Good Firms, Worker Flows and Productivity

Michel Serafinelli

University of Toronto

9. June 2013

Good Firms, Worker Flows and Productivity

Michel Serafinelli
University of Toronto [1]

August 2013

[1]Contact Information: Department of Economics, University of Toronto, 150 St George St Toronto, ON M5S 3G7, Canada, michel.serafinelli@utoronto.ca, +1 416-978-4622. I am very grateful to David Card, Enrico Moretti, Patrick Kline and Bronwyn Hall for invaluable guidance. I also thank Andrés Rodríguez-Clare, Yuriy Gorodnichenko, Miguel Almunia, Vladimir Asriyan, Audinga Baltrunaite, Pamela Campa, Frederico Finan, Francesco Devicienti, Tadeja Gracner, Jonas Hjort, Agata Maida, Steven Raphael, Ana Rocca, Jesse Rothstein, Fabio Schiantarelli, Victoria Vanasco, and seminar participants at Berkeley, Berkeley Haas, U Penn Wharton, Warwick, Pompeu Fabra, Paris School of Economics, U of Cambridge, Stockholm U, Universitat Autonoma Barcelona, U of Toronto, All-California Labor Conference at RAND and San Francisco Fed for comments. I acknowledge financial support from the Center for Equitable Growth at Berkeley. Thanks are due also to Federico Callegari, to Veneto firms' employees and officials of employers' associations who have generously given their time in interviews, and to Ishita Arora, Olivia Casey and Nitin Kohli for research assistance. I am indebted to Giuseppe Tattara for making available the Veneto Work History Data.

Abstract

I present direct evidence on the role of firm-to-firm labor mobility in enhancing the productivity of firms located near highly productive firms. Using matched employer-employee and balance sheet data for the Veneto region of Italy, I identify a set of high-wage firms (HWF) and show they are more productive than other firms. I then show that hiring a worker with HWF experience increases the productivity of other (non-HWF) firms. A simulation indicates that worker flows explain 10-15 percent of the productivity gains experienced by other firms when HWFs in the same industry are added to a local labor market.

JEL: J24; J31; J61; R23

Keywords: productivity, agglomeration advantages, linked employer-employee data, labor mobility.

1 Introduction

A prominent feature of the economic landscape in the most developed countries is the tendency for firms to locate near other firms producing similar products or services. In the United States, for example, biopharmaceutical firms are clustered in New York and Chicago and a sizeable share of the elevator and escalator industry is concentrated in the area around Bloomington, Indiana. In addition, the growth and diffusion of multinational corporations has led to the recent appearance of important industrial clusters in several emerging economies. Firms that originally agglomerated in Silicon Valley and Detroit now have subsidiaries clustered in Bangalore and Slovakia (Alfaro and Chen, 2010).

Researchers have long speculated that firms in industrial concentrations may benefit from agglomeration economies, and a growing body of work has been devoted to studying the importance of these economies. Despite the difficulties involved in estimating agglomeration effects, a consensus has emerged from the literature that significant productivity advantages of agglomeration exist for many industries (Rosenthal and Strange, 2003; Henderson, 2003; Ellison, Glaeser and Kerr, 2010; Greenstone, Hornbeck and Moretti, 2010; Combes et al., 2012). Localized knowledge spillovers are a common explanation for the productivity advantages of agglomeration. Nevertheless, as pointed out by Combes and Duranton (2006), if information can easily flow out of firms, the question of why the effects of spillovers are localized must be clarified.

This paper directly examines the role of labor mobility as a mechanism for the transfer of efficiency-enhancing knowledge and evaluates the extent to which labor mobility can explain the productivity advantages of firms located near other highly productive firms. The underlying idea is that knowledge is embedded in workers and diffuses when workers move between firms. The strong localized aspect of knowledge spillovers discussed in the agglomeration literature may thus arise from the propensity of workers to change jobs within the same local labor market.

In order to empirically assess the importance of labor-market based knowledge spillovers, I use a unique dataset from the Veneto region of Italy that combines Social Security earnings records and detailed financial information for firms. I begin by presenting a simple conceptual framework where some firms are more productive because they have some superior knowledge. Employees at these firms passively acquire some proportion of the firm's internal knowledge. For simplicity, I refer to these as "knowledgeable" workers. Other firms can gain access to the superior knowledge by hiring these workers. Empirically, I identify potential high-productivity firms as those that pay a relatively high firm-specific wage premium.[1] I show that these high-wage-firms (HWFs) have higher labor productivity, higher value added, and higher capital (in particular intangible capital) per worker, suggesting the presence of a firm-specific productivity advantage and thus a point of origin for the transfer of knowledge. Next, I evaluate the extent to which non HWFs benefit from hiring knowledgeable workers by studying the effect on productivity associated with hiring workers with recent experience at HWFs.

Productivity shocks that are correlated with the propensity to hire knowledgeable workers may give rise to an upward bias in the impact of knowledgeable workers. In order to address this potential endogeneity issue, I use control function methods from the recent productivity literature (Olley and Pakes, 1996; Levinsohn and Petrin, 2003). Another potential threat to identification is the fact that I do not observe labor quality. In particular, since the good firms pay a relatively high firm-specific wage premium, workers who separate from a good firm may be of lower quality. I refer to this potential adverse selection problem as "lemons bias" (Gibbons and Katz, 1991). Lemons bias will tend to work against the finding of a positive effect of knowledgeable workers. In order to address this issue, I obtain a proxy for worker ability

[1] This is consistent with many recent models of frictional labor markets (e.g., Christensen et al., 2005), in which higher-productivity firms pay higher wages for equivalent workers.

and I weight the number of workers in my OLS regression using the average ability to obtain effective labor input.

I conclude that hiring a worker with HWF experience significantly increases the productivity of other (non HWF) firms. A non HWF hiring at the mean gains 0.14-0.28 percent in productivity compared to an observationally identical firm that hired no knowledgeable workers. This gain is equivalent to moving 0.2-0.5 centiles up the productivity distribution for the median firm. The productivity effect of knowledgeable workers is not associated with recently hired workers in general; I do not find a similar productivity effect for recently hired workers without experience at good firms.

The number of knowledgeable workers may also be correlated with productivity shocks happening in the future if workers can foresee them and apply for jobs in firms with better growth prospects. To deal with this threat to identification, I adapt the control function methods to proxy for future productivity shocks. As an alternative approach, I instrument for the number of knowledgeable workers in a non HWF with the number of local good firms in the same industry that downsized in the previous period. Indeed, following a downsizing event at a HWF, it is more likely that a knowledgeable worker applies for job at local non HWFs because s/he is unemployed and does not want to relocate far away, and not because some particular non HWF offers better prospects than the HWF at which the worker is employed. This instrumental variable (IV) strategy also further guards against the possibility of lemons bias: the larger the number of workers laid off from HWFs, the lower, arguably, the extent of selection. The IV estimates return an economically and statistically significant effect of recruiting knowledgeable workers on non HWF productivity, with the point estimate larger than the OLS. While in principle this is consistent with the idea that the OLS coefficient is biased downward (lemons bias), in practice the IV standard errors are large and prevent me from drawing definitive conclusions.

In the last part of the paper, I assess the extent to which worker flows can explain the productivity advantages of firms located near other highly productive firms. I relate my findings to the existing evidence on the productivity advantages of agglomeration, focusing in particular on the study performed in Greenstone, Hornbeck and Moretti (2010, henceforth GHM). The authors find that after the opening of a large manufacturing establishment, total factor productivity (TFP) of incumbent plants in US counties that were able to attract one of these large plants increases significantly relative to the TFP of incumbent plants in counties that survived a long selection process but narrowly lost the competition. The observed effect on TFP is larger if incumbent plants are in the same industry as the large plant, and increases over time. These two facts are consistent with the presence of intellectual externalities that are embodied in workers who move from firm to firm. However, data limitations prevent GHM from drawing definitive conclusions regarding the driving mechanism. I evaluate the extent to which worker flows explain empirical evidence on the productivity advantages of agglomeration, by simulating an event similar to that studied by GHM but within the worker mobility framework described above. The change in productivity predicted within this framework equals 10-15 percent of the overall effect found in GHM, indicating that knowledge transfer through worker flows explain a significant portion of the productivity advantages through agglomeration.

2 Relation to Previous Research

This paper adds to the growing literature on productivity advantages through agglomeration, a literature critically surveyed in Duranton and Puga (2004), Rosenthal and Strange (2004) and Moretti (2011). The research relating most closely to this paper is the body of work on micro-foundations for agglomeration advantages based on knowledge spillovers. In Combes and Duranton (2006)'s theoretical analysis, firms clustering in the same locality face a trade-off between the advantages of labor pooling (i.e. access to knowledge carriers) and the costs of labor poaching (i.e. loss of some key employees to competitors along with higher wage bills to retain other key employees). In a case study of the British Motor Valley, Henry and Pinch (2000) conclude that

> as personnel move, they bring with them knowledge and ideas about how things are done in other firms helping to raise the knowledge throughout the industry...The crucial point is that whilst this process may not change the pecking order within the industry, this 'churning' of personnel raises the knowledge base of the industry *as a whole within the region.* The knowledge community is continually reinvigorated and, synonymous with this, so is production within Motor Sport Valley

In a similar vein, Saxenian (1994) maintains that the geographic proximity of high-tech firms in Silicon Valley is associated with a more efficient flow of new ideas. I contribute to the literature on micro-foundations for agglomeration advantages by showing direct evidence of productivity gains through worker flows. My results are consistent with the findings by Henry and Pinch (2000). Since worker flows in a local labor market are larger within an industry, and, as I shall show, the productivity effect is larger for workers moving within the same industry, my results may also help explain the findings in Henderson (2003), Cingano and Schivardi (2004) and Moretti (2004a) that local spillovers are increasing in economic proximity.[2]

Some research beyond the agglomeration literature has also emphasized the fact that new workers share ideas on how to organize production or information on new technologies that they learned with their previous employer. Balsvik (2011) uses matched employer-employee data from Norway and offers

[2]Measures of economic links include input and output flows and indicators of technological linkages.

a detailed account of productivity gains linked to worker flows from foreign multinational to domestic firms. Similarly, using linked worker-firm data, Parrotta and Pozzoli (2012) and Stoyanov and Zubanov (2012) show evidence from Denmark that is consistent with models of knowledge diffusion through labor mobility. My findings are consistent with those of these three recent papers. My empirical strategy, however, allows me to make progress on the identification of the causal effect of recruiting knowledgeable workers on productivity. I address the three main identification issues, namely (a) unobservable contemporaneous productivity shocks at time t, (b) unobservable worker quality and (c) unobservable future productivity shocks, using several approaches, including control function methods from the recent productivity literature and an IV strategy. Furthermore, while the above authors focus exclusively on the role of labor mobility for knowledge transfer, I seek to shed light on a broader question: the extent to which labor mobility can explain evidence on the productivity advantages through agglomeration.

