by the U.S. Department of Labor. Previous papers that have used the information on occupations found in O*NET include Acemoglu and Autor (2011) and Speer (2017). O*NET has been used in the Swedish context in Adermon and Gustavsson (2015), Black, Grönqvist, and Öckert (2018), and Johansson, Karimi, and Peter Nilsson (2019).

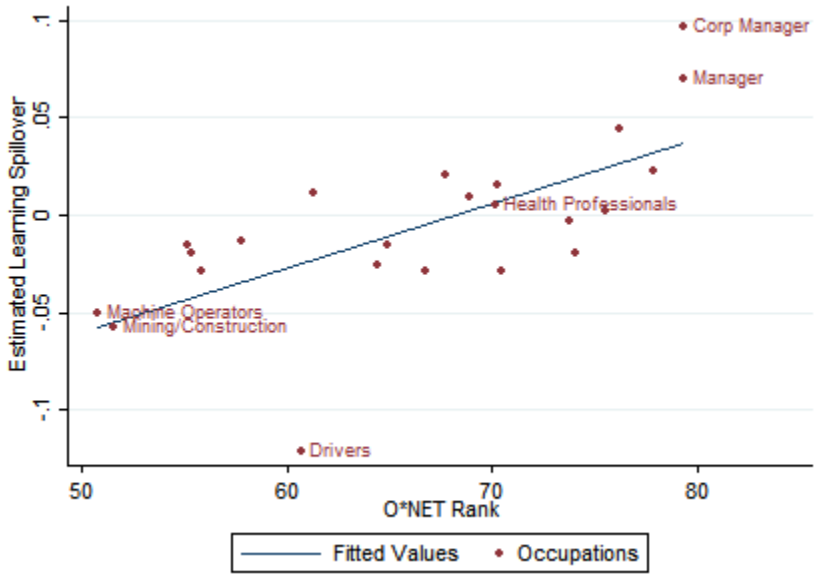
[34]O*NET uses the United States Standard Occupational Classification (SOC). The 26 major occupational groups in the SSYK variable are broadly comparable to the 23 major occupational groups in the SOC. However, they are not totally compatible. Furthermore, O*NET only provides rankings for the more detailed occupational categories. In Table D.3 in Appendix Section D I document how I construct a ranking using O*NET occupation categories, and then how I merge these categories into the SSYK categories. Note that in a small number of cases, the categories do not perfectly match, which means the matching will be imperfect and this could affect the results.

[35]The O*NET measure I use captures the degree to which an occupation involves "developing constructive and cooperative working relationships with others, and maintaining them over time."

of learning spillovers by occupation from Figure 4 against the ranking of occupations by amount of coworker interaction.

Figure 5 presents a scatterplot of the ranking of occupations using the O*NET measures and the estimates of learning spillovers from Figure 4. Occupations that experience higher learning spillovers also have higher average O*NET interaction rankings, suggesting that learning spillovers are a likely mechanism explaining the differences across occupations.

Figure 5: O*NET Rank of Occupation Potential for Learning Spillovers by Estimated Learning Spillovers



Notes: This figure shows the correlation between spillovers by occupation (on the y-axis) and the importance of interpersonal interactions in the occupation according to O*NET (on the x-axis). The relationship is shown through the scatterplot with an overlaid fitted line.

All together, these results make a strong case that the effects on workers' wages I have estimated are driven by learning spillovers. It is difficult to come up with an omitted variable that not only fits the four conditions outlined in Section 3 (time varying within county-by-industry and correlated with future wages and average education of coworkers), but also fits the distinctive age pattern in Figure 3 and the occupational patterns in Figure 5, is robust to all of the additional controls and alternative specifications, and provides a more compelling explanation for the estimated effects of the average education of coworkers in the previous year on log wages than learning spillovers.

6 Discussion and Implications

I now discuss the implications of my empirical estimates and theoretical results for the social returns to education. I use the estimates in Table 2 and an adaptation of the application from Altonji, Huang, and Taber (2015) to generate back-of-the-envelope calculations of the social returns to education and its components in order to answer the following questions: What is the impact in terms of income of adding an additional college worker? How much of this impact is due to learning spillovers and how much is due to the direct increase in productivity that comes from a college education? If learning spillovers are not fully internalized, how much larger are the social returns to college relative to the private returns?

The total return from a given individual obtaining a college education is equal to the direct return of that education (how much more productive the worker himself has become) plus the spillover effect of education (the increase in human capital provided by this worker to others, making them more productive in the future). If learning spillovers are fully internalized, then the social returns of adding a college-educated worker equal the private returns. This means that I do not need to know the effect of learning spillovers to estimate the total return of adding a college-educated worker. Table 2 shows that the private return to college is 0.195, which is also the social return when learning spillovers are fully internalized. However, knowing the effect of learning spillovers does allow me to answer the following question: If learning spillovers are fully internalized, how much of the total return of adding a college-educated worker is due to learning spillovers? To answer this, I can decompose the total return of adding a college-educated worker into the part due to learning spillovers and the part due to the direct increase in productivity of the worker who has obtained a college degree. To do this, I use the equilibrium wage equations when learning spillovers are fully internalized (see Proposition 2). I ignore the persistence of learning spillovers, assume there is no depreciation of spillovers, and assume that all workers have a discount rate of 1.[36]

In the equilibrium wage equations with full internalization of learning spillovers, the return to a college education includes not only the direct increase in productivity of the newly educated worker, but also the entire present discounted value of the learning spillovers the worker will provide for all of her coworkers. Under the assumptions

[36] Assuming no persistence will cause me to understate the percent of the total return due to learning spillovers. Assuming no depreciation and no discounting will cause me to overstate the percent of the total return due to learning spillovers.

used for this exercise, this implies the following equation:

$$\underbrace{x_1}_{direct\,return} + \underbrace{0.025}_{spillover\,return} \times \underbrace{\left(\frac{H_f+1}{N_f} - \frac{H_f}{N_f}\right)}_{\triangle spillover} \times \underbrace{N_f}_{\#\,colleagues} = \underbrace{0.195}_{total\,return} \tag{13}$$

which means that $\frac{0.195-0.025}{0.195} \times 100 = 87.18\%$ of the total return is due to the direct increase in productivity of the newly college-educated worker, while $\frac{0.025}{0.195} \times 100 = 12.82\%$ percent of the total return is due to learning spillovers.

In contrast, if learning spillovers are not fully internalized, then the social return of adding an additional college-educated workers exceeds the private return. If this is the case, it is necessary to know the effect of learning spillovers if one wishes to know the full social return from adding an additional college-educated worker.

If learning spillovers are not internalized at all, then Equation 13 becomes

$$\underbrace{0.195}_{direct\,return} + \underbrace{0.025}_{spillover\,return} \times \underbrace{\left(\frac{H_f+1}{N_f} - \frac{H_f}{N_f}\right)}_{\triangle spillover} \times \underbrace{N_f}_{\#\,colleagues} = \underbrace{x_2}_{total\,return} \tag{14}$$

Solving for x_2, the social return to an additional college-educated worker is 0.22. Decomposing the social return I find that $\frac{0.195}{0.22} \times 100 = 88.64\%$ of the total return is due to the direct increase in productivity of the newly college-educated worker, while $\frac{.028}{0.222} \times 100 = 11.36\%$ of the total return is due to learning spillovers.

From the results from the model in the first part of this paper, I cannot make a claim regarding how much internalization actually occurs. In fact, I showed three separate possibilities: no internalization occurs if firms ignore the spillovers (Proposition 1), full internalization occurs if firms know worker's learning types and are able to pay personalized wages (Proposition 2), and anything from no internalization to over-internalization could occur with asymmetric information (Proposition 3).

However, all together, the three Propositions (excluding the possibility of over-internalization for now) combined with the empirical results can at least provide bounds on both the social returns and the percentage of the social return attributable to learning spillovers. These bounds are summarized in Table 4.

Table 4: Private and Social Returns to Education with Learning Spillovers

	No Internalization	Full Internalization
Private return to education	0.195	0.1945
Social return to education	0.195	0.22
Amount by which social return exceeds private	0	0.025
Percent due to own productivity	87.18%	88.64%
Percent due to learning spillovers	12.82%	11.36%

Notes: Calculations that produce the estimates are described in the text.

7 Conclusion

There is a large literature broadly concerned with human capital spillovers, but a smaller literature on spillovers between coworkers within firms. In this paper I start with a simple point: if learning spillovers occur as a by-product of production and depend on average education within the firm, coworkers impose important externalities on each other. Applying insights from the theoretical literature on externalities, I show that in this context firms may not internalize learning spillovers. If firms fail to properly internalize learning spillovers into wages, individuals make inefficient investments in education.

With this result in hand, I turn to the second focus of this paper, an empirical assessment of learning spillovers. Using wage equations predicted by the model, I show that while the effect of average education of current coworkers on current wages is ambiguous, the effect of average education of past coworkers on current wages is unambiguous. For this reason I focus on estimating the effect of average education of a worker's coworkers in the previous year on current wages.

To deal with unobserved firm heterogeneity and worker sorting, I include plant and worker fixed effects in my empirical strategy. I estimate the effect of average education of coworkers in the previous year on current wages using both plant$\times$worker fixed effects and separate plant and worker fixed effects. To address time-varying omitted variables, I include county$\times$time and industry$\times$time dummies. To bring the empirical strategy to the data, I require a long panel on all workers and their peers. To meet these data requirements, I construct a unique data set using administrative data from Sweden.

I find that a 10 percentage point increase in the average education of coworkers within a worker's firm increases wages in the following year by 0.28%. Furthermore, I show that several additional results support the conclusion that the effects are due to learning spillovers. First, the results are robust to numerous alternative specifications. Different

specifications and the inclusion of additional controls suggest that, if anything, the main estimates understate the effect. Second, the effects are heterogeneous by age and occupation in ways that are consistent with learning spillovers. In the last section of the paper, I explore the broader implications of the main results. My findings suggest that the social returns of adding an additional college worker in terms of total income ranges from 0.195 to 0.22, with 11.36% to 12.82% of this total increase attributable to learning spillovers.

Having established that learning spillovers in the firm are both theoretically and empirically important, I see a number of areas for future research. Starting with the model, in the interest of simplicity I excluded the possibility of sorting driven by the learning spillovers. Relaxing this assumption could have interesting implications for sorting and employment. In particular, heterogeneity on the firm side may allow for sorting that makes it possible to support an outcome that is a Pareto improvement over what is possible with homogeneous firms. Second, firms may learn about worker's types over time, which could have both theoretical and empirical implications similar in spirit to the results from Kahn and Lange (2014).

On the empirical side, a great deal remains to be known about learning spillovers in the firm. For example, do learning spillovers occur based on other traits of coworkers, such as experience? Do some workers learn more from certain types of workers compared with others–for example, do men learn more from men and women learn more from women? How important are learning spillovers in other contexts? To what degree are workers aware of and selecting jobs based on learning spillovers? Another interesting question is whether skills obtained through spillovers also produce further learning spillovers for coworkers, leading to social multipliers. If this turns out to be the case, then the estimates presented here may understate the total impact of learning spillovers on individual wages and the economy as a whole.

This paper shows that learning spillovers are empirically meaningful and may not be fully internalized, causing inefficient investments in human capital. A natural next question is how the incentives to obtain an education can be set right. There are a few possible solutions suggested by this paper. First, if policy or some other intervention were able to resolve the information asymmetry that I show leads to inefficiency, this would allow firms to pay personalized wages that account for learning spillovers. Alternatively, a second-best solution could increase subsidies for education which lower the cost of education and will increase the number of individuals who go to school. Of course, without additional empirical evidence on learning externalities it will be difficult, if not impossible, to get these subsidies exactly correct, but my results suggest that at least small subsidies for education may be warranted.

References

ABOWD, J. M., AND F. KRAMARZ (1999): "The Analysis of Labor Markets Using Matched Employer-Employee Data," *Handbook of Labor Economics*, 3, 2629–2710.

