Learning Spillovers in the Firm

JOB MARKET PAPER

Emily Nix*

Yale University

December 11, 2015

Click here for the latest version

Abstract

To produce output for a firm, colleagues inevitably interact. This paper examines the possibility that as a by-product of these interactions, there are learning spillovers: colleagues learn general skills from each other that increase future productivity. The first half of the paper establishes a novel theoretical result. Learning spillovers imply an externality in the return to human capital investment, which firms may not internalize and will result in inefficient investments in education under realistic conditions. The second half of the paper shows that learning spillovers are empirically relevant. I match Swedish data on workers, their peers, and their firms from 1985-2012. I use a combination of fixed effects and controls to address bias from worker sorting and firm heterogeneity. I find that increasing average education of a given worker's colleagues 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 also persistent, in that average education of colleagues impacts wages at least five years in the future, although the impact decreases somewhat over time. In addition, I document interesting heterogeneity consistent with learning spillovers. I show that 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. I also find that the effect varies in expected ways across occupations. For example, professionals and managers obtain the largest spillovers from their coworkers, while drivers, who interact little with colleagues, experience the smallest impact.

*Telephone: 504-352-4072. Email: emily.nix@yale.edu. I am extremely grateful to my advisors, Joseph Altonji, Costas Meghir, Lisa Kahn, and Nancy Qian, for their invaluable guidance and support. This research is based upon work supported by the Institute for Evaluation of Labour Market Education Policy in Uppsala, Sweden, and I am particularly indebted to Lisa Laun, Anders Forslund, and Jörgen Moen. I am also thankful to Noriko Amano, Vittorio Bassi, Benjamin Friedrich, Ran Gu, Rebecca McKibbon, Corina Mommaerts, Ana Reynoso, Larry Samuelson, Franklin Williams, and participants at the Yale Labor Workshops for helpful comments. 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. In order to produce output colleagues often interact, possibly learning from and teaching one another. In this paper, I focus on these potential "learning spillovers" in firms. While there is little doubt that peers shape individual outcomes, prior research has primarily focused on the impact of peers in schools, neighborhoods, cities, and states. In contrast, much less is known about learning spillovers from coworkers in firms.

This paper provides a theoretical and empirical analysis of learning spillovers in firms. I first define a theoretical 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 of a given worker's colleagues 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 worker ranges from 0.194-0.222, with 12.61%-14.43% of the total increase attributable to learning spillovers.

In my theoretical model workers increase their stock of general skills as a by-product 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.[1] 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 type. Third, colleagues are non-excludable and (partially) non-rival inputs in the production of

[1]For example, Acemoglu discusses these spillovers saying "excluding education and R&D, major human capital interactions happen among employees within a firm: for example, young workers learn from their more experienced colleagues. But these interactions should be internalized within the firm, and no economy wide human capital externalities should be observed." (Acemoglu (1996)). Moretti states that "potential spillovers that occur within a plant...are likely to be internalized" (Moretti (2004b)). Barro says "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" (Barro (1996)). Topel and Lange summarize the literature saying "when productive interactions occur within firms they are merely complementarities that will be internalized and priced" (Lange and Topel (2006)).

learning spillovers. Under these conditions, colleagues 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 equlibrium may be inefficient.

I test for learning spillovers using Swedish administrative data. I construct a unique data set covering the universe of workers, their peers, and firms in Sweden from 1985-2012. Motivated by the theoretical model, I test for learning spillovers by looking at the relationship between the education level of past colleagues 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 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 colleagues impacts wages at least five years in the future, although the impact decreases somewhat over time. Compared to average wage growth in Sweden over this time period of roughly 1.7%-2% per year, my 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 opportunity for interactions with colleagues. 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 colleagues, experience the smallest impact.

These findings have important implications for the returns to education. Using the theoretical results, 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 of a worker with a college education. My findings suggest that the social return of adding an additional college worker ranges from 0.194-0.222, with 12.61%-14.43% of the total increase attributable to learning spillovers.

This paper is related to several literatures. First, this paper is closely related to the literature on human capital accumulation on the job. In his seminal paper, Becker (2009) shows that investments in general skills on the job are efficient. Workers are willing to take pay cuts to finance investments in general training.[2] This overturned the prior conclusion in Pigou (1912). Acemoglu (1997) and Acemoglu and Pischke (1999) extend the traditional general training model to include various frictions, and explore how these frictions affect the efficiency of general training investments. In this paper, I extend the general training literature to consider general skills obtained on the job through learning spillovers.

Second, this paper contributes to the peer effects literature. There is little doubt that peers shape individual outcomes. However, while there is a large body of evidence on peer effects at schools and within neighborhoods (Ammermueller and Pischke (2009), Angrist and Lang (2004), and Sacerdote (2001)), there is much less evidence on peer effects at work. 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. Waldinger (2012) finds no evidence of peer effects in academic departments using variation induced by the Nazi government's expulsion of scientists in Germany in 1933. Jackson and Bruegmann (2009) find that students in classrooms led by teachers exposed to better colleagues experience larger test gains. Most related to this paper are Martins and Jin (2010), who estimate contemporaneous social returns to education in firms in Portugal and find large social returns, between 14% and 23%, and Cornelissen et al. (2013) who estimate contemporaneous social returns to peer fixed effects in firms in Germany and find small evidence of peer effects in wages.

Third, this paper contributes to the literature on education externalities.[3] The education externalities literature has focused on across firm education externalities, with mixed results. Acemoglu and Angrist (2001) look at externalities from an increase in high school workers and find modest returns of around 1-3%, while Moretti (2004b) focuses on college educated workers and finds that a 1% increase in city share of college workers increases output by 0.5-0.6 percentage points. In contrast, this paper looks at education spillovers within the firm.

The paper is organized as follows. In Section 2 I define a theoretical model with learning spillovers. I then use a general equilibrium framework to solve for wages and discuss the implications for efficiency. 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. Section 5

[2] See also Becker (1962), Ben-Porath (1967), and Heckman et al. (1998).

[3] For example, see Lucas Jr (1988), Nelson and Phelps (1966), and Moretti (2005).

presents the main results, with additional results that address remaining threats to identification presented in 6. In Section 7 I summarize my findings and discuss the broader implications before concluding in Section 8.

2 A Model of Learning Spillovers in the Firm

In this section I present a theoretical model of learning spillovers in the firm. I start by defining a simple environment that includes learning spillovers based on the average education of workers in the firm. I then use a basic general equilibrium framework to explore the theoretical implications of a model that includes learning spillovers in the firm.

2.1 The Environment

The hypothesis underlying this entire paper is that workers learn general skills from their colleagues. More specifically, in this paper I focus on a specific, narrow question based on this general hypothesis: do workers learn general skills based on the fraction of workers in the firm who are college educated?

Formally, the economy consists of J firms and a continuum of individuals in I. The amount of general skills a given individual i learns at a firm f depends on the fraction of college workers in the firm and the individual's type.

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 are 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

$$\bar{S_f} = \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \tag{1}$$

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

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

and B types receive learning spillovers

$$s_f^B = \alpha^B \bar{S_f} \tag{3}$$

where α is the learning parameter and $\alpha^A > \alpha^B$.[4]

The J firms are all identical. This assumption combined with assumptions on total production (outlined in Appendix B.1[5]) allows me to rule out sorting driven by learning spillovers. I rule out such sorting so that I can start with the simplest possible theoretical framework in order to provide some initial implications of learning spillovers in the firm.

In particular, the solution is much simpler since in an equilibrium without sorting or firm heterogeneity, all firms demand the same average education. However, ruling out sorting ignores some interesting and important implications. In work in progress I am analyzing both the theoretical and empirical implications of allowing for sorting. I revisit this point in the conclusion.

There are three periods. In the first period, individuals choose to go to college or not. Their choice depends on their personalized cost of college, θ^i, and the relative return to college versus high school, which they take as given. The individual costs to college have a uniform distribution over the interval $[0, 1]$. These costs are uncorrelated with the learning parameters.[6]

In the second period, firms demand workers in order to produce consumption goods. Consumption goods are produced 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 by-product of hiring college and high school workers to produce consumption goods, these same workers also gain learning spillovers from each other, as given in equations 2 and 3. These learning spillovers enter the problem in two ways.

First, they impact total production of consumption goods in the second period. I assume that each worker's marginal productivity increases by exactly the amount of his learning spillover. Thus, with spillovers, total production of second period consumption goods at a firm f is

$$F\left(H_f, L_f\right) + \underbrace{\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\bar{S}_f}_{second\ period\ spillovers} \tag{4}$$

which is also constant returns to scale, given that F is constant returns to scale.[7]

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

[5] In brief, I assume that total production (of both consumption goods and spillovers) is increasing in each education-learning type, but at a decreasing rate.

[6] Allowing for correlation between individual's learning parameters and costs to education is another interesting extension that I leave to future research.

[7] Note that an alternative way of incorporating the spillovers would be to write:

Second, they 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 period due to learning spillovers is given by:

$$\underbrace{\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\bar{S_f}}_{\textit{second period spillovers}}+\underbrace{\delta\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\bar{S_f}}_{\textit{third period spillovers}} \quad (5)$$

The fact that learning spillovers impact future as well as present productivity of workers is key for the possibility of inefficiency.[8]

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. Directly modeling wage increases in the third period due to second period learning spillovers does not change the results.

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

$$U^i = c_1^i + c_2^i + c_3^i \quad (6)$$

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

2.2 Competitive Equilibrium with Learning Spillovers and Implications for Efficiency

In this subsection, I solve for a competitive equilibrium with learning spillovers under three possible scenarios. First, I present a worst case scenario where the externalities are ignored. As expected, I show that no internalization occurs and workers underinvest in education.

I then move on to a more interesting question. Will the competitive equilibrium be efficient when firms know that learning spillovers occur and attempt to internalize them by effectively "charging" workers for the spillovers? I find that if firms know workers' types, and are able to pay personalized wages, then the competitive equilibrium fully internalizes learning spillovers.

However, the conditions required for this result are not plausible, for reasons I discuss

---

$F\left(H_f+\alpha^A H_f^A\bar{S_f}+\alpha^B H_f^B\bar{S_f}, L_f+\alpha^A L_f^A\bar{S_f}+\alpha^B L_f^B\bar{S_f}\right)$. This does not change the results, so for simplicity, I use the current specification.

[8] As I discuss in Subsection 2.3, when spillovers only occur in the second period, the outcome is efficient.

in detail below. Given that, I close this subsection by presenting a solution under more realistic conditions. In particular, I solve for an equilibrium with asymmetric information. I find that a competitive equilibrium with learning spillovers is no longer guaranteed to be efficient.

To provide some intuition for the results, it is useful to start from the fact that the existence of learning spillovers means that workers impose externalities on each other. As a by-product of consumption good production, workers also obtain general skills based on the average education of the firm. Thus, a college education provides two benefits to the economy.

First, it increases the total amount of consumption goods due to the direct increase in productivity of the worker with the college education. Second, it increase the total amount of consumption goods by increasing the total average education in the economy, which in turn increases learning spillovers in the firm. In the Pareto efficient conditions for the optimal number of college educated workers of each type (see Appendix A.1), these two benefits are made explicit.

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. For college workers, this should imply an increase in wages. College workers increase average education in the firm, and thus impose positive externalities on colleagues. For high school workers, this should imply a decrease in wages. High school workers decrease average education in the firm, and thus impose negative externalities on colleagues.

Suppose instead that workers are not paid their marginal products in terms of producing learning spillovers. For example, suppose that firms are unaware the learning spillovers occur. In the language of the literature on externalities, this means that the externality is not priced. As is well known from the literature on externalities, if the externality is not priced, then inefficiency will result. I show this formally in the following proposition:

Proposition 1. *Suppose that firms are not aware of the learning spillovers provided for workers, 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:*

$$\begin{aligned} w_f^{H^K} &= F_1 + \alpha^K \bar{S_f}^* \\ &+ \left(\alpha^A \left(H_f^{A*} + L_f^{A*}\right) + \alpha^B \left(H_f^{B*} + L_f^{B*}\right)\right) \left(\frac{1}{H_f^* + L_f^*} - \frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}\right) \end{aligned} \tag{7}$$

$$
\begin{aligned}
w_f^{L^K} &= F_2 + \alpha^K \bar{S_f}^* \\
&\quad - \left(\alpha^A \left(H_f^{A*} + L_f^{A*}\right) + \alpha^B \left(H_f^{B*} + L_f^{B*}\right)\right) \left(\frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}\right) \qquad (8) \\
K &= A, B \qquad (9)
\end{aligned}
$$

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

As expected, equilibrium wages fail to fully internalize the externality. Equilibrium wages do not include the marginal productivity of each worker in terms of producing future learning spillovers, although they do include the marginal productivity of each worker in terms of producing current period learning spillovers. As a result, the outcome is inefficient. Workers underinvest in education.

It is worth stressing that the lack of internalization of learning spillovers in wages is not by itself sufficient for the outcome to be inefficient. Rather, it is the fact that the lack of internalization provides the wrong incentives for education, which is endogenously chosen, that makes the outcome inefficient. Suppose that education were actually exogenous. In that case, the equilibrium would be efficient, even though wages do not internalize learning spillovers.

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.

In summary, Proposition 1 shows that when the spillovers are ignored (and depend on endogenous choices of workers), the outcome is inefficient. I now move on to a more interesting question: will the competitive equilibrium be efficient when firms attempt to "charge" workers for learning spillovers?