3 Conceptual Framework

Assume there exists a finite number of locations, each constituting a separate local labor market. To fix ideas, assume that these labor markets are completely segmented with workers being immobile between them. There exists a finite collection $\mathcal{J} = \{\mathcal{J}_0, \mathcal{J}_1\}$ of firms consisting of the set $\mathcal{J}_1$ *good* firms, which are more productive because they have some superior knowledge and set $\mathcal{J}_0$ other firms which have no access to the superior knowledge. The superior knowledge is exogenously given and could include information about export markets, physical capital, process innovations, new managerial techniques, new organizational forms and intermediate inputs. Workers employed by good firms acquire some proportion of the firms' internal knowledge. For simplicity, I assume that this acquisition of internal knowledge takes place immediately after the workers join the good firm. Workers are *knowledge-*

able if they have knowledge of the relevant information and *unknowledgeable* otherwise. All workers employed by good firms, then, are knowledgeable. Additionally, some proportion of this knowledge can be *transferred* to a $j \in \mathcal{J}_0$ firm if the workers switch employers.[3] The production function of firm $j \in \mathcal{J}_0$ is

$$Y_j = F(L_j, K_j, M_j) = A_j[(\overline{\theta}_j L_j)^\alpha K_j^\gamma M_j^\lambda]^\delta \tag{1}$$

where $L = H + N$, i.e. the sum of knowledgeable workers (H,who moved at some point from a good firm to a non-good firm) and unknowledgeable workers (N); $\overline{\theta}$ is the quality of the workforce, K is total capital inputs, M is material inputs, and $\delta < 1$ represent an element of diminishing return to scale, or to "span of control" in the managerial technology (Lucas, 1978).[4] I allow for knowledge transfer by:

$$A_j = D_j e^{\beta_H H_j} \tag{2}$$

4 Econometric Framework

I obtain the regression equation that forms the basis of my empirical analysis, by combining equation (1) and (2) and taking logs:

$$\ln(Y_{jslt}) = \beta_L \ln(\overline{\theta}_{jslt} L_{jslt}) + \beta_K \ln(K_{jslt}) + \beta_M \ln(M_{jslt}) + \beta_H H_{jslt} + \beta_0 + \zeta_{jslt} \tag{3}$$

The dependent variable is the real value of total firm production, s denotes industry, l denotes locality and t denotes year.[5] The variable of interest, H is constructed from head counts in the matched employer-employee data. The term $\ln(D_j)$ is decomposed into two elements, β_0 and ζ_{jslt}. The constant β_0

[3] I assume that this type of knowledge cannot all be patented and that exclusive labor contracts are not available.

[4] This is in line with the large presence, that I document below, of small and medium size firms in the non-HWF sample.

[5] Notice that $\beta_L = \delta\alpha, \beta_K = \delta\gamma, \beta_M = \delta\lambda$.

denotes mean efficiency across all firms in $\mathcal{J}_0$ that is due to factors others than H. The time-variant ζ_{jslt} represents deviations from this mean efficiency level and captures (a) unobserved factors affecting firm output, (b) measurement error in inputs and output, and (c) random noise. Estimating the effect of recruiting a knowledgeable worker on a firm's productivity is difficult in the presence of unobservable contemporaneous productivity shocks, unobserved labor quality and unobservable future productivity shocks. I turn now to describing what type of biases these unobservables may introduce and how I deal with them in the empirical work.

4.1 Productivity shocks at time t

Express ζ_{jslt} , the deviations from mean firm efficiency not resulting from knowledge transfer, as

$$\zeta_{jslt} = \omega^*_{jslt} + \nu_{jslt} = \omega_{jslt} + \mu_{st} + \varpi_{lt} + \nu_{jslt} \tag{4}$$

which specifies that ζ_{jslt} contains measurement error ν_{jslt} and a productivity component ω^*_{jslt} (TFP) known to the firm but unobserved by the econometrician. The productivity component can be further divided into a firm-specific term, a term common to all firms in a given industry (μ_{st}) and a term common to all firms in a given locality (ϖ_{lt}). Equation (3) now becomes:

$$\ln(Y_{jslt}) = \beta_0 + \beta_L \ln(\overline{\theta}_{jslt} L_{jslt}) + \beta_K \ln(K_{jslt}) + \beta_M \ln(M_{jslt}) + \beta_H H_{jslt} + \mu_{st} + \varpi_{lt} + \omega_{jslt} + \upsilon_{jslt} \tag{5}$$

One major difficulty in estimating β_H in Equation (5) is that non HWFs may decide on their choice of H based on the realized firm-specific productivity shock (ω_{jslt}) unknown to the researcher. When employing OLS to estimate Equation (5) without accounting for the existence of ω_{jslt}, the bias induced by endogeneity between H and ω_{jslt} is likely positive (positive productivity shocks translate into higher probability to hire from HWFs), implying that the coefficient estimate will be biased upward ($\widehat{\beta_H} > \beta_H$).

I employ the productivity literature's techniques to control for the endogeneity of inputs in order to assess the relevance of this issue in my setting. In particular, I apply the Olley and Pakes (1996, henceforth OP) and the Levinsohn and Petrin (2003, henceforth LP) approaches. OP construct an explicit model for the firm's optimization problem in order to obtain their production function estimator. Essentially, the authors address the issue of endogeneity of inputs by inverting the investment function to back out—and thus control for—productivity. Building on OP, LP suggest the use of intermediate input demand in place of investment demand as a proxy for unobserved productivity. The results are shown in Section 7.1. See Eberhardt and Helmers (2010) for an in-depth discussion of these 'structural' estimators.

4.2 Unobserved Worker Quality

Another potential threat to identification is the fact that I do not observe labor quality. In particular, since the good firms pay a relatively high firm-specific wage premium, workers who separate from a good firm may be of lower quality. This lemons bias may work against the finding of a positive effect of knowledgeable workers. In order to address this issue, I obtain a proxy for worker ability and I weight the number of workers in my OLS regression using the average ability to obtain effective labor input. Specifically, I weight the total number of workers L by firm j's average worker ability level $\overline{\theta}_{jslt} = \frac{1}{L_{jslt}} \sum_{i=1}^{L_{jslt}} \theta_i$, to obtain effective labor input. $\overline{\theta}_{jslt}$ is time-varying at the firm level, given that the number and composition of workers change. In order to obtain the individual θ_i, I procure estimates of worker fixed effects from wage equations where both firm and worker effects can be identified. Section 4.5 describes this estimation in detail.

The IV strategy based on the events of downsizing at good firms (described in Section 4.4) further guards against the possibility of lemons bias:

the larger the number of workers laid off from HWFs, the lower, arguably, the extent of selection.

4.3 Productivity shocks at time t+1

The number of knowledgeable workers may be correlated with productivity shocks happening in the future if workers can foresee them and apply for jobs at firms with better growth prospects. If such firms prefer to hire workers from good firms, these workers will have a higher probability of being chosen. To the extent that preferring workers from good firms can be explained through knowledge transfer from these firms, a positive correlation between H and the receiving firm's productivity shocks in $t+1$ does suggest a role for labor mobility as a channel for knowledge transfer, even though it will overestimate its importance (Stoyanov and Zubanov, 2012).

In an effort to proxy for productivity shocks in $t+1$ that may be anticipated by the workers, I add polynomial functions of capital and investment and of capital and materials in both t and $t+1$. This is in the spirit of the OP and LP approaches and assumes that hiring firms are also able to anticipate their productivity shocks and adjust their inputs accordingly. As an alternative approach to deal with this issue, I adopt an IV strategy that I now describe.

4.4 Using the number of downsizing firms as IV

In Section 7.2 I present estimates where I instrument for the number of knowledgeable workers in a non HWF with the number of *local* good firms in the *same 5-digit industry* that downsized in the previous period. The IV strategy is an alternative approach to deal with the strategic mobility issue discussed in the previous section. Indeed following a downsizing event at a HWF, it is more likely that a knowledgeable worker applies for job at local non HWFs because s/he is unemployed and does not want to relocate far away, and not because some particular non HWFs offer better prospects than the HWF at which the worker is employed. Put differently, in the scenario captured by the IV approach, the strategic mobility explanation is less likely to play a major role.

One can think of two main reasons why good firms may downsize in a particular year. First, good firms may get a bad draw from the distribution of product-market conditions. Even though an inherent productivity advantage partly insulates the good firms from output shocks, sufficiently large shocks will pierce this insulation and induce the good firm to layoff workers. Alternatively, good firms may downsize in a particular year due to offshoring.

The basic intuition behind the IV approach is to consider moves from workers whose former employer downsized due to demand shocks or offshoring. While the timing of these moves is arguably exogenous, these workers may still decide which new employer to join among the set of non-HWFs. However, in small labor markets and specialized industries, workers may have a limited set of alternatives (Tecu, 2012).

The choice of the instrument is based on the notion that geographic proximity plays an important role in determining worker mobility. In January 2012, I visited several Veneto firms and interviewed employees about the history of their enterprises and their current operations. I also conducted phone interviews with officials of employers' associations and chambers of commerce. My anecdotal evidence supports the idea that distance acts as a barrier for job mobility.[6] In Appendix III I further discuss the role of geographic proximity.

In the presence of product demand shocks or offshoring, using the number of downsizing firms as an instrument is invalid if it cannot be excluded from

[6] In a phone interview, Federico Callegari of the Treviso Chamber of Commerce, reasoned out the role of geographic proximity: "I think distance matters a lot for workers' job mobility. When losing their job, workers tend to look for another job with a commuting time of maximum 20-30 minutes. Why? Because they want to go home during the lunch break!"

the causal model of interest (Equation 3). The identifying assumption of my IV strategy is therefore that the number of downsizing good firms is correlated with the causal variable of interest, H, but uncorrelated with any other unobserved determinants of productivity.

4.5 Identification of Good Firms

Empirically, I identify potential high-productivity firms as high-wage firms (HWFs), i.e. those that pay a relatively high firm-specific wage premium. This is consistent with many recent models of frictional labor markets (e.g., Christensen et al., 2005), in which higher-productivity firms pay higher wages for equivalent workers. As I shall show below using balance sheet data, HWFs have significant higher output per worker and value added per worker than other firms in my sample.

There are three reasons why I define the good firms as HWFs and detect them using Social Security data rather than define the good firms directly as the highly productive ones and detect them using balance sheet data. First, the availability of worker-level Social Security data allows the introduction of measured individual characteristics and worker effects, something impossible to capture with firm level data from balance sheets. Second, Social Security data are available for a longer period of time than the balance sheets, and therefore increase the precision of the categorization of firms into good and non-good groups. Third, since Social Security records are administrative data, measurement error is lower than in balance sheets.

Following Abowd, Kramarz and Margolis (1999, henceforth AKM), I specify a loglinear statistical model of wages as follows:

$$w_{ijt} \;=\; X_{it}'\beta + \theta_i + \psi_j + \upsilon_t + \varepsilon_{ijt} \tag{6}$$

where the dependent variable, the log of the average daily wage earned by worker i in firm j in year t, is expressed as a function of individual heterogeneity, firm heterogeneity, and measured time-varying characteristics.[7] The assumptions for the statistical residual ε_{ijt} are (a) $E[\varepsilon_{ijt}|i, j, t, x] = 0$, (b) $Var[\varepsilon_{ijt}|i, j, t, x] < \infty$ and (c) orthogonality to all other effects in the model. The presence of labor mobility in matched worker-firm data sets enables the identification of worker and firm effects.[8] I identify good firms as those whose estimated firm fixed effects fall within the top third of all estimated firm effects. Section 6 reports more details on the estimation procedure.

5 Data

The data set is for Veneto, an administrative region in the Northeast of Italy with a population of around 5 million people (8 percent of the country's total). Since the mid-1980s, the labor market in Veneto has been characterized by nearly full employment, a positive rate of job creation in manufacturing and positive migration flows (Tattara and Valentini, 2010). The dynamic regional economy features a large presence of flexible firms, frequently organized in districts with a level of industrial value added greatly exceeding the national average.[9] Manufacturing firms in Veneto specialize in metal-engineering, goldsmithing, plastics, furniture, garments, textiles, leather and shoes.[10] The manufacture of food and beverage, and wine and baked goods in particular, is also a prominent subsector.

