NBER WORKING PAPER SERIES

FIRM WAGE EFFECTS

Patrick M. Kline

Working Paper 33084
http://www.nber.org/papers/w33084

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2024, Revised February 2025

This is the first part of a larger chapter on the topic of "wage setting power" that was initially prepared for the Handbook of Labor Economics conference in Berlin, which was generously funded by the Rockwool Foundation Berlin (RFBerlin). Based on discussions with the editors it was determined to be better for pedagogical reasons to publish the two parts as separate chapters. I thank David Card, Raffaele Saggio, Ben Scuderi, Isaiah Andrews, and Sophie Sun for helpful comments on an earlier draft that substantially improved the paper. Jordan Cammarota provided outstanding research assistance on this project. Many of the ideas in this chapter grew out of past conversations with collaborators including David Card, Raffaele Saggio, and Mikkel Sølvsten. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Firm Wage Effects
Patrick M. Kline
NBER Working Paper No. 33084
October 2024, Revised February 2025
JEL No. J3

ABSTRACT

This paper reviews the literature on firm wage differences and the fixed effects methods typically used to measure these differences. High wage firms tend to be more productive, larger, more sought after by workers, and to employ more credentialed and higher wage workers. The latest evidence suggests high wage firms also tend to offer better amenities and are prone to outsourcing and mass layoffs. Reviewing the requirements of the "AKM model" of Abowd, Kramarz, and Margolis (1999), I provide a graph theoretic interpretation of the restrictions this model places on the wage changes of workers who switch employers and examine the extent to which they are satisfied in a benchmark dataset. Assumptions are provided that give these wage changes a causal interpretation and I discuss some difficulties that arise in aggregating them into a global ranking of firm wage levels. In reviewing the econometrics of variance decompositions, I argue that attention ought to focus on effect sizes rather than variance shares, which can be difficult to compare across datasets with different noise levels. Cross-fitting and clustering methods for addressing limited mobility bias are reviewed. A series of bounding and imputation exercises suggest the network pruning typically used in conjunction with cross-fitting methods has little effect on estimands of interest. A review of the latest international evidence finds that the bias-corrected standard deviation of firm effects tends to be substantially elevated in less developed countries. Variance estimation methods for second step regressions of firm effects on covariates are reviewed and illustrated with an empirical application to the firm size wage premium. Finally, I discuss connections between the AKM model and the celebrated sequential auction framework of Postel-Vinay and Robin (2002a), concluding with some areas for future work at this intersection.

Patrick M. Kline
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
pkline@econ.berkeley.edu

Contents

Nearly a century of empirical study supports the view that employers offer different wages for identical work. Fueled by the dissemination of linked employer-employee datasets, a rapidly advancing literature seeks to quantify the role of firms in generating wage inequality using high dimensional fixed effects methods. This paper provides an overview of the literature on firm wage effects, summarizing the evidence base that has been accumulated on which firms pay high wages, their contribution to inequality, and econometric issues that arise in working with models of firm wage fixed effects.

The paper begins with a survey of early empirical investigations of firm and industry components of wage dispersion. Slichter (1950) and Stigler (1962) pioneered the measurement of wage dispersion across employers, providing estimates of the variability of posted wages within narrowly defined job categories and the variability of wage offers within the same worker. Generations later, Krueger and Summers (1988) used the panel structure of large surveys to study the wage changes accompanying worker mobility between industries, concluding that substantial across industry dispersion is present in average pay for equivalent work. These findings renewed interest in deviations from competitive labor market models and foreshadowed many of the economic and econometric debates surrounding the use of fixed effects methods today (Katz et al. 1989; Murphy and Topel 1990; Gibbons and Katz 1992). A related literature on firm size wage premia and intra-industry dispersion documented sizable wage differences across firms and plants in the same industry (Brown and Medoff 1989; Brown, Hamilton, and Medoff 1990; Groshen 1991; Cappelli and Chauvin 1991). Abowd, Kramarz, and Margolis (1999)'s landmark study provided a unified framework for studying these phenomena by applying high dimensional fixed effects methods to matched employer-employee data.

A large empirical literature has refined and extended many of the conclusions from Abowd, Kramarz, and Margolis (1999)'s paper. Five notable patterns stand out from this literature. First, consistent with standard job ladder models, firm wage fixed effects have been found to be positively related to proxies of firm productivity, firm size, and revealed preference measures of firm desirability (Card, Cardoso, and Kline 2016; Bloom et al. 2018; Sorkin 2018; Crane, Hyatt, and Murray 2023). Second, high wage firms tend to employ high wage workers, men, and workers with greater educational attainment (Card, Heining, and Kline 2013; Card, Cardoso, and Kline 2016). Third, firm wage effects are highly temporally persistent (Lachowska, Mas, Saggio, et al. 2023; Engbom, Moser, and Sauermann 2023), and a handful of studies suggest changes in labor market institutions can alter the mix of firm effects in an economy (Card, Heining, and Kline 2013; Dustmann, Lindner, et al. 2022). Fourth, high wage firms are more likely to "fissure", or outsource jobs, and to conduct mass layoffs, both of which may indicate that firms face horizontal equity constraints in wage setting (Goldschmidt and Schmieder 2017; Bertheau, Acabbi, Barceló, et al. 2023). Fifth, the latest research suggests that the most productive firms also provide the best amenities, aligning with revealed preference evidence that high wage firms tend to be more desirable than low wage firms (Sorkin 2018; Lamadon, Mogstad, and Setzler 2022; Sockin 2022; Roussille and Scuderi 2023; Maestas et al. 2023; Lehmann 2023; Caldwell, Haegele, and Heining 2024b).

Delving into the econometric assumptions underlying many of these studies, we review "the AKM model": a two-way fixed effects model of wage determination allowing for unrestricted worker-firm sorting patterns. After discussing the standard identification requirements of two-way fixed effects estimators in matched employer-employee data, a graph theoretic interpretation of the model is introduced where firms serve as vertices and the wage changes between employers constitute directed edges. The restrictions that the AKM model places on these edges are explained, highlighting the special role played by cycles in the mobility network. These restrictions yield a complex mapping between wage changes and firm effects; however, pruning the mobility graph to a spanning tree yields a just-identified set of firm effects with a particularly simple structure. The plausibility of the AKM model restrictions is then evaluated empirically in a benchmark dataset. After accounting for noise in the edge specific wage changes, I find that the least squares estimates provide a remarkably accurate (albeit imperfect) summary of the wage changes associated with moving between particular pairs of firms.

Building on the graph theoretic interpretation, I introduce non-parametric assumptions that endow the wage changes accompanying worker mobility with a causal interpretation. Difficulties arise with aggregating these causal effects into a global ranking of firm wage levels. Least squares estimates of firm effects are shown to rely on "indirect contrasts" involving mobility between other firm pairs than those under consideration, a phenomenon that has been found to also arise in other settings with multiple treatments (Goldsmith-Pinkham, Hull, and Kolesár 2022). Indirect contrasts can be avoided when the network is pruned to a tree but least squares estimates of firm effects do not automatically allow global comparison of wage levels across firms without further assumptions. The section concludes by proposing an assumption that ensures a transitive ranking of firm wage levels and discussing how this assumption might be usefully weakened in future research.

Abowd, Kramarz, and Margolis (1999) proposed a now canonical variance decomposition of log wages into components attributable to worker and firm heterogeneity and sorting. Plugging estimated fixed effects into variance decompositions has long been understood to generate important biases (Krueger and Summers 1988; Andrews et al. 2008). I review approaches to circumventing these biases, including the cross-fitting based bias-correction of Kline, Saggio, and Sølvsten (2020) and recently proposed clustering methods that assume the firm heterogeneity possesses a lower dimensional structure (Bonhomme, Lamadon, and Manresa 2019; Bonhomme, Holzheu, et al. 2023). Cross-fitting approaches require a substantial amount of worker mobility, which researchers typically enforce by pruning to a set of "leave-out connected" firms. A simple approach to bounding the influence of this pruning step on the estimand is proposed and applied to a well known benchmark dataset. I also discuss imputation strategies that can be used to address concerns about biases arising either from pruning or neglected serial correlation. An empirical investigation suggests that the selection biases associated with pruning and serial correlation are likely minimal in large administrative datasets.

Reviewing the empirical literature on bias-corrected variance decompositions, I argue that interest ought to center on the magnitude of these variance components themselves rather than variance shares, which are difficult to compare across datasets with different intrinsic noise levels. Bias-corrected estimates of the economic magnitude of the variability in firm fixed effects are typically sizable, relative both to the dispersion in person effects and to the effect sizes of human capital interventions. A review of recent studies yields estimated standard deviations of firm fixed effects ranging from 15 to 60 log points, with estimates in the US and European countries clustering around 20 log points. In line with a growing literature on labor market misallocation (e.g., Hsieh and Klenow 2009), dispersion in firm effects appears to be most pronounced in the least developed countries. Investigating the factors driving this relationship between dispersion and development is a fruitful area for future research.

A virtue of fixed effects methods is that the estimates can be shared with other research teams who can explore other hypotheses about the relationship between the latent effects and observables. I review the logic of "two-step" regressions of estimated fixed effects on observables, contrasting it with one-step approaches predicated on stronger random effects assumptions. While second step regressions of estimated firm fixed effects on firm and worker level covariates are unbiased, inference is complicated by correlation across the fixed effects estimates, a problem that is well understood theoretically but has largely been ignored in applied work. Kline, Saggio, and Sølvsten (2020) proposed an approach to obtaining heteroscedasticity robust standard errors reflecting the uncertainty stemming from the error underlying the linear fixed effects model. I illustrate this approach with an application to the firm size wage premium, which is found to vary in complex ways across Italian regions. Naive two-step standard errors, of the sort that currently pervade the empirical literature, are shown to significantly understate the true uncertainty present in averages of firm fixed effects in this example.

Finally, I discuss connections between the AKM model and the influential class of search models pioneered by Postel-Vinay and Robin (2002a) and Postel-Vinay and Robin (2002b). While these "sequential auction" models have traditionally been assessed based on their ability to jointly explain job mobility and wage dynamics within firm matches, we discuss the theory's implications for hiring wages. Di Addario et al. (2023) showed that a simple linear specification allowing fixed effects for hiring origins nests the reduced form of hiring wages in the sequential auction model of Bagger, Fontaine, et al. (2014). Dispersion in these hiring origin fixed effects can be viewed as capturing a contribution of search frictions (or equivalently, "luck") to wage inequality. While recent evidence suggests that hiring origins are less influential than these models predict, bilateral competition between firms undoubtedly plays an important role in wage determination for some types of jobs. I discuss the importance for future work of allowing departures from the full information benchmark underpinning canonical variants of this competition framework and conclude with some directions for future research on the econometrics and economics of firm wage setting.

1 Background

Economists have long been aware that employers differ in the pay offered to equivalent workers. Slichter (1950) showed in survey data that the hourly wages of narrowly defined manual occupations varied widely across employers in Boston. Studying industry data from the 1950 Economic Census, he found that industry value added and profits were important drivers of average pay, leading him to conclude that managerial practices were an important determinant of industry pay setting. A decade later, Stigler (1962) collected data on the job offers of business school graduates. In one of the earliest analyses of matched employer-employee data, he documented that within occupational categories, the dispersion of wage offers across companies was of the same order of magnitude as dispersion of wage offers within individual. Moreover, these company pay differences were found to be persistent across years. He concluded from this evidence that wage dispersion for equivalent workers "is of the order of magnitude of 5-10 per cent even in so well organized a market as that of college graduates at a single university." (Stigler 1962, p. 96)

Generations later, Krueger and Summers (1988) examined the extent to which industry differences in pay reflected the sorting of high ability workers to high paying sectors. Using the 1984 Current Population Survey, they fit linear models with worker quality controls and industry fixed effects, finding a bias-corrected standard deviation across two-digit industries of industry fixed effects in wages of 14 log points and a standard deviation in total compensation of roughly 18 log points. To account for unobserved differences in worker quality, they fit longitudinal models to the 1984 displaced workers survey, finding that including worker fixed effects had little impact on estimates of one digit industry fixed effects, suggesting a limited role for selection on unobserved worker quality. Corroborating this view, Gibbons and Katz (1992) found sizable industry wage differentials even after restricting to transitions induced by mass layoffs or plant closures. A large literature debated the interpretation of these findings and whether they can be attributed to compensating differentials, efficiency wages, or employer learning (Katz et al. 1989; Murphy and Topel 1990; Holzer, Katz, and Krueger 1991; Gibbons, Katz, et al. 2005).

Several authors also studied wage differences between establishments and firms of different size (Oi and Idson 1999). Brown and Medoff (1989), Brown, Hamilton, and Medoff (1990), and Oi and Idson

(1999) showed that larger firms, and larger plants within large firms, paid higher wages. Studying worker switches between establishments again confirmed that these differences were generally not attributable to unobserved worker characteristics. Adjustments for workplace amenities were also found to have little impact on the firm size wage premium. Corroborating evidence from Groshen (1991) and Cappelli and Chauvin (1991) documented large wage dispersion across establishments within industry that could not be explained by differences in measured human capital. These intra-industry employer differentials were shown to be comparable in magnitude to inter-industry wage differences.

Seeking to unify these findings, Abowd, Kramarz, and Margolis (1999) – henceforth, AKM – studied employer wage differences in large administrative panels from France and the United States featuring worker and firm identifiers. In what may have been the first high dimensional regression in labor economics, they fit linear models allowing a separate fixed effect for each worker and each firm, along with firm specific trends intended to capture heterogeneity in firm seniority trajectories. AKM found that estimated firm wage effects varied substantially across firms and were correlated with observable measures of firm productivity. However, the estimates suggested that worker and firm fixed effects were only modestly positively correlated and that industry and firm size wage premia were largely accounted for by differences in person effects. Unfortunately, shortly after their study was published, subsequent work revealed that some of these empirical conclusions were artifacts of an inaccurate approximation to the full least squares solution (Abowd, Creecy, and Kramarz 2002; Abowd, Lengermann, and McKinney 2003).

Despite these early stumbles, the work of Abowd, Kramarz, and Margolis (1999) heralded an important transition in empirical labor economics towards interest in the development of econometric methods for the study of matched employer-employee data. While the literature on panel data econometrics traditionally treated fixed effects as nuisance parameters (Chamberlain 1984), AKM viewed these effects as objects of direct interest. This perspective permeates the literature today. Rather than focus attention on the relationship between wages and a handful of observable firm characteristics such as size, sector, or productivity, labor economists now routinely apply fixed effects estimators to enormous administrative datasets in an attempt to "let the data speak" directly about which employers offer high or low wages. The relationship between employer wage fixed effects and low dimensional worker and firm observables can then be scrutinized in a second step, perhaps even by a different research team. While similar transitions from structured to unstructured data analysis have occurred in many other areas of empirical economics – see the chapter in this Handbook by Walters (2024) for some examples – the change has arguably been most dramatic in the literature on wage determination, where it has long been understood that wages vary meaningfully across employers in ways that are difficult to capture with the worker and firm characteristics measured in standard datasets.

2 What sorts of firms pay high wages?

Before delving into the econometrics of fixed effects models, it is useful to provide an overview of what has been learned about the types of firms that offer high wages from empirical research utilizing matched employer-employee data. This body of work has refined our empirical understanding of traditional regularities such as the firm size and industry wage premiums, while also offering new insights into how labor market institutions, outsourcing practices, and job displacement contribute to wage inequality.

2.1 Productivity, worker flows, and firm size

The empirical literature finds that firm wage fixed effects are strongly associated both with observable measures of firm productivity and desirability. AKM's original study documented that firm wage effects were positively correlated with value added per worker and capital share. An updated analysis by Abowd, Kramarz, Lengermann, et al. (2012) utilizing exact least squares solutions finds qualitatively similar patterns in more recent panels of French and US administrative data. Using Portuguese data on hourly wages merged to firm accounting data from Bureau Van Dijk, Card, Cardoso, and Kline (2016) documented that firm wage effects exhibit a "hockey stick" like relationship with log value added per worker, exhibiting a slope of essentially zero at very low levels of value added followed by a nearly constant elasticity relationship at higher levels. Subsequent work documents similar nonlinearities in Germany (Bruns 2019), France (Coudin, Maillard, and Tô 2018), Canada (Li, Dostie, and Simard-Duplain 2023), Hungary (Boza and Reizer 2024), and Italy (Di Addario et al. 2023). Possible explanations for the hockey stick shape include the presence of binding wage floors that prohibit very low firm effects, the existence of a "competitive fringe" of less productive firms that engage in essentially competitive wage setting, and non-classical measurement error in value added per worker.

Sorkin (2018) devised a revealed preference measure of firm desirability based on the idea that a desirable firm hires workers from other desirable firms. The proposed measure, which is motivated by a wage posting model in the spirit of Burdett and Mortensen (1998), involves applying the Google PageRank algorithm (Page et al. 1999) to the network of job to job flows. Sorkin (2018) reports that his measure of firm desirability exhibits a correlation of roughly 0.54 with firm wage effects derived from quarterly earnings in Longitudinal Employer Household Dynamics (LEHD) data. Crane, Hyatt, and Murray (2023) also use LEHD data to show that firm wage fixed effects are strongly positively related to the "poaching rank" index of Bagger and Lentz (2019), which provides another revealed preference measure of firm desirability consistent with a class of sequential auction models that will be discussed below.

Firm wage fixed effects have been shown to be positively related to firm size and negatively related to quit rates (Card, Heining, and Kline 2013; Bassier, Dube, and Naidu 2022). Bloom et al. (2018) study the changing nature of the firm size wage premium by fitting separate fixed effects models to the US Social Security Administration's Master Earning File in each of three time periods: 1980-1986, 1994-2000, and 2007-2013. In the first period, firm wage fixed effects are monotonically increasing in firm size, with an enormous 55 log point gap in average firm effects between companies with 15,000 or more employees and those with 1-10 employees. In later periods, the relationship between wages and firm size grows more concave. In the final 2007-2013 sample, monotonicity appears to break down, with mean firm fixed effects estimated to be slightly higher among firms with 1,000-2,500 employees than at the largest firms. The pay gap between the largest and smallest firms falls to roughly 22 log points in this period. To date, little evidence is available regarding whether similar transitions have occurred in other countries.