How much can firms deduct from worker's wages? Any worker employed by a firm f is exposed to the same average education within the firm, $\bar{S_f}$. However, workers with different learning parameters receive different benefits from the same average education exposure. In order for wages to fully internalize learning spillovers, firms must account not only for the fact that learning spillovers exist, but also for the total amount of learning spillovers that occur.

To give a competitive equilibrium the best shot at meeting this requirement and generating an efficient outcome, I start by assuming firms observe types and can pay personalized wages to account for the total amount learned by each type. This is similar to the conditions for a Lindahl equilibrium for public goods (Lindahl (1919)).[9] Similar to the approach described in Milleron (1972), in order to sustain a competitive equilibrium with personalized wages, I redefine the spillover.

Specifically, let s_f^A be the spillover experienced by the A types with learning parameter α^A at firm f, and let s_f^B be the spillover experienced by the B types with learning parameter α^B at firm f. This extends the "public" good, $\bar{S}_f$, into I private goods, the exposure as experienced by each individual in the economy.

Then, firms maximize profits relative to each worker's participation constraint. The participation constraints are determined by the workers' problem. 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 type are:

$$\begin{aligned} w_f^{H^K} + \delta\alpha^K \bar{S}_f &\geq w^{H^K} && (10) \\ w_f^{L^K} + \delta\alpha^K \bar{S}_f &\geq w^{L^K} && (11) \\ K &= A, B && (12) \end{aligned}$$

These conditions make explicit the trade-off between wages and the spillover that in turn affect 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:*

$$\begin{aligned} w_f^{H^K} &= F_1 + \alpha^K \bar{S}_f^{\,*} && (13) \\ &\quad + (1+\delta)\left(\alpha^A\left(H_f^{A*} + L_f^{A*}\right) + \alpha^B\left(H_f^{B*} + L_f^{B*}\right)\right)\left(\frac{1}{H_f^* + L_f^*} - \frac{H_f^*}{\left(H_f^* + L_f^*\right)^2}\right) \\ w_f^{L^K} &= F_2 + \alpha^K \bar{S}_f^{\,*} && (14) \end{aligned}$$

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

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

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 of college workers relative to high school workers. This is due to the fact that college workers increase learning spillovers while high school workers decrease learning spillovers.

In addition, by effectively restricting each type of worker to purchase only the spillover as experienced by that 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 of 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 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, this result is driven by the assumption that it is possible to effectively charge workers different amounts for exposure to the same colleagues, through type dependent reductions in wages. In practice, this translates to an assumption that paying personalized wages is feasible. The assumption that personalized wages are feasible will generally not be true.

The major challenge is imperfect information. Personalized wages require more information than is usually needed for a competitive equilibrium to be efficient. Normally, all a firm needs to know is its own technology and the price of labor and all a worker needs to know is his own type and the price of labor. In contrast, here firms must also know individual's types. Either this information must be general knowledge, which is unlikely, or workers must voluntarily reveal their learning parameters.

Suppose there is asymmetric information. Will workers voluntary reveal their types? To answer this question, I ask if there is any incentive compatible set of wages that pay high learning and low learning types different amounts. 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:

$$\begin{aligned}
w_f^{H^B} + \delta\alpha^A \bar{S_f} &\geq w_f^{H^A} + \delta\alpha^A \bar{S_f} && (15)\\
w_f^{H^A} + \delta\alpha^B \bar{S_f} &\geq w_f^{H^B} + \delta\alpha^B \bar{S_f} && (16)\\
w_f^{L^B} + \delta\alpha^A \bar{S_f} &\geq w_f^{L^A} + \delta\alpha^A \bar{S_f} && (17)\\
w_f^{L^A} + \delta\alpha^B \bar{S_f} &\geq w_f^{L^B} + \delta\alpha^B \bar{S_f} && (18)
\end{aligned}$$

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 types 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, irregardless 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:

$$\begin{aligned}
w_f^{H} &\geq w^{H^K} - \delta\alpha^A \bar{S_f} && (19)\\
w_f^{H} &\geq w^{H^B} - \delta\alpha^B \bar{S_f} && (20)\\
w_f^{L} &\geq w^{L^A} - \delta\alpha^A \bar{S_f} && (21)\\
w_f^{L} &\geq w^{L^B} - \delta\alpha^B \bar{S_f} && (22)
\end{aligned}$$

In summary, the personalized wages that allow for an efficient outcome in Proposition 2 are similar to personalized prices required for a Lindahl equilibrium for public goods. Naturally, the concerns are also similar (asymmetric information and thin markets, see Arrow (1970)). My conclusion, then, is that insofar as a Lindahl equilibrium is a realistic solution to the public goods problem, firms are able to fully internalize learning spillovers by paying personalized wages in my setting.

This naturally leads to the following question. Under more realistic assumptions, are learning spillovers fully internalized and is the outcome efficient? In Proposition 3, I introduce asymmetric information by assuming that while workers know their learning parameters, firms do not. As I showed above, under these conditions it is not possible for firms to pay different wages to different learning types. As a result, I find that the efficient outcome is no longer guaranteed.

Proposition 3. *Suppose there is asymmetric information such that workers know their learning parameters and firms do not. Then:*

1. *Multiple prices are compatible with a competitive equilibrium.*
2. *Only one of the possible set of prices is Pareto efficient.*
3. *If the equilibrium is chosen at random, the equilibrium is efficient with probability zero.*

The set of possible equilibrium wages by education are:

$$w_f^H = F_1 + E[\alpha] + \delta\alpha^B \frac{L_f^*}{H_f^* + L_f^*} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L_f^*}{\left(H_f^* + L_f^*\right)^2} \tag{23}$$

$$w_f^L = F_2 - \delta\alpha^B \frac{H_f^*}{H_f^* + L_f^*} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H_f^*}{\left(H_f^* + L_f^*\right)^2} \tag{24}$$

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, high learning type participation constraint, λ_2 is the Lagrange multiplier on the college, low learning type participation constraint, λ_3 is the Lagrange multiplier on the high school, high learning type participation constraint, and λ_4 is the Lagrange multiplier on the high school, low learning type participation constraint. In addition, workers receive their type specific learning spillovers in the third period.

Proof. See Appendix A.4.

These results are preliminary, and the analysis of a richer class of models is needed so I will not spend too much time on them. Instead I refer the interested reader to the proof and accompanying discussion in the Appendix. In work in progress I am examining these results in more detail.

2.3 Discussion of the Theoretical Results and Comparison to Existing Literature

In summary, the conclusions from the theoretical model 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). There are three additional points that are worth discussing before turning to the main focus of this paper, an empirical assessment of learning spillovers.

First, the challenges in markets with learning spillovers do not occur with traditional training inputs. This is the standard result in the literature, which 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 types. Thus, with traditional inputs the equilibrium exists, is unique, and is Pareto efficient.

This highlights the unique challenges learning spillovers pose to firms attempting to internalize these spillovers. The challenge with learning spillovers is that since all workers are exposed to the same average education within the firm, it is simply not possible to get individuals to voluntarily receive different payments for that exposure.

A second important point is that learning spillovers must impact future productivity. Otherwise, the result is always efficient. To see this, consider a setting where learning spillovers do not affect future productivity. In that case, the participation constraints are simply

$$w_f^H \geq w^H \tag{25}$$
$$w_f^L \geq w^L \tag{26}$$

As I've discussed above and is shown in the proofs of the Propositions, the inefficiency is driven by the differences in participation constraints by type. In a setting where the effects of learning spillovers do not persist, the participation constraints no longer pose challenges, and the outcome is efficient.

Third, the theory provides important insights for my empirical approach. Given the possibility that 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 theory and my empirical results in Section 7.

Additionally, the possibility that learning spillovers may be partially or fully internalized has important implications for the empirical model. On that note and with the theoretical predictions in hand, I now turn to the main focus of this paper, an empirical assessment of learning spillovers in the firm.

3 Empirical Framework

The goal of the empirical section is to test for the prevalence of learning spillovers. All else equal, does a given worker exposed to more educated colleagues learn more general skills relative to an identical worker exposed to less educated colleagues? The theoretical model informs us that we should look at the effect of average education in a worker's previous firm on his current wage. This follows from equations 19-22 and 23-24, which show that the effect of current colleagues on current wages is ambiguous. 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 colleagues could be consistent with a model of learning spillovers.

In contrast, the predictions of the theoretical model is unambiguous regarding the impact of past colleagues. All else equal, the theoretical model predicts that a worker exposed to a firm with higher average education last 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

$$w_{it} = \pi_1 \bar{H}_{it-1} + \pi_0 h_i + d_t + \varepsilon_{it} \tag{27}$$

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

An estimate of $\hat{\pi}_0 > 0$ implies that college workers are more productive than high school workers ($F_1 > F_2$), but may also capture the positive externality high educated workers impose on colleagues relative to the negative externality low educated workers impose on colleagues.[10] In contrast, the coefficient on average education the worker was exposed to in his firm(s) last year only captures learning spillovers. If the estimator of π_1 is unbiased, $\hat{\pi}_1 = 0$ implies there are no learning spillovers and $\hat{\pi}_1 > 0$ implies there are learning spillovers.

There are two reasons why OLS estimates of 27 will be biased. First, there could be

[10] $\hat{\pi}_0$ will only capture learning spillovers insofar as these spillovers are internalized. I return to this point in Section 7.

time invariant omitted variables, such as worker sorting and unobserved firm heterogeneity. I use either firm×worker or firm and worker fixed effects to deal with any time invariant omitted variables. For example, individual fixed effects deals 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 migrants 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 (provided the increase in demand for local goods is strong enough to counteract the decrease in college worker wages from the exogenous supply shift). 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×time fixed effects, d_{ct}, and industry×time fixed effects, d_{kt}. This leads to the following regressions.

$$w_{it} = \pi_1 \bar{H}_{it-1} + \pi_x X_{it} + d_t + d_{i\mathbf{f}(i,t-1)} + d_{ct} + d_{kt} + \varepsilon_{cfikt} \quad (28)$$

$$w_{it} = \pi_1 \bar{H}_{it-1} + \pi_x X_{it} + d_t + d_i + d_{\mathbf{f}(i,t-1)} + d_{ct} + d_{kt} + \varepsilon_{cfikt} \quad (29)$$

X_{it} is a vector of time varying individual controls consisting of number of children and marital status.[11] 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 28, $d_{i\mathbf{f}(i,t-1)}$ is a firm (where the worker was employed last year) by worker fixed effect. In equation 29, 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 et al. (1999) and I estimate these fixed effects in equation 29 similarly to other papers in this literature. More details can be found in Appendix C.2.

I report estimates using both firm×worker fixed effects and firm and worker fixed

[11] In robustness checks, I have also included a quadratic in experience and own education interacted with year and county. Those estimates are available upon request, and are very similar to the main results.

effects because the two approaches identify the coefficient using different variation in the data.[12] Controlling for firm×worker matches restricts the identifying variation to movement of colleagues 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 firm average education from workers who move firms in addition to changes in colleagues among workers who stay in a given 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.[13]

The inclusion of firm, worker, time, county×time, and industry×time fixed effects naturally limits the scope for omitted variable bias.[14] An omitted variable must meet all of the following conditions at the same time in order to bias my estimates.

1. Time-varying
2. Correlated with changes in future wages
3. Correlated with changes in current average education in workers' firms
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 workers than high school workers relative to its existing ratio of college versus high school workers, increasing the average education within the firm. However, due to labor market frictions the firm can't hire as many college workers as it would like. This in turn causes the firm to increase training of existing workers to increase their productivity. To address this story and other possible stories

[12] 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, time and firm fixed effects.

[13] 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 education because they move firms compared to a worker who experiences a similar increase from a change in a few colleagues 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 colleagues with different skills.

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

like it, in Section 6 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 the worker's own education.[15] 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 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.[16] See appendix C.1 for more details.[17]

4 Data Construction and Descriptive Statistics

To estimate learning spillovers in the firm, I require data on workers and their current and past colleagues. In order to meet these requirements, I build on work by Lisa Laun and co-authors (see Friedrich et al. (2015) for more details) to construct a unique data set linking ten separate administrative and survey data sources.[18] The raw data is compiled by Statistics Sweden. I link the data for the entire population from 1985-2012.

The data on employers comes from two sources. First, there is registry data which covers all companies. Second, there is the Structural Business Statistics (SBS) which consists of accounting and balance sheet data. From 1997 onward, data is provided for all non-financial firms. From 1985-1996, 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 data 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

[15]Measurement error in average education of past colleagues 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.

[16]Individual fixed effects control perfectly for time invariant characteristics. Thus, to the extent that own education is time invariant, individual fixed effects control for it perfectly. As a result, measurement error in 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.

[17]Given the possible broader applicability of this solution to the upward bias in social returns estimates, in Appendix C.1 I derive the results formally, outline the conditions when it can be used successfully, and also demonstrate its usefulness though a simple simulation exercise.

[18]The major differences between the two uses for the data required important changes and additions to meet the needs of my particular application, so much so that the matching, cleaning, and variable selection was entirely redone. In particular, additional years were added, wage data from 5 additional data sets (although subject to population selection issues outlined below) was added, and variables used for specific controls were all incorporated. Furthermore, many changes to the structure of the data were necessary for my application and construction of the past exposure to colleagues posed some unique challenges. Given all of the changes made, in this section I describe the data selection and construction in detail.

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, the wage data is only available for all non-financial firms from 1997 and onwards. 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 provided in RAMS, but provide robustness checks in the appendix using the wage data.[19]

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 dividing total annual income by total months worked in the year.[20] 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.