My data set pools three sources of information: individual earnings records,

[7] The vector X'_{it} includes tenure, tenure squared, age, age squared, a dummy variable for manager and white collar status, and interaction terms between gender and other individual characteristics.

[8] The identification relies on the assumption that mobility is exogenous to the included regressors. Bias in the estimated firm effects arises when errors predict specific firm-to-firm transitions. Card, Heining and Kline (2012) conduct a series of checks for patterns of endogenous mobility that could lead to systematic bias in AKM's additive worker and firm effects model. The authors find little evidence of such biases in German data.

[9] The most famous industrial concentration is the eyewear district in the province of Belluno, where Luxottica, the world's largest manufacturer of eyeglasses, has production plants.

[10] Benetton, Sisley, Geox, Diesel, and Replay are Venetian brands.

firm balance sheets, and information on local labor markets (LLMs).[11] The earnings records come from the Veneto Workers History (VWH) dataset. The VWH has data on all private sector personnel in the Veneto region over the period 1975-2001. Specifically, it contains register-based information for virtually any job lasting at least one day. A complete employment history has been reconstructed for each worker.

Balance sheets starting from 1995 were obtained from AIDA (Analisi Informatizzata delle Aziende), a database circulated by Bureau Van Dijk containing official records of all incorporated nonfinancial Italian firms with annual revenues of at least 500,000 Euros. AIDA's balance sheets include firms' location, revenues, total wage bill, the book value of capital (broken into subgroups), value added, number of employees, value of materials and industry code. I use firm identifiers to match job-year observations for workers aged 16-64 in the VWH with firm financial data in AIDA for the period 1995-2001. Further details on the match and data restrictions I make are provided in Appendix I.

Information on LLMs is obtained from the National Institute of Statistics (ISTAT). The LLMs are territorial groupings of municipalities characterized by a certain degree of working-day commuting by the resident population. In 1991 the 518 municipalities or *comuni* in Veneto are divided into 51 LLMs.

6 AKM Estimation and Characterization of Good Firms

The method in Abowd, Creecy and Kramarz (2002) identifies separate groups of workers and firms that are connected via labor mobility in matched employer-employee data. When a group of workers and firms is connected, the group contains all persons who ever worked for any firm within the group and all

[11] The first two sources, combined for the period 1995-2001, have been used in the study on rent-sharing, hold-up and wages by Card, Devicienti and Maida (2010).

firms at which any of the persons were ever employed. I run the grouping algorithm separately using VWH data from 1987 to 2000 for firms that could be matched in AIDA. I then use the created group variable to choose the largest group as a sample for my fixed-effects estimation - Equation (6). Details on sample restrictions and descriptive statistics are provided in Appendix I. Figure 1 shows the distribution of estimated firm effects.

I identify HWFs as those firms whose firm effects rank in the top third of the sample. Figure 2 shows the geographical variation in the number of HWFs across LLMs for the most recent year (2001).

For labor mobility to generate productivity benefits of agglomeration, a firm-specific advantage should be observed at good firms that could be the basis for knowledge transfer to other firms in the region. Therefore, once I have categorized firms into HWF and non HWF groups, I estimate:

$$\ln O_{jst} = \beta_0 + \beta_1 HWF_{js} + \mu_s + \upsilon_t + e_{jst} \tag{7}$$

where the dummy HWF takes the value of 1 if firm j is classified as high-wage and O_{jst} represents different firm-level outcomes. Table 1 shows the results of estimating Equation (7).

[TABLE 1 HERE]

In the Veneto manufacturing sector clear differences between HWFs and non HWFs emerge in labor productivity (measured as output per worker, Column 1), value added per worker (Column 2) and capital per worker (Column 3), including both tangible capital (Column 4) and, most remarkably, intangible fixed assets (Column 5). This evidence is important for establishing the potential for knowledge transfer in the region. Since labor productivity is on average 15 percent higher in HWFs, and intangible capital per worker (intellectual property, accumulated research and development investments and goodwill) is 27 percent larger, we can also think of HWFs as high-productivity firms, or high-intangible-capital firms.

For labor mobility to be a mechanism for transfer of knowledge, we must also observe some workers moving from HWFs to other firms. Appendix II discusses the extent of labor mobility from HWF to non HWF. It also presents descriptive statistics on individual characteristics of the movers in my sample.

7 Evidence on Worker Flows and Productivity

7.1 Main Estimates

In this section I present the main result from regression analysis in this paper. Specifically, I evaluate the extent to which non HWFs benefit from hiring workers from HWFs. Estimation of Equation (3) is performed over the period for which balance sheet data are available (1995-2001). Details on sample restrictions and descriptive statistics for the variables used in the regression analysis are provided in Appendix I. Table 2 shows the estimation results. I cluster standard errors at the firm level. Coefficients associated with the H measure in Table 2 represent semielasticities because my variable of interest is not in logarithms. This choice for the baseline specification, which directly follows from Equation (2), is founded on the fact that H takes on the value 0 for the majority of observations. Thus, any possible transformation of the H measure could possibly affect the associated estimated parameters. In any case, Appendix IV I show results using different functional forms.

[TABLE 2 HERE]

Column 1 reports estimate from the baseline OLS specification: the coefficient on H_{jst} is positive (0.039) and significant.[12] Column 2 and 3 of Table

[12] All inputs are positive and statistically significant, and the labor coefficient is an expected 71% of the summed coefficients for labor and capital. The overall production

2 employ the productivity literature's techniques to control for the endogeneity of inputs. H_{jst} is treated as a freely variable input. Column 2 reports results using the OP estimator: the coefficient for H_{jst} is positive (0.037) and significant.[13] Column 3 reports the results for LP estimator: the coefficient for H_{jst} is positive (0.020) and significant; it is lower than the OLS estimate, confirming the theoretical and empirical results on variable inputs discussed in LP.[14]

Although the estimate of the coefficients for H_{jst} in the OP and LP specification are smaller than the baseline estimate, none of the specifications is qualitatively inconsistent with the empirical finding that labor mobility works as a channel of knowledge transfer. The point estimates suggest that the average effect of recruiting a knowledgeable worker on a non HWF's productivity is an increase of between 2 and 3.9 percent. This seems like a very large effect. However, I discuss in Appendix II, the mean number of knowledgeable workers is 0.071 and as many as 95.7 percent of non HWFs in a given year do not employ any worker of this type. Hiring one worker of this type therefore implies a large change for most firms in our data (Notice also that non-HWFs are quite small: the median number of employees at non HWFs is 33).

---

function has mild decreasing returns to scale, with a 1 percent increase in all inputs leading to a 0.9 percent increase in output.

[13] I use the opreg Stata routine developed by Yasar, Raciborski and Poi (2008). I do not observe investment, and hence derived a proxy variable in t as the difference between the reported book value of capital at time $t+1$ and its value in t. The way I constructed the proxy variable somehow exacerbates the measurement error problems typically associated with the proxy variable approach. In addition, augmenting my specification with this proxy variable reduces my sample size substantially, as (a) 3871 firm-year observations are lost when I take the difference in reported book values and (b) the OP approach requires positive values for the proxy variable, eliminating an additional 7174 firm-year observations. (The estimation routine will truncate firms' non-positive proxy variable observations because the monotonicity condition necessary to invert the investment function, and hence back out productivity, does not hold for these observations.)

[14] I use the *levpet* Stata routine developed by Petrin, Poi and Levinsohn (2004).

As a further illustration of Table 2's estimates, given the mean value of H (0.07) and its slope coefficient in Column 1, $\widehat{\beta_H}$ =0.039, a non-HWF hiring at the mean H gains $0.039 \cdot 0.07 = 0.28$ percent in productivity compared to an observationally identical firm that hired no-one. This gain is equivalent to moving 0.5 centiles up the productivity distribution for the median firm. If one uses the LP estimate instead of the OLS, the gain is equal to 0.14 percent (0.2 centiles up in the distribution).

In an effort to proxy for productivity shocks in t+1 that may be anticipated by the workers (recall the discussion in Section 4.3), in Column 4 and 5 I add polynomial functions of capital and investments or capital and materials in t and $t+1$. These estimates also suggest that non HWFs benefit from knowledgeable workers by experiencing increased productivity.[15] In Section 7.2 I show results from the IV strategy, an alternative approach to deal with the strategic mobility issue that may arise as a result of the presence of unobservable shocks in $t+1$.

Next, I address the questions of whether the knowledge embedded in workers is general enough to be applied in different industries: Column 6 of Table 2 differentiates between workers with HWF experience moving within the same two-digit industry and workers moving between industries. The coefficient of knowledgeable workers moving within industry is highly significant and positive (0.072). The coefficient of knowledgeable workers moving between industries is significant and positive but smaller (0.024). The difference in the two coefficients is significant at conventional levels.

Overall, the main empirical result in this Section is that labor mobility from HWFs to other firms in the region works as a mechanism for the transfer of efficiency-enhancing knowledge. Hiring within the same industry brings more relevant new knowledge than that which can be acquired from workers

[15]That said, most of the components in the polynomial approximations are statistically significant, implying that these extra terms contribute in explaining the variation in productivity among firms. Notice the drop in observations due to the fact that we are using the leads of inputs (polynomials in $t+1$).

previously employed outside.

Appendix IV investigates the robustness of these findings to different specifications and explores potential alternative explanations of the estimated productivity effects. In particular, I investigate the role of the selection of movers based on observable characteristics and unobserved firm heterogeneity. I also evaluate the role of functional form assumptions, and I explore the importance of time-varying unobservables correlated with the number of recent hires.

7.2 IV Estimates

In this section I instrument for the number of knowledgeable workers using the lagged number of good local firms in the same 5-digit industry that downsized in the previous period. This exercise is motivated by the possibility of strategic mobility and lemons bias that I discussed above. The exclusion restriction is violated and $\widehat{\beta_H}^{IV}$ is biased upward if there are localized unobservable industry shocks that lead good firms to downsize and *positively* affect productivity at non HWFs. Below, I further discuss the validity of the IV strategy.

Turning to the details of the instrument, a downsizing firm must see an employment reduction larger than 3 percent *compared to the previous year's level*. The division of good firms into downsizing and non-downsizing firms according to this criterion is less sensible for small firms. Accordingly, I impose the additional condition that the decrease in the labor force is greater than or equal to three individuals.

Table 3 shows the results from the IV estimation of Equation (3). Standard errors are clustered at the level of the LLM.

[TABLE 3 HERE]

The F test of excluded instruments in Column 1 gives a statistic of 23.1[16].

[16] The coefficient of the number of downsizing firm in the first-stage regression is equal

The estimated coefficient of H is large: $\widehat{\beta_H}^{IV} = 0.268$. However, the standard error is also large (0.154). The coefficient is significant at the 10 percent level.

A concern for the validity of the exclusion restriction arises from the observation that the dependent variable in my econometric model is the *value* of output.[17] Unobserved shifts in local demand from HWFs to non HWFs might simultaneously lead to (a) higher output prices for non HWFs, (b) downsizing by HWFs and (c) hiring of HWF employees by non HWFs. The LLM-year effects control for local demand shocks, but localized unobservable *industry* shocks may still play a role. Consequently, in principle, it is possible that $\widehat{\beta_H}^{IV} > 0$ reflects higher output prices, rather than higher productivity due to knowledge transfer. I do not expect this to be a major factor in my context; manufacturing firms in my sample generally produce goods traded outside the LLM.[18] To further explore this possibility in Column 2 I add a dummy taking value one if the industry produces goods that are not widely traded outside the LLM.[19] The results in Column 2 are very similar to those in Column 1.