2.2 Entry, reallocation, and dynamics

The distribution of firm effects has been shown to respond to changes in labor market institutions. Card, Heining, and Kline (2013) fit separate models to four overlapping 6-7 year intervals of German data spanning the period from 1985 to 2009. They find that the variance of firm wage effects roughly doubles over the course of their study. Most of the growth in dispersion of firm effects occurs in the latter two intervals, a period that saw a rapid liberalization of the German labor market. Analyzing cohorts of firms, they find that within cohort inequality in firm wage effects is roughly stable over time but newer cohorts of firms are more unequal.[1] Tying these cohort trends to the breakdown of the German collective bargaining system, they document that firms not covered by bargaining agreements are more likely to exhibit very low wage fixed effects.

Song et al. (2019) conduct a similar "rolling AKM" analysis making use of US social security records over the period 1978-2013. While they find that inequality increased dramatically across firms over this period, firm effect variances were surprisingly stable, suggesting the rise in between firm inequality was a consequence of increased worker-firm sorting. This discrepancy between the German and US results may have to do with differences in the institutional environment of these labor markets. The US has enjoyed a relatively stable regulatory environment over the period studied by Song et al. (2019), while post-unification Germany faced enormous pressure on its sectoral bargaining system that plausibly paved the way for the entry of very low wage firms (Dustmann, Fitzenberger, et al. 2014). Dustmann, Lindner, et al. (2022) show that the enactment of a German minimum wage led low wage workers to reallocate to firms with higher wage fixed effects, and that German regions differentially exposed to the minimum wage hike experienced an increase in the average AKM fixed effect of surviving establishments.

[1] Sorkin and Wallskog (2023) find a similar pattern in US data, albeit without controlling for person effects.

The temporal stability of the firm effect variances among cohorts of German firms documented by Card, Heining, and Kline (2013) suggests that firm wage effects are persistent. Lachowska, Mas, Saggio, et al. (2023) used hourly wage data derived from Washington state UI records to measure this persistence more carefully. They estimate unrestricted firm fixed effects over pairs of adjacent years, yielding a sequence of fixed effects for each firm. Fitting an AR1 model to these estimates, they find a bias-corrected autocorrelation of firm wage effects of 0.98. Contemporaneous work by Engbom, Moser, and Sauermann (2023) finds that projecting firm wage effects derived from 8 year intervals onto the effects derived from pooling 32 years of Swedish wage data yields a slope of roughly 0.95, suggesting that wage fixed effects are highly stable among long lived firms.

2.3 Sorting, outsourcing, and displacement

High wage firms employ high wage workers. This pattern has been repeatedly documented in the form of positive bias-corrected correlations between worker and firm fixed effects (Andrews et al. 2008; Kline, Saggio, and Sølvsten 2020; Bonhomme, Holzheu, et al. 2023). However, the pattern is usually evident (albeit attenuated) from uncorrected estimates fit to population level administrative records. Card, Heining, and Kline (2013) and Song et al. (2019) both find that the uncorrected correlation between worker and firm fixed effects has increased in recent decades. Observable worker characteristics are also predictive of firm effects. Low wage firms tend to disproportionately employ women (Card, Cardoso, and Kline 2016), immigrants (Dostie et al. 2023), minorities (Gerard et al. 2021), younger workers (Kline, Saggio, and Sølvsten 2020), and workers with lower educational attainment (Card, Heining, and Kline 2013). Low wage firms are also typically intensive in jobs involving low wage occupations (Card, Heining, and Kline 2013; Goldschmidt and Schmieder 2017) and tend to exhibit less complex job hierarchies (Huitfeldt et al. 2023).

Goldschmidt and Schmieder (2017) find that German firms with high wage fixed effects are more likely to outsource workers in food services, cleaning, security, and logistics (FCSL) occupations. One interpretation of this pattern is that firms face horizontal equity constraints making it difficult to tailor wages to the match surplus of individual workers. Consistent with this view, they estimate separate firm fixed effects for FCSL and non-FCSL workers at each employer and find that firms paying 10% higher wages to non-FCSL workers tend to pay FCSL worker roughly 8% higher wages (Goldschmidt and Schmieder 2017, Figure A-8).[2] Rather than share a large wage premium with workers at all layers of the organization, firms tend to spin off jobs lying outside their area of core competency in order to economize on wage costs. In the wake of an outsourcing event, measured as a setting where many FCSL workers move from a "mother" establishment to

[2] Conducting a similar exercise in Argentine data, Drenik et al. (2023) find that firms paying regular workers 10% higher wages pay temporary workers roughly 5% higher wages.

the same "daughter" establishment specializing in FCSL services, the wage of outsourced workers plummet. This drop turns out to be almost entirely explained by the low fixed effects of establishments specializing in FCSL services. Goldschmidt and Schmieder (2017) argue that the growth of firms specializing in FCSL services is an important driver of German inequality consistent with the firm cohort patterns documented by Card, Heining, and Kline (2013).

In line with the German evidence on outsourcing, Lachowska, Mas, and Woodbury (2020) document in Washington state UI records that firms in the top quintile of firm fixed effects account for a disproportionate share of displaced workers. However, they find that 70% of displaced workers move to employers with similar or better firm effects despite suffering wage losses. As a result, they estimate that firm fixed effects account for only 17% of the earnings losses associated with job displacement; however, this share rises to roughly two thirds among the workers who move to lower wage employers upon displacement. Schmieder, Von Wachter, and Heining (2023) find in German administrative records that nearly all of the average daily wage losses associated with displacement are explained by differences in firm effects. Bertheau, Acabbi, Barceló, et al. (2023) study a harmonized panel of seven European countries and find that between 35% (in Spain) and 100% (in Portugal) of the daily wage losses of job displacement after five years are explained by the loss of firm fixed effects. In a longer working paper (Bertheau, Acabbi, Barcelo, et al. 2022), they conjecture that this variation across countries may be attributable to the intensity of active labor market policies, which they show turns out to strongly predict the magnitude of country specific wage losses. Like Lachowska, Mas, and Woodbury (2020), Bertheau, Acabbi, Barceló, et al. (2023) find in all seven countries that job displacement is most common among firms with estimated firm fixed effects in the top quintiles.

2.4 Industry structure and amenities

A headline finding of Abowd, Kramarz, and Margolis (1999)'s original study was that industry wage differentials are largely explained by person effects. This conclusion turned out to have been driven by the computational method used in their analysis to approximate the least squares solution in the largest samples of firms (Abowd, Creecy, and Kramarz 2002). Subsequent analysis of early LEHD data from four states found substantial differences in average firm effects across sectors (Abowd, Lengermann, and McKinney 2003, Table 11). Sorkin (2018, Table V) finds in a broader LEHD dataset comprised of large employers in 27 states that four digit industry codes account for roughly 45% of the variation in firm fixed effects.

More recently, Card, Rothstein, and Yi (2024) analyze LEHD data covering all 50 states for the years 2010-2018. They find that roughly one third of the variance in firm wage effects is explained by four digit NAICS industry codes. Remarkably, the average industry premiums are nearly identical for workers who have, and have not, obtained a college degree. They estimate that the highest paying industry is coal mining, while the lowest paying industry is drinking places. Perhaps surprisingly, their industry wage premia estimates turn out to be positively correlated with production function based estimates of industry wage markdowns from Yeh, Macaluso, and Hershbein (2022), which may indicate that variation in industry averages of firm wage effects reflect productivity more than market power.

Card, Heining, and Kline (2013) find in German data that between industry dispersion of firm effects rose between 1985 and 2009 and that high wage workers increasingly sort to high wage industries. In contrast, Haltiwanger, Hyatt, and Spletzer (2024), fitting AKM models to three intervals of LEHD data covering the period 1996-2018, find that the contribution of industry averages of firm wage fixed effects to wage inequality has been relatively stable. Like Card, Heining, and Kline (2013), however, they find that the sorting of high wage workers to high wage industries increased substantially.

Sorkin (2018, Table V) reports that nearly half of the variation in his flows based measure of firm desirability is between 4 digit industries. He argues that elevated wages in sectors such as mining primarily reflect compensating differentials. Relating the firm wage effects to measures of firm desirability, he concludes that as much as two thirds of the variation in firm wage fixed effects could reflect compensating differentials. Subsequent work by Lamadon, Mogstad, and Setzler (2022) concurs that compensating differentials are an important determinant of firm wage fixed effects; however, they also find that high wage firms tend to have the best amenities. This view is corroborated by Sockin (2022), who documents that higher wage firms list more job amenities in job advertisements. Likewise, Maestas et al. (2023) find that adjusting for valuations of observed amenities derived from stated preference experiments actually widens inter-industry wage differentials.

An emerging consensus is that the most desirable firms tend to offer both the highest wages and the best amenities, making firms with large wage fixed effects highly desirable on average. Roussille and Scuderi (2023) provide revealed preference evidence from an online job board for software engineers that higher wage firms offer better observed and unobserved amenities. Similar conclusions are reached by Lehmann (2023) and Lagos (2024) utilizing administrative records from Austria and Brazil, respectively. Caldwell, Haegele, and Heining (2024b) provide survey evidence from German workers that perceptions of the wages available at other firms are strongly correlated both with firm effect estimates from administrative data and with workers' perceptions of the non-wage amenities at those firms. See Mas (2024) for a comprehensive analysis of the recent literature on compensating differentials.

3 The AKM model

The fixed effects model considered by Abowd, Kramarz, and Margolis (1999) can be written:

$$Y_{it} = \alpha_i + \psi_{\mathbf{j}(i,t)} + X_{it}'\beta + \varepsilon_{it}, \tag{1}$$

where Y_{it} is the logarithm of worker i's wages in year t and $\mathbf{j}(i,t) \in \{1,\ldots,J\} \equiv [J]$ is a function returning the identity of the firm employing worker i in year t. In their original application to an unbalanced panel of French administrative data, J was on the order of five hundred thousand, two million workers were studied, and the panel consisted of roughly five million person-year observations. Subsequent work has considered much larger samples. For instance, Song et al. (2019) fit models with over 79 million person effects and 5.8 million firm effects to a five year panel with 220 million person-year observations. To avoid notational clutter, it will be useful to restrict attention to the case where the panel is balanced in what follows such that $t \in \{1,\ldots,T\} \equiv [T]$.

The person effect α_i is a portable component of wages that a worker can take with them to other employers. This parameter can capture skills, as well as a worker's reputation, bargaining prowess, or discrimination at the market level. The firm effect ψ_j is a non-portable component of wages enjoyed only when a worker is employed at firm j. This effect can be a function of both the firm's productivity, some of which is shared with the worker in the form of higher wages, and its unobserved amenities, which may yield compensating differentials. The firm effect may also reflect the degree to which effort is monitorable at the firm, which can generate variation in efficiency wages (Shapiro and Stiglitz 1984; Akerlof and Yellen 1990). The vector X_{it} includes year fixed effects and measures of labor market experience.[3]

The time varying error ε_{it} captures innovations to the portable component of the worker's wage along with any measurement errors. These errors are assumed to obey a strict exogeneity restriction, requiring that $\mathbb{E}[\varepsilon_{it}|\mathbf{j}(i,s) = j, X_{is} = x] = 0$ for all workers $i \in \{1,\ldots,N\} \equiv [N]$, all time periods $(s,t) \in [T]^2$, and all possible firm assignments $j \in [J]$ and covariate values $x \in \mathcal{X}$. From a statistical perspective, ε_{it} provides the "noise" that creates slippage between firm effect estimates and the true fixed effects. Thinking carefully about how to account for this noise is the core contribution of much of the recent econometrics literature studying these models.

The strict exogeneity condition embeds both the requirement that worker mobility between firms is not driven by time varying wage fluctuations (often described as "exogenous mobility") and that the mapping from worker and firm heterogeneity to expected log wages is additively separable. However, it does not

[3] See Card, Cardoso, Heining, et al. (2018) for discussion of identification issues posed by introducing age and year effects. In their original study, Abowd, Kramarz, and Margolis (1999) included firm specific seniority trends, which introduces additional identification challenges that I will not consider here.

restrict, in any way, the joint distribution of worker and firm effects. Therefore, workers may sort to firms based on any function of their own α_i and the vector ψ of firm wage effects. The pairing of (1) with the strict exogeneity restriction has come to be known as "the AKM model" and I will follow convention in using this eponym as a shorthand. It is worth noting, however, that closely related assumptions are now employed in several literatures exploiting the switching of units between groups (e.g., Finkelstein, Gentzkow, and Williams 2016; Chetty and Hendren 2018).

In the AKM model, movements between firms reveal differences in firm wage setting. In the case where $T = 2$, for any two firms $j \neq k$ between which workers move, we have

$$\mathbb{E}\left[Y_{i2} - Y_{i1} \mid \mathbf{j}(i,1) = j, \mathbf{j}(i,2) = k, X_{i1}, X_{i2}\right] = \psi_k - \psi_j + \left(X_{i2} - X_{i1}\right)' \beta. \tag{2}$$

As Abowd, Creecy, and Kramarz (2002) detail, the firm effect levels are only identified up to a constant within the largest "connected set" of employers: that is, the set of firms connected, directly or indirectly, via worker moves. Intuitively, if there are two collections of firms between which workers never move, the difference in their wage levels will not be identified. A single restriction on the firm effects – typically a normalization that one of them is zero – within each connected set is required for the design matrix of worker and firm dummies to have full rank, enabling least squares estimation of (1).

In the German social security records analyzed by Card, Heining, and Kline (2013), the largest connected set captured around 97-98% of person year observations depending on the period analyzed. These shares can be lower when studying subpopulations. For example, fitting models separately by gender to Portuguese data, Card, Cardoso, and Kline (2016) find that the largest connected set comprises 88% of person-year observations for male workers and 91% of observations for female workers. In both settings, the wage distributions and worker characteristics in the largest connected set tend to be similar to those in the broader population.

Our discussion so far of the connectedness and normalization requirements for estimation of the firm effects has been a bit vague. The next subsection delves deeper into these subjects by providing a graph theoretic interpretation of the AKM model. I focus there on the properties of the *mobility network,* defined as a directed graph where vertices correspond to firms and edges represent worker moves between firms. At the cost of some additional notation, this network based lens will allow us to develop an interpretation of the AKM model as a restricted model of "edge effects." This interpretation motivates a corresponding representation of the least squares estimator of firm effects as a linear combination of estimated edge effects. A closely related representation was explored by Jochmans and Weidner (2019). My exposition differs from theirs primarily in clarifying how the presence of cycles in the mobility network influence the algebraic mapping between the wage changes of movers and the firm fixed effects estimates. Section 3.2 investigates the extent to which the restrictions motivating the firm fixed effects estimator are satisfied in a benchmark dataset. Section 3.3 discusses causal interpretations of edge and firm effects, concluding with some directions for future research.

3.1 An edgy interpretation of firm effects

We begin with some definitions. A *graph* is a collection of vertices and edges joining those vertices. The graph we are considering is *directed*, which means that each edge starts at one vertex and ends at another. Here, the vertices correspond to the set of firms $[J]$. An *edge* is an ordered pair of vertices $(j,k) \in [J]^2$, with the first entry denoting an origin firm from which a worker moved and the second entry denoting the destination of the move. To simplify the analysis, we will continue to assume $T = 2$, in which case the set of all edges in the graph can be defined as:

$$E = \left\{ (j,k) \mid (j,k) \in [J]^2, j \neq k, \sum_{i \in [N]} 1\left\{\mathbf{j}(i,1) = j, \mathbf{j}(i,2) = k\right\} > 0 \right\}.$$

Denoting the total number of edges by $|E|$, I will index the edges by $\ell \in \{1, \ldots, |E|\} \equiv [E]$, referring to individual edges by $\{e_\ell\}_{\ell \in [E]}$.

A *walk* is a sequence of edges that join a set of firms. A *trail* is a walk with no repeated edges. A *path* is a trail with no repeated firms. The mobility graph is *connected* if there is a path from any firm to any other firm. A *tree* is a connected graph for which there is a unique path between any pair of firms. A *spanning tree* is any subset of a connected graph that contains all firms and is a tree.

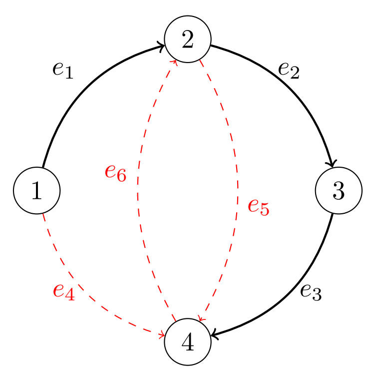

Figure 1: A mobility network $(J = 4, |E| = 6)$

Figure 1 depicts a connected graph with four firms and six edges. Arrows indicate the directions in which workers move between firms. A spanning tree of this network is given by the solid edges. The dashed edges depart from the tree by generating alternative paths of moving between firms. These alternate paths yield *cycles:* that is, trails that lead us back to where we started. For example, using a minus sign to denote traversal of an edge in reverse, the trail $\{e_1, e_2, e_3, -e_4\}$ is a cycle. A *fundamental cycle* is a cycle formed by departing from the spanning tree using a single edge not in the tree. There are $|E| - J + 1$ distinct fundamental cycles in a connected graph. The other fundamental cycles in this graph are $\{e_2, e_3, e_6\}$ and $\{e_2, e_3, -e_5\}$.

The *incidence matrix* $\boldsymbol{B}$ provides a mathematical representation of the graph's edges.[4] Every row of $\boldsymbol{B}$ represents a firm, while every column represents an edge. A single entry in each column equals 1, denoting that edge's destination firm, and a single entry equals -1, capturing that edge's origin firm. The remaining entries equal zero. In the graph above $\boldsymbol{B}$ takes the form:

	e_1	e_2	e_3	e_4	e_5	e_6
1	-1	0	0	-1	0	0
2	1	-1	0	0	-1	1
3	0	1	-1	0	0	0
4	0	0	1	1	1	-1

An important property of $\boldsymbol{B}$, to which we will return, is that its rows are orthogonal to the graph's cycles. For instance, the cycle $\{e_1, e_2, e_3, -e_4\}$ can be represented by the vector $c_1 = [1, 1, 1, -1, 0, 0]'$. Likewise, the cycles $\{e_2, e_3, e_6\}$ and $\{e_2, e_3, -e_5\}$ are captured by the vectors $c_2 = [0, 1, 1, 0, 0, 1]'$ and $c_3 = [0, 1, 1, 0, -1, 0]'$ respectively. It is easy to verify that $\boldsymbol{B}c_1 = \boldsymbol{B}c_2 = \boldsymbol{B}c_3 = 0$. More generally, $\boldsymbol{B}c = 0$ for any $|E| \times 1$ vector c in the linear span (also known as the "cycle space") of the fundamental cycles. For example, the trail $\{e_5, e_6\}$, which can be represented by $c_2 - c_3$, is in this graph's cycle space. In a connected graph, the cycle space is the *nullspace* of $\boldsymbol{B}$, meaning it contains the set of all vectors $c \in \mathbb{R}^{|E|}$ such that $\boldsymbol{B}c = 0$.