A second concern is that many 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.[21] 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 work-

[19] See Table 10 in Appendix E.

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

[21] 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 (which comes to $3,000 per month). From January through March, when the employment spells overlap, Tom counts as .75 units of college educated workers in plant B and .25 units of college educated workers in plant A. Then, from April-December, Tom counts as 1 unit of college educated workers in plan B.

ers 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 up over all months and divide by 12. This gives me monthly exposure to college and high school workers. Last, I divide monthly exposure to college 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 colleagues is that the finest level of interaction I can get with this data is at the plant/work site level. The plant is defined as "every address, property, or group of neighboring property units in which a company operates". While this is relatively detailed, it is still limited. Not all workers in the same plant may interact, and I have no way of identifying which workers do interact. While such data is rarely available in conventional data, it can be quite helpful, as shown in Mas and Moretti (2009). This limitation introduces a specific source of measurement error - the average education I use may not be an accurate portrayal of the average education of individuals with whom a given worker actually interacts. This likely introduces classical measurement error which will attenuate the coefficient on lagged average education.

In Table 1 I present summary statistics of my main variables of interest. While I used the entire population to construct the data and all variables in the analysis, due to computational constraints, I restrict estimation of learning spillovers in the firm to a 5% sample.[22] I also restrict to men age 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.

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

[23]I restrict to men since I am unable to adequately account for part time work, which is much more prevalent among women as compared to men.

Table 1: Summary Statistics

	All	≤High School	College
Average Year-Worker Observations	20.69	21.08	19.79
Real monthly earnings, 2012 SEK	27,685	24,654	34,320
Age	43.80	44.18	42.89
Married	0.50	0.47	0.56
Number of children aged 0-3	0.16	0.14	0.20
Number of children aged 4-6	0.12	0.11	0.15
Employed, of which	0.80	0.78	0.84
Job Stayer	0.82	0.82	0.81
Job Mover	0.14	0.14	0.14
Re-entrant	0.03	0.03	0.02
Industry			
Construction	0.12	0.16	0.04
Manufacturing	0.28	0.32	0.20
Retail Trade	0.13	0.15	0.08
Services	0.47	0.37	0.68
Lagged Average College Share	0.31	0.18	0.58
Observations	2,312,509	1,622,930	689,579

Notes: Based on the 5 percent sample used in estimation. Monetary values are in 2012 SEK.

Figure 1 provides suggestive evidence on learning spillovers. I present a binned scatterplot of current wages and average education of colleagues last year and the overlaid regression line.[24] It shows a strongly positive relationship. Figure 2 depicts the relationship conditional on the following controls: 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. Clearly, both of these figures are merely descriptive. Both are subject to the selection issues and endogeneity bias described in Section 3. To address these issues, I now move to the results from the estimation strategy outlined in Section 3.

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

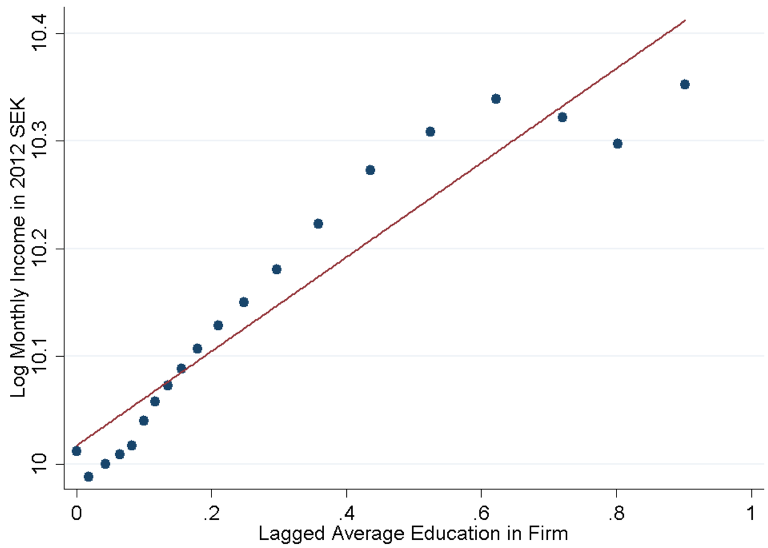


Figure 1: Binned Scatterplot of Current Wages and Average Education of Colleagues Last Year

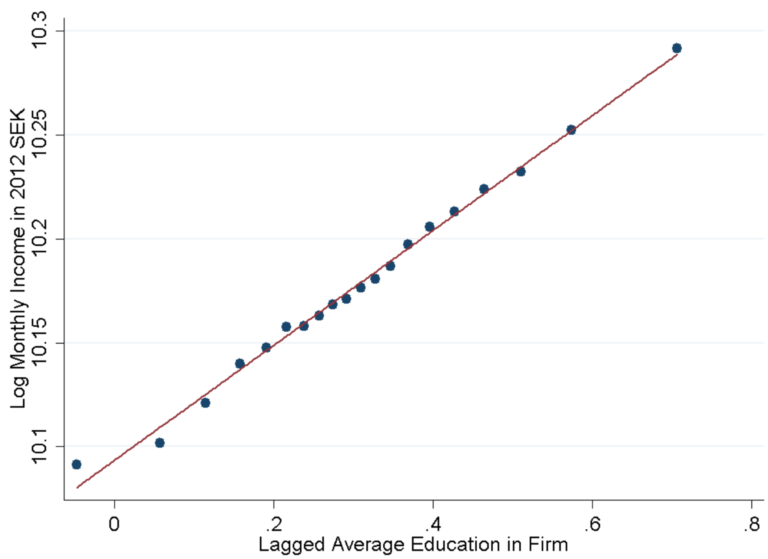


Figure 2: Binned Scatterplot of Current Wages and Average Education of Colleagues Last Year, Conditional on Controls

5 Estimates of Learning Spillovers in the Firm

In column 1 of Table 2 I report estimates from the OLS regression (equation 27).

Table 2: Learning Spillovers

	(1)	(2)	(3)	(4)	(5)
Own education	0.194***				
	(0.0010)				
Lagged average education	0.191***	0.049***	0.032***	0.028***	0.031***
	(0.0016)	(0.0035)	(0.0049)	(0.0049)	(0.0050)
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 parenthesis.

The coefficient on lagged average education, i.e. the learning spillover, is 0.191. It is almost identical to the estimated return to college, which is 0.194. Naturally, this raw correlation suffers from many sources of bias. Column 2 adds individual fixed effects. The estimate of learning spillovers drops substantially, to 0.049. 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 0.028.

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%. Recall that the remaining identification that identifies this effect is from movement of colleagues in and/or out of a given worker's firm that is not captured by industry×time and county×time trends.

In Table 3, I report the results from estimating equation 29 with separate firm and plant fixed effects. 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 variation in the data.[25] The estimates are similar to Table 2, although the coefficient on lagged average

[25] I present the estimates in 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.

education is larger at 0.058. The interpretation is that a 10 percentage point increase in average education of 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 parenthesis. In work in progress I am increasing the number of bootstrap replications. See Appendix C.2 for more details on the estimation procedure.

Figure 3 documents the persistence of learning spillovers over time since exposure to colleagues.[26] The graph plots the impact of lagged average education on wages by time since exposure. The effect appears to be persistent. It does decrease somewhat over time and once it has been six years since exposure, the effect is no longer statistically significant. However, the lack of significance for deeper lags could be driven by the reduction in sample size that accompanies deeper lags. The substantial loss in sample size with deeper lags is also why I do not present results that go even further back in time.

[26] Table of the estimates can be found in the Appendix.

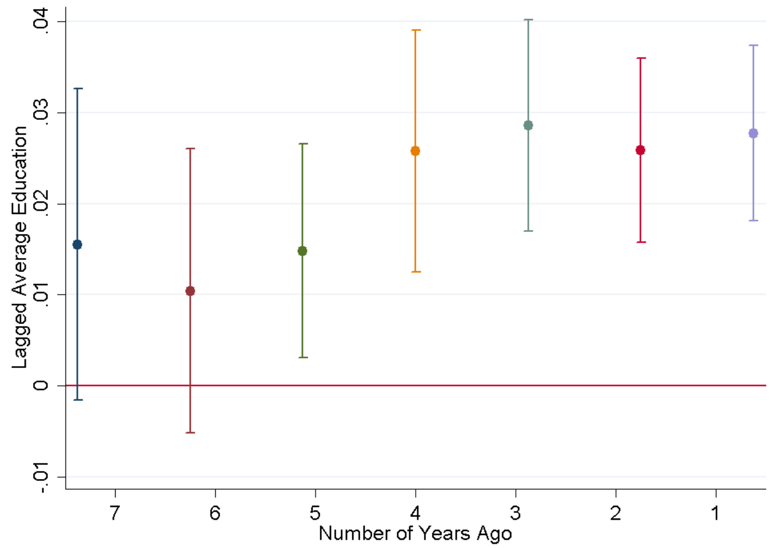


Figure 3: Persistence of Spillovers over Time

In Appendix E, I report estimates using additional controls and alternative data samples to deal with specific 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 Bartick shocks at the municipality level, and include these shocks as controls. For more information on Bartik shocks and how I construct the shocks for my setting, see Appendix E. The estimates of learning spillovers are slightly smaller with the inclusion of Bartik shocks (see Table 8). Second, I control for average education at the municipality level. The estimate does not change when I control for average education at the municipality level (see Table 9).[27] 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.

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

[27] The results do not yet include the worker and worker×firm fixed effect specifications. Those results are in progress and will be added shortly. However, the results without the fixed effects are identical.

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

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. Estimates using the wage data are identical (see Table 10).

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 assess this concern is to compare the estimates in Table 2 to estimates restricting the sample to workers employed by private sector firms. Given the relatively stronger joint wage bargaining in public jobs in Sweden versus private jobs, I would expect the 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 11, columns 1-5). Using the reported montly wages for private sector workers increases the estimates even more: I find that a 10 percentage point increase in average education of a worker's firm increases wages in the following year by 0.53% (see Table 11, columns 6-7).[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 7. First, though, I document interesting heterogeneity in the effects by age and occupation that provides further evidence in favor of learning spillovers.

6 Additional Results

The estimates presented in Section 5 show that workers exposed to colleagues with higher average education experience higher wages in future years. Increasing average education of a worker's colleagues this year by 10 percentage points increases his wages 0.3% in the following year. Compared to average wage growth in Sweden over this time period of

[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 are covered by collective bargaining in the U.S. in 2013 compared to 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 closer to the full impact of learning spillovers on productivity.

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.

An alternative explanation for these results must meet four conditions at the same time. It must be time-varying, correlated with changes in future wages, correlated with changes in current average education in workers' firms, and not captured by the industry×time and county×time, or county×industry×time fixed effects. In addition, it must be consistent with all of the robustness exercises in Section 5. While it is not impossible to come up with such a story, it has to be fairly stylized. Furthermore, in this section I document heterogeneity in the effects by age and occupation that are consistent with learning spillovers. I argue that learning spillovers are the most plausible explanation for both the main results and the heterogeneity in the effects.

6.1 Heterogeneous Effects over the Life Cycle

Consider learning spillovers by age. A reasonable prediction is that workers learn the most early in their careers, but 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 colleagues. 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 colleagues.

In Figure 4, 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 appears to be increasing at the earliest ages, and then decreases steadily until it is no longer statistically significantly different from zero, starting at the point where I estimate the effect over ages 38-48. In contrast, with idiosyncratic demand shocks I would expect the impact to be similar across ages.

[31] I produced Figure 4 by estimating the effect of learning spillovers on overlapping 10 year intervals, starting with ages 24-34, then 26-36, then 28-38, and so on. An alternative approach is to estimate the effects for non-overlapping 10 year age bins. I do so in Figure 8 in the appendix. The pattern is the same.

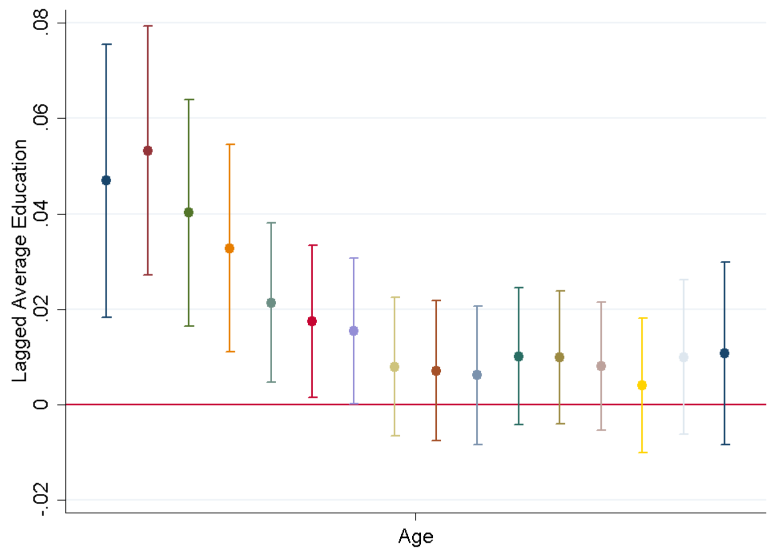


Figure 4: Age Profile of Learning Spillovers: Overlapping 10 Year Increments

6.2 Heterogeneous Effects by Occupation

The mechanism I have in mind for learning spillovers generates the natural prediction that 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 drivers, presumably offer fewer opportunities for interactions with colleagues 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 5 graphs the effect by occupation. All estimates are relative to the omitted occupational category, legislators and senior officials. Figure 5 shows that occupations that are likely more isolated, such as drivers, farmers, fisherman, machine operators, and elementary occupations (includes janitors, garbage collectors, deliverers, and street vendors), experience the smallest effects. In contrast, occupations that likely have more opportunities for learning spillovers, like 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 5 only includes data from 2000-2010. 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. In work in progress I am expanding these estimates to cover additional years and more detailed occupational categories.