Even when the level of tradability is controlled for, product demand effects might still be relevant and $\widehat{\beta_H}^{IV}$ might therefore be biased if an industry is strongly localized. In such a scenario the negative shock to the local HWF may lead to increased demand for the non HWF firm j even though the HWF and the non HWF produce a tradable good. This is because, since most of the firms producing that particular good in Italy are in the same Veneto LLM, the non HWF may experience an increase in demand, and hence in price, after the negative shock to a local HWF that is a direct competitor

to .017 (standard error is 0.003). A one standard deviation increase in the instrumental variable is associated with an increase in H of 0.02.

[17] The theoretically correct dependent variable in a productivity study is the *quantity* of output, but, due to data limitations, this study (and virtually all the empirical literature on productivity) uses price multiplied by quantity.

[18] Imagine the extreme case of a non-HWF that produces a nationally traded good in a perfectly competitive industry. Its output prices would not increase disproportionately if the LLM experienced an increased demand for its good.

[19] See Appendix V for details.

on the national market. To address this concern, I construct an index of industry localization as follows $r_s = (Italian\ Firms\ in\ s)/(Veneto\ Firms\ in\ s)$. Industries with low r have a relatively small number of firms outside the Veneto area. In Column 3 I enter r_s as additional regressor: the F test gives a slightly larger statistic (24.4) and $\widehat{\beta_H}^{IV}$ is estimated slightly more precisely (standard error is 0.145). The point estimate is very similar.

Finally, in column 4 I use a stricter definition of downsizing firms: a downsizing firm must see an employment reduction larger than 5 percent *compared to the previous year's level.*[20] The F test gives a slightly lower statistic (21.6), the standard error is larger (0.164) and $\widehat{\beta_H}$ is no longer significant at 10 percent level. However, the point estimate is quite similar to that in the previous columns (0.231).

Recall the OLS estimates: (a) the coefficient on knowledgeable workers is 0.039 and (b) the coefficient on knowledgeable workers moving within the same two-digit industry is 0.072. In principle, the IV estimates (that are likely to be driven by flows *within* industries, given the way the instrument is designed) are consistent with the idea that the OLS coefficient is biased downward because of negative selection (lemons bias). In practice, however, the IV standard errors are large and prevent me from drawing definitive conclusions.

Another tentative explanation for the magnitude of the IV results is that the effect of knowledgeable workers may be heterogeneous across firms. If there are indeed heterogeneous effects of H on productivity, then consistent OLS measures the average effect of H on productivity across all firms, while Two Stage Least Squares (TSLS) estimates the average effect in the subset of firms that are marginal in the recruitment decision, in the sense that they recruit knowledgeable workers if and only if there exists excess local supply.[21]

[20] I keep the additional condition that the decrease in the labor force is greater than or equal to three individuals. Both the baseline instrumental variable and this alternative one are summarized in Table A.5.

[21] See Imbens and Angrist (1994) for a discussion. For a recent example, see Eisensee

If the effect of knowledgeable workers on productivity is larger for non HWFs that are marginal in the recruitment decision, the TSLS estimates will exceed those of consistent OLS.

8 Worker flows and agglomeration advantages

In this Section I assess the extent to which worker flows can explain the productivity advantages of firms located near other highly productive firms. In order to do so, I simulate an event analogous to that studied by GHM but within my framework, and I predict the change in local productivity that is due to labor mobility. The event I simulate is an increase in the number of good firms such that the change in local output is comparable to the output of the average large plant whose opening is considered by GHM.[22]

An overview of my procedure is as follows. Denote the number of knowledgeable workers *moving within industry* observed at firm j with H^{ind}. As a first step, I estimate the effect on H_j^{ind} of a change in the number of good local firms within the same industry as j. If a worker is hired from a HWF in the same industry at time $t-g$, she contributes to H_j^{ind} from year $t-g$ until t.[23] This implies that H^{ind} exhibits a certain degree of persistence and suggests estimation of a dynamic model for the number of workers observed at firm j who have HWF experience in the same industry.

In the second step, I predict the change in H^{ind} that each of the non HWFs

and Strömberg (2007).

[22] The large plants in GHM generated bidding from local governments, almost certainly because there was a belief of important positive effects on the local economy. GHM observe that the mean increase in TFP after the opening is (a) increasing over time and (b) larger if incumbent plants have the same industrial classification as the large plant. These two facts are consistent with the presence of intellectual externalities that are embodied in workers who move from firm to firm. I think of the plants considered by GHM as "good" plants, and in order to simulate their experiment I consider a change in the number of Veneto good firms such that the change in local output is comparable.

[23] It may be instructive to consider a practical example. Consider a worker who joins HWF j in 1995 after separating from a HWF in 1992. If the worker remains in j until 2000, she will contributes to H_j^{ind} count for every year from 1995 to 2000.

in a LLM would experience if an output increase similar to the one considered by GHM were to occur, and I multiply the predicted change in H^{ind} by $\widehat{\beta_H}^{ind}$, the estimated coefficient on H^{ind} in my productivity regression. This product yields the predicted change in productivity due to worker flows for a given Veneto firm if its locality and industry were to experience an increase in output analogous to that considered by GHM.

In the final step, I compare my estimate of the predicted contribution of worker flows to productivity changes with GHM's estimate of the overall productivity effect. This comparison allows me to have a sense of the extent to which worker flows can explain the productivity gains experienced by other firms when high-productivity firms in the same industry are added to a local labor market.

I will now discuss the issues related to the implementation of the first step, i.e. the estimation of the dynamic effect on H_j^{ind} of a change in the number of good firms in the same locality and industry.

8.1 A dynamic model for the number of knowledgeable workers

Consider a model of the form

$$H_{jlst}^{ind} = aH_{jsl,t-1}^{ind} + bGood\_Firms_{ls(j)t} + e_{jlst} \tag{8}$$

$$e_{jlst} = m_j + v_{jlst}$$

$$E[m_j] = E[v_{jlst}] = E[m_j v_{jlst}] = 0 \tag{9}$$

where $Good\_Firms_{ls(j)t}$ is the number of local good firms in the same industry of firm j. Recall that the subscript ind represent workers moving within industry. The disturbance term e_{jlst} has two orthogonal components: the firm effect, m_j and the idiosyncratic shock, v_{jlst}. Using OLS to estimate Equation (8) is problematic because the correlation between $H_{jsl,t-1}^{ind}$ and the firm effect in the error term gives rise to "dynamic panel bias" (Nickell, 1981). Application of the Within Groups estimator would draw the firm effects out of the error term, but dynamic panel bias would remain (Bond, 2002). Therefore I employ the first-difference transform, proposed by Arellano and Bond (1991):

$$\Delta H_{jlst}^{ind} = a\Delta H_{jsl,t-1}^{ind} + b\Delta Good\_Firms_{ls(j)t} + \Delta v_{jlst} \tag{10}$$

The firm effects have now disappeared, but the lagged dependent variable is still potentially endogenous as the $H_{jsl,t-1}^{ind}$ in $\Delta H_{jsl,t-1}^{ind} = H_{jsl,t-1}^{ind} - H_{jsl,t-2}^{ind}$ is correlated with the $v_{jls,t-1}$ in $\Delta v_{jlst} = v_{jls,t} - v_{jls,t-1}$. However, longer lags of the regressors remain orthogonal to the error and are available for use as instruments. Natural candidate instruments for $\Delta H_{jsl,t-1}^{ind}$ are $H_{jsl,t-2}^{ind}$ and $\Delta H_{jsl,t-2}^{ind}$. Both $H_{jsl,t-2}^{ind}$ and $\Delta H_{jsl,t-2}^{ind}$ are mathematically related to $\Delta H_{jsl,t-1}^{ind} = H_{jsl,t}^{ind} - H_{jsl,t-1}^{ind}$ but not to the error term $\Delta v_{jlst} = v_{jls,t} - v_{jls,t-1}$, provided that the v_{jlst} are not serially correlated.[24]

In principle, another challenge in estimating (10) is that firms in a given industry do not select their location randomly. Firms maximize profits and decide to locate where their expectation of the present discounted value of future profits is greatest. This net present value differs across locations depending on many factors, including transportation infrastructure, subsidies, etc. These factors, whose value may be different for firms in different industries, are unobserved, and they may be correlated with ΔH_{jlst}^{ind}. It should be noted, however, that a positive shock in LLM j and industry s such that there is entry of HWFs (i.e. an increase in $\Delta Good\_Firms_{ls(j)t}$) makes it *less* likely that a non HWFs is going to hire from a good firm in the same industry. This is because the shock is *good news* for good firms, so in principle it should make it less likely for the labor force at the good firms to experience a decrease, and in turn, it should make it less likely for a non HWF to

[24] Arellano and Bond (1991) develop a test for autocorrelation in the idiosyncratic disturbance term v_{jlst}. It checks for serial correlation of order l in levels by looking for correlation of order $l+1$ in differences. I employ this test below.

hire from a good firm. The bias introduced by the fact that good firms do not choose their location randomly is therefore likely to be downwards, and thus working against the finding of a positive effect of $\Delta Good\_firms_{ls(j)t}$ on ΔH^{ind}. In any case, $\Delta Good\_Firms_{ls(j)t}$ is treated as endogenous in the estimation.

Table 4 gives the results of estimating Equation (10) for the period 1989-2001.[25]

[TABLE 4 HERE]

Column 1 uses the classic Arellano-Bond Difference GMM estimator and shows a positive (0.004) and significant coefficient of the number of good local firms. This is in line with the idea discussed above of an important role of geographic and economic proximity in determining worker mobility. Column 1 also shows a positive (0.231) and significant coefficient for the lagged dependent variable. The p-value of the Hansen test for overidentifying restrictions does not suggest misspecification. The Arellano-Bond test for serial correlation fails to indicate that the v_{jlst} are serially correlated.

Columns 2 to 4 investigate the robustness of these estimates to different specifications. I begin by using a different transform, proposed by Arellano and Bover (1995), namely the "forward orthogonal deviations" transform.[26] I then estimate the model with two-step GMM and Windmeijer (2005)-corrected cluster-robust errors.[27] Finally I estimate the model with

[25] I include time dummies in order to remove universal time-related shocks from the errors. Since these specifications do not require information collected from AIDA balance sheets, the sample period is not restricted to post-1995 observations.

[26] Rather than subtracting the observation in $t-1$ from the observation in t, the orthogonal deviations transform subtracts the average of all future available observations of a variable. This has the advantage of reducing data loss because, no matter how many gaps, it is computable for each firm. Since I remove all firm-year observations with remarkably high or low values for the number of employees, my estimation panel indeed has some gaps, which are magnified by the first-difference transform. (If some H_{jlst} is missing, for example, then both ΔH_{jlst} and ΔH_{jlst+1} are missing in the first-differenced data.)

[27] See Roodman (2009) for a detailed discussion of two-step GMM and Windmeijer-correction.

two-step GMM, Windmeijer-corrected standard errors and orthogonal deviations. The estimates in Columns 2 to 4 are similar to those in Column 1.