3.1.1 Firm effects as restricted edge effects

Returning to (2), we can now rewrite the AKM model in a notation directly linked to the structure of the graph. To simplify the analysis, suppose that the vector β is known and define $R = Y_2 - Y_1 - (\boldsymbol{X}_2 - \boldsymbol{X}_1)\,\beta$ as the $N \times 1$ vector of worker wage changes adjusted for the change in time varying covariates. Let $\boldsymbol{F}_t$ denote the $N \times J$ matrix of firm assignment indicators in period t, the i'th row of which can be written

[4] Jochmans and Weidner (2019) work with a weighted definition of the incidence matrix. I rely here on an unweighted definition in order to highlight connections to cycles in the graph. Weights are introduced below in section 3.1.2.

$(1\{\mathbf{j}(i,t)=1\}, 1\{\mathbf{j}(i,t)=2\}, \ldots, 1\{\mathbf{j}(i,t)=J\})$. The AKM model implies

$$R = (\boldsymbol{F}_2 - \boldsymbol{F}_1)\,\psi + \varepsilon,$$

where $\psi = (\psi_1, \ldots, \psi_J)'$ is the $J \times 1$ vector of firm fixed effects and $\varepsilon = (\varepsilon_{12} - \varepsilon_{11}, \ldots, \varepsilon_{N2} - \varepsilon_{N1})'$ is the $N \times 1$ vector of differences in wage errors obeying $\mathbb{E}[\varepsilon \mid \boldsymbol{F}_1, \boldsymbol{F}_2] = 0$.

We can write the matrix of first differenced firm indicators in terms of the edge dummies via the relation $\boldsymbol{F}_2 - \boldsymbol{F}_1 = \boldsymbol{E}\boldsymbol{B}'$ where $\boldsymbol{E}$ is an $N \times |E|$ matrix of (directed) edge indicators – i.e., dummies of the form $1\{\mathbf{j}(i,2)=k\} \cdot 1\{\mathbf{j}(i,1)=j\}$ for all origin-destination firm pairs (j,k) traversed by movers. Hence, the AKM model is equivalently expressed in terms of the incidence matrix as

$$R = \boldsymbol{E}\boldsymbol{B}'\psi + \varepsilon.$$

Here, strict exogeneity can be represented as the requirement that $\mathbb{E}[\varepsilon \mid \boldsymbol{E}] = 0$.

It is instructive to contrast the AKM model with a model of unrestricted edge fixed effects:

$$R = \boldsymbol{E}\Delta + u,$$

where Δ is an $|E| \times 1$ vector of edge effects and the $N \times 1$ error vector u obeys $\mathbb{E}[u|\boldsymbol{E}] = 0$. Section 3.3 introduces assumptions giving these edge effects a causal interpretation. The AKM model imposes $\Delta = \boldsymbol{B}'\psi$, which entails $|E| - J + 1$ linear restrictions on the edge effects. When these restrictions hold, the two error terms are identical ($u = \varepsilon$). Hence, the AKM model can be thought of as projecting the $|E| \leq J(J-1)$ edge effects down to only $J-1$ linearly independent firm effects.

To understand the nature of the edge restrictions entailed by the AKM model, note that for any vector c in the cycle space, $c'\Delta = c'\boldsymbol{B}'\psi = 0$, which follows from the cyclic orthogonality properties of $\boldsymbol{B}$ discussed earlier. For example, the AKM model imposes that wage changes should be symmetric across origin firm - destination firm pairs, a property that was emphasized by Card, Heining, and Kline (2013) and is reflected in our example of the cycle $c_2 - c_3$. However, the AKM restrictions go far beyond pairwise symmetry, restricting network dependent tuples of edge effects. For example, the fundamental cycle c_1 involves four edges. Though directly visualizing the restrictions pertaining to such *4-cycles* is challenging, their logic mirrors the restrictions pertaining to the *2-cycles* studied by Card, Heining, and Kline (2013): that "taking a walk" along the graph should have no effect on wages so long as one ends up back at the same firm where the walk began.

It may be helpful here to illustrate this reasoning with a simple thought experiment. Consider two workers of the same age, both of whom are employed at firm j in 2010, where they earn the same wage. In subsequent years, each worker switches employers twice before returning to firm j in 2020. The AKM model stipulates that we should expect these two workers to earn the same wages in 2020, regardless of the identity of their two intermediate employers. Indeed, if our two time periods were 2010 and 2020, these workers would be viewed as "stayers" and equation (2) predicts their wage change depends only on the change in time varying covariates.

Mathematically, these cyclic restrictions exhaust the empirical restrictions of the AKM model on edge effects within a connected set of firms. That is, if $c'\Delta = 0$ for any cycle in the graph space, then there must exist a set of firm effects capable of rationalizing the edges exactly. To understand why, recall that $\boldsymbol{B}$'s nullspace coincides with the cycle space of the graph, which implies we can decompose the edge effects as

$$\Delta = \boldsymbol{B}'\dot{\psi} + \boldsymbol{C}\dot{\eta},$$

where $\dot{\psi}$ is a vector of coefficients from a linear projection of Δ onto $\boldsymbol{B}'$, $\boldsymbol{C}$ is an $|E| \times |E| - J + 1$ matrix collecting the graph's fundamental cycles, and $\dot{\eta}$ is an $|E| - J + 1$ vector of "cycle effects" that serve as residuals. Plugging this decomposition into the edge effects model yields,

$$R = \boldsymbol{E}\boldsymbol{B}'\dot{\psi} + \boldsymbol{E}\boldsymbol{C}\dot{\eta} + u.$$

The AKM model amounts to assuming that $\dot{\eta} = 0$, in which case $\dot{\psi} = \psi$ and $\varepsilon = u$. When there are no cycle effects, then the true dimension of the edge effects is much lower than it appears: the AKM model reduces the $|E|$ edges to $J - 1$ linearly independent firm effects.

While $|E| - J + 1 = 3$ in the graph depicted in Figure 1, large scale empirical applications can feature hundreds of thousands (or even millions) of restrictions. As with any economic or statistical model, these restrictions are unlikely to be satisfied exactly. When the restrictions do not hold, the firm effects can be thought of as a linear projection that provides a lower dimensional summary of the edge effects. We will examine the quality of this summary in section 3.2.

3.1.2 Estimators

Let $\hat{\Delta} = \left(\boldsymbol{E}'\boldsymbol{E}\right)^{-1}\boldsymbol{E}'R$ denote the $|E| \times 1$ vector of estimated edge effects. The normal equations defining the least squares estimator of ψ can be written

$$\boldsymbol{B}\boldsymbol{W}\hat{\Delta} = \boldsymbol{L}\psi,$$

where $\boldsymbol{W} = \boldsymbol{E}'\boldsymbol{E}$ is a diagonal weighting matrix recording the number of workers moving along each edge and $\boldsymbol{L} = \boldsymbol{B}\boldsymbol{W}\boldsymbol{B}' = (\boldsymbol{F}_2 - \boldsymbol{F}_1)' (\boldsymbol{F}_2 - \boldsymbol{F}_1)$ is a symmetric $J \times J$ matrix known in graph theory as the *Laplacian.* The Laplacian encodes information about each's firm's role in the mobility network. The jth row and kth column of $\boldsymbol{L}$ equals the negative of the total number of workers moving (in either direction) between firms j and k when $j \neq k$, while the jth diagonal entry of $\boldsymbol{L}$ gives the total number of workers moving to or from firm j.

$\boldsymbol{L}$ is singular, which implies there are an infinite number of solutions to the normal equations. Jochmans and Weidner (2019) study the properties of the solution $\boldsymbol{L}^{\dagger}\boldsymbol{B}\boldsymbol{W}\hat{\Delta}$, where $\boldsymbol{L}^{\dagger}$ denotes the Moore-Penrose inverse of $\boldsymbol{L}$. I will take a slightly different approach by studying the solution that results when one of the firms is taken as the "reference firm" with zero firm effect. While both solutions yield the same predicted edge effects, the reference firm solution is typically used in practice and happens to also simplify the subsequent theoretical analysis. Bozzo (2013) provides some useful results on connections between the two approaches.

Define $\boldsymbol{B}_{(1)}$ as the submatrix leaving out the first row of $\boldsymbol{B}$ and let $\boldsymbol{L}_{(11)} = \boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{B}'_{(1)}$ denote the submatrix of $\boldsymbol{L}$ leaving out its first row and column. If we impose the restriction $\psi_1 = 0$, then we obtain the constrained normal equations

$$\boldsymbol{B}_{(1)}\boldsymbol{W}\hat{\Delta} = \boldsymbol{L}_{(11)}\psi_{(1)},$$

where $\psi_{(1)}$ is ψ omitting its first entry. A classic result in graph theory, *Kirchhoff's matrix tree theorem*, states that any cofactor of the unweighted Laplacian matrix gives the number of spanning trees in the graph. When the edges are weighted, generalizations of the theorem (e.g., Spielman 2019, Theorem 13.4.1) establish that any cofactor of $\boldsymbol{L}$ gives the total edge weight of the graph's spanning trees, where the weight of each tree is given by the product of the edge weights it contains. A connected graph must have at least one spanning tree. Hence, when the mobility graph is connected, it follows that $\det\left(\boldsymbol{L}_{(11)}\right) > 0$, implying that $\boldsymbol{L}_{(11)}$ has full rank.

The least squares estimator that results from treating the first firm as the reference can therefore be written

$$\hat{\psi}_{(1)} = \boldsymbol{L}_{(11)}^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}\hat{\Delta}. \tag{3}$$

Variants of this estimator are heavily used in applied research; however, computation is typically implemented by iterative conjugate gradient (CG) methods rather than direct inversion of $\boldsymbol{L}_{(11)}$.[5] CG routines are available in most scientific computing packages including MATLAB and SciPy. The efficiency of these routines is greatly aided by "preconditioning" the problem with an approximate Cholesky factorization of $\boldsymbol{L}_{(11)}$. In

[5] We have glossed over the issue of how to form R – and consequently $\hat{\Delta}$ – in the first place. Typically, one estimates the coefficient vector β on the time varying covariates X_{it} in a first step and subtracts them off. This initial adjustment step is also greatly accelerated with CG methods.

the empirical examples below, I rely on the combinatorial multigrid solver package of Koutis, Miller, and Tolliver (2011) as a preconditioner.

3.1.3 Combination weights and smoothing

Equation (3) reveals that the estimated firm effects are linear combinations of the average wage changes associated with each edge. In general, the combination weights are such that each firm effect can depend on each element of $\hat{\Delta}$. For example, when the edges in the graph depicted in Figure 1 each represent a single mover, the firm effect estimates can be written:

$$\hat{\psi}_{(1)} = \left(\boldsymbol{B}_{(1)}\boldsymbol{B}'_{(1)}\right)^{-1}\boldsymbol{B}_{(1)}\hat{\Delta} = \begin{pmatrix} \frac{7}{12} & -\frac{1}{12} & -\frac{1}{12} & \frac{5}{12} & -\frac{1}{6} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{5}{12} & \frac{1}{12} & \frac{1}{12} & \frac{7}{12} & \frac{1}{6} & -\frac{1}{6} \end{pmatrix}\hat{\Delta}. \tag{4}$$

Recall that $\boldsymbol{B}_{(1)}c = \boldsymbol{0}$ for any vector c in the graph's cycle space. It is easy to verify in the above example that perturbing $\hat{\Delta}$ by adding to it any vector $c \in \{c_1, c_2, c_3\}$ yields no change in the estimated firm effects $\hat{\psi}_{(1)}$. This cyclic invariance property can be thought as offering a form of robustness to certain types of confounding trends in the error ε. For example, a trend shared by the movers traversing edges e_5 and e_6 (i.e., movers between between Firms 2 and 4) will "difference out." Likewise, $\hat{\psi}_{(1)}$ is unaffected by adding a constant to the wage changes of the movers traversing each of the edges e_2, e_3, and e_5.[6] Adding a constant to the wage changes of movers on edges e_2 and e_3 while subtracting that constant from movers on e_5 also has no effect.

Whether confounding cyclic trends of this nature tend to be present in economic data is an interesting question for future research. Mobility cycles are common among employers in the same industry and region. Suppose the error ε takes the form $\varepsilon = \boldsymbol{C}\eta + u$ where η is a vector of cycle effects driven by demand shocks to those industry-regions. Though this error structure violates the AKM edge restrictions, unweighted firm effect estimates remain unbiased because $\left(\boldsymbol{B}_{(1)}\boldsymbol{B}'_{(1)}\right)^{-1}\boldsymbol{B}_{(1)}\hat{\Delta} = \psi + \left(\boldsymbol{B}_{(1)}\boldsymbol{B}'_{(1)}\right)^{-1}\boldsymbol{B}_{(1)}u$.

An important simplification of (3) arises in the case where the mobility graph is a tree. By definition, a tree has J vertices and $J-1$ edges, which implies the submatrix $\boldsymbol{B}_{(1)}$ is square. Recall that $\boldsymbol{L}_{(11)} = \boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{B}'_{(1)}$. Since $\boldsymbol{L}_{(11)}$ has full rank, $\boldsymbol{B}_{(1)}$ must also have full rank. Hence, we can write:

$$\hat{\psi}_{(1)} = \left(\boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{B}'_{(1)}\right)^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}\hat{\Delta} = \left(\boldsymbol{B}'_{(1)}\right)^{-1}\boldsymbol{W}^{-1}\boldsymbol{B}_{(1)}^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}\hat{\Delta} = \left(\boldsymbol{B}'_{(1)}\right)^{-1}\hat{\Delta}.$$

[6] When the number of movers differs across edges in a cycle, then the magnitude of a trend shared across the cycle's edges would need to be inversely proportional to the number of movers along each edge in order to difference out. That is, the firm effect estimates become invariant to perturbing $\hat{\Delta}$ in the direction $\boldsymbol{W}^{-1}c$ where c is a vector in the cycle space. However, fitting the AKM model directly to the edge effects by unweighted least squares (i.e., setting $\boldsymbol{W} = \boldsymbol{I}$ in estimation) restores invariance to cycle specific trends.

The predicted edge effects implied by the estimated firm effects are given by $\boldsymbol{B}'_{(1)}\hat{\psi}_{(1)} = \hat{\Delta}$, indicating that the firm effects rationalize the adjusted wage changes $\hat{\Delta}$ with no error. This phenomenon reflects that the firm effects are just-identified by (i.e., "they saturate") the edge specific wage changes.

Consider the spanning tree depicted in Figure 1, which is comprised of the graph's first three edges. The $\boldsymbol{B}'_{(1)}$ associated with this tree and its inverse are depicted below:

$$\boldsymbol{B}'_{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}, \qquad \left(\boldsymbol{B}'_{(1)}\right)^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

In any spanning tree, one can always ensure that $\left(\boldsymbol{B}'_{(1)}\right)^{-1}$ is triangular by ordering the edges of $\boldsymbol{B}_{(1)}$ according to their distance from Firm 1. However, some of the entries in such a triangle may possess a negative sign if reaching the reference firm requires traversing an edge in reverse.[7] The triangular structure of $\left(\boldsymbol{B}'_{(1)}\right)^{-1}$ ensures that each firm effect is simply the sum of the (oriented) edge effects on the path connecting it to the reference firm. Consequently, the difference in firm effect estimates for any two firms j and k connected by an edge must equal the average wage change of the workers moving directly between them. We will return to this property when discussing causal interpretations of firm effects.

A closely related property can be shown to hold when the graph is a *polytree*, meaning that the undirected graph is a tree but some firm pairs may be connected by edges in both directions. For example, adding an edge from Firm 4 to Firm 3 to the spanning tree depicted in Figure 1 yields a polytree. Any polytree can be transformed into a simple tree by transferring the weight from one edge to the other in each pair of edges connecting the same firms. This transformation can be represented by an $|E| \times J - 1$ matrix $\boldsymbol{T}$ that differences the relevant edge pairs in the incidence matrix. For example,

$$\underbrace{\begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}}_{\boldsymbol{B}} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}}_{\boldsymbol{T}} = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 2 \end{pmatrix}.$$

Remarkably, the firm effect estimates are invariant to such transformations. Specifically, when $\boldsymbol{B}$ represents a polytree, the least squares weights for the transformed graph $\left(\boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{W}\boldsymbol{B}'_{(1)}\right)^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{W}$ equal

[7] For example, choosing Firm 3 as the reference in this spanning tree yields $\left(\boldsymbol{B}'_{(3)}\right)^{-1} = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

the weights $\boldsymbol{L}_{(11)}^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}$ for the untransformed graph.[8] Since $\boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{T}$ is a square invertible matrix representing a simple tree, the firm effects derived from fitting the AKM model to a polytree can equivalently be written $\left(\boldsymbol{T}'\boldsymbol{W}\boldsymbol{B}_{(1)}'\right)^{-1}\boldsymbol{T}'\boldsymbol{W}\hat{\Delta}$. Consequently, in any polytree, the difference in estimated firm effects between any pair of firms joined by a pair of edges will equal a mover weighted average of the two oriented edge effects connecting them. As in a simple tree, the difference in firm effects for any pair of firms joined by a single edge will depend only on that estimated edge effect.

When the graph is not a tree, the firm effects become over-identified and the vector of predicted wage changes is $\tilde{\Delta} \equiv \boldsymbol{B}_{(1)}'\hat{\psi}_{(1)} = \boldsymbol{H}\hat{\Delta}$, where $\boldsymbol{H} = \boldsymbol{B}_{(1)}'\boldsymbol{L}_{(11)}^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}$ is an $|E| \times |E|$ weighted projection matrix that is invariant to the choice of reference firm. Like the usual "hat" matrix (Hoaglin and Welsch 1978), $\boldsymbol{H}$'s diagonal entries $\{h_{\ell\ell}\}_{\ell\in[E]}$ give the *leverage* of each observation (in this case each edge effect) on the predicted value. One can write the ℓ'th leverage:

$$h_{\ell\ell} = b_\ell'\boldsymbol{L}_{(11)}^{-1}b_\ell n_\ell,$$

where b_ℓ is the ℓ'th column of $\boldsymbol{B}_{(1)}$ and n_ℓ is the ℓ'th diagonal entry of $\boldsymbol{W}$. In large systems, costly inversion of $\boldsymbol{L}_{(11)}$ can be avoided by breaking computation into a CG step that solves a linear system and a subsequent matrix multiplication step.[9] Leverages lie in the interval $[0, 1]$, with larger values indicating that dropping that edge from the data would lead to a greater change in the estimated firm effects. Any edge that is part of a cycle has $h_{\ell\ell} < 1$. An edge with $h_{\ell\ell} = 1$ is known as a *bridge*. Dropping a bridge breaks the graph into two or more connected components, in which case at least one firm effect can no longer be estimated.