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–333.

ACEMOGLU, D. (1996): "A Microfoundation for Social Increasing Returns in Human Capital Accumulation," *The Quarterly Journal of Economics*, 111(3), 779–804.

ACEMOGLU, D., AND J. ANGRIST (2001): "How Large Are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws," in *NBER Macroeconomics Annual 2000, Volume 15*, pp. 9–74. MIT Press.

ACEMOGLU, D., AND D. AUTOR (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics*, 4, 1043–1171.

ADERMON, A., AND M. GUSTAVSSON (2015): "Job Polarization and Task-Biased Technological Change: Evidence from Sweden, 1975-2005," *The Scandinavian Journal of Economics*, 117(3), 878–917.

ALTONJI, J. G., C.-I. HUANG, AND C. R. TABER (2015): "Estimating the Cream Skimming Effect of School Choice," *Journal of Political Economy*, 123(2), pp. 266–324.

AMODIO, F., AND M. A. MARTINEZ-CARRASCO (2018): "Input Allocation, Workforce Management and Productivity Spillovers: Evidence from Personnel Data," *The Review of Economic Studies*, 85(4), 1937–1970.

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

ARCIDIACONO, P., J. KINSLER, AND J. PRICE (2017): "Productivity Spillovers in Team Production: Evidence from Professional Basketball," *Journal of Labor Economics*, 35(1), 191–225.

AZOULAY, P., J. S. GRAFF ZIVIN, AND J. WANG (2010): "Superstar Extinction," *The Quarterly Journal of Economics*, 125(2), 549–589.

BANDIERA, O., I. BARANKAY, AND I. RASUL (2010): "Social Incentives in the Workplace," *The Review of Economic Studies*, 77(2), 417–458.

BARRO, R. J. (1996): "Institutions and Growth, An Introductory Essay," *Journal of Economic Growth*, 1(2), 145–148.

BECKER, G. S. (2009): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.

BLACK, S. E., E. GRÖNQVIST, AND B. ÖCKERT (2018): "Born to Lead? The Effect of Birth Order on Non-Cognitive Skills," *Review of Economics and Statistics*, 100(2), 274–286.

BRUNE, L., E. CHYN, AND J. KERWIN (2020): "Peers and Motivation at Work Evidence from a Firm Experiment in Malawi," *Journal of Human Resources*, pp. 0919–10416.

CARD, D., J. HEINING, AND P. KLINE (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128(3), 967–1015.

COASE, R. H. (1960): "The Problem of Social Cost," *The Journal of Law & Economics*.

CORNELISSEN, T., C. DUSTMANN, AND U. SCHÖNBERG (2017): "Peer Effects in the Workplace," *American Economic Review*, 107(2), 425–56.

DIAMOND, R. (2016): "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 106(3), 479–524.

FRIEDRICH, B., L. LAUN, C. MEGHIR, AND L. PISTAFERRI (2019): "Earnings Dynamics and Firm-Level Shocks," Working Paper 25786, National Bureau of Economic Research.

GRILICHES, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, pp. 1–22.

GURYAN, J., K. KROFT, AND M. J. NOTOWIDIGDO (2009): "Peer Effects in the Workplace Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal Applied Economics*, 1(4), 34–68.

HAGGAG, K., B. MCMANUS, AND G. PACI (2017): "Learning by Driving: Productivity Improvements by New York City Taxi Drivers," *American Economic Journal: Applied Economics*, 9(1), 70–95.

HERBST, D., AND A. MAS (2015): "Peer Effects on Worker Output in the Laboratory Generalize to the Field," *Science*, 350(6260), 545–549.

HERKENHOFF, K., J. LISE, G. MENZIO, AND G. M. PHILLIPS (2018): "Production and Learning in Teams," Working Paper 25179, National Bureau of Economic Research.

JACKSON, C. K., AND E. BRUEGMANN (2009): "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers," *American Economic Journal: Applied Economics*, 1(4), 85–108.

JAROSCH, G., E. OBERFIELD, AND E. ROSSI-HANSBERG (2019): "Learning from Coworkers," Working Paper 25418, National Bureau of Economic Research.

JOHANSSON, P., A. KARIMI, AND J. PETER NILSSON (2019): "Worker Absenteeism: Peer Influences, Monitoring and Job Flexibility," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2), 605–621.

KAHN, L. B., AND F. LANGE (2014): "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures," *The Review of Economic Studies*, 81(4), 1575–1613.

KJELLBERG, A. (2011): "The Decline in Swedish Union Density Since 2007," *Nordic Journal of Working Life Studies*, 1(1), pp–67.

LANGE, F., AND R. TOPEL (2006): "The Social Value of Education and Human Capital," *Handbook of the Economics of Education*, 1, 459–509.

LEVITT, S. D., J. A. LIST, AND C. SYVERSON (2013): "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant," *Journal of Political Economy*, 121(4), 643–681.

LINDAHL, E. (1919): "Just Taxation - A Positive Solution," *Classics in the Theory of Public Finance*, 134, 168–76.

LUCAS JR, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3–42.

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

MARTINS, P. S., AND J. Y. JIN (2010): "Firm-Level Social Returns to Education," *Journal of Population Economics*, 23(2), 539–558.

MAS, A., AND E. MORETTI (2009): "Peers at Work," *American Economic Review*, 99(1), 112–45.

MORETTI, E. (2004a): "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data," *Journal of Econometrics*, 121(1), 175–212.

——— (2004b): "Workers Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions," *The American Economic Review*, 94(3), 656–690.

PAPAY, J. P., E. S. TAYLOR, J. H. TYLER, AND M. E. LASKI (2020): "Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data," *American Economic Journal: Economic Policy*, 12(1), 359–88.

RAUCH, J. (1993): "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," *Journal of Urban Economics*, 34(3), 380–400.

SPEER, J. D. (2017): "Pre-Market Skills, Occupational Choice, and Career Progression," *Journal of Human Resources*, 52(1), 187–246.

THORNTON, R. A., AND P. THOMPSON (2001): "Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding," *American Economic Review*, 91(5), 1350–1368.

WALDINGER, F. (2012): "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany," *The Review of Economic Studies*, 79(2), 838–861.

Appendix: For Online Publication Only

A Proofs of the Propositions

A.1 Pareto Efficient Solution

The Pareto efficient problem solves for the optimal number of A types who go to college, denoted M^A, and the optimal number of B types who go to college, denoted M^B.

$$\underset{M^A, M^B}{Max} \quad -\int_0^{M^A} 1di - \int_0^{M^B} 1di \\ JF\left(\frac{M^A + M^B}{J}, \frac{I - M^A - M^B}{J}\right)$$

$$
\begin{aligned}
&+\alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\
&+\alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\
&+\delta\alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\
&+\delta\alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2}
\end{aligned}
$$

The conditions defining the optimal number of college A types and college B types are

$$
\begin{aligned}
M^A &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) && (15) \\
M^B &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) && (16)
\end{aligned}
$$

What equation 15 and 16 show is that wages must reflect worker productivity in two dimensions in order to fully internalize learning spillover. First, workers must be paid their marginal productivities in producing consumption goods (F_1 and F_2). Second, workers must be paid their marginal productivities in terms of producing learning spillovers for their coworkers.

A.2 Proofs of Propositions 1

A.2.1 Consumer Problem

Consumers maximize utility subject to their budget constraint. A type consumers solve

$$
\underset{c_1,c_2,c_3,h^i}{Max} \quad c_1 + c_2 + c_3
$$

subject to

$$
c_1 + c_2 + c_3 \leq -\theta^i h^i + h^i w_f^{H^A} + \left(1 - h^i\right) w_f^{L^A} + \delta s_f^A
$$

B type consumers solve

$$\underset{c_1,c_2,c_3,h^i}{Max} \qquad c_1 + c_2 + c_3$$

subject to

$$c_1 + c_2 + c_3 \;\; \leq \;\; -\theta^i h^i + h^i w_f^{H^B} + \left(1 - h^i\right) w_f^{L^B} + \delta s_f^B$$

The right-hand side of the budget constraint is equal to the cost of college education a given worker incurs if he chooses to go to college in the first period, the wage the worker receives based on his learning type and education choice in the second period, and the skills from second period learning spillovers he consumes in the third period.

The separation theorem holds here. To maximize utility, it is sufficient to maximize total income, through the worker's choice of college education ($h^i = 1$) or not ($h^i = 0$). Given this fact, in what follows and in the remainder of the proofs, I simply maximize the budget constraint in the consumer problem. Specifically, in the first period, consumers choose whether or not to go to college (where $h^i = 1$ if the individual goes to college), taking wages, the spillover, and their own costs of college, θ^i, as given.

$$\underset{h^i \in \{0,1\}}{Max} \qquad -\theta^i h^i + h^i w_f^{H^A} + \left(1 - h^i\right) w_f^{L^A} + \delta s_f^A$$

$$\underset{h^i \in \{0,1\}}{Max} \qquad -\theta^i h^i + h^i w_f^{H^B} + \left(1 - h^i\right) w_f^{L^B} + \delta s_f^B$$

Thus, A types choose to go to college if and only if

$$\theta^i \leq w_f^{H^A} - w_f^{L^A}$$

and B types choose to go to college if and only if

$$\theta^i \leq w_f^{H^B} - w_f^{L^B}$$

For the last individual of each type to go to college, these constraints hold with equality. Thus, the last A type to go to college, M^A, solves

$$M^A = w_f^{H^A} - w_f^{L^A}$$

and the last B type to go to college, M^B, solves

$$M^B = w_f^{H^B} - w_f^{L^B}$$

A.2.2 Firm Problem

Each firm demands an amount of each of the four types of workers (high learning-high educated, low learning-high educated, high learning-low educated, low learning-low educated) in order to maximizes its profits. Firms ignore future learning spillovers provided for workers, but do take into account the current period effects on consumption good production from the spillovers. Recall that s_f^A and s_f^B are defined in equations 1 and 2 in the main text. Thus, firms solve

$$\begin{aligned} \underset{H_f^A, H_f^B, L_f^A, L_f^B}{Max} \quad & F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) \\ & + s_f^A\left(H_f^A + L_f^A\right) + s_f^B\left(H_f^B + L_f^B\right) \\ & - w_f^{H^A} H_f^A - w_f^{H^B} H_f^B - w_f^{L^A} L_f^A - w_f^{L^B} L_f^B \end{aligned}$$

Taking first-order conditions defines the firm's demand for each type of worker by education level:

$$\begin{aligned} w_f^{H^A} &= F_1 + \alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\ &\quad + \left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\ w_f^{L^A} &= F_2 + \alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\ &\quad - \left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \end{aligned}$$

$$
\begin{aligned}
w_f^{H^B} &= F_1 + \alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
w_f^{L^B} &= F_2 + \alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad - \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)
\end{aligned}
$$

A.2.3 Equilibrium Definition

A Walrasian equilibrium consists of type and education specific wages, $w_f^{H^A}$, $w_f^{L^A}$, $w_f^{H^B}$, $w_f^{L^B}$, and consumption bundles and a choice of human capital for each individual, $\left(c_1^i, c_2^i, c_3^i, h^i\right)_{i \in I}$ such that:

1. Firms maximize profits given equilibrium compensation and worker's participation constraints;
2. Individuals maximize utility given wages and learning spillovers; and
3. Markets Clear:

$$
\begin{aligned}
\int_{i=0}^{I} c_1^i + \int_{i=0}^{I} c_2^i + \int_{i=0}^{I} c_3^i &= -\int_0^{M^A} i di - \int_0^{M^B} i di \\
&\quad + JF\left(\frac{M^A + M^B}{J}, \frac{I - M^A - M^B}{J}\right) + \alpha^A \frac{\frac{M^A + M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\
&\quad + \alpha^B \frac{\frac{M^A + M^B}{J}}{\frac{I}{J}} \frac{I}{2} + \delta \alpha^A \frac{\frac{M^A + M^B}{J}}{\frac{I}{J}} \frac{I}{2} + \delta \alpha^B \frac{\frac{M^A + M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\
JH_f^A &= M^A \\
JL_f^A &= \frac{I}{2} - M^A \\
JH_f^B &= M^B \\
JL_f^B &= \frac{I}{2} - M^B
\end{aligned}
$$