8.2 Simulation Results

Having estimated the dynamic effect on H_j^{ind} of a change in $Good\_Firms_{ls(j)t}$, I can predict the changes in H, and hence in productivity, that a given non-HWF in Veneto would experience after an output increase similar to the one considered by GHM. As it turns out, the large manufacturing plants whose openings are studied by GHM are much larger than the typical good firm in Veneto.[28] In order to observe a change in local output comparable to the typical output increase caused by the opening of one large plant in GHM, a Veneto locality must experience an increase of 56 HWFs. This is the shock in my simulation.

The predicted change in H that each non-HWF would experience after 5 years, the time horizon considered in GHM, is then $\widehat{\Delta H}^{ind,5\ years} = 56 \cdot (b + ab + a^2b + a^3b + a^4b + a^5b)$. This change in H can be obtained using the estimates for a and b from Table 4.

In order to obtain the predicted change in productivity, I first obtain $\widehat{\beta_H}^{ind}$ by estimating Equation (3) after replacing H_j with H_j^{ind}. The results using the different approaches (baseline OLS, OP, LP, polynomial functions of capital and investments or capital and materials in t and $t+1$) are shown in Table A.7. Using the baseline OLS productivity regression, estimated in Column 1 of Table A.7, the predicted change in productivity attributable to worker flows five years the local output increase is equal to $\Delta\widehat{TFP}^{ind,5\ years} = \widehat{\Delta H}^{ind,5\ years} \cdot \widehat{\beta_H}^{ind,OLS} = 0.022$.

The final step is to compare the magnitude of $\Delta\widehat{TFP}^{ind,5\ years}$ with GHM's

[28] This is due both to the fact that new entrants in GHM are significantly larger than the average new plant in the United States and the fact that the Veneto region is characterized by the presence of small and medium-sized businesses, whose size is smaller than the typical firm in United States. See Appendix VI for descriptive statistics.

estimate of the overall productivity effect caused by a local output increase. The increase in productivity estimated by GHM five years after the opening for incumbent plants in the same two-digit industry equals 17 percent. Hence, my back-of-the-envelope calculations suggest that worker flows explain 13.3 percent of the agglomeration advantages estimated by GHM. Replacing $\widehat{\beta_H}^{ind,OLS}$ with $\widehat{\beta_H}^{ind,LP}$, the average effect of recruiting a knowledgeable worker with experience in the same industry estimated in the LP specification (Column 3 of Table A.7), the contribution of worker flows to the agglomeration advantages estimated by GHM is equal to 8.1 percent.

Overall, the results in this section of the paper suggest that worker flows explain an economically relevant proportion of the productivity gains experienced by other firms when HWFs in the same industry are added to a local labor market.

9 Conclusions

Identifying the microeconomic mechanisms underlying localized productivity spillovers is crucial for understanding agglomeration economies. Without knowing the precise nature of the interactions between firms and workers that generate agglomeration advantages, it is difficult to be confident about the existence of any such advantages. Additionally, pinpointing the ultimate causes of agglomeration advantages is helpful for understanding differences in productivity across industry clusters and localities. Finally, better knowledge of the sources of the productivity advantages of agglomeration is important for determining the optimal design of location-based policies.

This paper directly examined the role of labor mobility as a mechanism for the transfer of efficiency-enhancing knowledge and evaluated the extent to which labor mobility can explain the productivity advantages of firms located near other highly productive firms. In order to empirically assess the importance of labor-market based knowledge spillovers, I used Social Security earnings records and detailed financial information for firms from the Veneto region of Italy.

While the issues analyzed in this paper are of general interest, the case of Veneto is important because this region is part of a larger economic area of Italy where, as in the Silicon Valley, networks of specialized small and medium-sized firms, frequently organized in districts, have been effective in promoting and adapting to technological change during the last three decades. This so called "Third Italy" region has received a good deal of attention by researchers, in the United States as well as in Europe (Brusco, 1983; Piore and Sabel, 1984; Trigilia, 1990; Whitford, 2001; Piore, 2009).[29]

The empirical evidence presented using the unique dataset from Veneto points to the concrete possibility that agglomeration of economic activity creates important productivity advantages at the local level. The productivity benefits of a non-HWF from being located in a cluster with a large number of good firms rest with the opportunities to hire workers whose knowledge was gained in good firms. Such knowledge can be successfully adapted internally. More specifically, the regression analysis showed that hiring a worker with HWF experience increases the productivity of other (non-HWF) firms. A simulation indicated that worker flows explain 10-15 percent of the productivity gains experienced by other firms when HWFs in the same industry are added to a local labor market.

[29] Germany's Baden-Wurttermberg is also known "for its mix of small and medium-sized makers of machine tools, textile equipment, and automobile components alongside giant electronics corporations" (Saxenian, 1994, p. 7)

References

Abowd, J.M., F. Kramarz, and D.N. Margolis. 1999. "High-wage Workers and High-wage Firms." *Econometrica*, 67(2): 251–333.

Abowd, J.M., R.H. Creecy, and F. Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." Center for Economic Studies, U.S. Census Bureau Longitudinal Employer-Household Dynamics Technical Papers.

Alfaro, L., and M.X. Chen. 2010. "The Global Agglomeration of Multinational Firms." *The George Washington University Institute for International Economic Policy Working Paper.*

Arellano, M., and S. Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies*, 58(2): 277–297.

Arellano, Manuel, and Olympia Bover. 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of econometrics*, 68(1): 29–51.

Balsvik, R. 2011. "Is Labor Mobility a Channel for Spillovers from Multinationals? Evidence from Norwegian Manufacturing." *The Review of Economics and Statistics*, 93(1): 285–297.

Bond, S.R. 2002. "Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice." *Portuguese Economic Journal*, 1(2): 141–162.

Brusco, S. 1983. "The Emilian Model: Productive Decentralisation and Social Integration." *Cambridge Journal of Economics*, 6(2): 167–184.

Card, D., F. Devicienti, and A. Maida. 2010. "Rent-sharing, Holdup, and Wages: Evidence from Matched Panel Data." National Bureau of Economic Research Working Paper 16192.

Card, D., J. Heining, and P. Kline. 2012. "Workplace Heterogeneity and the Rise of German Wage Inequality." *Mimeograph UC Berkeley.*

Christiansen, B.J., R. Lentz, G. Mortensen, G. Neumann, and A. Werweck. 2005. "Job Separations and the Distribution of Wages."

Journal of Labor Economics, 23: 31–58.

Cingano, F., and F. Schivardi. 2004. "Identifying the Sources of Local Productivity Growth." *Journal of the European Economic Association*, 2(4): 720–744.

Combes, P. P., and G. Duranton. 2006. "Labour Pooling, labour Poaching, and Spatial Clustering." *Regional Science and Urban Economics*, 36(1): 1–28.

Combes, P.P., G. Duranton, L. Gobillon, D. Puga, and S. Roux. 2012. "The productivity advantages of large cities: Distinguishing agglomeration from firm selection." *Econometrica*.

Duranton, G., and D. Puga. 2004. "Micro-foundations of Urban Agglomeration Economies." *Handbook of Regional and Urban Economics*, 4: 2063–2117.

Eberhardt, M., and C. Helmers. 2010. *Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimators.* Department of Economics, University of Oxford.

Eisensee, T., and D. Strömberg. 2007. "News Droughts, News Floods, and U.S. Disaster Relief." *The Quarterly Journal of Economics*, 122(2): 693–728.

Ellison, G., E. L. Glaeser, and W. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review*, 3: 1195–1213.

Gibbons, R., and L.F. Katz. 1991. "Layoffs and Lemons." *Journal of Labor Economics*, 9(4): 351–380.

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

Henderson, J.V. 2003. "Marshall's Scale Economies." *Journal of Urban Economics*, 53(1): 1–28.

Henry, Nick, and Steven Pinch. 2000. "Spatialising knowledge: placing the knowledge community of Motor Sport Valley." *Geoforum*, 31(2): 191–208.

Imbens, G.W., and J.D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(4): 467–476.

Levinsohn, J., and A. Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70(2): 317–42.

Moretti, E. 2004a. "Estimating the External Return to Higher Education: Evidence from Cross-sectional and Longitudinal Data." *Journal of Econometrics*, 120: 175–212.

Moretti, E. 2011. "Local Labor Markets." *Handbook of Labor Economics*, 4: 1237–1313.

Nickell, S. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49: 1417–1426.

Olley, G.S., and A. Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–97.

Parrotta, P., and D. Pozzoli. 2012. "The Effect of Learning by Hiring on Productivity." *The RAND Journal of Economics*, 43(1): 167–185.

Petrin, A., B.P. Poi, and J. Levinsohn. 2004. "Production Function Estimation in STATA Using Inputs to Control for Unobservables." *Stata Journal*, 4: 113–123.

Piore, MJ. 2009. "Conceptualizing the Dynamics of Industrial Districts." In *The Handbook of Industrial Districts*. Cheltenham, Edward Elgar.

Piore, M.J., and C.F. Sabel. 1984. "Italian Small Business Development: Lessons for U.S. Industrial Policy." *American Industry in International Competition. Cornell University Press*, John Zysman and Laura Tyson (eds.).

Roodman, D. 2009. "How To Do Xtabond2: An Introduction to Difference and System GMM in Stata." *Stata Journal*, 9(1): 86–136.

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

Rosenthal, S.S., and W.C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." *Handbook of Regional and Urban Economics*, 4: 2119–2171.

Saxenian, A. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128.* Harvard University Press.

Stoyanov, A., and N. Zubanov. 2012. "Productivity Spillovers Across Firms through Worker Mobility." *American Economic Journal: Applied Economics*, 4(2): 168–198.

Tattara, G., and M. Valentini. 2010. "Turnover and Excess Worker Reallocation. The Veneto Labour Market between 1982 and 1996." *Labour*, 24(4): 474–500.

Tecu, Isabel. 2012. "Knowledge Diffusion and Labor Mobility in Agglomerations: Empirical Evidence from Inventors." PhD diss. Brown University.