When the graph is a tree, all edges are bridges and $\boldsymbol{H}$ is the identity matrix. However, when the graph exhibits cycles, $\boldsymbol{H}$ departs from identity and some "smoothing" across edges takes place. The rows of $\boldsymbol{H}$ give the smoothing weights used to form the prediction for each edge. Each row's weights sum to one but the entries can be negative. For example, if we assume a single mover traverses each edge of the graph depicted

[8]The transformation $\boldsymbol{T}$ maps the edges back into the span of the weighted incidence matrix, implying the weighted orthogonality condition $\boldsymbol{T}'\boldsymbol{W}\left(\boldsymbol{I} - \boldsymbol{B}_{(1)}'\boldsymbol{L}_{(11)}^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}\right) = 0$. Expanding this condition yields $\boldsymbol{T}'\boldsymbol{W} = \boldsymbol{T}'\boldsymbol{W}\boldsymbol{B}_{(1)}'\boldsymbol{L}_{(11)}^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}$. Pre-multiplying by $\boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{T}$ gives $\boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{W} = \boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{W}\boldsymbol{B}_{(1)}'\boldsymbol{L}_{(11)}^{-1}\boldsymbol{B}_{(1)}\boldsymbol{W}$. Dividing both sides by $\boldsymbol{B}_{(1)}\boldsymbol{W}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{W}\boldsymbol{B}_{(1)}'$ yields the result.

[9]Note that we can rewrite the ℓ'th leverage $h_{\ell\ell} = b_\ell' z_\ell$, where $z_\ell = \boldsymbol{L}_{(11)}^{-1}b_\ell n_\ell$. The first step solves the equation $\boldsymbol{L}_{(11)}z_\ell = b_\ell n_\ell$ for the vector z_ℓ via CG methods. The second step computes $h_{\ell\ell} = b_\ell' z_\ell$ by vector multiplication. This process can be parallelized across edges to recover all of the leverages.

in Figure 1 then the hat matrix takes the following form:

$$\boldsymbol{H} = \begin{pmatrix} \frac{7}{12} & -\frac{1}{12} & -\frac{1}{12} & \frac{5}{12} & -\frac{1}{6} & \frac{1}{6} \\ -\frac{1}{12} & \frac{7}{12} & -\frac{5}{12} & \frac{1}{12} & \frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{12} & -\frac{5}{12} & \frac{7}{12} & \frac{1}{12} & \frac{1}{6} & -\frac{1}{6} \\ \frac{5}{12} & \frac{1}{12} & \frac{1}{12} & \frac{7}{12} & \frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Inheriting the properties of $\boldsymbol{B}_{(1)}$, these smoothing weights are orthogonal to any vector in the cycle space but are otherwise widely dispersed across the edges. Placing weight on edges throughout the network is efficient when the AKM model restrictions hold. Otherwise, $\tilde{\Delta}$ may provide a poor estimate of Δ.

3.2 Evaluating the AKM restrictions

To evaluate whether the AKM model provides an accurate summary of the wage changes associated with (directed) moves between firm pairs, we study two years of the Veneto Workers History (VHW) data. This dataset has emerged as a popular benchmark in the literature due to the low barriers associated with obtaining access to it.[10] We work with an extract of 1,859,459 person-year observations from the years 1999 and 2001 that was studied previously by Kline, Saggio, and Sølvsten (2020). The largest connected set contains $73,933$ firms and $747,205$ workers, $197,572$ of whom switch employers between the two years. These $197,572$ "movers" are spread across $150,417$ edges. Hence, the AKM model implies $76,485$ restrictions on the edge effects.

The AKM model is fit to the log daily wage changes of workers by solving the normal equations using MATLAB's preconditioned conjugate gradient routine. The only time varying covariate included is an indicator for the year being 2001. Job stayers contribute to the firm effect estimates only indirectly via estimation of the year fixed effect $\hat{\beta}$. We use this same year effect estimate to preadjust wage changes before collapsing them to estimated edge effects $\hat{\Delta}$.

3.2.1 Visualizing goodness of fit

Figure 2 summarizes how the conditional distribution of estimated edge effects varies with the AKM predictions. Each dot depicts the mean edge effect within a bin of predicted edge effects $\left(\tilde{\Delta}\right)$. The bands around the dots give a sense of dispersion within each bin: the upper limit of each band gives the 75th percentile of

[10]The data can be requested at `https://www.frdb.org/en/dati/dati-inps-carriere-lavorative-in-veneto/`.

estimated edge effects in that bin, while the lower limit gives the 25th percentile.

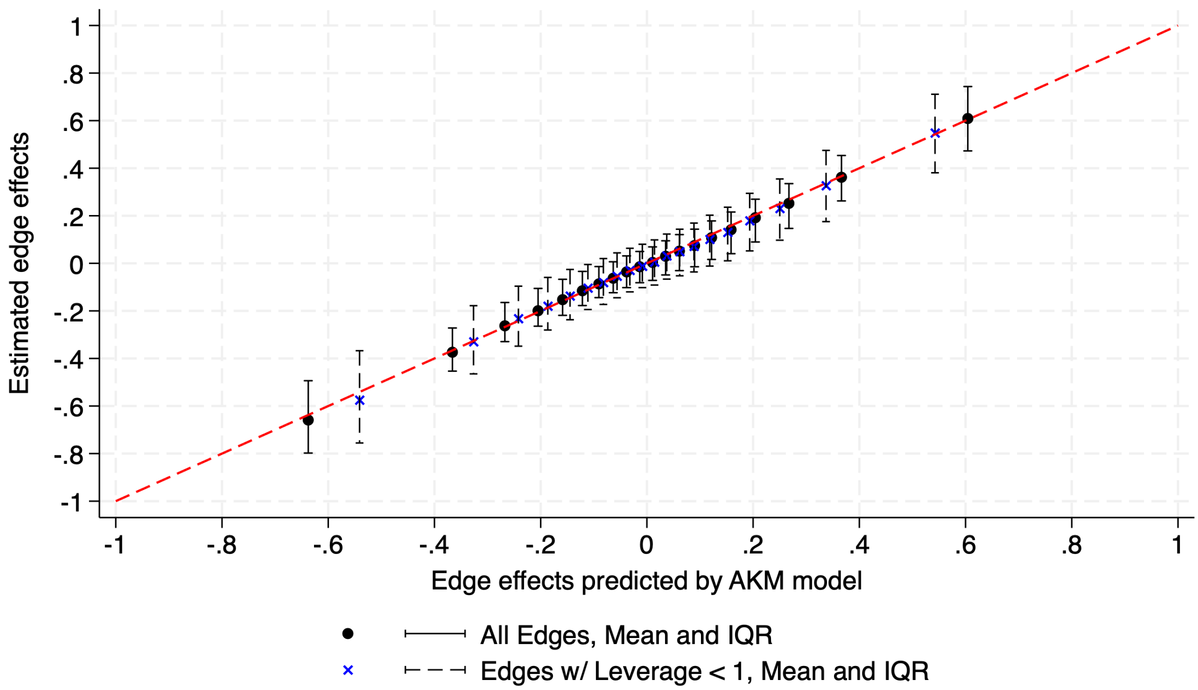

Figure 2: Log daily wage change of edge $\left(\hat{\Delta}\right)$ versus AKM prediction $\left(\tilde{\Delta}\right)$

Notes: The vertical axis depicts binned averages of the elements of $\hat{\Delta}$: the average adjusted log daily wage changes associated with each origin-destination firm edge. The horizontal axis gives bins of $\tilde{\Delta}$: the wage change predicted by the least squares estimates of firm effects. Panel comprised of the 1999 and 2001 waves of the Veneto Work Histories dataset developed by the Economics Department in Università Ca' Foscari Venezia under the supervision of Giuseppe Tattara.

The AKM model stipulates that, in the absence of noise, the dots should all lie on the dashed 45 degree line. On average, the edge effects do tend to lie remarkably close to the 45 degree line. Moreover, the bands around the dots reveal only modest dispersion around the averages. However, the AKM model was fit to the same data as the edge effects, which induces a mechanical dependence between the two sets of estimates. Indeed, if the graph had been a tree, the edge predictions would all lie exactly on the 45 degree line.

Looping over edges to compute the leverages $\{h_{\ell\ell}\}_{\ell\in[E]}$ reveals that about 22% of the edges are bridges that must mechanically lie on the 45 degree line. Roughly 44% of the firm effects are just-identified by one of these bridges. Dropping the bridges leaves 117,657 edges with $h_{\ell\ell} < 1$ that connect 41,195 firms. The x's in Figure 2 depict the mean predictions in this subpopulation, which still track the 45 degree line closely. However, the interquartile range of deviations is amplified. To evaluate whether these deviations are larger than we should expect under the AKM model requires accounting for noise in the estimated edge effects.

3.2.2 Accounting for noise

The noise in the edge effects that concerns us derives from the vector u of wage change errors. One can think of these errors as capturing the idea that if a different worker happened to traverse the same edge, a different wage change would likely result. In what follows, I will use the expectation and variance operators $\mathbb{E}_u[\cdot]$ and $\mathbb{V}_u[\cdot]$ to convey that integration is ultimately being conducted with respect to the edge effects error u introduced in section 3.1.1. Hence, the expected value of the AKM prediction is $\mathbb{E}_u\left[\tilde{\Delta}\right] = \boldsymbol{H}\Delta$ and the variance matrix of the estimated edge effects is $\mathbb{V}_u\left[\hat{\Delta}\right] = \left(\boldsymbol{E}'\boldsymbol{E}\right)^{-1}\boldsymbol{E}'\mathbb{E}\left[uu'\right]\boldsymbol{E}\left(\boldsymbol{E}'\boldsymbol{E}\right)^{-1}$.

Denote the vector of differences between the predicted and estimated edge fixed effects by $\hat{\Delta} - \tilde{\Delta} = \boldsymbol{M}\hat{\Delta}$, where $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{H})$ is the "residual maker" matrix. Let $\hat{\Delta}_\ell$ denote the ℓ'th entry of $\hat{\Delta}$ and $\tilde{\Delta}_\ell$ the ℓ'th entry of $\tilde{\Delta}$. A standard goodness of fit statistic is the sum of squared residuals. We will work with a mover-weighted version of this statistic: $\sum_\ell n_\ell\left(\hat{\Delta}_\ell - \tilde{\Delta}_\ell\right)^2 = \left(\hat{\Delta} - \tilde{\Delta}\right)'\boldsymbol{W}\left(\hat{\Delta} - \tilde{\Delta}\right) = \hat{\Delta}'\boldsymbol{M}'\boldsymbol{W}\boldsymbol{M}\hat{\Delta}$.[11] So long as the wage change errors have finite variance, we can write the expectation of this sum as

$$\mathbb{E}_u\left[\hat{\Delta}'\boldsymbol{M}'\boldsymbol{W}\boldsymbol{M}\hat{\Delta}\right] = \underbrace{\Delta'\boldsymbol{M}'\boldsymbol{W}\boldsymbol{M}\Delta}_{\text{squared bias}} + \underbrace{trace\left(\boldsymbol{M}'\boldsymbol{W}\boldsymbol{M}\mathbb{V}_u\left[\hat{\Delta}\right]\right)}_{\text{noise}}.$$

The AKM model stipulates that $\boldsymbol{M}\Delta = 0$, which implies $\Delta'\boldsymbol{M}'\boldsymbol{W}\boldsymbol{M}\Delta = \sum_\ell n_\ell\left(\Delta_\ell - \mathbb{E}_u\left[\tilde{\Delta}_\ell\right]\right)^2 = 0$. However, the model doesn't restrict the trace term, which captures the expected contribution of noise. If the wage change errors are independent across movers, then $\mathbb{V}_u\left[\hat{\Delta}\right]$ is a diagonal matrix and the trace expression simplifies to

$$trace\left(\boldsymbol{M}'\boldsymbol{W}\boldsymbol{M}\mathbb{V}_u\left[\hat{\Delta}\right]\right) = \sum_{\ell\in[E]} n_\ell\left(1 - h_{\ell\ell}\right)\mathbb{V}_u\left[\hat{\Delta}_\ell\right],$$

where $\mathbb{V}_u\left[\hat{\Delta}_\ell\right]$ is the ℓth diagonal entry of $\mathbb{V}_u\left[\hat{\Delta}\right]$. This formula captures the intuition that high leverage edges are expected to yield smaller residuals because of overfitting. Conversely, edges with higher noise levels $\mathbb{V}_u\left[\hat{\Delta}_\ell\right]$ should yield larger squared residuals.

For edges with more than a single mover, a simple unbiased estimator of $\mathbb{V}_u\left[\hat{\Delta}_\ell\right]$ is available: the squared standard error, $\widehat{\mathbb{V}}_u\left[\hat{\Delta}_\ell\right] = \frac{1}{n_\ell}s_\ell^2$, where s_ℓ is the standard deviation of adjusted wage changes along edge ℓ. Reflecting the sparsity of the mobility network, only 9,459 of the edges that are not bridges have 2 or more movers. Denote this set of edges by $\mathcal{E}_{2+}$. Combining the leverages with the edge specific standard errors yields an expected sum of squared residuals under the null hypothesis that the AKM model holds of $\sum_{\ell\in\mathcal{E}_{2+}} n_\ell\left(1 - h_{\ell\ell}\right)\widehat{\mathbb{V}}_u\left[\hat{\Delta}_\ell\right] = 207.58$.

[11] The residual maker matrix will not, in general, be symmetric when the number of movers varies across edges. Fortunately, $\boldsymbol{M}'\boldsymbol{W}\boldsymbol{M} = \boldsymbol{W}\boldsymbol{M}$ for any distribution of mover weights, which significantly simplifies the calculations below.

Empirically, the residual sum of squares is $\sum_{\ell\in\mathcal{E}_{2+}} n_\ell \left(\hat{\Delta}_\ell - \tilde{\Delta}_\ell\right)^2 = 360.51$. The difference, 360.51-207.58=152.92, between the actual and expected sum of squares provides an unbiased estimate of the sum of squared approximation errors: $\sum_{\ell\in\mathcal{E}_{2+}} n_\ell \left(\Delta_\ell - \mathbb{E}_u\left[\tilde{\Delta}_\ell\right]\right)^2$. A natural benchmark for these approximation errors is the (mover-weighted) sum of squared edge effects $\sum_{\ell\in\mathcal{E}_{2+}} n_\ell \Delta_\ell^2$, an unbiased estimate of which is $\sum_{\ell\in\mathcal{E}_{2+}} n_\ell \left(\hat{\Delta}_\ell^2 - \widehat{\mathbb{V}}_u\left[\hat{\Delta}_\ell\right]\right) = 978.67$. The ratio of these two numbers can be thought of as one minus the (uncentered) R^2 from an infeasible mover-weighted regression of the true edge effects in $\{\Delta_\ell\}_{\ell\in\mathcal{E}_{2+}}$ onto the matrix of first differenced firm dummies.[12] Hence, the data suggest the AKM approximation captures roughly $(1 - 152.92/978.67) \times 100 \approx 84\%$ of the variation in true edge effects. Equivalently, the estimated correlation between the edge effects and the AKM predictions is 0.92.

Table 1: Goodness of fit by edge sample

Sample	**Number of Movers**	**Number of Edges**	**Root Mean Squared Model Error**	**Root Mean Squared Edge Effect**	$\mathbf{R^2}$	**Root Mean Noise Level**
At least 2 movers	50,254	9,459	0.055	0.140	84.37	0.124
At least 3 movers	39,048	3,856	0.041	0.117	87.58	0.092
At least 4 movers	34,941	2,487	0.038	0.109	87.95	0.076
At least 1 mover						
Singleton noise level twice edges w/ 2 movers	158,452	117,657	0.119	0.225	71.85	0.195
Singleton noise imputed via log-log regression	158,452	117,657	0.112	0.219	73.86	0.204
Including bridges						
Singleton noise imputed via log-log regression	197,572	150,417	0.100	0.232	81.38	0.200

Notes: All samples but the last are comprised of edges that are not bridges. Letting $\mathcal{E}$ denote the set of edges under consideration and $|\mathcal{E}| = \sum_{\ell\in\mathcal{E}} n_\ell$ the number of movers across such edges, the "root mean squared model error" is computed as the square root of $\frac{1}{|\mathcal{E}|}\sum_{\ell\in\mathcal{E}} n_\ell \left\{\left(\hat{\Delta}_\ell - \tilde{\Delta}_\ell\right)^2 - (1-h_{\ell\ell})\widehat{\mathbb{V}}_u\left[\hat{\Delta}_\ell\right]\right\}$. "Root mean squared edge effects" is given by the square root of $\frac{1}{|\mathcal{E}|}\sum_{\ell\in\mathcal{E}} n_\ell \left(\hat{\Delta}_\ell^2 - \widehat{\mathbb{V}}_u\left[\hat{\Delta}_\ell\right]\right)$. The R^2 is the square of the ratio of these two quantities. Root mean noise level gives the square root of $\frac{1}{|\mathcal{E}|}\sum_{\ell\in\mathcal{E}} n_\ell \widehat{\mathbb{V}}_u\left[\hat{\Delta}_\ell\right]$. The row labeled "Singleton noise level twice edges w/ 2 movers" imputes $\widehat{\mathbb{V}}_u\left[\hat{\Delta}_\ell\right]$ for each edge with a single mover as twice the average squared standard error of edges with exactly two movers. The rows labeled "Singleton noise imputed via log-log regression" impute the noise level of edges with a single mover based upon a linear regression among edges with 2-10 movers of the log of the average noise level against an intercept and the log of the number of movers. Imputations conducted separately for bridges and non-bridges.

Of course, this R^2 estimate is itself subject to sampling uncertainty and applies only to a particular population of edges. Table 1 repeats this goodness of fit exercise restricting to edges with more movers. The R^2 estimates are remarkably stable, suggesting that these findings are unlikely to be an artifact of noise. The final column of the table reports the square root of the noise level within edges due to irreducible uncertainty across movers. Depending on the sample of edges considered, the average noise level is four to five times

[12]The centered R^2 is nearly identical because the mean edge effect in the $\mathcal{E}_{2+}$ sample is 0.01 and the mean AKM prediction in this sample is 0.01. In the broader sample of 117,657 edges that are not bridges, the mean edge effect is 0.001 and the mean AKM prediction is 0.005.

greater than the average squared model error.