A.2.4 Equilibrium Solution

Consider the following equilibrium wages:

$$\begin{aligned}
w_f^{H^K} &= F_1 + \alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad + \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
w_f^{L^K} &= F_2 + \alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad - \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
K &= A, B
\end{aligned}$$

Imposing these prices individuals go to college provided the following conditions hold:

$$\begin{aligned}
\theta^i &\leq F_1 - F_2 + \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B}\right) \\
\theta^i &\leq F_1 - F_2 + \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B}\right)
\end{aligned}$$

Imposing market clearing gives

$$\begin{aligned}
\theta^i &\leq F_1 - F_2 + \left(\alpha^A \frac{I}{2J} + \alpha^B \frac{I}{2J}\right) \frac{J}{I} \\
&= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) \\
\theta^i &\leq F_1 - F_2 + \left(\alpha^A \frac{I}{2J} + \alpha^B \frac{I}{2J}\right) \frac{J}{I} \\
&= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right)
\end{aligned}$$

For the last individual to get education, these conditions hold with equality:

$$\begin{aligned} M^A &= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) \\ M^B &= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) \end{aligned}$$

which is not identical to the Pareto efficient condition for education investments:

$$\begin{aligned} M^A &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) \\ M^B &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) \end{aligned}$$

Since $(1+\delta)\left(\alpha^A + \alpha^B\right) > \left(\alpha^A + \alpha^B\right)$, individuals underinvest in education.

A.3 Proof of Proposition 2

Here, I solve for a competitive equilibrium where types are known, learning spillovers are known, and firms can pay personalized wages by education and type. These assumptions are unrealistic, but illustrate conditions required for the learning spillovers to be fully internalized.

A.3.1 Consumer Problem

In the first period, consumers choose whether or not to go to college, taking wages, the spillover, and their own costs of college as given.

$$\underset{h^i \in \{0,1\}}{Max} \quad -\theta^i h^i + h^i w_f^{H^A} + \left(1 - h^i\right) w_f^{L^A} + \delta s_f^A$$

$$\underset{h^i \in \{0,1\}}{Max} \quad -\theta^i h^i + h^i w_f^{H^B} + \left(1 - h^i\right) w_f^{L^B} + \delta s_f^B$$

Thus, A types choose to go to college if and only if

$$\theta^i \leq w_f^{H^A} - w_f^{L^A}$$

and B types choose to go to college if and only if

$$\theta^i \leq w_f^{H^B} - w_f^{L^B}$$

For the last individual of each type to go to college, these constraints hold with equality. Thus, the last A type to go to college, M^A, solves

$$M^A = w_f^{H^A} - w_f^{L^A}$$

and the last B type to go to college, M^B, solves

$$M^B = w_f^{H^B} - w_f^{L^B}$$

In the second period, workers work at a given firm f if the total compensation provided by that firm exceeds their reservation compensation level, w^{H^A}, w^{H^B}, w^{L^A}, and w^{L^B}, which they take as given. These reservation compensations are determined in equilibrium. Total compensation provided by a given firm includes wages paid plus the learning spillovers workers receive and consume in the third period. Learning spillovers are subject to depreciation, given by δ. Thus the workers' participation constraints are given by:

$$\begin{aligned} w_f^{H^A} + \delta s_f^A &\geq w^{H^A} \\ w_f^{H^B} + \delta s_f^B &\geq w^{H^B} \\ w_f^{L^A} + \delta s_f^A &\geq w^{L^A} \\ w_f^{L^B} + \delta s_f^B &\geq w^{L^B} \end{aligned}$$

A.3.2 Firm Problem

Each firm demands an amount of each of the four types of workers in order to maximize its profits. However, firms can now also trade off the wages they pay for learning spillovers, provided they meet workers' type-specific participation constraints. For example, suppose equilibrium compensation for high educated-high learning types, w^{H^A}, is equal to \$20. If a given firm has average education such that the high learning types get \$5 in spillovers, the firm only has to pay \$15 in wages in order to meet the worker's \$20 participation constraint. Thus, firms solve

$$\underset{H_f^A,H_f^B,L_f^A,L_f^B,w_f^{H^A},w_f^{L^A},w_f^{H^B},w_f^{L^B}}{Max} \quad F\left(H_f^A+H_f^B,L_f^A+L_f^B\right) + s_f^A\left(H_f^A+L_f^A\right)+s_f^B\left(H_f^B+L_f^B\right) - w_f^{H^A}H_f^A - w_f^{H^B}H_f^B - w_f^{L^A}L_f^A - w_f^{L^B}L_f^B$$

subject to the workers' participation constraints (see equations 1 and 2 in the main text for the definition of s_f^A and s_f^B) :

$$\begin{aligned} w_f^{H^A}+\delta s_f^A &\geq w^{H^A} \\ w_f^{H^B}+\delta s_f^B &\geq w^{H^B} \\ w_f^{L^A}+\delta s_f^A &\geq w^{L^A} \\ w_f^{L^B}+\delta s_f^B &\geq w^{L^B} \end{aligned}$$

Plugging in the participation constraints, the firm problem simplifies to:

$$\begin{aligned} \underset{H_f^A,H_f^B,L_f^A,L_f^B}{Max} \quad & F\left(H_f^A+H_f^B,L_f^A+L_f^B\right) - w^{H^A}H_f^A - w^{H^B}H_f^B - w^{L^A}L_f^A - w^{L^B}L_f^B \\ & + \alpha^A(1+\delta)\frac{H_f^A+H_f^B}{H_f^A+H_f^B+L_f^A+L_f^B}\left(H_f^A+L_f^A\right) \\ & + \alpha^B(1+\delta)\frac{H_f^A+H_f^B}{H_f^A+H_f^B+L_f^A+L_f^B}\left(H_f^B+L_f^B\right) \end{aligned}$$

Taking first order conditions defines the firm's demand for each type of worker by education level:

$$\begin{aligned} w^{H^A} &= F_1 + (1+\delta)\alpha^A\frac{H_f^A+H_f^B}{H_f^A+H_f^B+L_f^A+L_f^B} \\ &\quad + (1+\delta)\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\left(\frac{1}{H_f^A+H_f^B+L_f^A+L_f^B} - \frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}\right) \\ w^{L^A} &= F_2 + (1+\delta)\alpha^A\frac{H_f^A+H_f^B}{H_f^A+H_f^B+L_f^A+L_f^B} \end{aligned}$$

$$
\begin{aligned}
& - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
w^{H^B} &= F_1 + (1+\delta)\,\alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
& (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
w^{L^B} &= F_2 + (1+\delta)\,\alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
& - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)
\end{aligned}
$$

A.3.3 Equilibrium Definition

A Walrasian equilibrium consists of type and education specific total compensation, w^{H^A}, w^{L^A}, w^{H^B}, w^{L^B}, and consumption bundles and a choice of human capital for each individual, $\left(c_1^i, c_2^i, c_3^i, h^i\right)_{i\in I}$ such that:

1. Individuals maximize utility given wages and learning spillovers, meeting the conditions in Subsection A.3.1;
2. Firms maximize profits given equilibrium compensation and worker's participation constraints, meeting the conditions in Subsection A.3.2; and
3. Markets Clear:

$$
\begin{aligned}
\int_{i=0}^{I} c_1^i + \int_{i=0}^{I} c_2^i + \int_{i=0}^{I} c_3^i &= -\int_0^{M^A} idi - \int_0^{M^B} idi \\
&\quad + JF\left(\frac{M^A + M^B}{J}, \frac{I - M^A - M^B}{J}\right) + \alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2} \\
&\quad + \alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2} + \delta\alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2} + \delta\alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2} \\
JH_f^A &= M^A \\
JL_f^A &= \frac{I}{2} - M^A \\
JH_f^B &= M^B
\end{aligned}
$$

$$JL_f^B = \frac{I}{2} - M^B$$

A.3.4 Equilibrium Solution

Consider the following equilibrium compensation amounts:

$$\begin{aligned}
w^{H^K} &= F_1 + (1+\delta)\,\alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad + (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
w^{L^K} &= F_2 + (1+\delta)\,\alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
K &= A, B
\end{aligned}$$

The associated equilibrium wages are

$$\begin{aligned}
w_f^{H^K} &= F_1 + \alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad + (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
w_f^{L^K} &= F_2 + \alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \\
K &= A, B
\end{aligned}$$

Imposing these prices, individuals go to college provided the following conditions hold:

$$\theta^i \leq F_1 - F_2$$

$$+ (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B}\right)$$

Imposing market clearing gives

$$\begin{aligned}\theta^i &\leq F_1 - F_2 + (1+\delta)\left(\alpha^A\frac{I}{2J} + \alpha^B\frac{I}{2J}\right)\frac{J}{I} \\ &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right)\end{aligned}$$

For the last individual to get education, this condition holds with equality for each learning type so that

$$\begin{aligned}M^A &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) \\ M^B &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right)\end{aligned}$$

and this condition is identical to the Pareto efficient condition for education investments (see Section A.1).

This is an equilibrium. First, it is an equilibrium by definition–wages satisfy the firm and consumer first-order conditions and markets clear. Second, there is no profitable deviation. At these prices, profits are zero.

$$\begin{aligned}\pi &= F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) - w^{H^A}H_f^A - w^{H^B}H_f^B - w^{L^A}L_f^A - w^{L^B}L_f^B \\ &\quad + (1+\delta)\,\alpha^A\frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B}\left(H_f^A + L_f^A\right) \\ &\quad + (1+\delta)\,\alpha^B\frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B}\left(H_f^B + L_f^B\right) \\ &= F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) - F_1H_f^A - F_1H_f^B - F_2L_f^A - F_2L_f^B \\ &\quad - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)H_f^A \\ &\quad - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)H_f^B \\ &\quad + (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)L_f^A\end{aligned}$$

$$
\begin{aligned}
&+ (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) L_f^B \\
= \; &- (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H^A + H^B + L^A + L^B}\right)\left(H_f^A + H_f^B\right) \\
&+ (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)\left(H_f^A + H_f^B + L_f^A + L_f^B\right) \\
= \; &0
\end{aligned}
$$

Note that the second-to-last equality follows from the preceding line in part due to the constant returns to scale assumption on the production function. What this means is that firms will not choose to raise total compensation to any education-type worker as such a deviation would yield negative profits. Lowering total compensation to any education type worker would lower profits, since in that case the firm would lose all workers of that education-type. Thus, there is no profitable deviation for firms.[37]

A.4 Proof of Proposition 3

More realistically, each individual's learning type is known only to him or her. This makes the efficient outcome in Proposition 2 impossible to implement. In this section, I instead solve for a competitive equilibrium where the worker's learning type is unobserved by firms.

[37]There is an interesting corollary to first-degree price discrimination with a monopolist. There, price discrimination leads to the monopolist extracting the entire social surplus. Here, wage discrimination leads to the high-educated workers extracting the entire social surplus from learning. This provides the correct incentives for education, but it is arguably not a happy outcome for low-educated workers, who do not receive any gains from learning spillovers. In fact, low-learning, low-educated workers could even end up worse off than if they didn't learn from their coworkers at all. If no one received any spillovers, they would simply get their marginal product in terms of consumption good production, F_2^{NS}. Instead, with spillovers low-learning, low-educated workers receive

$$
F_2^S + (1+\delta)\left(\alpha^B - \alpha^A\right)\left(H_f^A + L_f^A\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)
$$

Since $\alpha^B - \alpha^A < 0$, F_2^S must be sufficiently higher than F_2^{NS} in order for low-learning, low-educated workers to not be worse off despite the fact that they are more productive in producing consumption goods. This is the case because they are penalized for the negative externality they have on coworkers in the production of learning spillovers.