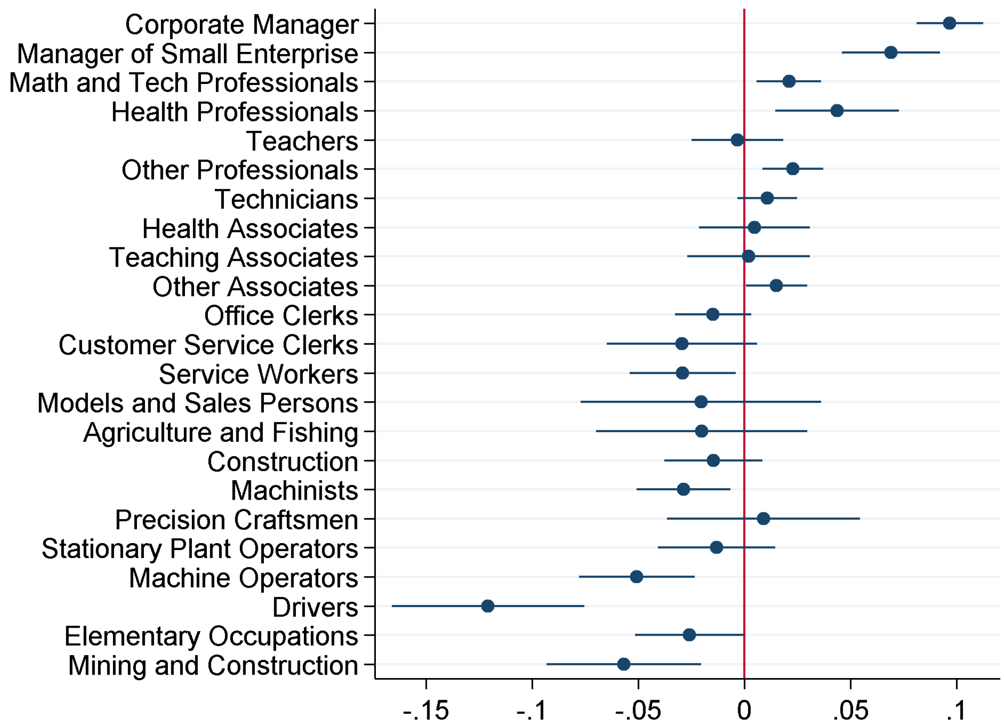


Figure 5: Occupational Profile of Learning Spillovers

To further explore the heterogeneity 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 the O*NET occupation categories and rank occupations according to their average O*NET rank of the importance of establishing and maintaining interpersonal relationships with colleagues.[34][35] Using this ranking, I compare the amount of learning spillovers by occupation from Figure 5 against the ranking of occupations by interaction.

[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 (2015). O*NET has been used in the Swedish context in Adermon and Gustavsson (2015), Black et al. (2015), and Johansson et al. (2015).

[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 15 in Appendix F.2 I describe how I construct a ranking using O*NET occupation categories, and then how I merge these categories into the SSYK categories.

[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".

In Figure 6 I present a scatterplot of the ranking of occupations using the O*NET measures and the estimates of learning spillovers from Figure 5. Occupations that experience higher learning spillovers also have higher average O*NET interaction ranks, suggesting that learning spillovers are a likely mechanism explaining the differences across occupations.

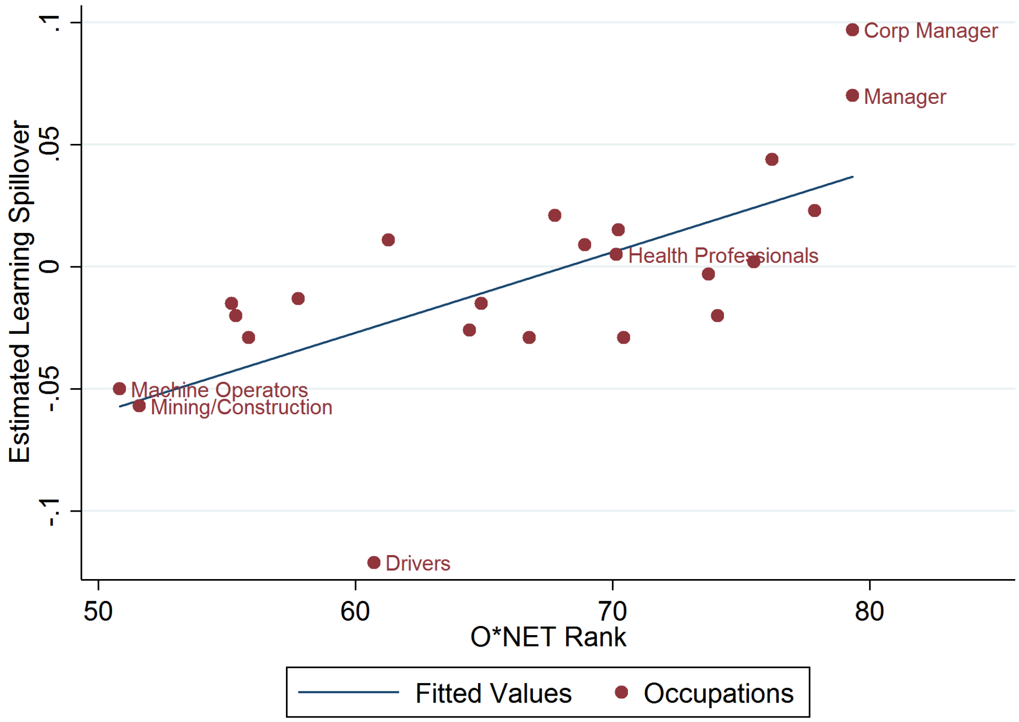


Figure 6: O*NET Rank of Occupation Potential for Learning Spillovers by Estimated Learning Spillovers

7 Discussion of Empirical Results and Broader Implications

To summarize the main results, I find that a 10 percentage point increase in average education of a worker's firm increases wages in the following year by 0.3%-0.6%. I also used the richness of the data to show heterogeneity by age and occupation consistent with learning spillovers.

I argue that the body of evidence in this paper makes a strong case that the effects 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 colleagues), but also fits the distinctive age pattern in Figure 4, the occupational patterns in Figure 6, is robust to all of the additional controls and alternative specifications, and provides a more compelling explanation than learning spillovers.[36]

Given the evidence strongly supports learning spillovers, I now discuss the implications for the social returns to education. I use the estimates in Table 2 and work by Altonji et al. (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 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 (as the theoretical results in this paper suggest may be the case), how much larger are the social returns to college relative to the private returns?

If learning spillovers are fully internalized, then the social returns to 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 to a college educated worker. From Table 2, I have that the private return to college is 0.194, 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 to adding a college worker into the part due to learning spillovers and the part due to the direct increase in productivity of the worker who obtains 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.[37]

In the equilibrium wage equations with full internalization of learning spillovers the return to college education includes not only the direct increase in productivity of the

[36] Furthermore, while randomized experiments provide the gold standard for identifying causal effects, it is particularly difficult to come up with a natural experiment or feasible randomized experiment that cleanly identifies learning spillovers. For example, suppose one could either randomize workers across existing firms or had random variation in average education at some local level. Both of these sources of random variation are insufficient to identify learning spillovers within the firm. Neither approach controls for unobserved firm heterogeneity, and the latter also fails to control for worker sorting.

[37] 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.

newly educated worker, but also the entire present discounted value of the learning spillovers the worker will provide for all of her colleagues. Under the assumptions used for this exercise, this implies the following equation

$$\underbrace{x_1}_{direct\ return} + \underbrace{0.028}_{spillover\ return} \times \underbrace{\left(\frac{H_f+1}{N_f} - \frac{H_f}{N_f}\right)}_{\triangle spillover} \times \underbrace{N_f}_{\#\, colleagues} = \underbrace{0.194}_{total\ return} \tag{30}$$

which means that $\frac{0.194-0.028}{0.194} \times 100 = 85.57\%$ of the total return is due to the direct increase in productivity of the newly college educated worker, while $\frac{0.028}{0.194} \times 100 = 14.43\%$ 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 to adding an additional college educated worker.

If learning spillovers are not internalized whatsoever, then Equation 30 becomes

$$\underbrace{0.194}_{direct\ return} + \underbrace{0.028}_{spillover\ return} \times \underbrace{\left(\frac{H_f+1}{N_f} - \frac{H_f}{N_f}\right)}_{\triangle spillover} \times \underbrace{N_f}_{\#\, colleagues} = \underbrace{x_2}_{total\ return} \tag{31}$$

Solving for x_2, the social return to an additional college educated worker is 0.222. Decomposing the social return I find that $\frac{0.194}{0.222} \times 100 = 87.39\%$ of the total return is due to the direct increase in productivity of the newly college educated worker, while $\frac{.028}{0.222} \times 100 = 12.61\%$ of the total return is due to learning spillovers.

From the theoretical results in 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 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, altogether 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	0.194	0.194
Social return to education	0.194	0.222
Amount by which social return exceeds private return	0	0.028
Percent due to own productivity	85.57%	87.39%
Percent due to learning spillovers	14.43%	12.61%

Notes: Calculations that produce the estimates are described in the text, and are based off of the estimates in Table 2.

8 Conclusion

There is a large literature on the possibility of human capital spillovers. Much has been written about human capital spillovers outside of firms. However, before this paper there was almost no existing work on the theoretical implications and empirical importance of human capital spillovers within the firm. In this paper, I address this gap in the existing literature. I provide one of the first theoretical and empirical assessments of learning spillovers in the firm. I start with a simple insight: if learning spillovers occur as a by-product of production and depend on average education within the firm, colleagues impose important externalities on each other. Applying existing results from the theoretical literature on externalities, I show that this fact makes it difficult for firms to 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 main focus of this paper, an empirical assessment of learning spillovers. Using wage equations predicted by the theory, I show that while the effect of average education of current colleagues on current wages is ambiguous, the effect of average education of past colleagues on current wages is unambiguous. For this reason I focus on estimating the effect of average education of a worker's colleagues 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 colleagues last 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 average education of 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 explored the broader implications of the main results. My findings suggest that the social returns of adding an additional college worker ranges from 0.194-0.222, with 12.61%-14.43% of the total increase attributable to learning spillovers.

Having established that learning spillovers in the firm are both theoretically and empirically important, there are a number of areas for future research. Starting with the theory, in the interest of simplicity I excluded the possibility of sorting driven by the learning spillovers. I excluded this possibility both through the assumption that firms are homogeneous and by my assumptions on aggregate production. 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. It would also likely generate some interesting testable predictions for the data.

On the empirical side, there is much that can be done building on the existing results. In work in progress I supplement the existing empirical strategy with exogenous variation in average education at the municipality level caused by policy changes in Sweden. I am also exploring the relative impacts of across firm spillovers versus within firm spillovers.

More generally, a great deal remains to be known empirically about learning spillovers in the firm. For example, do learning spillovers occur based on other traits of colleagues, such as experience? 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 learning spillovers for colleagues, 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.

References

Abowd, John M and Francis Kramarz, "The analysis of labor markets using matched employer-employee data," *Handbook of labor economics*, 1999, *3*, 2629–2710.

— , — , **and David N Margolis**, "High wage workers and high wage firms," *Econometrica*, 1999, *67* (2), 251–333.

Acemoglu, Daron, "A microfoundation for social increasing returns in human capital accumulation," *The Quarterly Journal of Economics*, 1996, *111* (3), 779–804.

— , "Training and innovation in an imperfect labour market," *The Review of Economic Studies*, 1997, *64* (3), 445–464.

— **and David Autor**, "Skills, tasks and technologies: Implications for employment and earnings," *Handbook of Labor Economics*, 2011, *4*, 1043–1171.

— **and Jorn-Steffen Pischke**, "Beyond Becker: training in imperfect labour markets," *The Economic Journal*, 1999, *109* (453), 112–142.

— **and Joshua Angrist**, "How large are human-capital externalities? Evidence from compulsory-schooling laws," in "NBER Macroeconomics Annual 2000, Volume 15," MIT Press, 2001, pp. 9–74.

Adermon, Adrian and Magnus Gustavsson, "Job Polarization and Task-Biased Technological Change: Evidence from Sweden, 1975-2005," *The Scandinavian Journal of Economics*, 2015, *117* (3), 878–917.

Altonji, Joseph G., Ching-I Huang, and Christopher R. Taber, "Estimating the Cream Skimming Effect of School Choice," *Journal of Political Economy*, 2015, *123* (2), pp. 266–324.

Ammermueller, Andreas and Jörn-Steffen Pischke, "Peer effects in European primary schools: Evidence from the progress in international reading literacy study," *Journal of Labor Economics*, 2009, *27* (3), 315–348.

Angrist, Joshua D, "The perils of peer effects," *Labour Economics*, 2014, *30*, 98–108.

— **and Kevin Lang**, "Does school integration generate peer effects? Evidence from Boston's Metco Program," *American Economic Review*, 2004, pp. 1613–1634.