The two rows in the second panel of Table 1 impute the noise levels of the edges with a single mover and recompute the relevant quadratic forms over all edges that are not bridges. The first of these rows sets the noise level of the singleton edges equal to twice the average noise level of edges with exactly 2 movers, an imputation that would be valid under homoscedasticity. The second row relaxes the homoscedasticity assumption by allowing an arbitrary linear relationship between the log of the average noise level and the log of the number of movers. This linear relationship, estimates of which are depicted in Appendix Figure A.1, fits the data well and suggests slightly higher noise levels for the singleton edges. Under both imputations, the R^2 falls modestly to just above 70%.

Finally, recall that nearly half of the firm effects are just-identified by a bridge, contributing no model error at all to the edge predictions. Applying a corresponding linear imputation of singleton noise levels to the bridges (depicted in Appendix Figure A.1) yields an estimated sum of squared edge effects across all edges $\sum_{\ell=1}^{150,417} n_\ell \left(\hat{\Delta}_\ell^2 - \widehat{\mathbb{V}}_u\left[\hat{\Delta}_\ell\right]\right)$ of roughly $10,671$. Hence, the estimated R^2 from an infeasible regression of all edge effects (inclusive of bridges) onto the first differenced firm dummies evaluates to $\left[1 - (0.112)^2 \times 158,452/10,671\right] \times 100 \approx 81\%$.

In sum, the AKM model provides a highly informative (albeit imperfect) summary of the expected wage effects of worker mobility. If we were using firm effect estimates to predict the wage changes associated with worker moves, these findings suggest that noise would be a greater hindrance than model error. The model errors that are present result from cycles in the mobility network among a highly concentrated subset of firms. One interpretation of these errors is that they reflect heterogeneity in the firm effects faced by different sorts of workers. We now turn to thinking about the conditions under which the estimated AKM firm effects retain a causal interpretation in the presence of such heterogeneous effects.

3.3 Causality

The AKM model bears a strong resemblance to a difference in differences specification with J treatment arms where firm effect differences $\psi_j - \psi_k$ represent average treatment effects and the exogenous mobility assumption ensures "parallel trends." It is natural then to ask whether least squares estimation of (1) can identify causal effects under non-parametric restrictions on potential outcomes and worker firm assignments. I will begin with the antecedent task of finding conditions under which the edge effects introduced in section 3.1.1 can be given a causal interpretation. To ease exposition, we will again confine attention to the case where $T = 2$ and ignore time varying covariates, which can be thought of as having been adjusted for in a previous step.

Let $Y_{it}(d_1, d_2)$ denote the potential log wage of worker i in year t who works at firm $d_1 \in [J]$ in period 1 and $d_2 \in [J]$ in period 2. To mimic conventional treatment effects notation, I will use the symbol $D_{it} = \mathbf{j}(i,t)$ to denote the firm employing worker i in period t. We now state three assumptions that endow the average wage changes of workers switching employers between the two periods with a causal interpretation. Our first assumption is an exclusion restriction:

Assumption 1 (Exclusion). $Y_{it}(d_1, d_2) = Y_{it}(d_t)$ *for* $t \in \{1, 2\}$.

This assumption rules out the possibility that past or future firm assignments affect wages. Assumption (1) is violated in sequential auction models (Postel-Vinay and Robin 2002a; Cahuc, Postel-Vinay, and Robin 2006), which posit that hiring wages are influenced by the firm from which a worker was poached. However, Di Addario et al. (2023) find in Italian data that past employers exhibit a negligible influence on hiring wages outside of the law and banking sectors, suggesting this assumption is likely to provide a reasonable approximation for most workers. When Assumption 1 holds, we can link observed wages to potential wages via the relation $Y_{it} = Y_{it}(D_{it})$.

The next assumption mimics the parallel trends assumption of standard difference in differences models:

Assumption 2 (Parallel trends). $\mathbb{E}[Y_{i2}(j) - Y_{i1}(j) | D_{i1} = j, D_{i2} = k] = 0 \quad \forall k \neq j \in [J]^2$.

Assumption 2 states that, among workers switching between any pair of firms, the average potential wages at their origin firms would not have changed between periods. As noted earlier, we should think of Y_{it} here as pre-adjusted for year and age / experience effects, in which case this amounts to a restriction that potential origin and destination wages exhibit a common time trend. Card, Heining, and Kline (2013) reported event study plots of the average earnings trajectories of workers who transitioned between groups of firms characterized by their leave-out wage quartile. These plots, which are now a standard diagnostic, indicate that workers moving to high wage firms do not experience faster wage growth before moving, nor does their wage trend change upon moving to a new firm, suggesting that Assumption 2 provides a reasonable approximation.

Finally, we make a stationarity assumption on average treatment effects among firm switchers:

Assumption 3 (Stationarity). $\mathbb{E}[Y_{i1}(k) - Y_{i1}(j) | D_{i1} = j, D_{i2} = k] = \mathbb{E}[Y_{i2}(k) - Y_{i2}(j) | D_{i1} = j, D_{i2} = k] \equiv \Delta_{jk}, \quad \forall k \neq j \in [J]^2$.

In a mild abuse of our earlier notation for edge effects, this last condition simply ensures that the average treatment effect Δ_{jk} of moving from firm j to firm k among those who make this transition is not time dependent. The plausibility of Assumption 3 will, of course, depend on the nature and length of the sample period under consideration. Lachowska, Mas, Saggio, et al. (2023) and Engbom, Moser, and Sauermann

(2023) provide empirical evidence that firm effects are quite stable over the five to seven year horizons typically studied in the literature.

The following proposition establishes that when these conditions are satisfied worker moves between pairs of firms identify average treatment effects on the wages of movers.

Proposition 1 (Firm switches identify average treatment effects on movers)**.** *If Assumptions 1, 2, and 3 hold, then*

$$\mathbb{E}\left[Y_{i2} - Y_{i1} \mid D_{i1} = j, D_{i2} = k\right] = \Delta_{jk}.$$

Proof. The assumptions used in each step of the below proof are listed above the equals sign:

$$\begin{aligned}
\mathbb{E}\left[Y_{i2} - Y_{i1} \mid D_{i1} = j, D_{i2} = k\right] &\overset{A1}{=} \mathbb{E}\left[Y_{i2}\left(k\right) - Y_{i1}\left(j\right) \mid D_{i1} = j, D_{i2} = k\right] \\
&= \mathbb{E}\left[Y_{i2}\left(k\right) - Y_{i2}\left(j\right) + Y_{i2}\left(j\right) - Y_{i1}\left(j\right) \mid D_{i1} = j, D_{i2} = k\right] \\
&\overset{A2}{=} \mathbb{E}\left[Y_{i2}\left(k\right) - Y_{i2}\left(j\right) \mid D_{i1} = j, D_{i2} = k\right] \\
&\overset{A3}{=} \Delta_{jk}.
\end{aligned}$$

□

Hence, contrasts of the form in (2) can identify causal estimands under plausible assumptions even if firm effects are heterogeneous. In particular, one does not need the process determining wages to be additively separable in unobserved worker and firm heterogeneity for these assumptions to hold.

While Proposition 1 endows the mean wage changes accompanying firm switches with a causal interpretation, these average causal effects are not sufficient to order firms in terms of their average wage levels. The Δ_{jk} represent average treatment effects for a potentially highly selected group of movers between firm j and firm k. Without further assumptions, this heterogeneity undermines our ability to rank the potential wages offered by firms because wage changes may be intransitive. For example, with three firms, we could have $\Delta_{12} > 0, \Delta_{23} > 0, \Delta_{13} < 0$ because the workers who move between Firm 1 and Firm 3 are different from those who move between Firm 2 and Firm 3 or Firm 1 and Firm 2.[13] Proposition 1 does not even rule out the possibility that wage changes between firm pairs are asymmetric – i.e., that $sign\left(\Delta_{jk}\right) = sign\left(\Delta_{kj}\right)$ – which is also a form of intransitivity.

[13] Patterns of this nature are familiar from the social choice literature, where pairwise elections have long been observed to exhibit intransitivities in the form of Condorcet (1785) cycles. Young (1995) provides an accessible introduction to the graph theoretic interpretation of these cycles.

3.3.1 Indirect contrasts and spanning trees

The AKM model enforces transitivity by imposing that $\Delta_{jk} = \psi_k - \psi_j$. We discussed in Section 3.1 how this assumption implies restrictions on edges forming a cycle. For example, if workers move from Firm 1 to Firm 2, Firm 2 to Firm 3, and Firm 3 to Firm 1, then the AKM model requires that $\Delta_{12} + \Delta_{23} + \Delta_{31} = 0$. When cyclic restrictions of this nature are violated, least squares estimation of (1) is not guaranteed to provide firm effect estimates that, when contrasted, yield a convex weighted average of treatment effects. This difficulty is familiar from both the difference in differences literature and recent work on least squares estimation in environments with multiple treatment arms (Goldsmith-Pinkham, Hull, and Kolesár 2022). As in those settings, the problem emerges, in part, from imposing over-identifying restrictions that are violated empirically. Unlike in randomized experiments, however, interpretation problems persist even when we saturate the model because the causal contrasts under study (i.e., the "edge effects") pertain to potentially non-comparable populations.

It was already mentioned in Section 3.1.3 that the firm effect estimates are, in general, a linear combination of all of the edge specific wage changes. The combination weights need not sum to one in each row and can be negative. These negative entries do not immediately undermine a causal interpretation of the firm effect estimates because the edge effects are directed. Returning to the graph depicted in Figure 1, in the case where a single mover is present along each edge, equation (4) implies that Firm 2's fixed effect estimate can be written:

$$
\begin{aligned}
\hat{\psi}_2 &= \underbrace{\frac{7}{12}\hat{\Delta}_{12}}_{\text{direct}} + \underbrace{\frac{3}{12}\hat{\Delta}_{14} + \frac{1}{12}\left(2\hat{\Delta}_{14} - \hat{\Delta}_{23} - \hat{\Delta}_{34}\right) + \frac{2}{12}\left(\hat{\Delta}_{42} - \hat{\Delta}_{24}\right)}_{\text{indirect}} \\
&= \underbrace{\hat{\Delta}_{12}}_{\text{direct}} + \frac{5}{12}\underbrace{\left(\hat{\Delta}_{14} - \hat{\Delta}_{12} - \hat{\Delta}_{23} - \hat{\Delta}_{34}\right)}_{-c_1'\hat{\Delta}} + \frac{2}{12}\underbrace{\left(2\hat{\Delta}_{23} + 2\hat{\Delta}_{34} + \hat{\Delta}_{42} - \hat{\Delta}_{24}\right)}_{(c_2+c_3)'\hat{\Delta}}.
\end{aligned}
$$

With the first firm effect normalized to zero, it is natural for $\hat{\psi}_2$ to place substantial weight on $\hat{\Delta}_{12}$, which offers a direct contrast of the wages at Firm 1 and Firm 2 for a well-defined subpopulation of movers. From the first line, we see a weight of 7/12 is placed on $\hat{\Delta}_{12}$. However, indirect contrasts measuring the effects of moving between other pairs of firms also contribute to $\hat{\psi}_2$, an example of what Goldsmith-Pinkham, Hull, and Kolesár (2022) term "contamination." The combination weights in this representation sum to 13/12, revealing that $\hat{\psi}_2$ cannot be written as a convex weighted average of direct and indirect contrasts.

Under the AKM model, the indirect contrasts contain additional information about Δ_{12}. To see this, note from the second line that by rearranging terms, we can write the firm effect as the direct contrast $\hat{\Delta}_{12}$ plus two terms that have mean zero under the AKM model because they correspond to cycles. With independent and identically distributed errors, adding these terms cuts the variance of $\hat{\psi}_2$ in half, which can be verified by summing the squares of the coefficients multiplying each edge effect in the first line. However, with unrestricted selection into edges and treatment effect heterogeneity, these indirect contrasts need not be informative about any individual's causal effect of moving from Firm 1 to Firm 2.

Indirect contrasts can be avoided by pruning the mobility network to a polytree. As discussed in Section 3.1.3, pruning the graph in Figure 1 to its first three edges yields estimates taking the form

$$\begin{bmatrix} \hat{\psi}_2 \\ \hat{\psi}_3 \\ \hat{\psi}_4 \end{bmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}}_{\boldsymbol{B}_{(1)}^{-1}} \begin{bmatrix} \hat{\Delta}_{12} \\ \hat{\Delta}_{23} \\ \hat{\Delta}_{34} \end{bmatrix}.$$

Importantly, this representation holds no matter how many workers traverse each edge. Here, each firm effect is a simple sum of contrasts $\hat{\Delta}_{jk}$, which provides a causal interpretation to the difference in estimated firm effects between any two firms that share an edge. For example $\hat{\psi}_3 - \hat{\psi}_2 = \hat{\Delta}_{23}$. However, the interpretation of differences in estimated firm effects between firms that do not share an edge is murky.

For example, the estimator $\hat{\psi}_4 = \hat{\Delta}_{12} + \hat{\Delta}_{23} + \hat{\Delta}_{34}$ is not guaranteed to reveal anything about the relative wage levels of Firm 4 and Firm 1. Fundamentally, without moves from Firm 4 to Firm 1 (which would introduce a cycle into the graph) there is no information in the data directly revealing these firms' relative wage levels for any given individual. For the wage changes of the workers moving between Firms 1 and 2 to even reveal the expected sign of the wage change associated with moving from Firm 1 to Firm 4, we need a transitivity restriction: e.g., that for any three firms $(j, k, m) \in [J]^3$, $\Delta_{jk} > 0, \Delta_{km} > 0 \Rightarrow \Delta_{jm} > 0$.

Ensuring transitivity requires either restricting the treatment effect heterogeneity or restricting selection. We will follow the tradition in the treatment effects literature of avoiding restrictions on the outcome equation and examine a restriction on selection that not only ensures a stable ordering of firms but allows cardinal comparison of firm wage levels.

3.3.2 Restricting selection

The following exogeneity assumption ensures comparability of firm wage levels based upon moves by assuming away selection on treatment effects:

Assumption 4 (No selection on treatment effects)**.** $Y_{i2}(k) - Y_{i2}(j) \perp D_{i1}, D_{i2} \quad \forall k \neq j \in [J]^2$.

Importantly, Assumption (4) permits mobility decisions to be related to *average* treatment effects $\mathbb{E}[Y_{i2}(k) - Y_{i2}(j)]$. For example, workers can gravitate towards high wage firms as in the Burdett and Mortensen (1998) model.

However, this assumption prohibits selection on "match" components of wages as arises in many models with comparative advantage (e.g., Gibbons, Katz, et al. 2005; Eeckhout and Kircher 2011; Haanwinckel 2023; Gottfries and Jarosch 2023).

When Assumption 4 does hold, worker mobility identifies unconditional average treatment effects. These average treatment effects necessarily obey transitivity because they pertain to the same population, allowing firm wage levels to be ranked on a common scale. Hence, we can write $\Delta_{jk} = \mathbb{E}\left[Y_{i2}(k) - Y_{i2}(j)\right] = \psi_k - \psi_j$, in which case least squares estimation of (1) identifies pairwise average treatment effects within the connected set of firms. We summarize this logic in the following result.

Proposition 2. *If Assumptions 1, 2, 3, and 4 hold, then* $\Delta_{jk} = \mathbb{E}\left[Y_{i2}(k) - Y_{i2}(j)\right]$ $\forall k \neq j \in [J]^2$. *Let* $\psi = (\psi_2, \ldots, \psi_J)'$, *where* $\psi_j = \mathbb{E}\left[Y_{i2}(j)\right] - \mathbb{E}\left[Y_{i2}(1)\right]$ *for* $j \in [2, \ldots, J]$, *and define the* $J-1 \times 1$ *vector* $F_{it} = \left(1\{D_{it} = 2\}, \ldots, 1\{D_{it} = J\}\right)'$. *If the worker mobility network is connected and* $\psi_1 = 0$, *then* $\mathbb{E}\left[\hat{\psi}_k - \hat{\psi}_j \mid \{F_{i2}, F_{i1}\}_{i\in[N]}\right] = \mathbb{E}\left[Y_{i2}(k) - Y_{i2}(j)\right]$ $\forall k \neq j \in [J]^2$, *where* $\hat{\psi} = \left(\hat{\psi}_2, \ldots, \hat{\psi}_J\right)' = \left[\sum_{i\in[N]} (F_{i2} - F_{i1})(F_{i2} - F_{i1})'\right]^{-1} \sum_{i\in[N]} (F_{i2} - F_{i1})(Y_{i2} - Y_{i1})$.

Proof. $\Delta_{jk} = \mathbb{E}\left[Y_{i2}(k) - Y_{i2}(j)\right]$ follows directly from Assumption 4. Using Proposition 1, the definition of ψ, and the assumption that $\psi_1 = 0$, we have

$$\mathbb{E}\left[Y_{i2} - Y_{i1} | D_{i2}, D_{i1}\right] = \sum_{(j,k)\in\{2,\ldots,J\}^2} (\psi_k - \psi_j)\left(1\{D_{i2} = k\} - 1\{D_{i1} = j\}\right) = (F_{i2} - F_{i1})'\psi.$$

Connectedness of the mobility network ensures the estimator $\hat{\psi}$ is well defined. It follows that $\mathbb{E}\left[\hat{\psi} \mid \{F_{i2}, F_{i1}\}_{i\in[N]}\right] = \psi$. The definition of ψ and assumption that $\psi_1 = 0$ imply $\mathbb{E}\left[\hat{\psi}_k - \hat{\psi}_j \mid \{F_{i2}, F_{i1}\}_{i\in[N]}\right] = \psi_k - \psi_j$ $\forall k \neq j \in [J]^2$. $\square$

This proposition implies that, in the absence of selection on treatment effects, the cyclic restrictions discussed in Section 3.1.1 should hold. While these restrictions offer a reasonable approximation to the edge effects, the empirical analysis in Section 3.2 indicated they are unlikely to hold exactly. Of course, the same could be said of most empirical work relying on quasi-experimental variation. Nonetheless, future researchers may find it fruitful to entertain some weakenings of Assumption 4.