A.4.1 Consumer Problem

The consumer problem in the first period is identical to Section A.3.1, with the result that the last A type to go to college, M^A, solves

$$M^A = w_f^{H^A} - w_f^{L^A}$$

and the last B type to go to college, M^B, solves

$$M^B = w_f^{H^B} - w_f^{L^B}$$

In the second period, workers work at a given firm f if the total compensation provided by that firm exceeds their reservation compensation level, w^{H^A}, w^{H^B}, w^{L^A}, and w^{L^B}, which they take as given. These reservation compensations are determined in equilibrium. Total compensation provided by a given firm includes wages paid plus the learning spillovers workers receive and consume in the third period. Learning spillovers are subject to depreciation, given by δ. Thus the workers' participation constraints are given by

$$w_f^{H^A} + \delta s_f^A \geq w^{H^A} \tag{17}$$
$$w_f^{H^B} + \delta s_f^B \geq w^{H^B} \tag{18}$$
$$w_f^{L^A} + \delta s_f^A \geq w^{L^A} \tag{19}$$
$$w_f^{L^B} + \delta s_f^B \geq w^{L^B} \tag{20}$$

A.4.2 Firm Problem

Workers provide their labor inelastically, subject to their second period participation constraints. Firms then maximize profits subject to these participation constraints. However, since firms cannot observe workers' learning type, they must also meet incentive compatibility constraints. Thus, firms solve

$$\begin{aligned} \underset{H_f^A, H_f^B, L_f^A, L_f^B, w_f^{H^A}, w_f^{L^A}, w_f^{H^B}, w_f^{L^B}}{Max} \quad & F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) \\ & + + s_f^A\left(H_f^A + L_f^A\right) + s_f^B\left(H_f^B + L_f^B\right) \\ & - w_f^{H^A} H_f^A - w_f^{H^B} H_f^B \\ & - w_f^{L^A} L_f^A - w_f^{L^B} L_f^B \end{aligned}$$

subject to the worker's participation constraints

$$\begin{aligned} w_f^{H^A} + \delta\alpha^A s_f &\geq w^{H^A} \\ w_f^{H^B} + \delta\alpha^B s_f &\geq w^{H^B} \\ w_f^{L^A} + \delta\alpha^A s_f &\geq w^{L^A} \\ w_f^{L^B} + \delta\alpha^B s_f &\geq w^{L^B} \\ s_f &= \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \end{aligned}$$

and incentive compatibility constraints

$$\begin{aligned} w_f^{H^B} + \delta\alpha^B s_f &\geq w_f^{H^A} + \delta\alpha^B s_f \\ w_f^{H^A} + \delta\alpha^A s_f &\geq w_f^{H^B} + \delta\alpha^A s_f \\ w_f^{L^A} + \delta\alpha^A s_f &\geq w_f^{L^B} + \delta\alpha^A s_f \\ w_f^{L^B} + \delta\alpha^B s_f &\geq w_f^{L^A} + \delta\alpha^B s_f \end{aligned}$$

The incentive compatibility constraints imply that

$$\begin{aligned} w_f^{H^A} &= w_f^{H^B} \\ w_f^{L^A} &= w_f^{L^B} \end{aligned}$$

which means that firms cannot induce workers to reveal their learning type by offering different wages. The reason a separating equilibrium is not possible is because all workers within a firm are exposed to the same average education, regardless of their type. This is due to the "public good" nature of average education within the firm. Given that, workers will always claim to be whatever type receives the highest wage.

This results in the following, updated firm problem:

$$\begin{aligned} \underset{H_f^A, H_f^B, L_f^A, L_f^B, w_f^H, w_f^L}{Max} \quad & F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) + E\left[\alpha\right]\left(H_f^A + H_f^B + L_f^A + L_f^B\right) \\ & -w_f^H\left(H_f^A + H_f^B\right) - w_f^L\left(L_f^A + L_f^B\right) \end{aligned}$$

subject to the worker's participation constraints:

$$\begin{aligned} w_f^H &\geq w^{H^A} - \delta\alpha^A s_f \\ w_f^H &\geq w^{H^B} - \delta\alpha^B s_f \\ w_f^L &\geq w^{L^A} - \delta\alpha^A s_f \\ w_f^L &\geq w^{L^B} - \delta\alpha^B s_f \\ s_f &= \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \end{aligned}$$

Unlike before, when profit maximization required all four participation constraints to bind with equality, that assumption no longer holds in this setting. Whether all four bind or only two bind depends on the equilibrium compensation amounts, which are determined in equilibrium.

Instead, I solve for the Kuhn-Tucker conditions.[38] The Lagrangian is

$$\begin{aligned} \mathcal{L}\left(H_f^A, H_f^B, L_f^A, L_f^B, w_f^H, w_f^L\right) &= F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) + E\left[\alpha\right]\left(H_f^A + H_f^B + L_f^A + L_f^B\right) \\ &\quad - w_f^H\left(H_f^A + H_f^B\right) - w_f^L\left(L_f^A + L_f^B\right) \\ &\quad + \lambda_1\left(w_f^H - w^{H^A} + \delta\alpha^A s_f\right) \\ &\quad + \lambda_2\left(w_f^H - w^{H^B} + \delta\alpha^B s_f\right) \\ &\quad + \lambda_3\left(w_f^L - w^{L^A} + \delta\alpha^A s_f\right) \\ &\quad + \lambda_4\left(w_f^L - w^{L^B} + \delta\alpha^B s_f\right) \end{aligned}$$

and the corresponding Kuhn-Tucker Conditions are

$$\begin{aligned} F_1 + E\left[\alpha\right] - w_f^H + \delta\left(\left(\lambda_1 + \lambda_3\right)\alpha^A + \left(\lambda_2 + \lambda_4\right)\alpha^B\right)\frac{L_f}{\left(H_f + L_f\right)^2} &\leq 0 \\ H_f^A &\geq 0 \\ H_f^A\left(F_1 + E\left[\alpha\right] - w_f^H + \delta\left(\left(\lambda_1 + \lambda_3\right)\alpha^A + \left(\lambda_2 + \lambda_4\right)\alpha^B\right)\frac{L_f}{\left(H_f + L_f\right)^2}\right) &= 0 \\ F_1 + E\left[\alpha\right] - w_f^H + \delta\left(\left(\lambda_1 + \lambda_3\right)\alpha^A + \left(\lambda_2 + \lambda_4\right)\alpha^B\right)\frac{L_f}{\left(H_f + L_f\right)^2} &\leq 0 \end{aligned}$$

[38] As you would expect, in the two preceding propositions the solution is identical if I use Kuhn-Tucker conditions.

$$
\begin{aligned}
H_f^B &\geq 0 \\
H_f^B \left(F_1 + E\left[\alpha\right] - w_f^H + \delta \left(\left(\lambda_1 + \lambda_3\right) \alpha^A + \left(\lambda_2 + \lambda_4\right) \alpha^B \right) \frac{L_f}{\left(H_f + L_f\right)^2} \right) &= 0 \\
F_2 + E\left[\alpha\right] - w_f^L - \delta \left(\left(\lambda_1 + \lambda_3\right) \alpha^A + \left(\lambda_2 + \lambda_4\right) \alpha^B \right) \frac{H_f}{\left(H_f + L_f\right)^2} &\leq 0 \\
L_f^A &\geq 0 \\
L_f^A \left(F_2 + E\left[\alpha\right] - w_f^L - \delta \left(\left(\lambda_1 + \lambda_3\right) \alpha^A + \left(\lambda_2 + \lambda_4\right) \alpha^B \right) \frac{H_f}{\left(H_f + L_f\right)^2} \right) &= 0 \\
F_2 + E\left[\alpha\right] - w_f^L - \delta \left(\left(\lambda_1 + \lambda_3\right) \alpha^A + \left(\lambda_2 + \lambda_4\right) \alpha^B \right) \frac{H_f}{\left(H_f + L_f\right)^2} &\leq 0 \\
L_f^B &\geq 0 \\
L_f^B \left(F_2 + E\left[\alpha\right] - w_f^L - \delta \left(\left(\lambda_1 + \lambda_3\right) \alpha^A + \left(\lambda_2 + \lambda_4\right) \alpha^B \right) \frac{H_f}{\left(H_f + L_f\right)^2} \right) &= 0 \\
\lambda_1 + \lambda_2 - H_f^A - H_f^B &\leq 0 \\
w_f^H \left(\lambda_1 + \lambda_2 - H_f^A - H_f^B \right) &= 0 \\
w_f^H &\geq 0 \\
\lambda_3 + \lambda_4 - L_f^A - L_f^B &\leq 0 \\
w_f^L \left(\lambda_3 + \lambda_4 - L_f^A - L_f^B \right) &= 0 \\
w_f^L &\geq 0 \\
w_f^H &\geq w^{H^A} - \delta \alpha^A s_f \\
w_f^H &\geq w^{H^B} - \delta \alpha^B s_f \\
w_f^L &\geq w^{L^A} - \delta \alpha^A s_f \\
w_f^L &\geq w^{L^B} - \delta \alpha^B s_f \\
\lambda_1 &\geq 0 \\
\lambda_2 &\geq 0 \\
\lambda_3 &\geq 0 \\
\lambda_4 &\geq 0 \\
\lambda_1 \left(w_f^H - w^{H^A} + \delta \alpha^A s_f \right) &= 0 \\
\lambda_2 \left(w_f^H - w^{H^B} + \delta \alpha^B s_f \right) &= 0 \\
\lambda_3 \left(w_f^L - w^{L^A} + \delta \alpha^A s_f \right) &= 0
\end{aligned}
$$

$$\lambda_4 \left(w_f^L - w^{L^B} + \delta \alpha^B s_f \right) = 0$$

A.4.3 Equilibrium Definition

A Walrasian equilibrium consists of type and education specific total compensation, w^{H^A}, w^{L^A}, w^{H^B}, w^{L^B}, and consumption bundles and a choice of human capital for each individual, $\left(c_1^i, c_2^i, c_3^i, h^i\right)_{i \in I}$ such that:

1. Individuals maximize utility given wages and learning spillovers, meeting the conditions in Subsection A.4.1;

2. Firms maximize profits given equilibrium compensation and worker's participation constraints, meeting the conditions in Subsection A.4.2; and

3. Markets Clear:

$$\begin{aligned}
\int_{i=0}^{I} c_1^i + \int_{i=0}^{I} c_2^i + \int_{i=0}^{I} c_3^i &= -\int_0^{M^A} i di - \int_0^{M^B} i di \\
&\quad + JF\left(\frac{M^A + M^B}{J}, \frac{I - M^A - M^B}{J}\right) + \alpha^A \frac{\frac{M^A + M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\
&\quad + \alpha^B \frac{\frac{M^A + M^B}{J}}{\frac{I}{J}} \frac{I}{2} + \delta \alpha^A \frac{\frac{M^A + M^B}{J}}{\frac{I}{J}} \frac{I}{2} + \delta \alpha^B \frac{\frac{M^A + M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\
JH_f^A &= M^A \\
JL_f^A &= \frac{I}{2} - M^A \\
JH_f^B &= M^B \\
JL_f^B &= \frac{I}{2} - M^B
\end{aligned}$$

A.4.4 Equilibrium Solution

I can rule out either $H_f^A = 0$ or $H_f^B = 0$ or $L_f^A = 0$ or $L_f^B = 0$, since these fail to be equilibria as markets will not clear. Then, if $H_f^A > 0$ and $H_f^B > 0$ and $L_f^A > 0$ and $L_f^B > 0$, from the Kuhn-Tucker conditions on H_f^A, H_f^B, L_f^A, and L_f^B I have that

$$\begin{aligned}
w_f^H &= F_1 + E\left[\alpha\right] + \delta \left(\left(\lambda_1 + \lambda_3\right) \alpha^A + \left(\lambda_2 + \lambda_4\right) \alpha^B \right) \frac{L_f}{\left(H_f + L_f\right)^2} \\
w_f^L &= F_2 + E\left[\alpha\right] - \delta \left(\left(\lambda_1 + \lambda_3\right) \alpha^A + \left(\lambda_2 + \lambda_4\right) \alpha^B \right) \frac{H_f}{\left(H_f + L_f\right)^2}
\end{aligned}$$

From the Kuhn-Tucker conditions that $\lambda_j \geq 0$, $j = 1,2,3,4$ and the fact that $F_1 > 0$, it follows that $w_f^H > 0$.