Arrow, Kenneth, "The Organization of Market Activity: Issues Pertinent to the Choice of Market versus Non-market Allocations," *Haveman and Margolis, eds., Public Expenditure and Policy Analysis, Markham*, 1970.

Barro, Robert J, "Institutions and growth, an introductory essay," *Journal of Economic Growth*, 1996, *1* (2), 145–148.

Becker, Gary S, "Investment in Human Capital: A Theoretical Analysis," *The Journal of Political Economy*, 1962, pp. 9–49.

— , *Human capital: A theoretical and empirical analysis, with special reference to education*, University of Chicago Press, 2009.

Ben-Porath, Yoram, "The production of human capital and the life cycle of earnings," *The Journal of Political Economy*, 1967, *75* (4), 352–365.

Black, Sandra, Erik Grönqvist, and Björn Öckert, "Born to Lead? The Effect of Birth Order on Non-Cognitive Skills," *Working Paper*, 2015.

Card, David, Jörg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality*," *The Quarterly Journal of Economics*, 2013, *128* (3), 967–1015.

Coase, Ronald Harry, "The Problem of Social Cost," *The Journal of Law & Economics*, 1960, *3*, 1.

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg, "Peer Effects in the Workplace," 2013.

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

Friedrich, Benjamin, Lisa Laun, Costas Meghir, and Luigi Pistaferri, "Earnings Dynamics and Firm-level Shocks," *Working Paper*, 2015.

Griliches, Zvi, "Estimating the returns to schooling: Some econometric problems," *Econometrica: Journal of the Econometric Society*, 1977, pp. 1–22.

Heckman, James J, Lance Lochner, and Christopher Taber, "Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents," *Review of Economic Dynamics*, 1998, *1* (1), 1–58.

Jackson, C Kirabo and Elias Bruegmann, "Teaching students and teaching each other: The importance of peer learning for teachers," *National Bureau of Economic Research*, 2009.

Johansson, Per, Arizo Karimi, and J. Peter Nilsson, "Other-regarding Preferences in the Workplace: Evidence from two Randomized Experiments Affecting Workers' Incentives to Shirk," *Working Paper*, 2015.

Jr, Robert E Lucas, "On the mechanics of economic development," *Journal of Monetary Economics*, 1988, *22* (1), 3–42.

Kjellberg, Anders, "The decline in Swedish union density since 2007," *Nordic journal of working life studies*, 2011, *1* (1), pp–67.

Lange, Fabian and Robert Topel, "The social value of education and human capital," *Handbook of the Economics of Education*, 2006, *1*, 459–509.

Lindahl, Erik, "Just taxation - a positive solution," *Classics in the Theory of Public Finance*, 1919, *134*, 168–76.

Manski, Charles F, "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

Martins, Pedro S and Jim Y Jin, "Firm-level social returns to education," *Journal of Population Economics*, 2010, *23* (2), 539–558.

Mas, Alexandre and Enrico Moretti, "Peers at Work," *American Economic Review*, 2009, *99* (1), 112–45.

Milleron, Jean-Claude, "Theory of value with public goods: A survey article," *Journal of Economic Theory*, 1972, *5* (3), 419–477.

Moretti, Enrico, "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data," *Journal of Econometrics*, 2004, *121* (1), 175–212.

_ , "Workers education, spillovers, and productivity: Evidence from plant-level production functions," *The American Economic Review*, 2004, *94* (3), 656–690.

_ , "Social returns to human capital," *NBER Reporter Online*, 2005, (Spring 2005), 13–15.

Nelson, Richard R and Edmund S Phelps, "Investment in humans, technological diffusion, and economic growth," *The American Economic Review*, 1966, *56* (1/2), 69–75.

Pigou, Arthur Cecil, *Wealth and welfare*, Macmillan and Company, limited, 1912.

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

Sacerdote, Bruce, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 2001, *116* (2), 681–704.

Speer, Jamin D, "Pre-Market Skills, Occupational Choice, and Career Progression," *Forthcoming, Journal of Human Resources*, 2015.

Waldinger, Fabian, "Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany," *The Review of Economic Studies*, 2012, *79* (2), 838–861.

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.

$$\begin{aligned}\underset{M^A,M^B}{Max} \quad & -\int_0^{M^A} 1di - \int_0^{M^B} 1di \\ & JF\left(\frac{M^A+M^B}{J}, \frac{I-M^A-M^B}{J}\right) \\ & +\alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\ & +\alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\ & +\delta\alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\ & +\delta\alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2}\end{aligned} \tag{32}$$

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

$$M^A = F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) \tag{33}$$

$$M^B = F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) \tag{34}$$

What equation 33 and 34 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 colleagues.

A.2 Proof of Proposition 1: Equilibrium without a Market for the Spillover

In this section, I show that if firms ignore future learning spillovers the competitive equilibrium is efficient if education is exogenous and is inefficient when education is endogenous. The second result is expected - when externalities are ignored, we expect the competitive equilibrium to be inefficient.

The first result is perhaps less obvious. The reason the outcome is efficient when education is exogenous is driven by the assumptions on total production, stated in Section B.1. These assumptions imply that with a fixed education mix, introducing learning spillovers to the environment does not change the optimal input combination.

As a result, when education is exogenous, the Pareto efficient outcome is trivial - all firms receive the same combination of each type of worker, and all workers are employed. Who is compensated for the spillover simply shifts the equilibrium along the Pareto frontier. When the spillover is internalized, high educated workers are better off while when the spillover is not internalized, low educated workers are better off.

A.2.1 Consumer problem

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

$$\underset{c_1,c_2,c_3,h^i}{Max} \qquad c_1 + c_2 + c_3 \tag{35}$$

subject to

$$c_1 + c_2 + c_3 \quad \leq \quad -\theta^i h^i + h^i \left(w_f^{H^A} - w_f^{L^A} \right) + \delta s_f^A \tag{36}$$

B type consumers solve:

$$\underset{c_1,c_2,c_3,h^i}{Max} \qquad c_1 + c_2 + c_3 \tag{37}$$

subject to

$$c_1 + c_2 + c_3 \quad \leq \quad -\theta^i h^i + h^i \left(w_f^{H^B} - w_f^{L^B} \right) + \delta s_f^B \tag{38}$$

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.

As you can see, 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 ($h^i = 0$), and the choice over firms. Given this fact, in what follows and in the remainder of the proofs, I simply maximize the budget constraint in the consumer problem.

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.

$$\underset{h^i \in \{0,1\}}{Max} \qquad -\theta^i h^i + h^i \left(w_f^{H^A} - w_f^{L^A} \right) + \delta s_f^A \tag{39}$$

$$\underset{h^i \in \{0,1\}}{Max} \qquad -\theta^i h^i + h^i \left(w_f^{H^B} - w_f^{L^B} \right) + \delta s_f^B \tag{40}$$

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

$$\theta^i \leq w_f^{H^A} - w_f^{L^A} \tag{41}$$

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

$$\theta^i \leq w_f^{H^B} - w_f^{L^B} \tag{42}$$

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

$$M^A = w_f^{H^A} - w_f^{L^A} \tag{43}$$

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

$$M^B = w_f^{H^B} - w_f^{L^B} \tag{44}$$

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 their 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.

Thus, firms solve:

$$\begin{aligned}\underset{H_f^A, H_f^B, L_f^A, L_f^B}{Max} \quad & F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) \\ & + s_f^A\left(H_f^A + L_f^A\right) + s_f^B\left(H_f^B + L_f^B\right) \\ & - w_f^{H^A} H_f^A - w_f^{H^B} H_f^B - w_f^{L^A} L_f^A - w_f^{L^B} L_f^B\end{aligned} \tag{45}$$

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

$$\begin{aligned} w_f^{H^A} &= F_1 + \alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\ & + \left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \end{aligned} \tag{46}$$

$$\begin{aligned} w_f^{L^A} &= F_2 + \alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\ & - \left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \end{aligned} \tag{47}$$

$$\begin{aligned} w_f^{H^B} &= F_1 + \alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\ & \left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \end{aligned} \tag{48}$$

$$w_f^{L^B} = F_2 + \alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \tag{49}$$
$$- \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)$$

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
3. Markets Clear

$$\int_{i=0}^{I} c_1^i + \int_{i=0}^{I} c_2^i + \int_{i=0}^{I} c_3^i = -\int_0^{M^A} idi - \int_0^{M^B} idi \tag{50}$$
$$+JF\left(\frac{M^A + M^B}{J}, \frac{I - M^A - M^B}{J}\right) + \alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} + \alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2}$$
$$+\delta\alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} + \delta\alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2}$$
$$JH_f^A = M^A \tag{51}$$
$$JL_f^A = \frac{I}{2} - M^A \tag{52}$$
$$JH_f^B = M^B \tag{53}$$
$$JL_f^B = \frac{I}{2} - M^B \tag{54}$$

A.2.4 Equilibrium Solution

Consider the following equilibrium wages:

$$w_f^{H^K} = F_1 + \alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \tag{55}$$

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

$$w_f^{L^K} = F_2+\alpha^K\frac{H_f^A+H_f^B}{H_f^A+H_f^B+L_f^A+L_f^B} \tag{56}$$

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

$$K = A, B \tag{57}$$

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

$$\theta^i \leq F_1-F_2+\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\left(\frac{1}{H_f^A+H_f^B+L_f^A+L_f^B}\right) \tag{58}$$

$$\theta^i \leq F_1-F_2+\left(\alpha^A\left(H_f^A+L_f^A\right)+\alpha^B\left(H_f^B+L_f^B\right)\right)\left(\frac{1}{H_f^A+H_f^B+L_f^A+L_f^B}\right) \tag{59}$$

Imposing market clearing gives:

$$\theta^i \leq F_1-F_2+\left(\alpha^A\frac{I}{2J}+\alpha^B\frac{I}{2J}\right)\frac{J}{I} \tag{60}$$

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

$$\theta^i \leq F_1-F_2+\left(\alpha^A\frac{I}{2J}+\alpha^B\frac{I}{2J}\right)\frac{J}{I} \tag{61}$$

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

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

$$M^A = F_1-F_2+\frac{1}{2}\left(\alpha^A+\alpha^B\right) \tag{62}$$

$$M^B = F_1-F_2+\frac{1}{2}\left(\alpha^A+\alpha^B\right) \tag{63}$$

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

$$M^A = F_1-F_2+\frac{1}{2}\left(1+\delta\right)\left(\alpha^A+\alpha^B\right) \tag{64}$$

$$M^B = F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A + \alpha^B\right) \tag{65}$$

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: Equilibrium with Personalized Prices

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.

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.

$$\underset{h^i \in \{0,1\}}{Max} \quad -\theta^i h^i + h^i \left(w_f^{H^A} - w_f^{L^A}\right) + \delta s_f^A \tag{66}$$

$$\underset{h^i \in \{0,1\}}{Max} \quad -\theta^i h^i + h^i \left(w_f^{H^B} - w_f^{L^B}\right) + \delta s_f^B \tag{67}$$

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

$$\theta^i \leq w_f^{H^A} - w_f^{L^A} \tag{68}$$

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

$$\theta^i \leq w_f^{H^B} - w_f^{L^B} \tag{69}$$

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

$$M^A = w_f^{H^A} - w_f^{L^A} \tag{70}$$

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

$$M^B = w_f^{H^B} - w_f^{L^B} \tag{71}$$

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.

$$
\begin{aligned}
w_f^{H^A} + \delta s_f^A &\geq w^{H^A} && (72)\\
w_f^{H^B} + \delta s_f^B &\geq w^{H^B} && (73)\\
w_f^{L^A} + \delta s_f^A &\geq w^{L^A} && (74)\\
w_f^{L^B} + \delta s_f^B &\geq w^{L^B} && (75)
\end{aligned}
$$

A.3.2 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 their 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:

$$
\begin{aligned}
\underset{H_f^A, H_f^B, L_f^A, L_f^B, w_f^{H^A}, w_f^{L^A}, w_f^{H^B}, w_f^{L^B}}{Max} \quad & F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) && (76)\\
& + s_f^A\left(H_f^A + L_f^A\right) + s_f^B\left(H_f^B + L_f^B\right) \\
& - w_f^{H^A} H_f^A - w_f^{H^B} H_f^B \\
& - w_f^{L^A} L_f^A - w_f^{L^B} L_f^B
\end{aligned}
$$

subject to the workers' participation constraints:

$$
w_f^{H^A} + \delta s_f^A \geq w^{H^A} \qquad (77)
$$

$$w_f^{H^B} + \delta s_f^B \geq w^{H^B} \tag{78}$$

$$w_f^{L^A} + \delta s_f^A \geq w^{L^A} \tag{79}$$

$$w_f^{L^B} + \delta s_f^B \geq w^{L^B} \tag{80}$$

$$s_f^A = \alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \tag{81}$$

$$s_f^B = \alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \tag{82}$$

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

$$\begin{aligned}\underset{H_f^A, H_f^B, L_f^A, L_f^B}{Max} \quad & F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) - w^{H^A} H_f^A - w^{H^B} H_f^B - w^{L^A} L_f^A - w^{L^B} L_f^B \\ & + \alpha^A (1+\delta) \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \left(H_f^A + L_f^A\right) \\ & + \alpha^B (1+\delta) \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \left(H_f^B + L_f^B\right)\end{aligned} \tag{83}$$