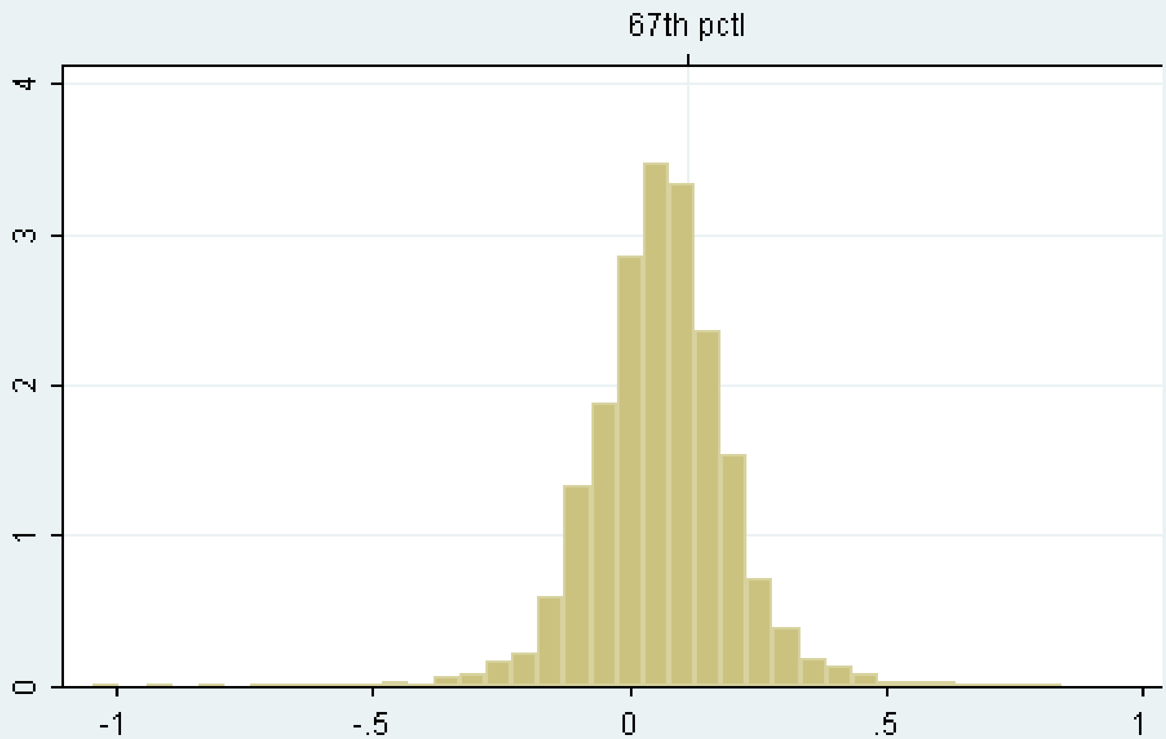
Trigilia, C. 1990. "Work and Politics in the Third Italy's Industrial Districts." *Industrial Districts and Inter-Firm Co-operation in Italy, Geneva: International Institute for Labor Studies*, F. Pyke, G. Becattini and W. Sengenberger (eds.): 160–184.

Whitford, J. 2001. "The Decline of a Model? Challenge and Response in the Italian Industrial Districts." *Economy and Society*, 30(1): 38–65.

Windmeijer, Frank. 2005. "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of econometrics*, 126(1): 25–51.

Yasar, M., R. Raciborski, and B. Poi. 2008. "Production Function Estimation in STATA Using the Olley and Pakes method." *Stata Journal*, 8(2): 221.

Figure 1: *Distribution of Firm Effects*



Firm Effects are obtained using the method in Abowd, Creecy and Kramarz (2002) for the period 1987-2000. I define High-Wage-Firms as those whose estimated firm fixed effects fall within the top tercile of all estimated firm effects

Figure 2: *Distribution of HWFs across Local Labor Markets (LLMs)*

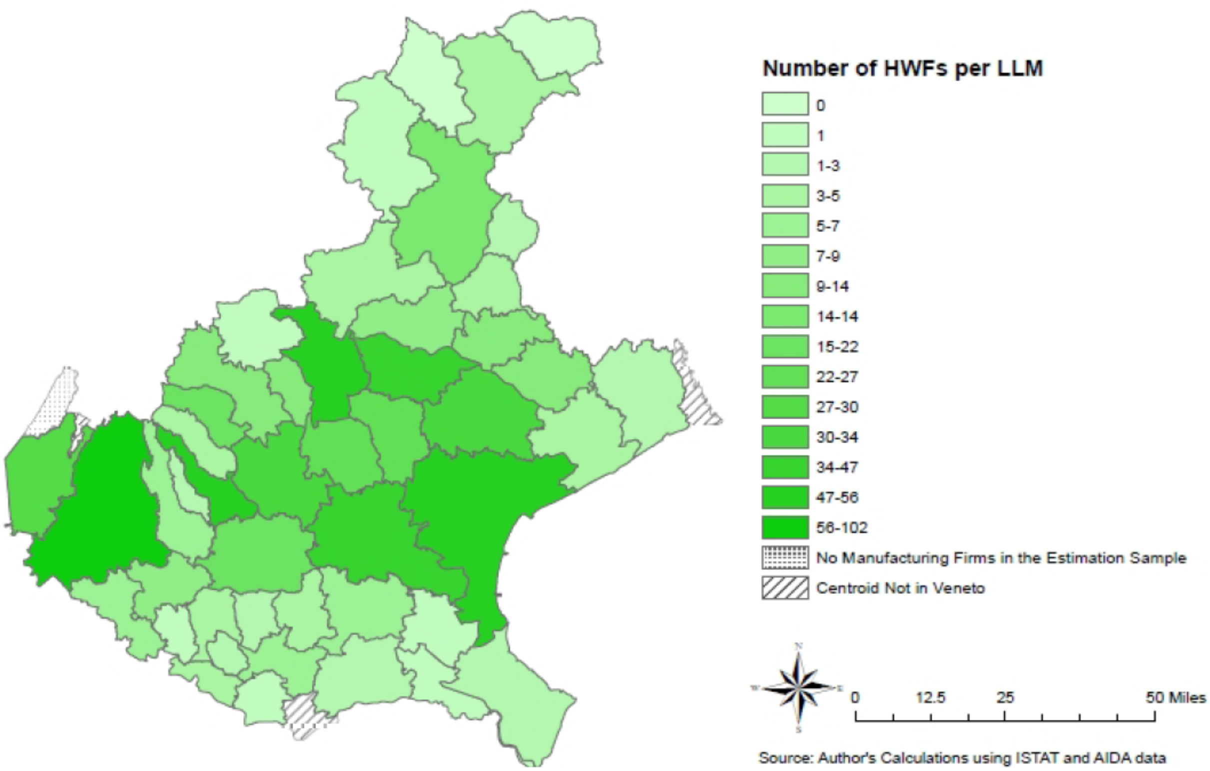


Table 1: Characteristics of HWFs, 1995-2001

	(1) Y/L	(2) VA/L	(3) K/L	(4) Tangible K/L	(5) Intangible K/L
HWF	0.150	0.113	0.104	0.066	0.270
	(0.017)	(0.012)	(0.025)	(0.027)	(0.042)
Observations	26041	26041	26041	26041	26041
Adj. R-squared	0.160	0.106	0.181	0.187	0.0644

Dependent Variables are in logs. All OLS regressions include year and 4-digit industry dummies. Output, Value Added and Capital variables are in 1000's of 2000 euros. Standard errors (in parentheses) clustered by firm. The dummy HWF takes value 1 if the firm is classified as high-wage after estimating the AKM model on the period 1987-2000.

Table 2: H Workers and Productivity in non-HWFs, Main Estimates, 1995-2001

	(1) OLS	(2) OP	(3) LP	(4) Inv-Cap Interact.	(5) Inv-Mat Interact.	(6) Same/Diff Industry
log(capital)	0.097	0.093	0.149	0.000	0.000	0.097
	(0.005)	(0.020)	(0.010)	(0.000)	(0.000)	(0.005)
log(materials)	0.571	0.576		0.591	-3.878	0.571
	(0.008)	(0.009)		(0.014)	(0.510)	(0.008)
log(employees)	0.235	0.235	0.204	0.212	0.181	0.235
	(0.008)	(0.009)	(0.006)	(0.014)	(0.006)	(0.008)
H workers	0.039	0.037	0.020	0.039	0.022	
	(0.008)	(0.011)	(0.006)	(0.020)	(0.009)	
H from same Ind						0.072
						(0.018)
H from diff Ind						0.024
						(0.009)
$\beta_H^{same} = \beta_H^{diff}$,pv						0.018
Observations	17937	6892	17937	3063	14120	17937
Adj. R-squared	0.924			0.930	0.948	0.924

Dependent variable: Log(Output). Standard errors (in parentheses) clustered by firm. H workers is the number of workers with HWF experience currently observed at non-HWFs. Column 1 reports estimates from the baseline specification. Column 2 implements the procedure in Olley and Pakes (1996). Column 3 implements the procedure in Levinsohn and Petrin (2003). Column 4 adds a third-degree polynomial function of log capital and log investment and the interaction of both functions in t and t+1. Column 5 includes the same controls as col. 5 but replaces log investment with log materials. Column 6 differentiates between workers moving within the same industry and between industries. $\beta_H^{same} = \beta_H^{diff}$,pv is the p-value of the equality of coefficients of the variable 'H from same Ind' and the variable 'H from diff Ind'.

Table 3: Knowledgeable Workers and Productivity in non-HWFs, IV Estimates 1995-2001

	(1) Baseline	(2) tradability	(3) localization	(4) 5 percent
H workers	0.268	0.269	0.278	0.231
	(0.154)	(0.154)	(0.145)	(0.164)
log(capital)	0.095	0.095	0.094	0.095
	(0.005)	(0.005)	(0.005)	(0.006)
log(materials)	0.568	0.568	0.568	0.568
	(0.010)	(0.010)	(0.010)	(0.010)
log(employees)	0.227	0.227	0.227	0.229
	(0.010)	(0.010)	(0.010)	(0.010)
Observations	17937	17937	17937	17937
Adj. R-squared	0.908	0.908	0.907	0.910
Fstat, instrum., 1st stage	23.06	23.14	24.41	21.55

Dependent variable: Log(Output). Standard errors (in parentheses) clustered by LLM (47). Regressions include industry-year interaction dummies and LLM-year interaction dummies. Column 1 reports IV estimates using the lagged number of downsizing local good firms in the same 5-digit industry. A good firm is considered as downsizing if the drop in L is larger than 3 percent. The decrease in the labor force must also be greater than or equal to three individuals. Column 2 adds an indicator of the importance of local demand, namely a dummy taking value 1 if the 4-digit industry produces goods that are not widely traded outside the LLM. Column 3 controls for an index of industry localization, namely the ratio between the number of firms in Veneto and total Italian firms in a given 4-digit industry. In Column 4 a good firm is considered as downsizing if the drop in L is larger than 5 percent.

Table 4: Number of local HWFs in same Industry and Knowledgeable Workers moving within industry, GMM Estimates, 1989-2001

	(1) Baseline	(2) Deviations	(3) Two-step	(4) Deviations/ Two-step
lag(H from same Ind)	0.231	0.355	0.150	0.208
	(0.079)	(0.122)	(0.081)	(0.115)
Local HWFs in same Ind	0.004	0.003	0.003	0.002
	(0.002)	(0.001)	(0.001)	(0.001)
Observations	29554	29933	29554	29933
AR(1)z	-6.164	-5.053	-5.244	-4.063
AR(2)z	0.109	0.458	-0.405	-0.237
HansPv	0.272	0.366	0.272	0.366

Dependent variable: 'H from same Ind', the number of H workers moving within Industry. Cluster-robust standard errors in parentheses. Regressions include year dummies. Column 1 reports the baseline Difference GMM results. Column 2 uses the 'forward orthogonal deviations' transform, proposed by Arellano and Bover (1995). Column 3 estimates the model with two-step GMM and Windmeijer-corrected standard errors. Column 4 estimates the model with two-step GMM, Windmeijer-corrected standard errors and orthogonal deviations. The variable 'Local HWFs in same industry' is treated as endogenous. AR(1)z and AR(2)z: Arelanno and Bond (1999) test of first and second order serial correlation, distributed as N(0,1). HansPv: p-value of Hansen test of overidentifying restrictions. For all variables only the shortest allowable lagged is used as instrument.

Appendix

I Data: Sample Restrictions and Descriptive Statistics

I use firm identifiers to match job-year observations for workers aged 16-64 in the VWH with firm financial data in AIDA for the period 1995-2001. The match rate is fairly high: at least one observation in the VHW was found for over 95 percent of the employers in the AIDA sample, and around 50 percent of employees observed in the VWH between 1995 and 2001 can be matched to an AIDA firm. Most of the nonmatches seem to be workers of small firms THAT are omitted from AIDA. In sum, I was able to match at least one employee for around 18,000 firms, or around 10 percent of the entire universe of employers contained in the VWH.[30] From this set of potential matches I execute two exclusions to obtain my estimation sample for Equation (6). First, I remove all workers outside manufacturing. Next, I exclude job-year observations with remarkably high or low values for wages (I trim observations outside the 1 percent - 99 percent range).