This in turn requires that

$$\begin{aligned} \lambda_1 + \lambda_2 &= H_f^A + H_f^B \\ &= H_f \end{aligned}$$

Similarly, if $w_f^L > 0$ then it must be that

$$\begin{aligned} \lambda_3 + \lambda_4 &= L_f^A + L_f^B \\ &= L_f \end{aligned}$$

Plugging these expressions back into the wage equations, I obtain

$$\begin{aligned} w_f^H &= F_1 + E[\alpha] + \delta\left(\left(\alpha^A - \alpha^B\right)\lambda_1 + H_f\alpha^B + \left(\alpha^A - \alpha^B\right)\lambda_3 + L_f\alpha^B\right)\frac{L_f}{\left(H_f + L_f\right)^2} \\ &= F_1 + E[\alpha] + \delta\alpha^B\frac{L_f}{H_f + L_f} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L_f}{\left(H_f + L_f\right)^2} \\ w_f^L &= F_2 + E[\alpha] - \delta\left(\left(\alpha^A - \alpha^B\right)\lambda_1 + H_f\alpha^B + \left(\alpha^A - \alpha^B\right)\lambda_3 + L_f\alpha^B\right)\frac{H_f}{\left(H_f + L_f\right)^2} \\ &= F_2 + E[\alpha] - \delta\alpha^B\frac{H_f}{H_f + L_f} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H_f}{\left(H_f + L_f\right)^2} \end{aligned}$$

and profits are

$$\begin{aligned} \pi &= F\left(H_f, L_f\right) + E[\alpha]H_f + E[\alpha]L_f \\ &\quad -F_1H_f - F_2L_f - E[\alpha]H_f - E[\alpha]L_f \\ &\quad -\delta\alpha^B\frac{L_fH_f - H_fL_f}{H_f + L_f} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L_fH_f - H_fL_f}{\left(H_f + L_f\right)^2} \\ &= 0 \end{aligned}$$

Since profits do not depend on the choices of λ's (and are always zero under the proposed wages), any of the following solutions for the λ's are all equally good (in terms of maximizing profits) and also all meet the Kuhn-Tucker conditions:

$$\begin{aligned} \lambda_1 &\in [0, H_f] \\ \lambda_2 &= H_f - \lambda_1 \end{aligned}$$

$$\begin{aligned}\lambda_3 &\in [0, L_f] \\ \lambda_4 &= L_f - \lambda_3\end{aligned}$$

Thus, the competitive equilibrium is given by:

$$w_f^H = F_1 + E[\alpha] + \delta\alpha^B \frac{L_f^*}{H_f^* + L_f^*} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L_f^*}{\left(H_f^* + L_f^*\right)^2} \tag{21}$$

$$w_f^L = F_2 + E[\alpha] - \delta\alpha^B \frac{H_f^*}{H_f^* + L_f^*} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H^*}{(H^* + L^*)^2} \tag{22}$$

$$w^{HA} = F_1 + E[\alpha] + \delta\alpha^B \frac{L_f^*}{H_f^* + L_f^*} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L_f^*}{\left(H_f^* + L_f^*\right)^2} + \delta\alpha^A \frac{H_f^*}{H_f^* + L_f^*}$$

$$w^{HB} = F_1 + E[\alpha] + \delta\alpha^B + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L_f^*}{\left(H_f^* + L_f^*\right)^2}$$

$$w^{LA} = F_2 + E[\alpha] + \delta\left(\alpha^A - \alpha^B\right)\frac{H_f^*}{H_f^* + L_f^*} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}$$

$$w^{LB} = F_2 + E[\alpha] - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}$$

with

$$\begin{aligned}\lambda_1 &\in \left[0, H_f^*\right] \\ \lambda_2 &= H_f^* - \lambda_1 \\ \lambda_3 &\in \left[0, I - H_f^*\right] \\ \lambda_4 &= I - H_f^* - \lambda_3\end{aligned}$$

What this means is that there are an infinite number of solutions that are competitive equilibria. If education is exogenous, all of the solutions are efficient, and which one actually occurs simply moves the solution along the Pareto frontier. However, if education is endogenous, only one solution out of the infinite possible solutions is efficient:

$$\begin{aligned}w^{HA} - w^{LA} &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) \\ w^{HB} - w^{LB} &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right)\end{aligned}$$

For the proposed equilibrium I have that

$$\begin{aligned}
w^{H^A} - w^{L^A} &= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) \\
&\quad +\delta\alpha^B \frac{I - H^*}{I} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{I - H^*}{I^2} + \delta\alpha^A \frac{H^*}{I} \\
&\quad -\delta\left(\alpha^A - \alpha^B\right)\frac{H^*}{I} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H^*}{I^2} \\
&= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) \\
&\quad +\delta\alpha^B + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{1}{I} \\
w^{H^B} - w^{L^B} &= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) \\
&\quad +\delta\alpha^B + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{1}{I}
\end{aligned}$$

which means that Pareto efficiency requires:

$$\begin{aligned}
\lambda_1 + \lambda_3 &= \frac{1}{2}\left(H^* + L^*\right) \\
&= \lambda_2 + \lambda_4
\end{aligned}$$

This is only consistent with one of the infinite number of possible equilibrium wage and compensation packages:

$$\begin{aligned}
w_f^H &= F_1 + E[\alpha] + \delta\alpha^B \frac{L^*}{H^* + L^*} + \delta\left(\alpha^A - \alpha^B\right)\frac{1}{2}\frac{L^*}{H^* + L^*} \\
w_f^L &= F_2 + E[\alpha] - \delta\alpha^B \frac{H^*}{H^* + L^*} - \delta\left(\alpha^A - \alpha^B\right)\frac{1}{2}\frac{H^*}{H^* + L^*} \\
w^{H^A} &= F_1 + E[\alpha] + \delta\alpha^B \frac{L^*}{H^* + L^*} + \delta\left(\alpha^A - \alpha^B\right)\frac{1}{2}\frac{L^*}{H^* + L^*} + \delta\alpha^A \frac{H^*}{H^* + L^*} \\
w^{H^B} &= F_1 + E[\alpha] + \delta\alpha^B + \delta\left(\alpha^A - \alpha^B\right)\frac{1}{2}\frac{L^*}{H^* + L^*} \\
w^{L^A} &= F_2 + E[\alpha] + \delta\left(\alpha^A - \alpha^B\right)\frac{H^*}{H^* + L^*} - \delta\left(\alpha^A - \alpha^B\right)\frac{1}{2}\frac{H^*}{H^* + L^*} \\
w^{L^B} &= F_2 + E[\alpha] - \delta\left(\alpha^A - \alpha^B\right)\frac{1}{2}\frac{H^*}{H^* + L^*}
\end{aligned}$$

If the equilibrium is chosen at random, the probability that the efficient solution occurs is 0.

This result is not particularly surprising. From the equilibrium definition, we can see that the problem is fundamentally under identified. Excluding consumption, we have the

following 16 unknowns:

$$\left(H_f^A, H_f^B, L_f^A, L_f^B, w_f^H, w_f^L, w^{H^A}, w^{H^B}, w^{L^A}, w^{L^B}, \lambda_1, \lambda_2, \lambda_3, \lambda_4, M^A, M^B\right)$$

with only 14 independent equations, combining consumer FOC, firm FOC, and market clearing:

$$\begin{aligned}
F_1 + E[\alpha] - w_f^H + \delta\left((\lambda_1+\lambda_3)\alpha^A + (\lambda_2+\lambda_4)\alpha^B\right)\frac{L_f^A + L_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2} &= 0 \\
F_2 + E[\alpha] - w_f^L - \delta\left((\lambda_1+\lambda_3)\alpha^A + (\lambda_2+\lambda_4)\alpha^B\right)\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2} &= 0 \\
\lambda_1 + \lambda_2 - H_f^A - H_f^B &= 0 \\
\lambda_3 + \lambda_4 - L_f^A - L_f^B &= 0 \\
w_f^H - w^{H^A} - \delta\alpha^A s_f &= 0 \\
w_f^H - w^{H^B} - \delta\alpha^B s_f &= 0 \\
w_f^L - w^{L^A} - \delta\alpha^A s_f &= 0 \\
w_f^L - w^{L^B} - \delta\alpha^B s_f &= 0 \\
M^A &= w^{H^A} - w^{L^A} \\
M^B &= w^{H^B} - w^{L^B} \\
JH_f^A &= M^A \\
JH_f^B &= M^B \\
JH_f^B &= M^B \\
JL_f^B &= \frac{I}{2} - M^B
\end{aligned}$$

Given that the number of unknowns exceeds the number of equations, we could have predicted that the solution would not be unique from the outset. Note that in the main text, I will focus on the first set of possible solutions.

It is also not surprising that inefficiency is a likely outcome. I discuss in the main text (see Section 2.1) and have also shown here that it is not incentive compatible for workers to reveal their learning type. Thus, when there is asymmetric information firms will only observe education (and not learning type). As a result firms must offer the same high-education wage to all high-educated workers and the same low-education wage to all low-educated workers. Under these conditions it is only logical that firms are unable to fully internalize learning spillovers. The above proofs serve only as a formalization of this logic.

B Additional Comments and Discussion on the Model

B.1 Conditions that Prevent Sorting

I assume that

$$F_1 + (1+\delta)\,\alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \tag{23}$$

$$+ (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) > 0$$

$$F_1 + (1+\delta)\,\alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B}$$

$$+ (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) > 0$$

$$F_2 + (1+\delta)\,\alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B}$$

$$- (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) > 0$$

$$F_2 + (1+\delta)\,\alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B}$$

$$- (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) > 0$$

This implies that full employment is optimal–the marginal product of adding an additional worker to production, in particular a low-educated worker, is always greater than 0.

I also assume that

$$F_{11} + (1+\delta)\,\alpha^A\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)$$

$$+ (1+\delta)\,\alpha^A\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)$$

$$
\begin{aligned}
-(1+\delta)\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\left(\frac{L_f^A+L_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}\right) &< 0\\
F_{11}+(1+\delta)\,\alpha^B\left(\frac{1}{H_f^A+H_f^B+L_f^A+L_f^B}-\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}\right)&\\
+(1+\delta)\,\alpha^B\left(\frac{1}{H_f^A+H_f^B+L_f^A+L_f^B}-\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}\right)&\\
-(1+\delta)\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\left(\frac{L_f^A+L_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}\right) &< 0\\
F_{22}-(1+\delta)\,\alpha^A\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}-(1+\delta)\,\alpha^A\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}&\\
+(1+\delta)\,2\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\left(\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^3}\right) &< 0\\
F_{22}-(1+\delta)\,\alpha^B\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}-(1+\delta)\,\alpha^B\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}&\\
+(1+\delta)\,2\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\left(\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^3}\right) &< 0
\end{aligned}
$$

These assumptions combined with the previous assumptions (equation 23) imply that total production is increasing in each input but at a decreasing rate, which means that unbalanced inputs are never optimal. Specifically, it is not optimal to put all the high-learning types in firms with higher average education and the low-learning types in firms with lower average education. One reason these assumptions would hold is that the loss in consumption good production from using unbalanced inputs (since high- and low-educated workers are complements in production in F) outweighs the gain in skill accumulation obtained from a production plan using unbalanced input combinations (such as some firms with high average education and some firms with low average education).