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

$$\begin{aligned} w^{H^A} = \; & F_1 + (1+\delta)\, \alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\ & + (1+\delta) \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)\end{aligned} \tag{84}$$

$$\begin{aligned} w^{L^A} = \; & F_2 + (1+\delta)\, \alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\ & - (1+\delta) \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)\end{aligned} \tag{85}$$

$$\begin{aligned} w^{H^B} = \; & F_1 + (1+\delta)\, \alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\ & (1+\delta) \left(\alpha^A \left(H_f^A + L_f^A\right) + \alpha^B \left(H_f^B + L_f^B\right)\right) \left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)\end{aligned} \tag{86}$$

$$
\begin{aligned}
w^{L^B} &= F_2 + (1+\delta)\,\alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \\
&\quad - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)
\end{aligned} \tag{87}
$$

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
3. Markets Clear

$$
\begin{aligned}
\int_{i=0}^{I} c_1^i + \int_{i=0}^{I} c_2^i + \int_{i=0}^{I} c_3^i &= -\int_0^{M^A} idi - \int_0^{M^B} idi \\
&\quad + JF\left(\frac{M^A + M^B}{J}, \frac{I - M^A - M^B}{J}\right) + \alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} + \alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} \\
&\quad + \delta\alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2} + \delta\alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}} \frac{I}{2}
\end{aligned} \tag{88}
$$

$$ JH_f^A = M^A \tag{89} $$

$$ JL_f^A = \frac{I}{2} - M^A \tag{90} $$

$$ JH_f^B = M^B \tag{91} $$

$$ JL_f^B = \frac{I}{2} - M^B \tag{92} $$

A.3.4 Equilibrium Solution

Consider the following equilibrium compensation amounts:

$$
\begin{aligned}
w^{H^K} &= F_1 + (1+\delta)\,\alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} && (93)\\
&\quad + (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)\\
w^{L^K} &= F_2 + (1+\delta)\,\alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} && (94)\\
&\quad - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)\\
K &= A, B && (95)
\end{aligned}
$$

The associated equilibrium wages are:

$$
\begin{aligned}
w_f^{H^K} &= F_1 + \alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} && (96)\\
&\quad + (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)\\
w_f^{L^K} &= F_2 + \alpha^K \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} && (97)\\
&\quad - (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)\\
K &= A, B && (98)
\end{aligned}
$$

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

$$
\begin{aligned}
\theta^i &\leq F_1 - F_2 && (99)\\
&\quad + (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B}\right)\\
\theta^i &\leq F_1 - F_2 && (100)\\
&\quad + (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B}\right)
\end{aligned}
$$

Imposing market clearing gives:

$$\begin{aligned}
\theta^{i} &\leq F_1 - F_2 + (1+\delta)\left(\alpha^{A}\frac{I}{2J} + \alpha^{B}\frac{I}{2J}\right)\frac{J}{I} && (101)\\
&= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^{A} + \alpha^{B}\right)\\
\theta^{i} &\leq F_1 - F_2 + (1+\delta)\left(\alpha^{A}\frac{I}{2J} + \alpha^{B}\frac{I}{2J}\right)\frac{J}{I} && (102)\\
&= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^{A} + \alpha^{B}\right)
\end{aligned}$$

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

$$\begin{aligned}
M^{A} &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^{A} + \alpha^{B}\right) && (103)\\
M^{B} &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^{A} + \alpha^{B}\right) && (104)
\end{aligned}$$

and this condition is identical to the Pareto efficient condition for education investments.

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.

$$\begin{aligned}
\pi &= F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) - w^{H^A}H_f^A - w^{H^B}H_f^B - w^{L^A}L_f^A - w^{L^B}L_f^B && (105)\\
&\quad + (1+\delta)\,\alpha^{A}\frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B}\left(H_f^A + L_f^A\right)\\
&\quad + (1+\delta)\,\alpha^{B}\frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B}\left(H_f^B + L_f^B\right)\\
&= F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) - F_1 H_f^A - F_1 H_f^B - F_2 L_f^A - F_2 L_f^B\\
&\quad - (1+\delta)\left(\alpha^{A}\left(H_f^A + L_f^A\right) + \alpha^{B}\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)H_f^A\\
&\quad - (1+\delta)\left(\alpha^{A}\left(H_f^A + L_f^A\right) + \alpha^{B}\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)H_f^B\\
&\quad + (1+\delta)\left(\alpha^{A}\left(H_f^A + L_f^A\right) + \alpha^{B}\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right)L_f^A
\end{aligned}$$

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

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.

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 unfair to 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 colleagues 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 they receive

$$
F_2^S + (1+\delta)\left(\alpha^B - \alpha^A\right)\left(H_f^A + L_f^A\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \tag{106}
$$

Since $\alpha^B - \alpha^A < 0$, F_2^S must be sufficiently higher than F_2^{NS} in order for low ability 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 colleagues in the production of learning spillovers.

A.4 Proof of Proposition 3: Equilibrium with Asymmetric Information

In practice, individual's types are known only to them. This makes the efficient outcome in Proposition 2 impossible to implement. In this section, I instead solve for a competitive equilibrium where worker's types are unobserved by firms.

A.4.1 Consumer problem

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

$$\underset{h^i \in \{0,1\}}{Max} \qquad -\theta^i h^i + h^i \left(w_f^{H^A} - w_f^{L^A} \right) + \delta s_f^A \tag{107}$$

$$\underset{h^i \in \{0,1\}}{Max} \qquad -\theta^i h^i + h^i \left(w_f^{H^B} - w_f^{L^B} \right) + \delta s_f^B \tag{108}$$

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

$$\theta^i \leq w_f^{H^A} - w_f^{L^A} \tag{109}$$

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

$$\theta^i \leq w_f^{H^B} - w_f^{L^B} \tag{110}$$

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

$$M^A = w_f^{H^A} - w_f^{L^A} \tag{111}$$

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

$$M^B = w_f^{H^B} - w_f^{L^B} \tag{112}$$

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.

$$w_f^{H^A} + \delta s_f^A \geq w^{H^A} \tag{113}$$
$$w_f^{H^B} + \delta s_f^B \geq w^{H^B} \tag{114}$$
$$w_f^{L^A} + \delta s_f^A \geq w^{L^A} \tag{115}$$
$$w_f^{L^B} + \delta s_f^B \geq w^{L^B} \tag{116}$$

A.4.2 Firm Problem

Workers provide their labor inelastically, subject to their second period participation constraints. Firms then maximize over production of consumption goods subject to these participation constraints. However, since firms no longer observe types, they must also meet incentive compatibility constraints. Thus, firms solve:

$$\underset{H_f^A, H_f^B, L_f^A, L_f^B, w_f^{H^A}, w_f^{L^A}, w_f^{H^B}, w_f^{L^B}}{Max} \quad F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) - w_f^{H^A} H_f^A - w_f^{H^B} H_f^B - w_f^{L^A} L_f^A - w_f^{L^B} L_f^B \tag{117}$$

subject to the worker's participation constraints:

$$w_f^{H^A} + \delta \alpha^A s_f \geq w^{H^A} \tag{118}$$
$$w_f^{H^B} + \delta \alpha^B s_f \geq w^{H^B} \tag{119}$$
$$w_f^{L^A} + \delta \alpha^A s_f \geq w^{L^A} \tag{120}$$
$$w_f^{L^B} + \delta \alpha^B s_f \geq w^{L^B} \tag{121}$$
$$s_f = \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} \tag{122}$$

and incentive compatibility constraints

$$w_f^{H^B} + \delta \alpha^B s_f \geq w_f^{H^A} + \delta \alpha^B s_f \tag{123}$$
$$w_f^{H^A} + \delta \alpha^A s_f \geq w_f^{H^B} + \delta \alpha^A s_f \tag{124}$$
$$w_f^{L^A} + \delta \alpha^A s_f \geq w_f^{L^B} + \delta \alpha^A s_f \tag{125}$$

$$w_f^{L^B} + \delta\alpha^B s_f \quad \geq \quad w_f^{L^A} + \delta\alpha^B s_f \tag{126}$$

The incentive compatibility constraints imply that

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

and firms cannot induce workers to reveal their types 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, irregardless of their type. This is due to the "public" 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.

$$\begin{aligned} \underset{H_f^A, H_f^B, L_f^A, L_f^B, w_f^H, w_f^L}{Max} \quad & F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) + E\left[\alpha\right]\left(H_f^A + H_f^B\right) \\ & - w_f^H\left(H_f^A + H_f^B\right) - w_f^L\left(L_f^A + L_f^B\right) \end{aligned} \tag{127}$$

subject to the worker's participation constraints:

$$\begin{aligned} w_f^H &\geq w^{H^A} - \delta\alpha^A s_f && (128) \\ w_f^H &\geq w^{H^B} - \delta\alpha^B s_f && (129) \\ w_f^L &\geq w^{L^A} - \delta\alpha^A s_f && (130) \\ w_f^L &\geq w^{L^B} - \delta\alpha^B s_f && (131) \\ s_f &= \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} && (132) \end{aligned}$$

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

$$\mathcal{L}\left(H_f^A, H_f^B, L_f^A, L_f^B, w_f^H, w_f^L\right) = F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) + E\left[\alpha\right]\left(H_f^A + H_f^B\right) \tag{133}$$

[38] I could have done the same in the previous setting with full information, and would have obtained the same solution as I get from plugging in directly.

$$
\begin{aligned}
& -w_f^H\left(H_f^A + H_f^B\right) - w_f^L\left(L_f^A + L_f^B\right) \\
& +\lambda_1\left(w_f^H - w^{H^A} - \delta\alpha^A s_f\right) && (134) \\
& +\lambda_2\left(w_f^H - w^{H^B} - \delta\alpha^B s_f\right) && (135) \\
& +\lambda_3\left(w_f^L - w^{L^A} - \delta\alpha^A s_f\right) && (136) \\
& +\lambda_4\left(w_f^L - w^{L^B} - \delta\alpha^B s_f\right) && (137)
\end{aligned}
$$

and the corresponding Kuhn Tucker Conditions are:

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

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

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

3. Markets Clear

4.

$$
\int_{i=0}^{I} c_1^i+\int_{i=0}^{I} c_2^i+\int_{i=0}^{I} c_3^i = -\int_0^{M^A} idi-\int_0^{M^B} idi \tag{138}
$$

$$
\begin{aligned}
& +JF\left(\frac{M^A+M^B}{J},\frac{I-M^A-M^B}{J}\right)+\alpha^A\frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2}+\alpha^B\frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2} \\
& +\delta\alpha^A\frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2}+\delta\alpha^B\frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2}
\end{aligned}
$$

$$
JH_f^A = M^A \tag{139}
$$

$$
JL_f^A = \frac{I}{2}-M^A \tag{140}
$$

$$
JH_f^B = M^B \tag{141}
$$

$$
JL_f^B = \frac{I}{2}-M^B \tag{142}
$$

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

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

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 $w_f^L>0$ (I discuss the alternative later), 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

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

And profits are:

$$
\begin{aligned}
\pi &= F\left(H_f, L_f\right) + E[\alpha]H_f \\
&\quad - F_1 H_f - E[\alpha]H_f \\
&\quad - \delta\alpha^B\frac{L_f H_f - H_f L_f}{H_f + L_f} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L_f H_f - H_f L_f}{\left(H_f + L_f\right)^2} \\
&= 0
\end{aligned}
$$

Since profits do not depend on the choices of λ's, 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:

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

Suppose instead that $w_f^L = 0$. Then it must be that

$$
\begin{aligned}
\lambda_3\left(\delta\alpha^A\frac{H_f}{H_f + L_f} - w^{L^A}\right) &= 0 \\
0 &\geq w^{L^A} - \delta\alpha^A\frac{H_f}{H_f + L_f} \\
\lambda_4\left(\delta\alpha^B\frac{H_f}{H_f + L_f} - w^{L^B}\right) &= 0
\end{aligned}
$$

$$
\begin{aligned}
0 &\geq w^{L^B} - \delta\alpha^B \frac{H_f}{H_f + L_f} \\
F_2 &= \delta\left((\lambda_1 + \lambda_3)\,\alpha^A + (\lambda_2 + \lambda_4)\,\alpha^B \right) \frac{H_f}{\left(H_f + L_f\right)^2}
\end{aligned}
$$

Which implies that

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

Profits are

$$
\begin{aligned}
\pi &= F\left(H_f, L_f\right) + E[\alpha]\,H_f \\
&\quad - F_1 H_f - E[\alpha]\,H_f - \delta\left((\lambda_1 + \lambda_3)\,\alpha^A + (\lambda_2 + \lambda_4)\,\alpha^B \right) \frac{L_f H_f}{\left(H_f + L_f\right)^2} \\
&= F\left(H_f, L_f\right) + E[\alpha]\,H_f \\
&\quad - F_1 H_f - E[\alpha]\,H_f - F_2 L_f \\
&= 0
\end{aligned}
$$

And again, since profits do not depend on the choices of λ's, any of the following solutions for the λ's are all equally good:

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

Thus, there are two sets of prices that are competitive equilibria:

The first set of equilibria is given by:

$$w_f^H = F_1 + E[\alpha] + \delta\alpha^B \frac{L^*}{H^* + L^*} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L^*}{(H^* + L^*)^2} \tag{143}$$

$$w_f^L = F_2 - \delta\alpha^B \frac{H^*}{H^* + L^*} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H^*}{(H^* + L^*)^2} \tag{144}$$

$$w^{H^A} = F_1 + E[\alpha] + \delta\alpha^B \frac{L^*}{H^* + L^*} + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L^*}{(H^* + L^*)^2} + \delta\alpha^A \frac{H^*}{H^* + L^*} \tag{145}$$

$$w^{H^B} = F_1 + E[\alpha] + \delta\alpha^B + \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{L^*}{(H^* + L^*)^2} \tag{146}$$

$$w^{L^A} = F_2 + \delta\left(\alpha^A - \alpha^B\right)\frac{H^*}{H^* + L^*} - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H^*}{(H^* + L^*)^2} \tag{147}$$

$$w^{L^B} = F_2 - \delta\left(\alpha^A - \alpha^B\right)(\lambda_1 + \lambda_3)\frac{H^*}{(H^* + L^*)^2} \tag{148}$$

with

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

Thus, any of the following wages are a competitive equilibrium:

$$\begin{aligned}
w_f^H &= F_1 + E[\alpha] + k\frac{L^*}{H^* + L^*} \\
w_f^L &= F_2 - k\frac{H^*}{H^* + L^*} \\
k &\in \left[\delta\alpha^B, \delta\alpha^A\right]
\end{aligned}$$

The second set of equilibria is given by:

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

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

with

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

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:

$$
\begin{aligned}
w^{H^A} - w^{L^A} &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A+\alpha^B\right) && (149) \\
w^{H^B} - w^{L^B} &= F_1 - F_2 + \frac{1}{2}(1+\delta)\left(\alpha^A+\alpha^B\right) && (150)
\end{aligned}
$$

For the first set of solutions, I have that

$$
\begin{aligned}
w^{H^A} - w^{L^A} &= F_1 - F_2 + \frac{1}{2}\left(\alpha^A+\alpha^B\right) && (151) \\
&\quad +\delta\alpha^B\frac{I-H^*}{I} + \delta\left(\alpha^A-\alpha^B\right)\left(\lambda_1+\lambda_3\right)\frac{I-H^*}{I^2} + \delta\alpha^A\frac{H^*}{I} \\
&\quad -\delta\left(\alpha^A-\alpha^B\right)\frac{H^*}{I} + \delta\left(\alpha^A-\alpha^B\right)\left(\lambda_1+\lambda_3\right)\frac{H^*}{I^2} \\
&= F_1 - F_2 + \frac{1}{2}\left(\alpha^A+\alpha^B\right) \\
&\quad +\delta\alpha^B + \delta\left(\alpha^A-\alpha^B\right)\left(\lambda_1+\lambda_3\right)\frac{1}{I} \\
w^{H^B} - w^{L^B} &= F_1 - F_2 + \frac{1}{2}\left(\alpha^A+\alpha^B\right) && (152) \\
&\quad +\delta\alpha^B + \delta\left(\alpha^A-\alpha^B\right)\left(\lambda_1+\lambda_3\right)\frac{1}{I}
\end{aligned}
$$

Which means that Pareto efficiency requires:

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

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

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

Similarly for the second set of solutions. For the second set of solutions, I have that:

$$\begin{aligned}
w^{H^A} - w^{L^A} &= F_1 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) + \delta\left(\left(\lambda_1 + \lambda_3\right)\alpha^A + \left(\lambda_2 + \lambda_4\right)\alpha^B\right)\frac{L^*}{\left(H^* + L^*\right)^2} \qquad (153) \\
&= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) + \delta\left(\left(\lambda_1 + \lambda_3\right)\alpha^A + \left(\lambda_2 + \lambda_4\right)\alpha^B\right)\frac{1}{H^* + L^*} \\
w^{H^B} - w^{L^B} &= F_1 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) + \delta\left(\left(\lambda_1 + \lambda_3\right)\alpha^A + \left(\lambda_2 + \lambda_4\right)\alpha^B\right)\frac{L^*}{\left(H^* + L^*\right)^2} \qquad (154) \\
&= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) + \delta\left(\left(\lambda_1 + \lambda_3\right)\alpha^A + \left(\lambda_2 + \lambda_4\right)\alpha^B\right)\frac{1}{H^* + L^*}
\end{aligned}$$

Which again, is only efficient for one of the infinite possible equilibrium wages and total compensation packages:

$$\begin{aligned}
w_f^L &= 0 \\
w_f^H &= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) + \delta\frac{1}{2}\left(\alpha^A + \alpha^B\right) \\
w^{H^A} &= F_1 - F_2 + \frac{1}{2}\left(\alpha^A + \alpha^B\right) + \delta\frac{1}{2}\left(\alpha^A + \alpha^B\right) + \delta\alpha^A \frac{H^*}{H^* + L^*}
\end{aligned}$$

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

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:

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

Given 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.

B Theory Appendix

B.1 Conditions that Prevent Sorting

I assume that:

$$\begin{aligned}
F_1 + (1+\delta)\,\alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} & \qquad (155) \\
+ (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) &> 0 \\
F_1 + (1+\delta)\,\alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} & \\
+ (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) &> 0 \\
F_2 + (1+\delta)\,\alpha^A \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} & \\
- (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) &> 0 \\
F_2 + (1+\delta)\,\alpha^B \frac{H_f^A + H_f^B}{H_f^A + H_f^B + L_f^A + L_f^B} & \\
- (1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right)\left(\frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) &> 0
\end{aligned}$$

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:

$$F_{11} + (1+\delta)\,\alpha^A\left(\frac{1}{H_f^A + H_f^B + L_f^A + L_f^B} - \frac{H_f^A + H_f^B}{\left(H_f^A + H_f^B + L_f^A + L_f^B\right)^2}\right) \qquad (156)$$

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

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

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

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

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

$$F_{22}-\left(1+\delta\right)\alpha^{A}\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}-\left(1+\delta\right)\alpha^{A}\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}$$

$$+\left(1+\delta\right)2\left(\alpha^{A}\left(H_f^A+L_f^A\right)+\alpha^{B}\left(H_f^B+L_f^B\right)\right)\left(\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^3}\right) < 0$$

$$F_{22}-\left(1+\delta\right)\alpha^{B}\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}-\left(1+\delta\right)\alpha^{B}\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^2}$$

$$+\left(1+\delta\right)2\left(\alpha^{A}\left(H_f^A+L_f^A\right)+\alpha^{B}\left(H_f^B+L_f^B\right)\right)\left(\frac{H_f^A+H_f^B}{\left(H_f^A+H_f^B+L_f^A+L_f^B\right)^3}\right) < 0$$

These assumptions combined with the previous assumptions (equations 155) imply that total production is increasing in each input but at a decreasing rate, which mean 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

Note that I chose this particular functional form for the spillover 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 0 returns to scale, the total amount of consumption goods produced by learning spillover in the firm is increasing returns to scale (assuming F is not decreasing returns to scale), since

$$(1+\delta)\left(\alpha^A\left(\lambda H_f^A + \lambda L_f^A\right) + \alpha^B\left(\lambda H_f^B + \lambda L_f^B\right)\right) G\left(\lambda H_f, \lambda L_f\right) = \tag{157}$$

$$\lambda(1+\delta)\left(\alpha^A\left(H_f^A + L_f^A\right) + \alpha^B\left(H_f^B + L_f^B\right)\right) G\left(\lambda H_f, \lambda L_f\right) \tag{158}$$

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.

Thus, if learning spillovers are not decreasing or zero returns to scale, the outcome is likely inefficient. However, in this paper I focus on a more general, and I believe a 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.[39]

The second motivation for this particular specification is empirical. This paper was originally inspired by the literature on education externalities across firms.[40] In order to make the empirical results in the second half of this paper more comparable to that literature, I chose the same specification used in that literature.

B.3 Equilibrium with Traditional Training Inputs

In this section, I show that if firms and workers are choosing traditional, rival training inputs that produce general skills, they do not face the same challenges. 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.

[39] In work in progress, I set up a model where learning spillovers are constant returns to scale, consumers have preferences for variety, and there is monopolistic competition.

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

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 worker's learning parameters, so that:

$$s^A = \alpha^A \tau^A \tag{159}$$
$$s^B = \alpha^B \tau^B \tag{160}$$

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.

B.3.1 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, and the optimal number of traditional training inputs, τ^A and τ^B.

$$\begin{aligned} \underset{M^A,M^B}{Max} \quad & -\int_0^{M^A} 1di - \int_0^{M^B} 1di \\ & JF\left(\frac{M^A+M^B}{J}, \frac{I-M^A-M^B}{J}\right) \\ & +J\left(\left(1+\delta\right)\alpha^A\tau^A - \nu\left(\tau^A\right)\right)\frac{I}{2J} \\ & +J\left(\left(1+\delta\right)\alpha^B\tau^B - \nu\left(\tau^B\right)\right)\frac{I}{2J} \end{aligned} \tag{161}$$

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

$$M^A = F_1 - F_2 \tag{162}$$
$$M^B = F_1 - F_2 \tag{163}$$
$$\left(1+\delta\right)\alpha^A = \nu'\left(\tau^A\right) \tag{164}$$
$$\left(1+\delta\right)\alpha^B = \nu'\left(\tau^B\right) \tag{165}$$

Competitive Equilibrium

B.3.2 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.

$$\underset{h^i \in \{0,1\}}{Max} \qquad -\theta^i h^i + h^i \left(w_f^{H^A} - w_f^{L^A} \right) \tag{166}$$

$$\underset{h^i \in \{0,1\}}{Max} \qquad -\theta^i h^i + h^i \left(w_f^{H^B} - w_f^{L^B} \right) \tag{167}$$

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

$$\theta^i \leq w_f^{H^A} - w_f^{L^A} \tag{168}$$

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

$$\theta^i \leq w_f^{H^B} - w_f^{L^B} \tag{169}$$

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

$$M^A = w_f^{H^A} - w_f^{L^A} \tag{170}$$

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

$$M^B = w_f^{H^B} - w_f^{L^B} \tag{171}$$

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.

$$w_f^{H^A} + \delta \alpha^A \tau^A \geq w^{H^A} \quad (172)$$
$$w_f^{H^B} + \delta \alpha^B \tau^B \geq w^{H^B} \quad (173)$$
$$w_f^{L^A} + \delta \alpha^A \tau^A \geq w^{L^A} \quad (174)$$
$$w_f^{L^B} + \delta \alpha^B \tau^B \geq w^{L^B} \quad (175)$$

B.3.3 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 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:

$$\underset{H_f^A, H_f^B, L_f^A, L_f^B, \tau^A, \tau^B}{Max} \quad F\left(H_f^A + H_f^B, L_f^A + L_f^B\right) - w^{H^A} H_f^A - w^{H^B} H_f^B - w^{L^A} L_f^A - w^{L^B} L_f^B \quad (176)$$
$$+ \left((1+\delta)\,\alpha^A \tau^A - \nu\left(\tau^A\right)\right)\left(H_f^A + L_f^A\right)$$
$$+ \left((1+\delta)\,\alpha^B \tau^B - \nu\left(\tau^B\right)\right)\left(H_f^B + L_f^B\right)$$

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

$$w^{H^A} = F_1 + (1+\delta)\,\alpha^A \tau^A - \nu\left(\tau^A\right) \quad (177)$$
$$w^{L^A} = F_2 + (1+\delta)\,\alpha^A \tau^A - \nu\left(\tau^A\right) \quad (178)$$
$$w^{H^B} = F_1 + (1+\delta)\,\alpha^B \tau^B - \nu\left(\tau^B\right) \quad (179)$$
$$w^{L^B} = F_2 + (1+\delta)\,\alpha^B \tau^B - \nu\left(\tau^B\right) \quad (180)$$
$$(1+\delta)\,\alpha^A = \nu'\left(\tau^A\right) \quad (181)$$
$$(1+\delta)\,\alpha^B = \nu'\left(\tau^B\right) \quad (182)$$

B.3.4 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 learning spillovers.
3. Markets Clear

$$\begin{aligned}
\int_{i=0}^{I} c_1^i + \int_{i=0}^{I} c_2^i + \int_{i=0}^{I} c_3^i &= -\int_0^{M^A} idi - \int_0^{M^B} idi && (183)\\
&+JF\left(\frac{M^A+M^B}{J}, \frac{I-M^A-M^B}{J}\right) + \alpha^A \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2} + \alpha^B \frac{\frac{M^A+M^B}{J}}{\frac{I}{J}}\frac{I}{2}\\
&+\left(\left(1+\delta\right)\alpha^B\tau^B - \nu\left(\tau^B\right)\right)\frac{I}{2} +\\
&\delta\left(\left(1+\delta\right)\alpha^A\tau^A - \nu\left(\tau^A\right)\right)\frac{I}{2}\\
JH_f^A &= M^A && (184)\\
JL_f^A &= \frac{I}{2} - M^A && (185)\\
JH_f^B &= M^B && (186)\\
JL_f^B &= \frac{I}{2} - M^B && (187)
\end{aligned}$$

B.3.5 Equilibrium Solution

Consider the following equilibrium compensation amounts:

$$\begin{aligned}
w^{H^A} &= F_1 + \left(1+\delta\right)\alpha^A\tau^A - \nu\left(\tau^A\right) && (188)\\
w^{L^A} &= F_2 + \left(1+\delta\right)\alpha^A\tau^A - \nu\left(\tau^A\right) && (189)\\
w^{H^B} &= F_1 + \left(1+\delta\right)\alpha^B\tau^B - \nu\left(\tau^B\right) && (190)\\
w^{L^B} &= F_2 + \left(1+\delta\right)\alpha^B\tau^B - \nu\left(\tau^B\right) && (191)\\
\left(1+\delta\right)\alpha^A &= \nu'\left(\tau^A\right) && (192)\\
\left(1+\delta\right)\alpha^B &= \nu'\left(\tau^B\right) && (193)
\end{aligned}$$

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

$$
\begin{aligned}
\theta^{i} &\leq F_1 - F_2 && (194)\\
\theta^{i} &\leq F_1 - F_2 && (195)
\end{aligned}
$$

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

$$
\begin{aligned}
M^{A} &= F_1 - F_2 && (196)\\
M^{B} &= F_1 - F_2 && (197)
\end{aligned}
$$

and the solution for the traditional training inputs is:

$$
\begin{aligned}
(1+\delta)\,\alpha^{A} &= \nu'\left(\tau^{A}\right) && (198)\\
(1+\delta)\,\alpha^{B} &= \nu'\left(\tau^{B}\right) && (199)
\end{aligned}
$$

This is identical to the Pareto efficient solution for education and traditional training inputs:

$$
\begin{aligned}
M^{A} &= F_1 - F_2 && (200)\\
M^{B} &= F_1 - F_2 && (201)\\
(1+\delta)\,\alpha^{A} &= \nu'\left(\tau^{A}\right) && (202)\\
(1+\delta)\,\alpha^{B} &= \nu'\left(\tau^{B}\right) && (203)
\end{aligned}
$$

and I conclude that the competitive equilibrium is efficient.