The method in Abowd, Creecy and Kramarz (2002) identifies separate groups of workers and firms that are connected via labor mobility in the data. I run the grouping algorithm separately using VHW data from 1987 to 2000 for firms that could be matched in AIDA and have more than 10 employees in VHW. I then use the created group variable to choose the largest group as the sample for my fixed-effects estimation. The largest group contains 99.1 percent of the woker-year observations (2,567,040 observations combining 457,763 individuals with 5,937 firms). I identify HWFs as those firms whose firm effects rank in the top third of the sample.[31] Table A.2 illustrates that,

[30] Card, Devicienti and Maida (2011) show that the average firm size for the matched jobs sample (36.0 workers) is considerably larger than that for total employers in the VWH (7.0 workers). Mean daily wages for the matched observations are also greater, while the fractions of under 30 and female employees are lower.

[31] In order to implement the approach in Abowd, Creecy and Kramarz (2002), I use the a2reg Stata routine developed by Ouazad (2007).

in contrast to firm characteristics, workforce characteristics of HWFs and non HWFs are not so different: the shares of white collar workers and managers are 1.8 and 0.3 percentage points higher, respectively, in HWFs; the share of female workers is 3.1 percentage points lower. No difference emerges in the share of workers younger than 30 or older than 45. [32]

The sample of non HWFs used in the main firm-level analysis – equation (3) - is summarized in Table A.1.[33] The main analysis is performed over the period for which balance sheet data are available (1995-2001). Notice the overlap with the period over which Equation (6) is estimated (1987-2000). In principle one would like to perform the two estimations - Equation (6) and (3) - on two different samples. However, in practice the AKM routine requires a large number of events of labor mobility in order for the firm and worker effects to be identified. Moreover, to precisely estimate $\widehat{\beta_H}$ one would like to exploit as much variation as possible in H, i.e. as many moves from good firms to other firms as possible. Choosing 2000 as the end period for the first estimation seems a good compromise because (a): it guarantees a long enough panel for the AKM estimation, (b) it allows consideration of all the possible moves from good firms to other firms (including in particular workers who separate from good firms in 2000 and are observed in other firms in 2001) and (c) still prevents a full overlapping between the two sample periods for the two different estimations. I experimented with other choices for the period of the AKM estimation, such as 1986-2000 or 1987-1999. Results are very

[32] Notice that since the specifications in Table A.2 do not require information collected from AIDA balance sheets, the sample period is not restricted to post-1995 observations.

[33] In order to obtain this estimation sample I first remove HWF observations from the sample of worker-firm matches. From this non-HWF sample I remove (a) firms that close during the calendar year and (b) firm-year observations with remarkably high or low values (outside the 1% - 99% range) for several key firm-level variables, such as total value of production, number of employees, capital stock and value of materials. (c) firms in LLM with centroids outside Veneto (3 LLMs). I then attempt to reduce the influence of false matches, particularly for larger firms, by implementing a strategy of Card, Devicienti and Maida (2011) to eliminate the "gross outliers", a minor number of matches (less than 1% of all employers) for which the absolute gap between the number of workers reported in a firm's AIDA balance sheet and the number found in the VWH is larger than 100.

similar and available upon request.

II The Extent of Labor Mobility

For labor mobility to be a mechanism for transfer of knowledge, we must observe some workers moving from HWFs to other firms. On average, between 1995 and 2001, 4.3 percent of non HWFs in a given year employ workers with HWF experience. Overall, 1187 workers switch from HWFs to non HWFs during my sample period.[34]

It is important to observe that these numbers do not imply that in a typical year 4.3 percent of Veneto firms are potentially affected by knowledge transfer. Recall that I only consider flows from firms in the top third of estimated firm fixed effects to firms in the bottom third. As a result, these numbers should be interpreted as implying that in a typical year about 4.3 percent of the firms in the bottom third of the distribution employ at least one worker with experience at a firm in the top third. There obviously exists significant labor mobility within the two groups that may also serve as a channel of knowledge transfer. To illustrate, one can intuitively imagine that a worker moving from a firm in the 1st percentile of the distribution to a firm in the 19th percentile may bring efficiency-enhancing knowledge to his or her new job[35], and the same can be imagined for a worker moving from a firm in the 21st percentile to a firm in the 99th percentile. However I focus solely on flows between the two groups.

It is also important to note that the percentage of firms that employ

[34] 787 are blue collar workers, 331 are white collar workers, 46 are managers and 23 are apprentices.

[35] Despite potential lawsuits due to violations of non-compete covenants and trade secret law, one frequently observes top firms poaching employees from competitors in an effort to acquire some of their internal knowledge. This poaching is sometimes so intense that companies may cut deals to refrain from competing for employees. In December 2010, the U.S. Justice Department settled an antitrust suit with Lucasfilm over a "no solicitation" agreement with rival Pixar. In September of the same year, the Justice Department had settled another suit over similar agreements involving Adobe Systems, Apple, Google, Intel, Intuit and Pixar (The New York Times, January 2, 2011).

workers with HWF experience varies with the threshold that I impose on the distribution. For instance, if I define HWFs as firms with fixed effects in the top half of the overall distribution, 8.4 percent of non HWFs employ workers with HWF experience, compared with 4.3 percent if HWFs are defined by falling in the top third of the fixed-effects distribution.

In the main text I show estimates of the extent to which non HWFs benefit from hiring workers from HWFs by entering an annual firm-level measure (H) of the number of workers with experience at HWFs into a production function. Since only a small subest of non HWFs in a given year employ workers with HWF experience, the mean value of H workers across the sample of non HWFs is small (0.071). The maximum value is 7. Notice that the mean number of employees at non HWFs is 48, and the median is 33.

As regards to individual characteristics of the movers in my sample, in all years movers from HWFs are significantly more likely to be young and male than non HWFs workers without experience at good firms. In most years, these movers are also significantly more likely to be white-collar workers and managers. Table A.3 and A.4 give descriptive statistics in the most recent year (2001) for movers from good firms to non HWFs and non HWFs workers without experience at good firms.

III The Role of Geographical Proximity

There exist at least two reasons why geographic proximity might be important for observed worker flows. First, distance may act as a barrier for workers' job mobility because of commuting costs or idiosyncratic preferences for location. Descriptive statistics in Combes and Duranton (2006) show that labor flows in France are mostly local: about 75% of skilled workers remain in the same employment area when they switch firms. The degree of geographical mobility implied by this figure is small, since the average French employment area is comparable to a circle of radius 23 kilometers. In Dal

Bo', Finan and Rossi (2013), randomized job offers produce causal estimates of the effect of commuting distance on job acceptance rates. Distance appears to be a very strong (and negative) determinant of job acceptance: applicants are 33% less likely to accept a job offer if the municipality to which they are assigned is more than 80 kilometers away from their home municipality. The estimates in Manning and Petrongolo (2013) also suggest a relatively fast decay of job utility with distance. Another reason geographical proximity may be an important determinant of job mobility is that the firm's informational cost of identifying the "right" employee are larger across localities than within them. A similar argument can be made for the informational costs for workers.

IV Sensitivity analysis

The main empirical result in the first part of the paper is that labor mobility from HWFs to other firms in the region works as a mechanism for the transfer of efficiency-enhancing knowledge. Table A.6 shows results from a series of specification checks. As a basis for comparison, Column 1 shows the estimates from the baseline specification in Column 1 of Table 2. Considering the differences in observable characteristics documented in Appendix II between movers from HWFs and other workers at non HWFs, in Column 2 I augment Equation (3) with the share of females, managers, blue-collar and white-collar workers, and differently aged workers at each firm. The results largely remained unchanged.

Column 3 shows estimates using the within-transformation. These estimates should be interpreted cautiously because the within estimator is known from practical experience to perform poorly in the context of production functions (Eberhardt and Helmer, 2010). Indeed, estimates in Column 3 indicate severely decreasing returns to scale, likely due to measurement error in the input variables, whose influence is exacerbated by the variable transformation. The problem of using the within-transformation is the removal of considerable information from the data, since only variation over time is left to identify parameters. Setting this concern aside, the results show a positive and significant coefficient on H (0.012) that is smaller than the baseline OLS coefficient, and the coefficients in other specifications reported in Table 2.

Columns 4-5 investigate the role of functional form assumption. Until now, I have presented results based on specifications where the intensity of potential knowledge transferred is measured by the number of H workers. In Column 4, I model this intensity as the share of workers with recent experience at good firms, dividing H by L. The coefficient is positive and significant: a one percentage point increase in h is associated with a change in productivity of 0.8percent.[36] In Column 5 I estimate:

$$\begin{aligned}\ln(Y_{jslt}) &= \beta_0 + \beta_L \ln(\overline{\theta}_{jslt} L_{jslt}) + \beta_K \ln(K_{jslt}) + \beta_M \ln(M_{jslt}) + \beta_{H_l} \log(H_{jslt}) + \\ +\delta 1(H &= 0)_{jslt} + \mu_{st} + \varpi_{lt} + v_{jslt}\end{aligned}$$

Compared to Equation (3) I replaced H_{jslt} with its logarithm, and I imposed $\log(H_{jslt}) = 0$ for the observations with $H_{jslt} = 0$. Plus, I added the dummy $1(H = 0)_{jslt}$ taking value 1 if the number of knowledgeable workers is equal to 0. The results using this alternative functional form are again consistent with those discussed in the main text.

Finally, I address the issue of unobservables related with new hires. If workers who recently changed firms are more productive than stayers, the effect of newly hired workers with HWF experience may equally apply to newly hired employees without HWF experience. In order to explore this possibility I first define medium-wage-firms (MWFs) as those whose estimated firm

[36] Since there may be measurement error in L, the number of employees in the AIDA data, a potential problem with such specification arises. Rewrite equation (3) as $\ln(\frac{Y_{jst}}{\overline{\theta} L_{jst}}) = \beta_K \ln(K_{jst}) + \beta_M \ln(M_{jst}) + \beta_h h_{jst} + \mu_{st} + \varpi_{lt} + v_{jslt}$. Since $h = H/L$, a mechanical relantionship between h and the dependent variable may arise at time t. To address this issue, I also use the lagged share of H workers obtained from head counts in the Social Security dataset. The coefficient estimate (not shown) is 0.650 (0.345).

fixed effects from the AKM model fall between the 33th percentile and the 67th percentile of all estimated firm effects, and low-wage-firms (LWFs) as those whose estimated firm fixed effects fall below the bottom third. I then construct two new variables. The first one, denoted with $\widetilde{H}$ is the number of workers with recent experience at HWFs currently observed at MWF m. I define a worker as having recent HWF-experience in year t, if he or she is observed working in a HWF for one or more of the years $t-5$ to $t-1$. If a worker is hired at time $t-g$, and has experience at a HWF between $t-g$ and $t-5$, she contributes to $\widetilde{H}$ count from year $t-g$ until t. [37] The second variable I construct, denoted with $\widetilde{N}$, is the number of workers with recent experience at LWFs currently observed at MWF m. I then estimate for the sample of MWFs:

$$\begin{aligned}\ln(Y_{mslt}) &= \beta_0 + \beta_L \ln(\overline{\theta}_{mslt} L_{mslt}) + \beta_K \ln(K_{mslt}) + \beta_M \ln(M_{mslt}) + \\ &\quad + \beta_{\widetilde{H}} \widetilde{H}_{mslt} + \beta_{\widetilde{N}} \widetilde{N}_{mslt} + \mu_{st} + \varpi_{lt} + \upsilon_{mslt}\end{aligned}$$

In this specification, the identification of knowledge transfer relies on the differential effect of hiring an employee with recent HWF experience over hiring an employee from a LWF. By including both $\widetilde{H}$ and $\widetilde{N}$, any potential bias caused by the correlation between unobservables and new hires is removed. Column 6 shows the results. The coefficient of $\widetilde{H}$ is positive (.041) and significant. The coefficient of $\widetilde{N}$ is positive but smaller (0.015) and not significant. The difference in productivity effects associated with each type of "movers" is significant at 10 percent level. The productivity effect attributed to knowledgeable workers, therefore, does not appear to be associated with recently hired workers in general. That large productivity gains linked to hiring seem to be realized only when new hires come from more productive firms is consistent with the knowledge spillovers hypothesis.