B.2 Functional Form of the Learning Spillovers

The particular functional form of the learning spillovers is chosen for two reasons. The first reason is theoretically motivated. Consider a more general specification of the spillover,

$G\left(H_f, L_f\right)$. Then, the total amount of consumption goods produced by learning spillovers is $S_f = (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right) G\left(H_f, L_f\right)$. Unless G exhibits decreasing or zero returns to scale, the total amount of consumption goods produced by the learning spillover in the firm is increasing returns to scale (assuming F is not decreasing returns to scale), since

$$\begin{aligned}(1+\delta)\left(\alpha^A\left(\lambda H_f^A + \lambda L_f^A\right) + \alpha^B\left(\lambda H_f^B + \lambda L_f^B\right)\right) G\left(\lambda H_f, \lambda L_f\right) &= \\ \lambda\,(1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right) G\left(\lambda H_f, \lambda L_f\right)&\end{aligned}$$

With increasing returns to scale in production of the learning spillovers, it is optimal to have a single firm. Under these conditions, inefficiency is the most likely outcome, but this is a less interesting case. In this paper I focus on a more general, and, I believe, more compelling result. I show that even when a competitive equilibrium is possible, inefficiency is the most likely outcome. To do so, I choose a zero returns to scale function for individual learning spillovers to make perfect competition possible. I leave further examination of the increasing returns case and its implications to future work.

The second motivation for this particular specification is empirical. This paper was originally inspired by the literature on education externalities across firms.[39] In order to make the empirical results in the second half of this paper more comparable to that literature, I have chosen the same specification that has often been used in that literature.

B.3 Equilibrium with Traditional Training Inputs

In this section, I show why traditional, rival training inputs that produce general skills do not have the same issues as learning spillovers. Suppose firms can choose a certain number of rival inputs into general training, given by τ^i. The firm must purchase these inputs separately for each and every worker it employs. I assume that the cost of these inputs, $\nu\left(\tau^i\right)$, is constant returns to scale and is increasing in τ^i but at a diminishing rate.

Any worker i employed at a firm f that spends $\nu\left(\tau^i\right)$ on that worker's rival on-the-job training inputs will accumulate additional human capital that depends on that worker's learning parameters, so that:

$$s^A = \alpha^A \tau^A \tag{24}$$

$$s^B = \alpha^B \tau^B \tag{25}$$

[39] See, for example, Rauch (1993), Acemoglu and Angrist (2001)Moretti (2004a), and Moretti (2004b).

As with learning spillovers, I assume the training increases productivity this period and also increases productivity next period, but subject to depreciation of skills given by δ.

Pareto Efficient Solution

The Pareto efficient problem solves for the optimal number of A types who go to college, denoted M^A, and the optimal number of B types who go to college, denoted M^B, and the optimal number of traditional training inputs, τ^A and τ^B.

$$\begin{aligned}\underset{M^A,M^B}{Max} \quad & -\int_0^{M^A} 1di - \int_0^{M^B} 1di \\ & JF\left(\frac{M^A+M^B}{J}, \frac{I-M^A-M^B}{J}\right) \\ & +J\left(\left(1+\delta\right)\alpha^A\tau^A - \nu\left(\tau^A\right)\right)\frac{I}{2J} \\ & +J\left(\left(1+\delta\right)\alpha^B\tau^B - \nu\left(\tau^B\right)\right)\frac{I}{2J}\end{aligned}$$

The conditions defining the optimal number of college A types and college B types and optimal traditional training inputs are:

$$\begin{aligned} M^A &= F_1 - F_2 \\ M^B &= F_1 - F_2 \\ \left(1+\delta\right)\alpha^A &= \nu'\left(\tau^A\right) \\ \left(1+\delta\right)\alpha^B &= \nu'\left(\tau^B\right)\end{aligned}$$

Competitive Equilibrium

B.3.1 Consumer Problem

In the first period, consumers choose whether or not to go to college, taking wages, training options, and their own costs of college as given.

$$\underset{h^i\in\{0,1\}}{Max} \quad -\theta^i h^i + h^i w_f^{H^A} + \left(1-h^i\right) w_f^{L^A} + \delta\alpha^A\tau^A$$

$$\underset{h^i \in \{0,1\}}{Max} \quad -\theta^i h^i + h^i w_f^{H^B} + \left(1 - h^i\right) w_f^{L^B} + \delta \alpha^B \tau^B$$

Thus, A types choose to go to college if and only if

$$\theta^i \leq w_f^{H^A} - w_f^{L^A}$$

and B types choose to go to college if and only if

$$\theta^i \leq w_f^{H^B} - w_f^{L^B}$$

For the last individual of each type to go to college, these constraints hold with equality. Thus, the last A type to go to college, M^A, solves

$$M^A = w_f^{H^A} - w_f^{L^A}$$

and the last B type to go to college, M^B, solves

$$M^B = w_f^{H^B} - w_f^{L^B}$$

In the second period, workers work at a given firm f if the total compensation provided by that firm exceeds their reservation compensation level, w^{H^A}, w^{H^B}, w^{L^A}, and w^{L^B}, which they take as given. These reservation compensations are determined in equilibrium. Total compensation provided by a given firm includes wages paid plus the training workers receive and consume in the third period. Training is subject to depreciation, given by δ.

$$\begin{aligned} w_f^{H^A} + \delta \alpha^A \tau^A &\geq w^{H^A} \\ w_f^{H^B} + \delta \alpha^B \tau^B &\geq w^{H^B} \\ w_f^{L^A} + \delta \alpha^A \tau^A &\geq w^{L^A} \\ w_f^{L^B} + \delta \alpha^B \tau^B &\geq w^{L^B} \end{aligned}$$

B.3.2 Firm Problem

Each firm demands an amount of each of the four types of workers in order to maximize their profits. They also account for the fact that they can trade off training inputs for wages, but that they incur a cost for the training inputs for each worker.

Thus, firms solve

$$\begin{aligned}\underset{H_f^A,H_f^B,L_f^A,L_f^B,\tau^A,\tau^B}{Max} \quad & F\left(H_f^A+H_f^B,L_f^A+L_f^B\right)-w^{H^A}H_f^A-w^{H^B}H_f^B-w^{L^A}L_f^A-w^{L^B}L_f^B \\ & +\left(\left(1+\delta\right)\alpha^A\tau^A-\nu\left(\tau^A\right)\right)\left(H_f^A+L_f^A\right) \\ & +\left(\left(1+\delta\right)\alpha^B\tau^B-\nu\left(\tau^B\right)\right)\left(H_f^B+L_f^B\right)\end{aligned}$$

Taking first-order conditions defines the firm's demand for each type of worker by education level:

$$\begin{aligned} w^{H^A} &= F_1+\left(1+\delta\right)\alpha^A\tau^A-\nu\left(\tau^A\right) \\ w^{L^A} &= F_2+\left(1+\delta\right)\alpha^A\tau^A-\nu\left(\tau^A\right) \\ w^{H^B} &= F_1+\left(1+\delta\right)\alpha^B\tau^B-\nu\left(\tau^B\right) \\ w^{L^B} &= F_2+\left(1+\delta\right)\alpha^B\tau^B-\nu\left(\tau^B\right) \\ \left(1+\delta\right)\alpha^A &= \nu'\left(\tau^A\right) \\ \left(1+\delta\right)\alpha^B &= \nu'\left(\tau^B\right)\end{aligned}$$

B.3.3 Equilibrium Definition

A Walrasian equilibrium consists of type and education specific total compensation, w^{H^A}, w^{L^A}, w^{H^B}, w^{L^B}, a choice of traditional training inputs by type, τ^A and τ^B, and consumption bundles and a choice of human capital for each individual, $\left(c_1^i,c_2^i,c_3^i,h^i\right)_{i\in I}$ such that:

1. Firms maximize profits given equilibrium compensation and worker's participation constraints;

2. Individuals maximize utility given wages and training inputs; and

3. Markets Clear:

$$\begin{aligned}\int_{i=0}^{I}c_1^i+\int_{i=0}^{I}c_2^i+\int_{i=0}^{I}c_3^i &= -\int_0^{M^A}idi-\int_0^{M^B}idi \\ &\quad +JF\left(\frac{M^A+M^B}{J},\frac{I-M^A-M^B}{J}\right)+\alpha^A\frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2} \\ &\quad +\alpha^B\frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2}+\left(\left(1+\delta\right)\alpha^B\tau^B-\nu\left(\tau^B\right)\right)\frac{I}{2}+\end{aligned}$$

$$\delta\left(\left(1+\delta\right)\alpha^{A}\tau^{A}-\nu\left(\tau^{A}\right)\right)\frac{I}{2}$$

$$\begin{aligned}
JH_f^A &= M^A \\
JL_f^A &= \frac{I}{2}-M^A \\
JH_f^B &= M^B \\
JL_f^B &= \frac{I}{2}-M^B
\end{aligned}$$

B.3.4 Equilibrium Solution

Consider the following equilibrium compensation amounts:

$$\begin{aligned}
w^{H^A} &= F_1+\left(1+\delta\right)\alpha^{A}\tau^{A}-\nu\left(\tau^{A}\right) \\
w^{L^A} &= F_2+\left(1+\delta\right)\alpha^{A}\tau^{A}-\nu\left(\tau^{A}\right) \\
w^{H^B} &= F_1+\left(1+\delta\right)\alpha^{B}\tau^{B}-\nu\left(\tau^{B}\right) \\
w^{L^B} &= F_2+\left(1+\delta\right)\alpha^{B}\tau^{B}-\nu\left(\tau^{B}\right) \\
\left(1+\delta\right)\alpha^{A} &= \nu^{'}\left(\tau^{A}\right) \\
\left(1+\delta\right)\alpha^{B} &= \nu^{'}\left(\tau^{B}\right)
\end{aligned}$$

Imposing these prices individuals go to college provided the following conditions hold:

$$\begin{aligned}
\theta^i &\leq F_1-F_2 \\
\theta^i &\leq F_1-F_2
\end{aligned}$$

For the last individual to get education, these conditions hold with equality:

$$\begin{aligned}
M^A &= F_1-F_2 \\
M^B &= F_1-F_2
\end{aligned}$$

and the solution for the traditional training inputs is

$$\left(1+\delta\right)\alpha^{A} = \nu^{'}\left(\tau^{A}\right)$$

$$(1+\delta)\,\alpha^B \;=\; \nu'\left(\tau^B\right)$$

These results are identical to the Pareto efficient solution for education and traditional training inputs, so the competitive equilibrium is efficient. Asymmetric information is no longer an issue with the traditional training inputs application because individuals will not choose to lie about their learning type. This is due to the fact that instead of effectively charging individuals different prices for the same quantity of exposure, here firms are charging different prices for different quantities of training inputs. For this reason, it is incentive compatible for individuals to select the appropriate package of training inputs and accompanying wage deductions. Thus, the competitive equilibrium with traditional inputs is efficient even when there is asymmetric information, as we would expect given the results in Becker (2009).

C Estimation Appendix

C.1 Upward Bias in Estimates of Social Return Functions and Solution

I start by briefly summarizing the problem.[40] I am trying to get an unbiased estimate of π_1 in

$$w_{it} \;=\; \pi_0 h_i + \pi_1 \bar{H}_{ft-1} \tag{26}$$

Recall that h_i represents the individual's education, while $\bar{H}_{ft-1}$ represents the average education in the firm.