One approach is to find richer time varying covariates that plausibly account for selection. Recent work by Vafa et al. (2022) demonstrates that low dimensional embeddings of employment histories can capture significant information about both potential wages and mobility, potentially restoring independence of adjusted wages. Similarly, conditioning on transition patterns less likely to be plagued by selection could improve the credibility of firm effect estimates. For example, Di Addario et al. (2023) show that the sequential auction model of Bagger, Fontaine, et al. (2014) predicts that the AKM model restrictions should hold for the subpopulation of workers displaced from their two previous jobs and provide evidence supporting this hypothesis. The scope for selection may also be diminished among subpopulations whose transitions are prompted by plant closures or mass layoffs (Gibbons and Katz 1992).

A second approach involves imposing a priori bounds on the maximal selection present in the network. For example, one could constrain $\max_{(j,k)\in[J]^2} |\Delta_{kj} - \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)]|$ and seek estimation and inference procedures that perform well subject to this bound, utilizing extensions of the methods discussed in Armstrong and Kolesár (2018), Armstrong and Kolesár (2021), and Rambachan and Roth (2023).[14] In some contexts it might be reasonable to consider asymmetric bounds on selection. For example, a static Roy (1951) selection model would posit that $\Delta_{jk} \geq \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)]$ for voluntary moves. While this sort of condition can be violated in sequential auction models and models with compensating differentials, it seems reasonable to expect positive selection on wage gains more often than negative selection.

A third approach is to develop network formation models that deliver dynamic propensity scores for mobility between firms that can be used to make semi-parametric adjustments to wage changes. While some early progress has been made in this direction (Abowd, McKinney, and Schmutte 2019), particularly with the use of stochastic block models (Nimczik 2017; Jarosch, Nimczik, and Sorkin 2024), this literature is still in its infancy. A challenge for future work in this area is evaluating the quality of propensity score models in networks that are extremely sparse.

Finally, a number of authors depart from "design based" assumptions on selection and work with interactive factor models of wage outcomes (Bonhomme, Lamadon, and Manresa 2019; Lei and Ross 2023). These models rationalize intransitivities in edge effects in terms of latent heterogeneity in the sorts of workers that transition between different edges. To date, however, most research in this vein has worked with lower dimensional representations of the mobility network, typically by clustering firms into a small set of groups, in order to circumvent the incidental parameter biases that emerge from fitting nonlinear models to sparse networks (Chen, Fernández-Val, and Weidner 2021). An interesting question for future research is whether the clustering step can be skipped and the structure of the underlying (uncoarsened) edge effects more fully rationalized with factor models of this nature.

[14] A major technical hurdle in this setting relative to conventional difference in differences problems is that most edges have very few movers, implying that normality of the estimated edge effects is not assured.

4 Variance decomposition

Abowd, Kramarz, and Margolis (1999) proposed summarizing the influence of firms on covariate-adjusted wage inequality via the finite sample variance decomposition

$$\mathbb{V}_n\left[Y_{it} - X_{it}'\beta\right] = \underbrace{\mathbb{V}_n\left[\alpha_i\right]}_{\text{person effect variance}} + \underbrace{\mathbb{V}_n\left[\psi_{\mathbf{j}(i,t)}\right]}_{\text{firm effect variance}} + 2\underbrace{\mathbb{C}_n\left[\alpha_i, \psi_{\mathbf{j}(i,t)}\right]}_{\text{sorting}} + \underbrace{\mathbb{V}_n\left[\varepsilon_{it}\right]}_{\text{noise}},$$

where n is the number of person-year observations in the sample, $\mathbb{V}_n\left[x_{it}\right] = n^{-1}\sum_{i,t}\left(x_{it} - \mathbb{E}_n\left[x_{it}\right]\right)^2$, $\mathbb{E}_n\left[x_{it}\right] = n^{-1}\sum_{i,t} x_{it}$, and $\mathbb{C}_n\left[\alpha_i, \psi_{\mathbf{j}(i,t)}\right] = n^{-1}\sum_{i,t}\alpha_i\left(\psi_{\mathbf{j}(i,t)} - \mathbb{E}_n\left[\psi_{\mathbf{j}(i,t)}\right]\right)$. Attention usually focuses on the firm effect variance $\mathbb{V}_n\left[\psi_{\mathbf{j}(i,t)}\right]$, which gives a first pass measure of the importance of firms in wage determination. Note that this quantity is person-year weighted, so that the firm effects of larger firms make a greater contribution to wage inequality. The covariance component $\mathbb{C}_n\left[\alpha_i, \psi_{\mathbf{j}(i,t)}\right]$, which is often converted into a correlation coefficient, measures the assortativeness of worker-firm matching.

It has become common to scale the variance and covariance components by $\mathbb{V}_n\left[Y_{it} - X_{it}'\beta\right]$ in order to give each component a share interpretation. While such exercises allow a complete decomposition of residual wage inequality, the variance shares depend critically on the noise level $\mathbb{V}_n\left[\varepsilon_{it}\right]$, which can vary depending on the volatility of earnings in the country being studied, the nature of the earnings measure (hourly, monthly, quarterly, or annual), and the demographics of the workers under study. As discussed below, cross-fitting and clustering methods both provide approaches to consistently estimating $\mathbb{V}_n\left[\varepsilon_{it}\right]$. To maximize comparability across studies then, it is advisable to scale decomposition exercises by the "signal variance" $\mathbb{V}_n\left[Y_{it} - X_{it}'\beta\right] - \mathbb{V}_n\left[\varepsilon_{it}\right]$, which captures the variability of long run expected wages of the worker-firm pairings observed in the data.

Variance shares measure the relative importance of variance components but say nothing about the absolute magnitude of variability present. Variance components are also difficult to interpret because they are measured in squared log points. Standard deviations allow a more direct assessment of the magnitude of worker and firm heterogeneity because they are measured in log points. For example, a finding that $\mathbb{V}_n\left[\psi_{\mathbf{j}(i,t)}\right]^{1/2} = 0.25$ implies that moving to a standard deviation higher paying firm yields a roughly $[\exp(.25) - 1] \times 100 \approx 28\%$ higher wage. Moreover, by Chebyshev's inequality, we know that the (employment-weighted) share of firms with firm effects more than k standard deviations above the mean is at most $1/k^2$. Hence, in this example, no more than 6.25% of person year observations can be at firms with wages 100 log points or more above the mean.

4.1 Limited mobility bias

The exogenous mobility assumption guarantees that least squares will produce unbiased estimates of each fixed effect. It is therefore tempting to plug the least squares estimates $\{\hat{\alpha}_i\}_{i=1}^N$and $\left\{\hat{\psi}_j\right\}_{j=1}^J$ into the $\mathbb{V}_n$ and $\mathbb{C}_n$ operators to form estimates of the relevant variance components. Unfortunately, doing so will produce biased estimates because these operators are quadratic functions. To understand the problem, observe that for any unbiased estimator $\hat{\psi}_j$ of ψ_j, we can write

$$\mathbb{E}_\varepsilon\left[\hat{\psi}_j^2\right] = \mathbb{E}_\varepsilon\left[\left(\hat{\psi}_j - \psi_j + \psi_j\right)^2\right] = \mathbb{E}_\varepsilon\left[\left(\hat{\psi}_j - \psi_j\right)^2 + 2\psi_j\left(\hat{\psi}_j - \psi_j\right) + \psi_j^2\right] = \mathbb{V}_\varepsilon\left[\hat{\psi}_j\right] + \psi_j^2 > \psi_j^2,$$

where $\mathbb{E}_\varepsilon[\cdot]$ denotes expectation with respect to the mean zero noise terms $\{\varepsilon_{it}\}_{i\in[N],t\in[T]}$ in (1) and $\mathbb{V}_\varepsilon[\cdot]$ gives the corresponding variance. Hence, estimation noise leads the square of the estimator to provide an upwardly biased estimate of the square of the parameter. A similar argument reveals that for any unbiased person effect estimator $\hat{\alpha}_i$ of α_i and any unbiased firm effect estimator $\hat{\psi}_j$ of ψ_j that $\mathbb{E}_\varepsilon\left[\hat{\alpha}_i\hat{\psi}_j\right] = \alpha_i\psi_j + \mathbb{C}_\varepsilon\left[\hat{\alpha}_i, \hat{\psi}_j\right]$. Abowd, Creecy, and Kramarz (2002) termed these biases in the context of fixed effects estimation (1) "limited mobility bias" on account of the observation that if the number of movers between each firm were to grow infinitely large, the noise would disappear and the bias along with it. Andrews et al. (2008) derived the nature of the bias in plugin estimates of the variance components in the AKM decomposition more formally and established that the covariance $\mathbb{C}_n\left[\hat{\alpha}_i, \hat{\psi}_{\mathbf{j}(i,t)}\right]$ between person and firm effects must be biased down.

When a consistent estimator $\hat{\mathbb{V}}_\varepsilon\left[\hat{\psi}_j\right]^{1/2}$ of the standard error $\mathbb{V}_\varepsilon\left[\hat{\psi}_j\right]^{1/2}$ is available, one can form a bias-corrected estimate of each ψ_j^2 with $\hat{\psi}_j^2 - \hat{\mathbb{V}}_\varepsilon\left[\hat{\psi}_j\right]$; that is, by subtracting off the squared standard error from the plugin estimate. Likewise, an unbiased estimate of the variance of firm wage effects $\theta_\psi = \mathbb{V}_n\left[\psi_{\mathbf{j}(i,t)}\right] = \mathbb{E}_n\left[\psi_{\mathbf{j}(i,t)}^2\right] - \mathbb{E}_n\left[\psi_{\mathbf{j}(i,t)}\right]^2$ can be obtained from its debiased analogue

$$\hat{\theta}_\psi = \mathbb{E}_n\left[\hat{\psi}_{\mathbf{j}(i,t)}^2 - \hat{\mathbb{V}}_\varepsilon\left[\hat{\psi}_{\mathbf{j}(i,t)}\right]\right] - \left\{\mathbb{E}_n\left[\hat{\psi}_{\mathbf{j}(i,t)}\right]^2 - \hat{\mathbb{V}}_\varepsilon\left[\mathbb{E}_n\left[\hat{\psi}_{\mathbf{j}(i,t)}\right]\right]\right\}. \tag{5}$$

Krueger and Summers (1988) implemented a bias correction of this form when computing the variance of industry wage fixed effects. Replacing $\hat{\psi}_{\mathbf{j}(i,t)}$ with $\hat{\alpha}_i$ in the above formula yields a bias-corrected variance of person effects $\hat{\theta}_\alpha$. The bias-corrected covariance between person and firm effects can be obtained from the formula $\hat{\theta}_{\alpha,\psi} = \mathbb{E}_n\left[\hat{\alpha}_i\hat{\psi}_{\mathbf{j}(i,t)} - \hat{\mathbb{C}}_\varepsilon\left[\hat{\alpha}_i, \hat{\psi}_{\mathbf{j}(i,t)}\right]\right]$.

Andrews et al. (2008) proposed a correction for the AKM variance and covariance components under the assumption that the ε_{it} are *iid*. However, these corrections yielded small changes in the variance components, corrections that appeared to be too small given the magnitude of the biases found in the subsampling exercises reported in Andrews et al. (2012). Card, Heining, and Kline (2013, Online Appendix 3) conjectured that this under-correction was likely a result of unmodeled heteroscedasticity and serial correlation in wage innovations, properties that had been well documented in the literature on earnings dynamics (e.g., MaCurdy 1982; Abowd and Card 1989; Meghir and Pistaferri 2004).

It is tempting to use conventional heteroscedasticity-consistent standard errors to estimate and remove the bias. However, these "standard standard errors" and their bootstrap analogues are known to exhibit bias when the number of parameters being estimated is proportional to the number of observations (Bickel and Freedman 1981; MacKinnon and White 1985; Mammen 1993; Cattaneo, Jansson, and Newey 2018; El Karoui and Purdom 2018). Kline, Saggio, and Sølvsten (2020) proposed replacing the usual heteroscedasticity consistent standard errors (e.g., White 1980; MacKinnon and White 1985) with heteroscedasticity *unbiased* variance estimates derived from cross-fitting that are robust to arbitrary heteroscedasticity.

4.2 Cross-fitting and bias correction

Cross-fitting can be thought of as a version of sample splitting designed to remove overfitting biases while making maximally efficient use of the data (Newey and Robins 2018).[15] To understand the logic behind this approach, it is useful to rewrite (1) in the notation

$$Y_m = \boldsymbol{D}_m \alpha + \boldsymbol{F}_m \psi + \varepsilon_m, \tag{6}$$

where Y_m is a vector of all of the wages in worker-firm match $m \in \{1, \dots, M\} \equiv [M]$. For example, if a worker spends three years at a job, then that match yields a 3×1 vector of wages. The matrix $\boldsymbol{D}_m$ is comprised of worker dummies; it has as many rows as there are time periods in match m and N columns, one for each worker in the sample. The vector α collects the person effects. The matrix $\boldsymbol{F}_m$ is comprised of firm dummies. We assume one firm effect has been normalized to zero so that $\boldsymbol{F}_m$ has $J-1$ columns and the vector ψ collects $J-1$ firm effects. I have again abstracted from the time varying covariates, which can be partialled out in a first stage. Throughout this discussion, we will treat the $\{\boldsymbol{D}_m, \boldsymbol{F}_m\}_{m \in [M]}$, along with (α, ψ) as fixed, leaving ε_m as the only source of randomness in the model. The $\{\varepsilon_m\}_{m \in [M]}$ are assumed to be mutually independent and to exhibit mean zero.

We will write $\mathbb{V}_\varepsilon [\varepsilon_m] = \boldsymbol{\Omega}_m$ which conveys both that the noise level may vary from match to match and that arbitrary within match correlation of the errors ε_m is permitted. The variance of least squares estimates

[15] Sorkin (2018), Drenik et al. (2023), and Card, Rothstein, and Yi (2024) estimate vectors of firm effects using two independent half samples of workers. The covariance between the two samples provided an unbiased (and transparent) estimate of the variance of the latent firm effects. Unfortunately, the connected set can grow much smaller when the sample is split and randomness in how the split was chosen contributes to the variability of the estimator.

of firm effects takes the usual “sandwich” form

$$\mathbb{V}_{\varepsilon}\left[\hat{\psi}\right]=\left(\sum_{m\in[M]}\tilde{\boldsymbol{F}}_m'\tilde{\boldsymbol{F}}_m\right)^{-1}\left(\sum_{m\in[M]}\tilde{\boldsymbol{F}}_m'\boldsymbol{\Omega}_m\tilde{\boldsymbol{F}}_m\right)\left(\sum_{m\in[M]}\tilde{\boldsymbol{F}}_m'\tilde{\boldsymbol{F}}_m\right)^{-1},$$

where $\tilde{\boldsymbol{F}}_m$ is the matrix of firm dummies that results after partialling out the worker dummies – i.e., after deviating the firm indicators from their worker specific means. Hence, if we knew the $\{\boldsymbol{\Omega}_m\}_{m\in[M]}$, we could compute “match clustered” standard errors that allow us to bias correct the square of each firm effect by subtracting off its squared standard error.

Let $\hat{\psi}_{-m}$ denote the vector of firm effects derived from fitting (1) by least squares when leaving out the observations for match m and $\hat{\alpha}_{-m}$ the corresponding vector of person effects. For $\hat{\alpha}_{-m}$ to exist, we need that every worker has at least two worker-firm matches. Assume for the moment then that the sample has been restricted to job switchers so that $\hat{\alpha}_{-m}$ exists. This assumption is without loss of generality since job stayers do not contribute to estimation of the firm effects but only to the firm weights used to define the variance of interest. Another option is to consider long differences – i.e., to omit all but the first and last periods – and to treat the first and last wage error of job stayers as independent, which may be plausible in longer panels.

Define the cross-fit residual as

$$\hat{\varepsilon}_m=Y_m-\boldsymbol{D}_m\hat{\alpha}_{-m}-\boldsymbol{F}_m\hat{\psi}_{-m}=\varepsilon_m+\xi_{-m},$$

where $\xi_{-m}\equiv\boldsymbol{D}_m\left(\alpha-\hat{\alpha}_{-m}\right)+\boldsymbol{F}_m\left(\psi-\hat{\psi}_{-m}\right)$ is a mean zero vector of noise arising from estimation error in the coefficients $\left(\hat{\alpha}_{-m},\hat{\psi}_{-m}\right)$. Note that $\mathbb{E}_{\varepsilon}\left[\varepsilon_m\xi_{-m}'\right]=0$ because the noise is independent across matches. Hence, unlike traditional regression residuals, which tend to be too small due to overfitting, the cross-fit residuals are generally too large, as $\mathbb{E}_{\varepsilon}\left[\hat{\varepsilon}_m\hat{\varepsilon}_m'\right]=\boldsymbol{\Omega}_m+\mathbb{E}_{\varepsilon}\left[\xi_{-m}\xi_{-m}'\right]$.

Kline, Saggio, and Sølvsten (2020) propose multiplying the cross-fit residual by the outcome, which yields the unbiased estimator

$$\hat{\boldsymbol{\Omega}}_m=Y_m\hat{\varepsilon}_m'=\left(\boldsymbol{D}_m\alpha+\boldsymbol{F}_m\psi+\varepsilon_m\right)\left(\varepsilon_m+\xi_{-m}\right)'.$$

Unbiasedness of $\hat{\boldsymbol{\Omega}}_m$ for $\boldsymbol{\Omega}_m$ follows from the observation that $\boldsymbol{D}_m\alpha+\boldsymbol{F}_m\psi$ is a matrix of constants and the presumed independence of ε_m from ξ_{-m}.[16] Hence, an unbiased estimator of $\mathbb{V}_\varepsilon\left[\hat{\psi}\right]$ is

$$\hat{\mathbb{V}}_\varepsilon\left[\hat{\psi}\right]=\left(\sum_{m\in[M]}\tilde{\boldsymbol{F}}_m'\tilde{\boldsymbol{F}}_m\right)^{-1}\left(\sum_{m\in[M]}\tilde{\boldsymbol{F}}_m'\hat{\boldsymbol{\Omega}}_m\tilde{\boldsymbol{F}}_m\right)\left(\sum_{m\in[M]}\tilde{\boldsymbol{F}}_m'\tilde{\boldsymbol{F}}_m\right)^{-1}. \tag{7}$$

Note that unlike the classic HC2 and HC3 estimators of MacKinnon and White (1985), the cross-fit variance estimator is unbiased for any sample size. A corresponding formula for the covariance between person and firm effects is provided in the appendix.