[37] It may be instructive to consider a practical example. Consider a worker who separates from a HWF in 1992 and joins MWF m in 1995. Provided that the worker remains in j, she will be counted as a knowledgeable worker for every year from 1995 to 1997.

V Non-Tradable Goods

In Column 2 I added a dummy taking value one if the industry produces goods that are not widely traded outside the LLM. Industries for which the dummy takes value one are those classified as SMSA industries by Weiss (1974): Bottled and Canned Soft Drinks and Carbonated, Mineral, and Plain Waters; Fluid Milk; Bread and Other Bakery Products, Except Cookies and Crackers; Manufactured Ice; Primary Forest Products; Newspapers; Commercial Printing (except Lithographic); Commercial Printing (Lithographic); Engraving and Plate Printing; Typesetting; Photo-Engraving; Electrotyping and Stereotyping; Ready-Mix Concrete.

VI Simulation details

Table 1 in GHM reports statistics for the sample of plants whose opening is considered in their study. These plants are quite large: they are more than twice the size of the average incumbent plant and account for roughly nine percent of the average county's total output one year prior to their opening. The mean output (five years after their assigned opening date) is 452,801, 000 of year-2006 dollars, or 395,476,000 of 2000 euros. Standard deviation is 901,690, 000 of year-2006 dollars. As explained in the notes of Table 1 in GHM, these statistics are for a subset of the 47 plant openings studied by the authors. In particular, a few very large outlier plants were dropped so that the mean would be more representative of the entire distribution (those dropped had output greater than half of their county's previous output and sometimes much more).

In order to establish the increase in the number of HWFs that a Veneto locality must experience to observe a change in local output comparable to the output increase caused by the opening of one large plant in GHM, I need to obtain the value of output for a typical HWF. Instead of dropping very large outlier plans as in GHM, I take the median of the distribution. The median value of output for HWFs in my sample is 7,028,000 of year-2000 euros. Therefore a Veneto locality must experience an increase of 395,476,000/7,028,000=56 HWFs. This is the shock in my simulation.

VII Additional References

Dal Bo E., F. Finan and A. Rossi, 2013 "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service", Forthcoming, Quarterly Journal of Economics.

Manning, A. and B. Petrongolo, 2013 "How local are labor markets? Evidence from a spatial job search model", Mimeo, LSE, http://personal.lse.ac.uk/petrongo/.

Ouazad A., 2007 "Program for the Estimation of Two-Way Fixed Effects", Mimeo, LSE, http://personal.lse.ac.uk/ouazad/.

Weiss, L., 1972. “The Geographic Size of Markets in Manufacturing.” Rev. Econ. and Statis. 54 (August): 245–57.

VIII Additional Tables and Figures

Table A.1: non-HWFs, Main Estimation Sample

Variable	Mean	(Std. Dev.)	Min.	Max.	N
Output	8205.547	(9085.215)	1086.012	83537.188	17937
Capital	1829.342	(2400.112)	57.222	20876.002	17937
materials	4148.033	(5403.845)	68.405	47337.867	17937
value added	2088.875	(2293.955)	-4082.134	34466.188	17937
Tangible Capital	1691.601	(2265.701)	2.833	20668	17937
Intangible Capital	137.741	(382.714)	0	11837.857	17937
employees from AIDA	48.069	(47.239)	2	420	17937
employees from VWH	49.173	(45.404)	11	458	17937
apprentices	1.033	(2.004)	0	47	17937
blue collars	30.178	(30.144)	0	348	17937
white collars	9.638	(11.927)	0	251	17937
managers	0.662	(1.855)	0	54	17937
female employees	13.157	(18.874)	0	309	17937
employees age$<$ 30	13.988	(13.499)	0	201	17937
employees age$>$ 45	9.128	(12.893)	0	199	17937
H workers	0.071	(0.302)	0	7	17937
H workers same Ind	0.021	(0.161)	0	5	17937
H workers diff Ind	0.051	(0.245)	0	7	17937
H managers	0.003	(0.053)	0	2	17937
H white collars	0.024	(0.164)	0	3	17937
H blue collars	0.044	(0.231)	0	6	17937

Sample includes 3661 Individual Firms in the period 1995-2001. Output, Capital, Materials, Value Added are in thousands of 2000 euros. Employees from AIDA refers to the values found in the AIDA balance sheet data. Employees from VWH refers to the values obtained from head count in the Veneto Worker History data from Social Security.

Table A.2: Characteristics of HWFs Workforce, 1987-2001

	(1) share white coll.	(2) share manager	(3) share female	(4) share age<30	(5) share age>45
HWF	0.018	0.003	-0.031	0.000	-0.006
	(0.004)	(0.001)	(0.005)	(0.005)	(0.004)
Observations	58102	58102	58102	58102	58102
Adj. R-squared	0.226	0.103	0.569	0.167	0.140

All OLS regressions include year and 4-digit industry dummies. Standard errors (in parentheses) clustered by firm. The dummy HWF takes value 1 if the firm is classified as high-wage after estimating the AKM model on the period 1987-2000.

Table A.3: Characteristics of Knowledgeable Workers observed at non-HWFs, 2001

Variable	**Mean**	**(Std. Dev.)**	**Min.**	**Max.**	**N**
age	33.813	(8.481)	18	62	407
female	0.251	(0.434)	0	1	407
blue collar	0.548	(0.498)	0	1	407
white collar	0.388	(0.488)	0	1	407
manager	0.049	(0.216)	0	1	407

Table A.4: Characteristics of Workers without HWF experience observed at non-HWFs, 2001

Variable	**Mean**	**(Std. Dev.)**	**Min.**	**Max.**	**N**
age	37.08	(9.538)	16	65	192588
female	0.32	(0.467)	0	1	192588
blue collar	0.71	(0.454)	0	1	192352
white collar	0.242	(0.428)	0	1	192352
manager	0.023	(0.15)	0	1	192352

Table A.5: Instrumental Variables, 1995-2001

Variable	**Mean**	**(Std. Dev.)**	**Min.**	**Max.**	**N**
lag (downsizing HWFs, $>$ 3 percent)	0.33	(0.973)	0	7	17937
lag (downsizing HWFs, $>$ 5 percent)	0.307	(0.909)	0	7	17937

The variable 'lag (downsizing HWFs, $>$ 3 percent) ' is the lagged number of downsizing local good firms in a given 5-digit industry. A good firm is considered as downsizing if the drop in L is larger than 3 percent. The decrease in the labor force must also be greater than or equal to three individuals. In constructing the variable 'downsizing HWFs, $>$ 5 percent' a good firm is considered as downsizing if the drop in L is larger than 5 percent.

Table A.6: Knowledgeable Workers and Productivity in non-HWFs, Robustness to Different Specifications

	(1) Baseline OLS	(2) Workforce Characteristics	(3) Within	(4) Share	(5) Log	(6) Recent Experience
log(capital)	0.097	0.093	0.066	0.097	0.097	0.095
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)
log(materials)	0.571	0.561	0.586	0.571	0.571	0.565
	(0.008)	(0.008)	(0.015)	(0.008)	(0.008)	(0.012)
log(employees)	0.235	0.243	0.064	0.238	0.235	0.251
	(0.008)	(0.008)	(0.005)	(0.008)	(0.008)	(0.012)
H workers	0.039	0.034	0.012			
	(0.008)	(0.008)	(0.005)			
share H workers				0.765		
				(0.171)		
log(H workers)					0.066	
					(0.030)	
No H workers					-0.040	
					(0.011)	
Recent HWF exp						0.041
						(0.010)
Recent LWF exp						0.015
						(0.011)
$\beta_{\widetilde{H}}^{HWF} = \beta_{\widetilde{N}}^{LWF}$,pv						0.092
Observations	17937	17937	17937	17937	17937	9269
Adj. R-squared	0.924	0.925	0.985	0.924	0.924	0.932

Dependent variable: Log(Output). Standard errors (in parentheses) clustered by firm. Regressions include industry-year interaction dummies and LLM-year interaction dummies. The variable 'H workers' is the number of knowledgeable workers currently observed at non-HWFs. The variable 'log(H workers)' is the logarithm of number of knowledgeable workers. The dummy 'No H workers' takes value 1 if the number of knowledgeable workers is equal to 0. The variable 'Recent HWF exp' is the number of workers currently observed at Column 1 reports estimates from the baseline specification. Column 2 adds the shares of managers, white collars, blue collars, females, and differently aged workers. Column 3 reports within estimates. Column 4 replaces the number of H workers with the share of H workers. Column 5 replaces the number of H workers with the log of H workers plus the dummy 'No H workers'. Column 6 is estimated on the sample of MWFs and includes workers with recent experience at HWF and Low-Wage-Firms (LWFs). $\beta_{\widetilde{H}}^{HWF} = \beta_{\widetilde{N}}^{LWF}$,pv is the p-value of the equality of coefficients of the variable 'Recent HWF exp' and the variable 'Recent LWF exp'

Table A.7: Knowledgeable Workers with experience in the same industry and Productivity in non-HWFs, 1995-2001

	(1) Baseline OLS	(2) OP	(3) LP	(4) Inv-Cap Interactions	(5) Inv-Mat Interactions
log(capital)	0.098	0.094	0.149	...	...
	(0.005)	(0.021)	(0.010)		
log(materials)	0.571	0.576		0.592	...
	(0.008)	(0.009)		(0.014)	
log(employees)	0.235	0.235	0.204	0.213	0.181
	(0.008)	(0.009)	(0.006)	(0.014)	(0.006)
H workers same Ind	0.073	0.078	0.044	0.094	0.058
	(0.018)	(0.025)	(0.015)	(0.045)	(0.016)
Observations	17937	6892	17937	3063	14120
Adj. R-squared	0.924			0.930	0.948

Dependent variable: Log(Output). Standard errors (in parentheses) clustered by firm. The variable 'H workers same Ind' is the number of workers with HWF experience in the same industry currently observed at non-HWFs. Column 1 reports estimates from the baseline specification. Column 2 implements the procedure in Olley and Pakes (1996). Column 3 implements the procedure in Levinsohn and Petrin (2003). Column 4 adds a third-degree polynomial function of log capital and log investment and the interaction of both functions in t and t+1. Column 5 includes the same controls as col. 5 but replaces log investment with log materials.