To start with, this equation, in the terminology of Manski (1993), identifies exogenous peer effects, and is not subject to all of the concerns that plague outcome on outcome regressions of peer effects. This follows since education is predetermined and the group average is assumed to affect later outcomes. However, as originally pointed out in Griliches (1977) and extended to the peer effects framework in Acemoglu and Angrist (2001), significant challenges remain. They show that a simple derivation yields the following solution for the coefficients:

$$\begin{aligned} \pi_0 &= \frac{\psi_0 - \psi_1 R^2}{1 - R^2} \\ \pi_1 &= \frac{\psi_1 - \psi_0}{1 - R^2} \end{aligned} \tag{27}$$

[40] For a more detailed description, see Angrist (2014).

Where R^2 is the first-stage R-squared from two-stage least squares estimation using average education in the firm$\times$year dummies as instruments for a given worker's own education, ψ_0 is the ordinary least squares (OLS) coefficient of education in equation 26, excluding average education, and ψ_1 is the 2SLS estimate of education, instrumented with the average education in the firm/year. Thus, I will find positive peer effects if the 2SLS estimate of the impact of h_i on w_{it} using $\bar{H}_{ft-1}$ as dummies for h_i differs for any reason from a simple OLS estimate of the impact of h_i on w_{it}. In particular, if there is measurement error in h_i, then I will find $\pi_1 > 0$ even in the absence of peer effects.

Angrist (2014) argues this concern is first-order in the peer effects literature. He proposes all papers on peer effects should meet two conditions: "the first is a clear distinction between the subjects of a peer effects investigation on the one hand and the peers who potentially provide the mechanism for causal effects on these subjects on the other. This distinction eliminates mechanical links between own and peer characteristics, making it easier to create or to isolate variation in peer characteristics that is independent of subject's own characteristics. The second is a set-up where fundamental OLS and 2SLS parameters [ψ_0 and ψ_1, in my notation] can be expected to produce the same results in the absence of peer effects" (p. 9).

Fixed Effects as a Solution

To formally show that fixed effects addresses this issue, I re-derive equation 27 with fixed effects for worker$\times$workplace spells.[41]

Rewrite equation 26 as follows:

$$w_{it} = \pi_0 \tau_i + (\pi_0 + \pi_1) \bar{H}_{ift-1} + \xi_i \tag{28}$$

where $\tau_i = h_i - \bar{H}_{ift-1}$. Now add fixed effects for worker$\times$workplace spells to equation 28:

$$w_{it} - \bar{w}_{it} = \pi_0 (\tau_i - \bar{\tau}_i) + (\pi_0 + \pi_1) \left(\bar{H}_{ift-1} - \bar{H}_{if}\right) + \xi_i \tag{29}$$

where

$$\begin{aligned} (\tau_i - \bar{\tau}_i) &= \left(h_i - \bar{H}_{ift-1}\right) - \left(\bar{h}_i - \bar{H}_{if}\right) \\ &= \bar{H}_{if} - \bar{H}_{ift-1} \end{aligned} \tag{30}$$

And equation 29 becomes

[41] The derivation is equivalent with individual fixed effects alone.

$$w_{it} - \bar{w}_{it} = \pi_0 \left(\bar{H}_{if} - \bar{H}_{ift-1} \right) + (\pi_0 + \pi_1) \left(\bar{H}_{ift-1} - \bar{H}_{if} \right) + \xi_i \quad (31)$$

$$= \pi_1 \left(\bar{H}_{ift-1} - \bar{H}_{if} \right) \quad (32)$$

Then,

$$\pi_1 = \frac{C\left(\left(\bar{H}_{ift-1} - \bar{H}_{if} \right), (w_{it} - \bar{w}_{it}) \right)}{V\left(\left(\bar{H}_{ift-1} - \bar{H}_{if} \right) \right)} \quad (33)$$

which is precisely the desired result. In the absence of peer effects, and excluding endogeneity concerns, I will find that $\pi_1 = 0$.

To show that this approach works using the data, I have replicated Table 3 from Angrist (2014). My columns 1–3 and 5–7 estimate identical regressions to those in Angrist's paper. Specifically, in the first column I estimate the effect of a college degree on wages. In the second column, I estimate the effect of average education at the municipality level on wages.[42] In the third column, I estimate the effect of both average education and given worker's own schooling (college degree or not) on wages.

In columns 5–7, I repeat the exercise in columns 1--3 but add in measurement error on own schooling. As in Angrist (2014), this biases the estimates of peer effects (the coefficient on average education$\times$municipality) upward. This demonstrates the purely mechanical positive effect (driven by measurement error) we expect to get when estimating peer effects.

I now draw your attention to the estimates with fixed effects in columns 4 and 8. In column 4, I estimate a fixed effects specification without measurement error. In column 8, I estimate the fixed effects specification with measurement error. In contrast to the original regression, the introduction of measurement error now biases the coefficient downward, as we would normally expect.

[42] I use average education in the municipality instead of average education in the firm because this variable makes my table more directly comparable to Angrist's table.

Table C.1: Empirical Support for Estimation Approach

	Reported schooling				With reliability 0.7			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own schooling	0.274		0.266	-	0.136		0.129	-
	(0.006)		(0.007)	-	(0.005)		(0.005)	-
Municipality average		0.454	0.173	0.350		0.404	0.275	0.113
schooling		(0.049)	(0.051)	(0.021)		(0.041)	(0.043)	(0.008)
First stage R2		0.1115				0.1113		

Notes: The dependent variable is log monthly wage. Standard errors, clustered on municipality, are reported in parentheses. All models include county of residence and year effects. Average education at the municipality level is computed using the sample (not the full population). The sample consists of 2,393,573 men from 1985 to 2012.

Note the conditions that must be met for this approach to work. First, h_i must be fixed within a worker$\times$workplace spell. This requirement will always hold in my setting, provided either work or school is full time. However, it may not hold in other settings, in which case the term does not drop out and the result no longer holds.

Second, the peer effect, $\bar{H}_{ift-1}$, must vary over time. Otherwise the right-hand side consists only of the error term (absent additional controls). This amounts to a requirement that there is sufficient variation in peers, holding the subject's characteristics constant. This also may not hold in many other settings. In particular, this does not generally hold in a school setting, where classes are assigned at the start of the year but there is generally no variation thereafter, conditional on holding the student$\times$class match fixed.

Third, one must have repeated observations on individuals, and also have corresponding repeated observations on all of their peers. This is obvious, but it is worth pointing out as it is arguably quite demanding in terms of data, and in some settings may not be available.

C.2 Estimation of Firm and Worker Fixed Effects

The identification and estimation of firm, worker, and time fixed effects was pioneered by Abowd, Kramarz, and Margolis (1999).[43] As pointed out in that paper, identification is obtained using a "connected set" of firms linked by workers who have moved between the firms. The major assumption underlying the identification result is that mobility is exogenous conditional on the controls, including time-invariant firm and worker character-

[43]More recently, Card, Heining, and Kline (2013) used the approach to decompose rising inequality in West Germany into the firm and worker specific components.

istics. However, estimation remains technically more challenging than simpler, two-way fixed effects models. The issue is that if one wishes to recover the fixed effects themselves, the number of parameters becomes very large (in particular, one must estimate fixed effects for every firm). Additionally, there is an issue with sparse matrices, since only a few workers (relative to the population) work for any given firm, resulting in a majority of 0 values for each firm dummy.

To estimate the results in this paper, I implement the user-written Stata command a2reg, which estimates the model as described in Abowd and Kramarz (1999).[44] I estimate the problem in two parts. First, I run a regression of log wages on dummies for year, county$\times$year, industry$\times$year, married, and number of children. I then save the residuals from this regression. Next, I use a2reg to estimate a regression of residualized wages this period on average education in the firm last period. Note that a2reg requires all variables to be non-missing. Thus, after the first step above, I drop all observations with missing values of either average education of coworkers last period, workplace, or residual wage this period.

To obtain standard errors, a2reg requires a user written bootstrap. I thus also program a bootstrap that runs over the entire procedure.[45]

[44] Amine Ouazad, Program for the Estimation of Two-Way Fixed Effects, available at http://personal.lse.ac.uk/ouazad/, 2007.

[45] I produced the standard errors using 50 bootstraps.

D Data Appendix

Table D.1: Variable Descriptions

Variable	Source	Notes
Income	RAMS, Statistics Sweden, 1985–2012	Described in detail in the main text (see Section 4).
Average education of coworkers	LOUISE, Statistics Sweden, 1985--2012	Described in detail in the main text (see Section 4).
Firm ID	RAMS, Statistics Sweden, 1985–2012	The firm ID comes in two levels: the firm ID and the workplace ID. I use the workplace ID for the main analysis, but also have used firm$\times$worker fixed effects in robustness checks.
Worker ID	RAMS, Statistics Sweden, 1985–2012	
Wages	Arb, Statistics Sweden, 1985–2011	Private employee wages for firms with over 50 employees. Includes people who had hourly wages and were employed at companies/organizations in the private sector
Wages	Tjm, Statistics Sweden, 1985–2011	Private official wages for firms with over 500 employees. Includes people who had a monthly salary and worked at a company/organization in the private sector
Wages	Kommun, Statistics Sweden, 1985--2011	Public employee wages at the local level. People employed in the primary sector who had local wage settlement
Wages	Landkomm, Statistics Sweden, 1985–2011	Public employee wages at the county council level. People employed in the county councils whose wages were governed by county councils' general provisions of the collective agreement for civil servants
Wages	Stat, Statistics Sweden, 1985–2011	Governmental public employee wages. People employed in the state sector who received state-regulated wages

Table D.2: Variable Descriptions

Variable	Source	Notes
Married	LOUISE, Statistics Sweden, 1985–2012	All years past 1990
Male	LOUISE, Statistics Sweden, 1985–2012	
Education	LOUISE, Statistics Sweden, 1985–2012	Only available in relatively coarse categories.
Number of children	LOUISE, Statistics Sweden, 1985–2012	
Industry	RAMS, Statistics Sweden, 1985–2012	17 industry categories in total.
County	LOUISE, Statistics Sweden, 1985–2012	There are 21 counties.
Municipality	LOUISE, Statistics Sweden, 1985–2012	There are currently 290 current municipalities in Sweden. However, there have been important revisions over time, which I accounted for when constructing the data.
CPI	Statistics Sweden	CPI is the deflation variable used to deflate monthly income and wages in the data. Throughout, I deflate the monthly income/wage variables so they are given in 2012 SEK.
Bartik shocks	Statistics Sweden	
Occupation ranking by interactions	O*Net	See Tables D.3 and D.4

Table D.3: Conversion from O*NET Categories to SSYK Categories: Part 1

SSYK Category Name	SSYK Code	SOC/O*NET Category Name	O*NET Codes
Legislators and senior officials	11	No comparable category found, used as omitted category	
Corporate managers	12	Management occupations	11
Managers of small enterprises	13	Management occupations (does not distinguish between large and small enterprises)	11
Physical, mathematical, and engineering science professionals	21	Computer and mathematical occupations; Architecture and engineering occupations	15, 17-1000, 17-2000
Life science and health professionals	22	Life, physical occupations; Health diagnosing and treating practitioners	19-1000, 19-2000, 29-1000
Teaching professionals	23	Education and training occupations	25-1000:25-3000 (excluding 25-1190)
Other professionals	24	Social science occupations; Community and social services occupations; Legal occupations; Library occupations; Entertainment, sports, and media occupations	19-3000, 21, 23-1000, 25-4000:25-9000, 27-2000:27-4000
Physical and engineering science associate professionals	31	Drafters, engineering technicians, and mapping technicians	17-3000
Life science and health associate professionals	32	Life, physical and social science technicians; Health diagnosing and treating practitioners; Other healthcare practitioners and technical occupations; Healthcare support occupations	19-4000, 29-2000, 29-9000, 31
Teaching associate professionals	33	Education and training occupations	25-1190
Other associate professionals	34	Legal support workers; Protective service occupations	23-2000, 33
Office clerks	41	Supervisors of office and administrative support workers; Material recording, scheduling, dispatching and distributing workers; Secretary and Administrative Assistants; Other office and administrative support workers	43-1000, 43-5000, 43-9000