Unbiasedness does not guarantee that the variance estimate for any particular firm effect will be accurate. In fact, a necessary consequence of unbiasedness is that there must be some probability that the realized variance estimate for each of the diagonal terms $\hat{\mathbb{V}}_\varepsilon\left[\hat{\psi}_j\right]$ is negative. However, Kline, Saggio, and Sølvsten (2020) show that weighted averages of the estimated variances, such as the average estimated noise level $\mathbb{E}_n\left[\hat{\mathbb{V}}_\varepsilon\left[\hat{\psi}_{\mathbf{j}(i,t)}\right]\right]$ are guaranteed to converge to $\mathbb{E}_n\left[\mathbb{V}_\varepsilon\left[\hat{\psi}_{\mathbf{j}(i,t)}\right]\right]$ as the sample size grows large. If we have restricted estimation to the firm movers, we can also compute the weighted average noise level, which reweights the firms according to their share of all person-year observations including the firm stayers. Consequently, unbiased estimation of the variance of firm effects does not require taking a stand on the serial correlation of the stayer wage errors. Bias-corrected estimates of firm variance components can often be measured quite precisely. For instance, Kline, Saggio, and Sølvsten (2020) obtain a bias-corrected point estimate of the person-year weighted variance of firm effects of 0.024 (which we will replicate shortly) with a corresponding standard error of only 0.0006.

4.2.1 Leave-out connectedness

An important requirement of cross-fitting methods is that the model must be estimable after leaving out any particular observation. Within a given connected set, many firms may be connected by only a single move, implying their ψ_j would not be estimable if that worker's wage observations were dropped. Kline, Saggio, and Sølvsten (2020) find when using only two periods of data that 43% of the firms in the largest connected set are "just-connected" in this manner. Workers who move to or from such firms have no residual associated with their wage change, prohibiting an assessment of the level of noise in their wages and consequently a bias correction. Fundamentally then, the variance of firm effects is only identified within the leave-out connected set that prunes the just-connected firms.

To assess how this pruning might change estimands, Kline, Saggio, and Sølvsten (2020, Table IV) report

[16] As mentioned earlier, one typically preadjusts log wages for time varying covariates in a first step, which introduces a small higher order bias due to estimation error $\hat{\beta}-\beta$ influencing both $\boldsymbol{y}_m$ and $\hat{\boldsymbol{\varepsilon}}_m$. Even so, it is often wise to ensure $\boldsymbol{y}_m$ has mean zero before applying cross-fitting in order to reduce the variability of $\hat{\boldsymbol{\Omega}}_m$ and we will do so in our empirical example below.

the results of further restricting the set of firms under study to be connected when any *two* matches are left out. Requiring that each firm effect be estimable when any two matches are left out further reduces the number of estimable firm effects by 43%. Surprisingly, the effect of this restriction turns out to be negligible, nudging the point estimate of the variance of firm effects from 0.240 to 0.238. Similar insensitivity to these leave out requirements is found for subsamples of older and younger workers. One reason for this insensitivity is that weakly connected firms typically employ few workers and hence make a small contribution to the overall (person-year weighted) variance of firm effects. Another is that in finite samples, there is a large degree of randomness in which firms happen to be connected, a phenomenon consistent with standard random search models exhibiting Poisson arrival of mobility events.

4.2.2 Bounding and imputation

Existing applications of the cross-fitting correction report variance components describing heterogeneity within the leave-out connected set of workers and firms. Moving the goalposts to estimate whatever target parameter is identified by a research design is standard fare in empirical economics (Crump et al. 2009; Imbens 2010). It is nonetheless prudent to examine the extent to which the leave-out connected set might differ from the broader population of workers and firms. Fortunately, it is relatively straightforward to compute bounds on variance components describing the broader connected set of firms.

The key insight that allows the construction of bounds is to note that the noise level $\boldsymbol{\Omega}_m$ in any match must obey the bound: $0 \leq \boldsymbol{\Omega}_m \leq \mathbb{E}_\varepsilon \left[Y_m Y_m'\right]$. The upper bound follows from observing that

$$\mathbb{E}_\varepsilon \left[Y_m Y_m'\right] = \left(\boldsymbol{D}_m \alpha + \boldsymbol{F}_m \psi\right) \left(\boldsymbol{D}_m \alpha + \boldsymbol{F}_m \psi\right)' + \boldsymbol{\Omega}_m.$$

The first term in this sum is the outer product of a vector and therefore must be positive semi-definite. Consequently, the upper bound is sharp, arising when $\boldsymbol{D}_m \alpha + \boldsymbol{F}_m \psi$ equals a vector of zeros.

Intuitively, the wages associated with a just-connected match could be pure noise, in which case $\boldsymbol{\Omega}_m = \mathbb{E}_\varepsilon \left[Y_m Y_m'\right]$, or they could be entirely noiseless, in which case $\boldsymbol{\Omega}_m = 0$. This observation suggests estimating bounds on $\mathbb{V}_\varepsilon \left[\hat{\psi}\right]$ using $\hat{\boldsymbol{\Omega}}_m{=}Y_m Y_m'$ as an upper bound and $\hat{\boldsymbol{\Omega}}_m{=}0$ as a lower bound on the noise contribution of just-connected matches. As before, we use the leave-out estimator $\hat{\boldsymbol{\Omega}}_m = Y_m \hat{\varepsilon}_m'$ for leave-out connected matches. Denote the resulting estimated upper and lower bounds by $\hat{\mathbb{V}}_\varepsilon^+ \left[\hat{\psi}\right]$ and $\hat{\mathbb{V}}_\varepsilon^- \left[\hat{\psi}\right]$ respectively. Plugging these bounds on the noise level into (5) yields a corresponding lower bound estimate $\hat{\theta}_\psi^-$ and upper bound estimate $\hat{\theta}_\psi^+$ on the variance of firm effects $\theta_\psi = \mathbb{V}_n \left[\psi_{\mathbf{j}(i,t)}\right]$:

$$\hat{\theta}_\psi^- = \mathbb{E}_n \left[\hat{\psi}_{\mathbf{j}(i,t)}^2 - \hat{\mathbb{V}}_\varepsilon^+ \left[\hat{\psi}_{\mathbf{j}(i,t)}\right]\right] - \left\{\mathbb{E}_n \left[\hat{\psi}_{\mathbf{j}(i,t)}\right]^2 - \hat{\mathbb{V}}_\varepsilon^+ \left[\mathbb{E}_n \left[\hat{\psi}_{\mathbf{j}(i,t)}\right]\right]\right\},$$

$$\hat{\theta}^{+}_{\psi} = \mathbb{E}_n \left[\hat{\psi}^2_{\mathbf{j}(i,t)} - \hat{\mathbb{V}}^{-}_{\varepsilon} \left[\hat{\psi}_{\mathbf{j}(i,t)} \right] \right] - \left\{ \mathbb{E}_n \left[\hat{\psi}_{\mathbf{j}(i,t)} \right]^2 - \hat{\mathbb{V}}^{-}_{\varepsilon} \left[\mathbb{E}_n \left[\hat{\psi}_{\mathbf{j}(i,t)} \right] \right] \right\}.$$

These bounds, which are consistent for the corresponding population bounds under conditions that parallel the case where all observations are leave-out connected, may be especially useful in thin samples or environments characterized by low mobility. A corresponding approach to bounding the covariance $\theta_{\alpha,\psi} = \mathbb{C}_n \left(\alpha_i, \psi_{\mathbf{j}(i,t)} \right)$ is detailed in the appendix.

4.2.3 An empirical example

To illustrate these ideas, we now return to the benchmark VHW sample introduced in Section 3.2. With two years of data, estimating the firm effects in levels and first differences is numerically equivalent, which implies that we can think of Y_m as a scalar measuring wage changes without loss of generality. As shown in the top panel of Table 2, roughly 83% of movers in the largest connected set are also in the leave-out connected set. These leave-out connected workers exhibit mildly higher average wages that are slightly less dispersed. The lower panels of the table report estimates of the AKM variance decomposition in each sample.

Squaring the plug-in and bias-corrected standard deviations of firm effects reported in the second column of Table 2 reproduces the firm effect variances reported in Table II of Kline, Saggio, and Sølvsten (2020). The cross-fitting bias correction has substantial bite in the leave-out sample, cutting the estimated standard deviation of firm effects from 18.9 to 15.5 log points. It is natural to worry, however, that this bias reduction constitutes a pyrrhic victory, as the roughly 40% of connected firms that are not leave-out connected may differ from those that are connected. How large of a bias correction would we have obtained if we knew the error variances in the original connected set?

The first column of the bottom panel of Table 2 sheds light on this question. The upper bound on the variance of firm effects assumes that the matches at just-connected firms have error variance zero, yielding an upper bound on the standard deviation of firm effects of 18.5 log points. Coincidentally, this upper bound is very near the plug-in estimate of the standard deviation of firm effects in the leave-out connected sample. Conversely, if we assume matches at each just-connected firm have error variance equal to the squared wage change involving that firm, then we attain a lower bound standard deviation of 14.2 log points. Finally, if the just-connected matches exhibit a noise level equal to the average of the leave-out connected matches – an assumption that I will term “connected at random” (CAR) – then we can form a bias correction by imputing for every just-connected match the average cross-fit noise level of wage changes of movers in the leave-out connected set.[17] The CAR imputation yields an estimated standard deviation of firm effects of 16.4 log points.

[17] One could argue that this assumption should be termed “connected completely at random” (CCAR) as the imputation is not conditioned on any covariates.

Table 2: Sample Composition and Variance Components in Veneto, Italy

	Connected Set	Leave-Out Connected Set
Number of Person Year Observations	1,859,459	1,319,972
Number of Movers	197,572	164,203
Number of Firms	73,933	42,489
Mean Log Wage	4.7507	4.8066
Standard Deviation of Log Wage	0.4455	0.4293
Standard Deviation of Firm Effects		
Plug-in	0.2161	0.1892
Bias-Corrected	[0.1421, 0.1847]	0.1549
Connected at Random	0.1643	
Standard Deviation of Person Effects		
Plug-in	0.3712	0.3634
Bias-Corrected	[0.3216, 0.3423]	0.3345
Connected at Random	0.3320	
Impute Stayer Noise Level	[0.3138, 0.3353]	0.3267
CAR (Impute Stayer Noise)	0.3245	
Covariance of Firm and Worker Effects		
Plug-in	-0.0053	0.0039
Bias-Corrected	[0.0063, 0.0188]	0.0146
Connected at Random	0.0128	

Notes: This table reports properties of the connected and leave-out connected sets in a panel comprised of the 1999 and 2001 waves of the Veneto Work Histories dataset developed by the Economics Department in Università Ca' Foscari Venezia under the supervision of Giuseppe Tattara. The standard deviation of firm effects refers to the person-year weighted standard deviation of firm effects in a regression of log daily wages on worker fixed effects, firm fixed effects, and a year fixed effect. "Plug-in" refers to the OLS estimates. "Bias-corrected" estimate uses the cross-fit bias correction. Intervals correspond to bounds on the variance component in question resulting from the assumption that the noise levels of just-connected movers equal either zero or that mover's squared wage change. "Connected at random" bias corrects by imputing the average error variance of leave-out connected movers to just-connected movers. "Impute Stayer Noise Level" bias corrects assuming that workers who don't switch jobs have the average noise level of leave-out connected movers. Intervals again correspond to bounds on the variance component in question resulting from the assumption that the noise levels of just-connected movers equal either zero or that mover's squared wage change. "CAR (Impute Stayer Noise)" bias corrects assuming that both just-connected movers and job stayers exhibit the average noise level of leave-out connected movers.

Evidently, the bias-corrected standard deviation estimate in the leave-out connected sample is almost exactly halfway between the lower bound and CAR estimates in the broader connected sample. Moreover, the range of estimates is relatively narrow. Little seems to have been lost here by restricting to the leave-out connected set. If the CAR estimate had been very different from the bias-corrected estimate, however, we might have come away more concerned about selection bias. Hence, the CAR estimate seems like a useful diagnostic to report in addition to the standard bias-corrected estimates describing the leave-out connected set.

An equivalent exercise can be conducted with the variance of person effects and the covariance between person and firm effects. Bias correcting the standard deviation of person effects in the leave-out connected set reduces its magnitude by about 3 log points, which is comparable to the effects of bias correction on the standard deviation of firm effects. In the broader sample of connected workers, the bounds on the standard deviation are quite narrow, ranging from 32.2 to 34.2 log points. Like the bias-corrected estimate in the leave-out connected set, the CAR estimate of the standard deviation of person effects in the broader connected set is 33 log points.

As was noted earlier, it is possible that the noise level of job stayers has been underestimated by neglecting serial correlation. While job stayers do not contribute to estimation of firm effects, they are essential for the estimation of person effects. Underestimation of job stayer noise levels could therefore lead to overestimation of the variance of person effects. To assess this possibility, we also report the person effect standard deviation that would result if job stayers had the same average noise level as job movers. Doing so in the leave-out connected set yields a marginally smaller person effect standard deviation of 32.7 log points. In the broader connected sample, this imputation lowers both the upper and lower bounds on the person effect standard deviation by slightly less than a log point. Likewise, the CAR estimate in the connected sample falls by nearly a log point and is essentially indistinguishable from the estimate in the leave-out connected sample.

Finally, bias correcting the covariance between worker and firm effects in the leave-out connected set yields small increases. Fortunately, bias correcting the covariance does not require recovering the noise level of stayers because the covariance must reflect estimation error in firm effects, which depend entirely on movers. In the broader connected sample, the bounds are again fairly narrow. Moreover, the CAR estimate of covariance is close to the bias-corrected covariance in the leave-out connected set.

Using the bias-corrected estimates in the leave-out connected sample yields a correlation coefficient of 0.28. If we ascribe to the stayers the noise level of the movers, the correlation rises negligibly to 0.29 because the person effect variance falls. In the broader connected set the correlation is an increasing function of the unknown noise level of the just-connected movers. Consequently, we can obtain lower bound under the assumption that the noise level is zero and an upper bound under the assumption that the noise level is given by the squared wage change. It turns out that this yields a non-trivial range of possible correlation coefficients [0.09, 0.40]. However, these bounds entertain the implausible possibility that the wage changes of just-connected movers are either all noise or all signal. The CAR estimate of correlation is 0.23 and imputing the mover noise level to the stayers raises this correlation negligibly to 0.24. These estimates are quite close to our bias-corrected estimate in the leave-out connected sample, suggesting that selection is probably not a major concern here.

In sum, we can be relatively confident that trimming has little effect on the person-year weighted variance of worker or firm effects. The influence of trimming on the correlation between worker and firm effects is less clear; however, the agreement between CAR estimates and estimates in the leave-out connected set suggests that selection bias in the correlation coefficient is also likely to be mild. Future research in this area could consider more sophisticated imputation schemes that allow noise levels of just-connected workers to be estimated based upon features of the worker-firm mobility network. Finally, our imputation experiments suggest that person effect variances are unlikely to be dramatically overstated by cross-fitting approaches neglecting the serial correlation of job stayers. In settings where serial correlation is a known concern, imputing the noise levels of job stayers based on the average estimated noise level of movers offers a potentially attractive way of mitigating this bias.

4.3 Clustering approaches

Bonhomme, Lamadon, and Manresa (2019) analyze a version of the AKM model in which firm heterogeneity is restricted to be discrete. They assume the firm effects can be represented in a lower dimensional space via the relation

$$\psi_j = \sum_{k=1}^{K} T_{jk}\bar{\psi}_k, \tag{8}$$

where the $\{T_{jk}\}_{k=1}^{K}$ are indicators for the latent type of the j'th firm effect obeying $\sum_{k=1}^{K} T_{jk} = 1$ and the $\{\bar{\psi}_k\}_{k=1}^{K}$ are the wage effects of those firm types. In their baseline specification, they work with $K = 10$ types, a choice that has been focal in the subsequent literature.

Directly imposing (8) and optimizing jointly over the indicators T_{jk} and the locations $\bar{\psi}_k$ via nonlinear least squares is a non-convex and often intractable computational problem. To circumvent this obstacle, Bonhomme, Lamadon, and Manresa (2019) propose a two step approach. First, they apply a variant of K-means clustering (Forgy 1965; Lloyd 1982) to firm wage distributions to obtain firm type assignments $\hat{T}_{jk}$. These type assignments are then treated as regressors in second step estimation of the model

$$Y_{it} = \alpha_i + \sum_{k=1}^{K} \hat{T}_{\mathbf{j}(i,t)k}\bar{\psi}_k + X_{it}'\beta + \varepsilon_{it}.$$

Rather than estimate this equation by OLS, they treat the α_i as normal mixtures with means that depend on $\hat{T}_{jk}$, which further reduces the number of parameters to be estimated, and maximize the likelihood via the EM algorithm (Dempster, Laird, and Rubin 1977). Once the type specific parameters have been estimated, the type estimates can (in principle) be updated, yielding reclassified firm and worker type assignments that provide approximations to one step maximum likelihood estimates of the full model.

In some respects, the clustering approach mirrors the earlier literature on industry wage differentials (e.g., Krueger and Summers 1988). Rather than using as regressors indicators for 20 or so 2-digit industries, the "industries" are treated as latent random variables to be reconstructed via clustering of firm wage distributions. By reducing the high dimensional AKM specification down to a low dimensional model, the clustering approach sidesteps the usual incidental parameters problem, substantially reducing the biases associated with squaring estimated parameters. Clustering also circumvents the requirement to limit the analysis to the largest connected set of firms, as one only needs the estimated firm types $\hat{T}_{jk}$, rather than each individual firm, to be connected by worker mobility for the second step model to be estimable. Interactions between estimated worker and firm types can also be treated as regressors, allowing estimation of non-separable models. These interactions turn out to be negligible in Swedish data, however, raising the estimated R^2 of the model from 74.8% to 75.8%, leading Bonhomme, Lamadon, and Manresa (2019) to conclude that "complementarities explain only a small part of the variance of log-earnings."[18]

The advantages of the clustering approach come at the cost of strong assumptions on the data generating process. For one thing, it seems implausible that there exist large groups of firms that offer exactly the same wage premiums. The restriction in (8) is at best an approximation and one that inevitably leads to understatement of firm effect variances by neglecting within-type variability. Neglected covariances between any within firm type employer heterogeneity and worker heterogeneity can also lead to bias in the estimates of worker-firm sorting. Card, Rothstein, and Yi (2023) note both of these problems when revisiting the industry wage differential literature, where they find substantial variation in employer wage premiums within industry along with significant worker-firm sorting. It seems unlikely that any partition of firms into 10 or even 10,000 groups would entirely resolve these problems.