B.3.6 Why the Solution with Traditional Inputs is Sustainable in a Competitive Equilibrium

Recall the challenges to sustaining the equilibrium with learning spillovers, in particular, thin markets and asymmetric information. First, thin markets is no longer an issue since all individuals of the same education type face the same wages.

Second, asymmetric information is no longer an issue since individuals will not choose to lie about their types. This is due to the fact that instead of effectively charging individuals different prices for the same quantity of exposure, here firms are effectively 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, 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.[41] I am trying to get an unbiased estimate of π_1 in:

$$w_{it} = \pi_0 h_i + \pi_1 \bar{H}_{ft-1} \tag{204}$$

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. Acemoglu and Angrist (2001) show that a simple derivation yields the following solution for the coefficients:

$$\begin{aligned} \pi_0 &= \frac{\psi_0 - \psi_1 R^2}{1 - R^2} \\ \pi_1 &= \frac{\psi_1 - \psi_0}{1 - R^2} \end{aligned} \tag{205}$$

Where R^2 is the first-stage R squared from 2SLS using average education in the firm×year dummies as instruments for own education, ψ_0 is the OLS coefficient of education in equation 204, 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.

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

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" (page 9).

Fixed Effects as a Solution

I described why fixed effects addresses this issue in the main text. Here, I formally derive the result and show that it works using a simple simulation exercise. To formally show this result, I re-derive equation 205 with fixed effects for worker×workplace spells.[42]

Rewrite equation 204 as follows:

$$w_{it} = \pi_0 \tau_i + (\pi_0 + \pi_1)\, \bar{H}_{ift-1} + \xi_i \tag{206}$$

where $\tau_i = h_i - \bar{H}_{ift-1}$. Now add fixed effects for worker×workplace spells to equation 206.

$$w_{it} - \bar{w}_{it} = \pi_0 \left(\tau_i - \bar{\tau}_i\right) + (\pi_0 + \pi_1) \left(\bar{H}_{ift-1} - \bar{H}_{if}\right) + \xi_i \tag{207}$$

where

$$\begin{aligned} \left(\tau_i - \bar{\tau}_i\right) &= \left(h_i - \bar{H}_{ift-1}\right) - \left(\bar{h}_i - \bar{H}_{if}\right) \\ &= \bar{H}_{if} - \bar{H}_{ift-1} \end{aligned} \tag{208}$$

And equation 207 becomes

$$w_{it} - \bar{w}_{it} = \pi_0 \left(\bar{H}_{if} - \bar{H}_{ift-1}\right) + (\pi_0 + \pi_1) \left(\bar{H}_{ift-1} - \bar{H}_{if}\right) + \xi_i \tag{209}$$

$$= \pi_1 \left(\bar{H}_{ift-1} - \bar{H}_{if}\right) \tag{210}$$

[42] The derivation is equivalent with just individual fixed effects.

Then,

$$\pi_1 = \frac{C\left(\left(\bar{H}_{ift-1} - \bar{H}_{if}\right), \left(w_{it} - \bar{w}_{it}\right)\right)}{V\left(\left(\bar{H}_{ift-1} - \bar{H}_{if}\right)\right)} \tag{211}$$

Which is precisely the result we want. 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). Columns 1-3 and 5-7 are identical to that 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.[43] In the third column, I estimate the effect of both average education and 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×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.

[43] I use average education in the municipality instead of average education in the firm for two reasons. First, it makes it more directly comparable to the original table. Second, I did not have time to recompute the average education within the firm with measurement error in own education - an exercise that involves the full population and takes substantial time.

Table 5: 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 parenthesis. All models include county of residence and year effects. Average education at municipality is computed using the sample (not using the full population). The sample consists of 2,393,573 men from 1985-2012.

Note the conditions that must be met for this approach to work. First, h_i must be fixed within a worker×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 only consists of the error term (absent additional controls). This amounts to a requirement that there is sufficient variation in peers, holding subject's characteristics constant. This also may not hold in many other settings. In particular, this does not generally hold in the schools setting, where classes are assigned at the start of the year, but there is generally no variation thereafter, conditional on holding the student×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 be impossible.

C.2 Estimation of Firm and Worker Fixed Effects

The identification and estimation of firm, worker, and time fixed effects was pioneered by Abowd et al. (1999).[44] 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 characteristics. However, estimation remains technically more challenging than simpler, two-way fixed effects

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

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).[45] I estimate the problem in two parts. First, I run a regression of log wages on dummies for year, county×year, industry×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 colleagues last period, workplace, or residual wage this period.

To obtain standard errors, a2reg requires a user written bootstrap. I thus also programmed a bootstrap that runs over the entire procedure.[46]

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

[46] I produced the standard errors using 50 bootstraps. In work in progress, I am increasing the number of bootstraps.

D Data Appendix

Table 6: Variable Descriptions

Variable	Source	Notes
Income	RAMS, Statistics Sweden, 1985-2012	Described in detail in the main text, see Section 4.
Average education of colleagues	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 the 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 and had local wage settlement
Wages	Landkomm, Statistics Sweden, 1985-2011	Public employee wages at the county council level. People employed in the county councils and 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 by state-regulated wages

Table 7: 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, in Swedish they are lans.
Municipality	LOUISE, Statistics Sweden, 1985-2012	There are currently 290 current municipalities in Sweden. However, there have been important revisions over time, which I account 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 Table 15 and Table 16

E Main Results Appendix

In this section of the appendix I report the results for a number of robustness checks, discussed in the main text.

E.1 Bartik Shocks

Bartik shocks introduce regional variation in labor demand based on changes in national demand for different industry's 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 able to construct Bartik shocks at the municipality level, allowing for control at a smaller level than the county controls.

Traditionally, Bartik shocks are used to instrument or control for shifts in labor demand.[47] Since I am interested in controlling for shifts in demand for average education, I adjust the traditional Bartik series accordingly. My series is given by:

$$\triangle B_{mt}^{\bar{S}} = \sum_k \frac{s_{mt-j}^k \left(1 + \frac{\triangle N_{(-m)t}^k}{N_{(-m)t-j}^k}\right)}{\sum_k s_{mt-j}^k \left(1 + \frac{\triangle N_{(-m)t}^k}{N_{(-m)t-j}^k}\right)} \bar{S}_{(-m)t}^k - \sum_k s_{mt-j}^k \bar{S}_{(-m)t-j}^k$$

I construct the Bartik shocks using data aggregated by Statistics Sweden.

Table 8: Controls for Bartik Shocks

	(1)	(2)	(3)	(4)	(5)
Own Education	0.199***				
	(0.0013)				
Lagged Average Education	0.179***	0.031***	0.016**	0.013*	0.016**
	(0.0021)	(0.0042)	(0.0054)	(0.0054)	(0.0055)
Bartik Shocks	0.564***	0.160***	0.011	0.001	0.022
	(0.0173)	(0.0466)	(0.0526)	(0.0542)	(0.0550)
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 parenthesis.

E.2 Controlling for Average Education in the Municipality

Note that this table does not yet include the worker and worker×plant fixed effects. Those results are in progress and will be added shortly.

[47] See, for example, Diamond (2012)

Table 9: Controlling for Average Education in the Municipality

	(1)	(2)	(3)	(4)	(5)
Own Education	0.191^{***}				
	(0.0010)				
Lagged Average Education	0.159^{***}	0.048^{***}	0.032^{***}	0.028^{***}	0.031^{***}
	(0.0016)	(0.0035)	(0.0049)	(0.0049)	(0.0050)
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 are reported in parenthesis.

E.3 Estimates Using Wages

Table 10: Using Wage Data

	(1)	(2)	(3)	(4)	(5)
Own Education	0.211^{***}				
	(0.0010)				
Lagged Average Education	0.106^{***}	0.042^{***}	0.034^{***}	0.026^{***}	0.033^{***}
	(0.0017)	(0.0034)	(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. 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 parenthesis.

E.4 Estimates Restricting to Selected Private Firms

Table 11: Restricting to Select Private Firms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own Education	0.193***						
	(0.0011)						
Lagged Average Education	0.378***	0.061***	0.046***	0.040***	0.043***	0.053***	0.055***
	(0.0020)	(0.0045)	(0.0094)	(0.0092)	(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 parenthesis.

E.5 Estimates Restricting to Plants with >20 Workers

Table 12: Restricting to Plants with >20 Workers

	(1)	(2)	(3)	(4)	(5)
Own Education	0.212***				
	(0.0011)				
Lagged Average Education	0.168***	0.065***	0.045***	0.038***	0.043***
	(0.0019)	(0.0046)	(0.0063)	(0.0062)	(0.0063)
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 parenthesis.

E.6 Persistence of Spillovers

Estimates with Deeper Lags

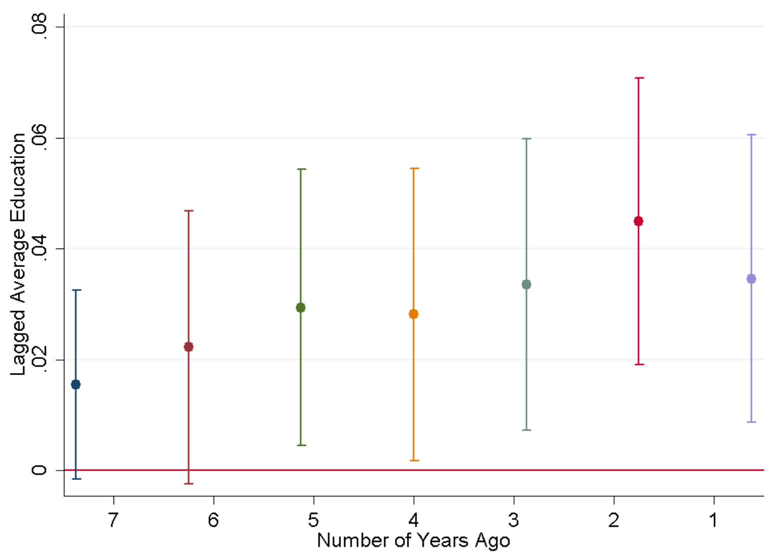
Table 13: Using Firm× Worker Spells

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lagged Average Education	0.028***	0.026***	0.029***	0.026***	0.015*	0.010	0.016
	(0.0049)	(0.0051)	(0.0059)	(0.0068)	(0.0060)	(0.0080)	(0.0087)
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 parenthesis.

Restricting to Same Sample In the graph below, I repeat the exercise in the main paper, but restrict to the same sample. Specifically, this graphs the persistence of spillovers, restricting every specification to individuals who remain at the same workplace for the past 7 years. The table with estimates is available upon request.

Figure 7: Persistence of Spillovers over Time, Robust



F Additional Results Appendix

F.1 Estimates by Age

Figure 8: Age Profile of Learning Spillovers: Non-Overlapping 10 Year Increments

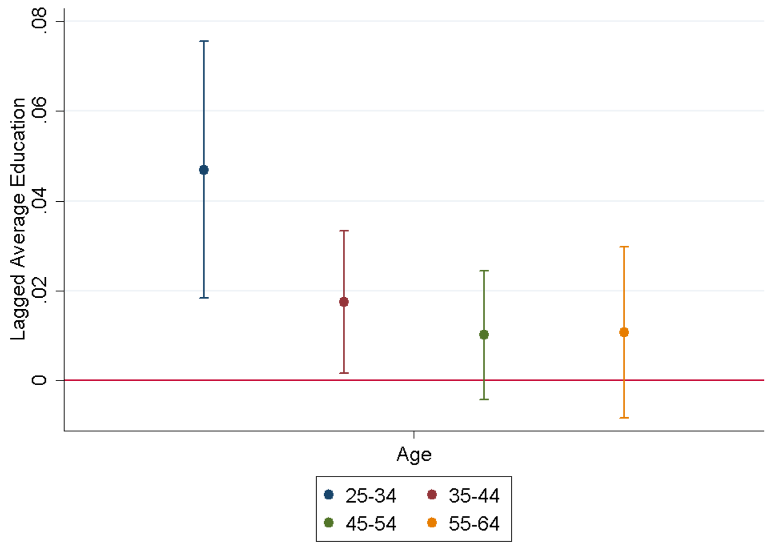


Table 14: Estimates by Age

	(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 parenthesis.

F.2 Construction of Occupation Ranks Using O*NET

Table 15: 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	
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 16: 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	