Table D.4: Conversion from O*NET Categories to SSYK Categories

SSYK Category Name	SSYK Code	SOC/O*NET Category Name	O*NET Codes
Customer service clerks	42	Communications equipment operators; Information and records clerks; Financial clerks	43-2000, 43-3000, 43-4000
Personal and protective service workers	51	Food preparation and serving related occupations; Personal care and service occupations	35, 37-1000, 39
Models, salespersons, and demonstrators	52	Sales and related occupations	41, excluding 41-9090
Skilled agricultural and fishery workers	61	Farming, fishery, and forestry occupations	45
Extraction and building trades workers	71	Construction and extraction occupations	47
Metal, machinery, and related trades workers	72	Installation, maintenance, and repair occupations	49
Precision, handicraft, craft printing, and related trades workers	73	Art and design workers	27-1000
Other craft and related trades workers	74	Food processing workers; Textile, apparel, and furnishing workers; Woodworkers	51-3000, 51-6000, 51-7000
Stationary-plant and related operators	81	Plant and system operators	51-8000
Machine operators and assemblers	82	Assemblers and fabricators; Metal workers and plastic workers	51-2000, 51-4000
Drivers and mobile-plant operators	83	Motor vehicle operators; Rail transportation workers; Water transportation workers	53-3000, 53-4000, 53-5000
Sales and services elementary occupations	91	Building and grounds cleaning and maintenance occupations; Miscellaneous sales and related workers	37-2000, 37-3000, 41-9090
Agricultural, fishery, and related labourers	92	Farming, fishery, and forestry occupations (Does not distinguish labourers from skilled workers)	45
Labourers in mining, construction, manufacturing, and transport	93	All miscellaneous construction, mining, manufacturing, and transport workers and helpers	

E Robustness Checks and Additional Results

In this section I report the results for a number of alternative specifications and robustness checks. See the main text for discussion of these results.

E.1 Women

Table E.1: Learning Spillovers for Women

	(1)	(2)	(3)	(4)	(5)
Own education	0.219***				
	(0.0000)				
Lagged average education	0.081***	0.021***	0.020***	0.016***	0.017***
	(0.0015)	(0.0033)	(0.0040)	(0.0040)	(0.0040)
Individual effects		Yes			
Worker×Plant effects			Yes	Yes	Yes
County×Year				Yes	
Industry×Year				Yes	
County×Industry×Year					Yes

Notes: Dependent variable is current log wage. All models include year effects, as well as controls for number of children and marital status. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation within individual (column 2) and worker-plant spells (columns 3-5) are reported in parentheses. Column 1 also includes a quadratic control for experience but this drops out in the other columns once individual fixed effects are added.

E.2 Bartik Shocks

Bartik shocks introduce regional variation in labor demand based on changes in national demand for different industries' products. I construct Bartik shocks at the county and municipality level for every 5 years. I then include the Bartik shocks as a control in a regression of five-year differences. Bartik shocks are included as a finer level control for time-varying local demand shocks. While I include industry×county×time dummies in the main results, I was also able to construct Bartik shocks at the municipality level, which is a finer level of control compared with the county controls.

Traditionally, Bartik shocks are used to instrument or control for shifts in labor demand.[46] Since I am interested in controlling for shifts in demand for average education, I

[46]See, for example, Diamond (2016).

adjust the traditional Bartik series accordingly. My series is given by

$$\triangle B^{\bar{S}}_{mt} = \sum_k \frac{s^k_{mt-j}\left(1+\frac{\triangle N^k_{(-m)t}}{N^k_{(-m)t-j}}\right)}{\sum_k s^k_{mt-j}\left(1+\frac{\triangle N^k_{(-m)t}}{N^k_{(-m)t-j}}\right)}\bar{S}^k_{(-m)t} - \sum_k s^k_{mt-j}\bar{S}^k_{(-m)t-j}$$

where m stands for municipality, and k stands for industry. In the series above, $\bar{S}^k_{(-m)t}$ is the average education level in industry k in municipalities excluding the given municipality m in year t. $\bar{S}^k_{(-m)t-j}$ is the same but five years prior to t. s^k_{mt-j} is the average education level in municipality m in industry k five years before the current year. $N^k_{(-m)t}$ gives the number of workers in industry k not in municipality m at time t. Similarly for $N^k_{(-m)t-j}$, except in this case it is the same thing but five years earlier ($j = 5$). Thus, $\frac{s^k_{mt-j}\left(1+\frac{\triangle N^k_{(-m)t}}{N^k_{(-m)t-j}}\right)}{\sum_k s^k_{mt-j}\left(1+\frac{\triangle N^k_{(-m)t}}{N^k_{(-m)t-j}}\right)}$ gives the share of total workers in a given industry k for every municipality excluding the specific municipality in q. I construct the Bartik shocks using data aggregated by Statistics Sweden.

Table E.2: Controls for Bartik Shocks

	(1)	(2)	(3)	(4)	(5)
Own education	0.199***				
	(0.0013)				
Lagged average education	0.183***	0.040***	0.015**	0.015**	0.018***
	(0.0021)	(0.0042)	(0.0052)	(0.0052)	(0.0053)
Bartik shocks	0.559***	0.158***	0.122*	0.108*	0.118*
	(0.0174)	(0.0461)	(0.0475)	(0.0493)	(0.0496)
Individual effects		Yes			
Worker×Plant effects			Yes	Yes	Yes
County×Year				Yes	
Industry×Year				Yes	
County×Industry×Year					Yes

Notes: Dependent variable is current log wage. All models include year, county-year, and industry-year effects and controls for number of children and marital status. Children refers to number of children under age 6. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation are reported in parentheses.

E.3 Controlling for Average Education in the Municipality

Table E.3: Controlling for Average Education in the Municipality

	(1)	(2)	(3)	(4)	(5)
Own education	0.192***				
	(0.0010)				
Lagged average education	0.161***	0.059***	0.027***	0.025***	0.028***
	(0.0016)	(0.0036)	(0.0044)	(0.0044)	(0.0044)
Individual effects		Yes			
Worker×Plant effects			Yes	Yes	Yes
County×Year				Yes	
Industry×Year				Yes	
County×Industry×Year					Yes

Notes: Dependent variable is current log wage. All models include a control for average education at the municipality level. All models also include year effects, as well as controls for number of children and marital status. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation are reported in parentheses.

E.4 Estimates Using Wages

Table E.4: Using Wage Data

	(1)	(2)	(3)	(4)	(5)
Own education	0.211***				
	(0.0010)				
Lagged average education	0.104***	0.043***	0.037***	0.027***	0.036***
	(0.0017)	(0.0035)	(0.0033)	(0.0032)	(0.0033)
Individual effects		Yes			
Worker×Plant effects			Yes	Yes	Yes
County×Year				Yes	
Industry×Year				Yes	
County×Industry×Year					Yes

Notes: Dependent variable is current log wage, but in this case using the actual wage data as opposed to the income data converted to "monthly wages." However, this data is only available for select individuals (see text for discussion). All models include year effects, as well as controls for number of children and marital status. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation within individual (column 2) and worker-plant spells (columns 3-5) are reported in parentheses.

E.5 Estimates Restricted to Selected Private Firms

Table E.5: Restricting to Select Private Firms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own education	0.195***						
	(0.0011)						
Lagged average education	0.418***	0.070***	0.043***	0.038***	0.041***	0.053***	0.055***
	(0.0020)	(0.0041)	(0.0093)	(0.0091)	(0.0093)	(0.0065)	(0.0066)
Individual effects		Yes					
Worker×Plant effects			Yes	Yes	Yes	Yes	Yes
County×Year				Yes		Yes	
Industry×Year				Yes		Yes	
County×Industry×Year					Yes		Yes
Using wage data						Yes	Yes

Notes: Dependent variable is current log wage. All models include year effects, as well as controls for number of children and marital status. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation within individual (column 2) and worker-plant spells (columns 3-7) are reported in parentheses.

E.6 Estimates Restricting to Plants with <50 Workers

Table E.6: Restricting to Plants with <50 Workers

	(1)	(2)	(3)	(4)	(5)
Own education	0.166***				
	(0.0013)				
Lagged average education	0.193***	0.040***	0.012*	0.010*	0.012*
	(0.0020)	(0.0040)	(0.0048)	(0.0048)	(0.0048)
Individual effects		Yes			
Worker×Plant effects			Yes	Yes	Yes
County×Year				Yes	
Industry×Year				Yes	
County×Industry×Year					Yes

Notes: Dependent variable is current log wage. All models include year effects, as well as controls for number of children and marital status. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation within individual (column 2) and worker-plant spells (columns 3-5) are reported in parentheses.

E.7 Persistence of Spillovers

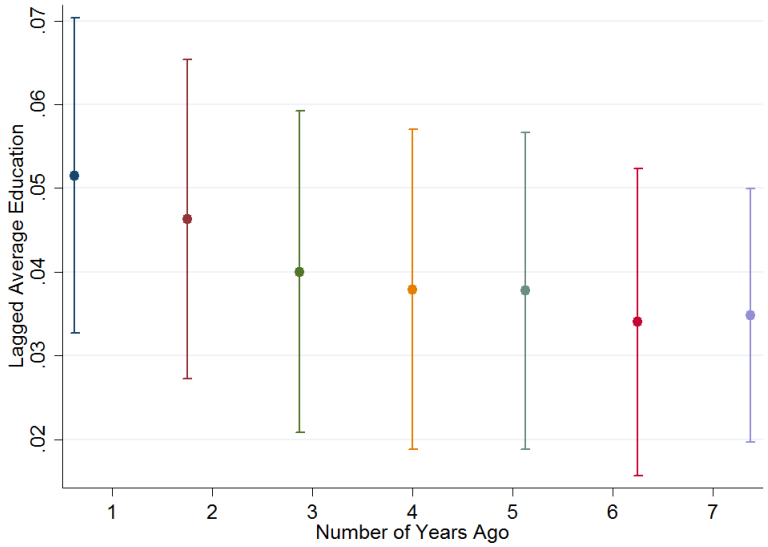
Table E.7: Persistence of learning spillovers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lagged average education	0.025***	0.029***	0.031***	0.031***	0.029***	0.028***	0.035***
	(0.0044)	(0.0049)	(0.0053)	(0.0059)	(0.0052)	(0.0072)	(0.0077)
Lag year	1	2	3	4	5	6	7

Notes: Dependent variable is current log wage. All models include year, county-year, industry-year and plant-worker fixed effects. All models also include controls for number of children and marital status. Children refers to number of children under age 6. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation within spells are reported in parentheses.

Restricting to Same Sample In the graph below, I repeat the exercise in Figure 2, but restrict to the same sample. Specifically, Figure E.1 shows the persistence of spillovers, but restricts every specification to individuals who have remained at the same workplace for the past seven years. The table with estimates is available upon request. The figure shows that, once again, the effect of learning spillovers is persistent. However, in this figure the effects decreases gradually over time.

Figure E.1: Persistence of Spillovers over Time, Robust



E.8 Estimates by Age

In the table below I estimate the learning spillovers by age, but unlike Figure 3 in the main text, I provide a robustness check where I do not use overlapping 10-year intervals. The main takeaway is the same, namely that learning spillovers are larger for younger workers, and still positive but no longer significant for older workers.

Table E.8: Estimates by Age

	(1)	(2)	(3)	(4)
Lagged average education	0.047**	0.017*	0.010	0.011
	(0.0146)	(0.0081)	(0.0073)	(0.0098)
Age	25-34	35-44	45-54	55-64

Notes: Dependent variable is current log wage. All models include year, county-year, industry-year and spell fixed effects. Children refers to number of children under age 6. County controls consist of dummies for each of the 21 counties. Industry controls consist of dummies for each of 17 industry categories. Full regression results are available upon request. Each column is a separate regression. Robust standard errors accounting for the serial correlation within spells are reported in parentheses.