Even if (8) were to hold exactly, the type assignments $\hat{T}_{jk}$ will be noisy for small firms, which can generate bias in the estimated locations parameters $\left\{\hat{\bar{\psi}}_k\right\}_{k=1}^{K}$. Indeed, the formal assumptions used by Bonhomme, Lamadon, and Manresa (2019) to establish consistency of the two step clustering approach require that the number of wage observations at the smallest firm grow with the number of firms. This potential for bias that arises with finite sized firms is a cost of having to estimate a regressor instead of relying on a predetermined grouping such as industry, firm size, or geography. Another cost concerns interpretability. While some judgement calls are involved in choosing industry and geographic categories, variation across them is substantially easier to interpret than variation across firm groups determined via K-means clustering of wage distributions.

A related conceptual difficulty is that the type assignments are determined based on cross-sectional wage distributions rather than worker mobility. However, any cross-sectional distribution of wages could be driven by worker sorting rather than firm heterogeneity. The ability to separate the two comes only from the assumption that the economy possesses a finite number of well separated firm types. Parametric identification of this nature is contrary to the ethos of the AKM approach, which relies entirely on worker

[18]In fact, their estimated interactions are smaller than those reported by Card, Heining, and Kline (2013, Table III) for German data, who find that allowing for unrestricted worker-firm match effects raises the adjusted R^2 by roughly 2 percentage points. In a recent analysis of US earnings data, Lamadon, Mogstad, and Setzler (2022, Table A6) report that adding worker-firm interactions to an additive group fixed effects model raises the R^2 by less than one percentage point.

mobility to separate worker and firm heterogeneity.

The robustness exercises reported in Bonhomme, Lamadon, and Manresa (2019, Table III) reveal that the estimated variance of firm effects can, in fact, be quite sensitive to the details of the procedure used to form the type assignments. They find that clustering firms into ten groups based on their cross-sectional wage distributions yields a variance of firm effects that accounts for 2.6% of the overall variance of earnings in Swedish administrative data. Splitting those groups by firm value added raises the share of wage variance explained by firms to 3.4%. Reclassifying the firm types – which can be thought of as choosing the firm groups to directly approximate the firm effects of the movers – raises the estimated contribution of firm effects to 4.1% of the variance.

A recent paper by Bonhomme, Holzheu, et al. (2023) relaxes (8) by assuming

$$\psi_j = \sum_{k=1}^{K} T_{jk} \left(\bar{\psi}_k + \upsilon_k \right), \tag{9}$$

where each $\{\upsilon_k\}_{k=1}^{K}$ is a mean zero normally distributed random effect with a different variance. By allowing for within firm type dispersion, this correlated random effects (CRE) approach generally picks up a greater degree of firm dispersion. For instance, Lamadon, Mogstad, and Setzler (2022, Table A6) find that firm effects explain only 3.2% of annual earnings variance in US tax data when using the two-step estimator imposing (8), whereas Bonhomme, Holzheu, et al. (2023, Table F2) estimate that share at 6.2% in a six year panel of the same data using the CRE estimator predicated on (9). However, the CRE estimator still relies on functional form assumptions to separate worker and firm types. In particular, the estimator is predicated on moment conditions imposing that $\alpha_i - \mathbb{E}\left[\alpha_i | T_{jk}\right]$ is independent of υ_k, which implies there is no worker firm sorting within firm types, while the type assignments $\hat{T}_{jk}$ are still based on a first step clustering routine applied to the cross-sectional wages of job stayers.

Bonhomme, Holzheu, et al. (2023) find in both Monte Carlo exercises and real datasets that both their CRE estimator and the cross-fitting estimator of Kline, Saggio, and Sølvsten (2020) successfully address limited mobility bias. On average the parametric CRE estimator yields modestly smaller firm effect estimates than the bias-corrected estimator based on cross-fitting. It is difficult to assess the extent to which these differences arise from violations of the functional form assumptions baked into the CRE model. A traditional justification for CRE methods is that, by exploiting additional restrictions, they can offer more efficient (albeit less robust) estimates (Chamberlain 1982; Angrist and Newey 1991). Monte Carlo evidence suggests that the CRE estimates of variance components are indeed likely to be more efficient than the cross-fitting estimator when the CRE assumptions hold. Hence, the CRE approach may be useful in small samples where precision is a practical concern. Another potentially important use case for the CRE estimator is settings with extremely limited mobility, where restricting to the leave-out connected set would drop an unacceptably large share of the units under study (e.g., Fenizia 2022). When using such approaches, it may be worthwhile to pursue iteratively updated versions of the estimator, which have been found to yield improved performance in some settings (Bonhomme, Lamadon, and Manresa 2019; Lentz, Piyapromdee, and Robin 2022).

4.4 How variable are worker and firm effects?

Bias-corrected estimates of worker and firm contributions to wage inequality have now been reported in many countries. The figure below depicts bias-corrected estimates of worker and firm effect variability drawn from nine recent studies utilizing the cross-fitting correction of Kline, Saggio, and Sølvsten (2020). Rather than focus on variances or variance shares, I compare the standard deviation of person effects to the standard deviation of firm effects, the units of which are directly interpretable in log points. When reported, multiple specifications from the same study are included to illustrate the sensitivity of estimates to the sample period and population. The list of studies depicted is provided in Appendix Table A.1. Some studies that used bias corrections could not be included because they failed to report the magnitude of the variance components, relying on variance shares without reporting the marginal variance.

The 45 degree line through the origin of Figure 3 gives what one should expect if worker and firm components are equally important and scale with the overall level of inequality in an economy. Perhaps surprisingly, many of the estimates lie very near this line. As expected, the scale of inequality appears most pronounced in middle income countries such as Mexico, South Africa, and Brazil, while Italy, the US, and Sweden are relatively more equal in both dimensions. The estimates falling below the 45 degree line come predominantly from high income countries and from Brazil. Interestingly, these studies all find comparable standard deviations of firm effects near 0.25. However, the standard deviations of worker effects vary widely from sample to sample.

To some extent, this variability of person effect variances is to be expected given that many of the estimates partition by race or sex, groups within which we expect person effects to be less dispersed. For example, four of the Brazilian estimates are from Gerard et al. (2021), who report estimates separately by race and sex using a 12 year panel, which accounts for some of the lowest worker effect standard deviations in that country. However, person effect variances also seem to vary with other features of the data including the time horizon studied.

Abowd and McKinney (2023), for example, find a nearly identical standard deviation of firm effects in 3 year and 24 year extracts of annualized earnings records from the LEHD. However, in the 3 year panel, the bias-corrected standard deviation of person effects is roughly 50% larger than the standard deviation of

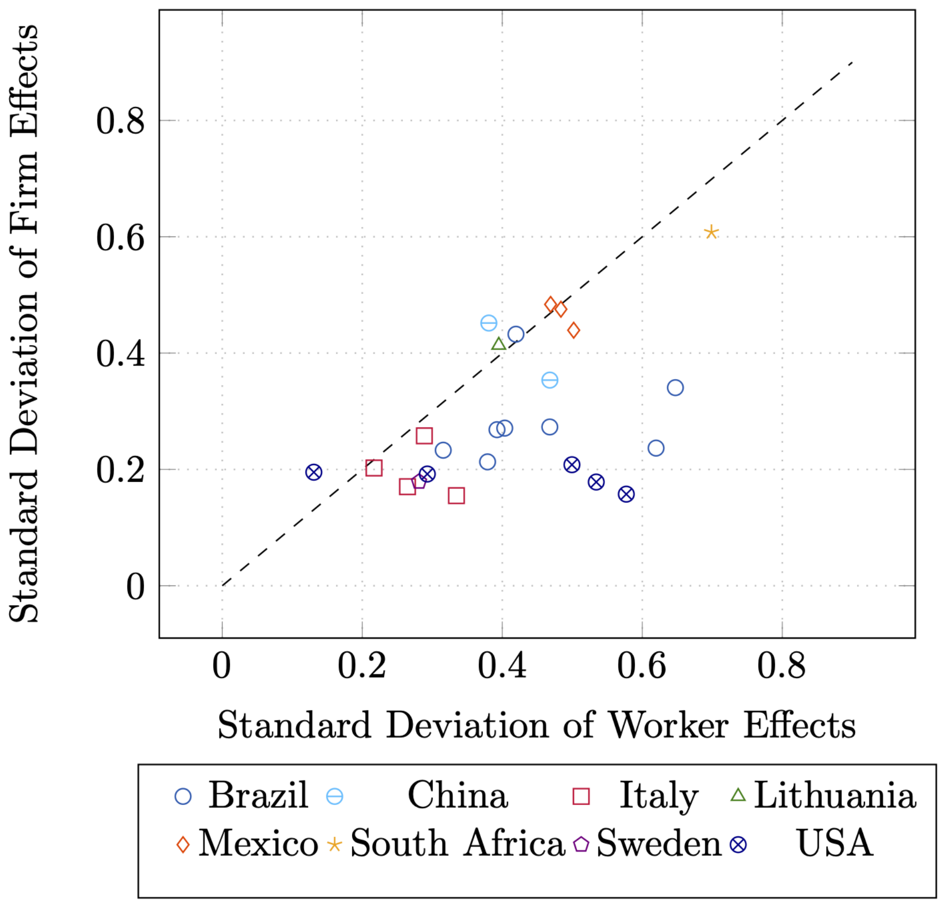

Figure 3: Bias-corrected standard deviations of firm and worker fixed effects by country

firm effects, while in the 24 year panel, the person effects exhibit a standard deviation roughly 33% below that of the firm effects. Likewise, Lachowska, Mas, Saggio, et al. (2023) find using hourly wage data from Washington state that person effects are substantially more dispersed in a 2 year panel than a 12 year panel. While it is tempting to conclude that this sensitivity to time scale reflects drift in the person effects, Lachowska, Mas, Saggio, et al. (2023) demonstrate that person effect estimates remain strongly correlated across decades.

Recall that the variance of person effects among firm stayers cannot be estimated by cross-fitting at the match level. A majority of the studies considered include firm stayers and it is reasonable to assume that these studies treat the errors of firm stayers as serially independent, as this is the default option provided in the most widely used software package used to implement the cross-fitting correction.[19] The estimated person effect variances may therefore be subject to an upward bias stemming from neglected serial correlation, albeit

[19]See https://github.com/rsaggio87/LeaveOutTwoWay for details.

a smaller one than if no correction were implemented. It seems likely then that the tendency for shorter panels to yield larger person effect variances reflects this tendency to under-correct, as adjacent observations are more strongly correlated. If the person effect variances are in fact upwardly biased due to serial correlation, then it is even more surprising that so many studies yield estimates near the 45 degree line.

Though the estimated person effect variances appear sensitive to sample composition, the firm effect standard deviations are remarkably resilient. Among the estimates depicted here, the firm effect standard deviations all exceed 0.15 and for high income countries cluster around 0.20. A potentially useful comparison comes from Bonhomme, Holzheu, et al. (2023), who estimate variance decompositions in five high income countries (Austria, Italy, Norway, Sweden, and the U.S.). While they do not report person effect variances, averaging their cross-fitting based estimates of the standard deviation of firm effects across countries and samples yields a mean value of roughly 0.14, which is a bit below the estimates reported for rich countries in the figure above.[20] Some of this discrepancy is likely attributable to their procedure for harmonizing samples across countries with different earnings measures. In the U.S., data limitations require them to study annual earnings. Bonhomme, Holzheu, et al. (2023, Appendix Figure F10) show in Norwegian data that using annual rather than hourly earnings substantially lowers estimated firm effect variances. To facilitate comparisons between the U.S. and European countries, they impose on all samples a minimum annual earnings threshold of 32.5% of the national average, which in the U.S. approximates the full time earnings of minimum wage workers. Selecting on the dependent variable reduces its variability and Bonhomme, Holzheu, et al. (2023, Appendix Figure F2) document in U.S. data that imposing higher minimum earnings thresholds yields lower firm effect variances.

A reasonably informed guess then, is that across a wide range of high income countries, the standard deviation of firm effects in average daily or hourly wages typically ranges between 15 and 20 log points. For middle income countries, the standard deviation of firm effects appears to be higher, perhaps as high as 0.4 in some cases. It is plausible that firm effects are more important in developing countries, where search frictions and misallocation have been argued to be more prevalent (Hsieh and Klenow 2009). However, many of these studies are very recent and have yet to clear peer review. It will be important to see estimates from more countries and research teams before drawing strong conclusions about the relationship between economic development and the dispersion in firm pay components.

The median firm effect standard deviation estimate among all those pictured in Figure 3 is 0.26 and an unweighted average of them is 0.30. If, in high income countries, the standard deviation of firm wage effects is somewhere between 15 and 20 log points, then switching to a standard deviation higher firm yields wages

[20]Bonhomme, Holzheu, et al. (2023) report the match weighted variance of firm effects rather than the person-year weighted variance.

16-22% higher – a very substantial effect size. For comparison, Chetty, Friedman, et al. (2011) estimate that a standard deviation increase in kindergarten classroom quality in the Project STAR experiment raises adult earnings by 13 percentage points. These findings suggest workplace heterogeneity is an important contributor to wage inequality.

5 Regressing firm effects on observables

Examining how fixed effects covary with observables can help to demystify the nature of these fundamentally unobservable objects. Many of the empirical findings summarized in Section 2 were derived from regressing estimated firm fixed effects on observed features of workers and firms. Besides greater robustness to modeling assumptions, an important advantage of fixed effects methods over more structured random effects approaches (e.g. Hanushek 1974; Amemiya 1978) is that fixed effect estimates can be shared with different research teams, who can subsequently use them to examine different downstream hypotheses via "second step" regressions. I will now review the logic of these downstream regressions and discuss the subtleties of inference on second step projection coefficients. These ideas will be illustrated with an example to the firm size wage premium in the VHW data.

5.1 One step vs two

Suppose we are interested in the relationship between the vector of population firm effects ψ and a set of firm covariates such as firm size and the average education level of the firm's employees. Descriptive relationships of this nature are often summarized with linear projections of the form

$$\psi = \boldsymbol{Z}\theta + v,$$

where $\boldsymbol{Z}$ is a matrix of firm covariates and the parameter of interest is $\theta = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\psi$. By construction, the projection error v obeys $\mathbf{Z}'v = 0$. Plugging this relationship into (6) yields:

$$Y_m = \boldsymbol{D}_m\alpha + \boldsymbol{F}_m\boldsymbol{Z}\theta + \boldsymbol{F}_m v + \varepsilon_m.$$

Since the projection error v is orthogonal to $\boldsymbol{Z}$, one might be tempted by this representation to estimate θ from a least squares regression of Y_m on $(\boldsymbol{D}_m, \boldsymbol{F}_m\boldsymbol{Z})$ – i.e., on person dummies plus the firm characteristics. There are two difficulties with this logic. The first objection, which is largely pedantic, has to do with weighting. The cross product $\sum_{m\in[M]} \left(\boldsymbol{F}_m\boldsymbol{Z}\right)' \boldsymbol{F}_m v = \sum_{m\in[M]} \boldsymbol{Z}'\boldsymbol{F}_m'\boldsymbol{F}_m v$ will not, in general, equal zero unless all firms are the same size. Hence, orthogonality need not hold in the microdata even if it holds across firms. Of course, if we had initially defined the estimand θ as the firm size weighted projection, then the relevant v would satisfy orthogonality in the microdata.

A more significant objection is that even if $\boldsymbol{F}_m v$ is uncorrelated with $\boldsymbol{F}_m \boldsymbol{Z}$, it is still likely to be correlated with $\boldsymbol{D}_m$. The fact that higher wage workers tend to work at higher wage firms suggests $\sum_{m\in[M]} \boldsymbol{D}'_m \boldsymbol{F}_m v > 0$, which violates the exogeneity requirements of least squares. This violation will not only tend to generate bias in the estimated person effects but also in estimates of θ because $\boldsymbol{F}_m \boldsymbol{Z}$ is correlated with $\boldsymbol{D}_m$. Hence, unless one has a strong reason to suspect that the elements of $\mathbf{Z}$ account for all of the correlation between worker and firm wage effects, dropping the firm dummies as controls (i.e., treating v as an uncorrelated random effect) will tend to generate bias.

The two step approach is to first compute the fixed effects $\hat{\psi}$ and then regress them on $\boldsymbol{Z}$ to obtain the projection coefficient $\hat{\theta} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\hat{\psi}$. Under strict exogeneity, the firm effects are unbiased. The projection coefficient, which is just a linear combination of the estimated firm effects, inherits this property, obeying $\mathbb{E}_\varepsilon \left[\hat{\theta}\right] = \theta$. Hence, the two step estimator provides robust estimates of the projection regardless of the dependence between worker and firm effects.

Another advantage of the two step estimator is that it can foster scientific cooperation: the research team that produces $\hat{\psi}$ need not be the team that has access to $\boldsymbol{Z}$. Fixed effects estimates are often computed once on population microdata by expert researchers and then made available to outside teams who do may not have access to the same microdata files (e.g., Bellmann et al. 2020). These sorts of data sharing arrangements enable a broader range of hypotheses and external data sources to be brought to bear on questions of scientific interest.

5.2 Variance estimation

The variance of the estimated projection coefficient is

$$\mathbb{V}_\varepsilon \left[\hat{\theta}\right] = \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\mathbb{V}_\varepsilon \left[\hat{\psi}\right] \boldsymbol{Z} \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}.$$

While second step regressions will yield unbiased estimates of linear projection coefficients, the standard errors produced by conventional software packages will mistakenly assume that the noise $\hat{\psi} - \psi$ in the second step regressand is independent across firms -- i.e., that $\mathbb{V}_\varepsilon \left[\hat{\psi}\right]$ is diagonal. Neglecting correlation between the estimated firm effects can lead to severe understatement (or overstatement) of the uncertainty in second step regression coefficients.

The sign of this bias in the estimated standard errors is theoretically ambiguous because the residuals from the second step regression will tend to overstate the intrinsic noise level of each estimated fixed effect. To take an extreme example, suppose that the wage disturbances ε_{it} in (1) are exactly zero. In such a case, the first step regression will fit perfectly, yielding $\hat{\psi} = \psi$. However, a second step regression of firm fixed effects on observed firm characteristics will nonetheless yield residuals capturing unexplained variation in the vector ψ of true firm effects. Consequently, conventional software packages will produce a positive standard error estimate despite the fact that the true firm effects are fixed and exhibit no uncertainty.

Standard errors reflecting only the uncertainty associated with the ε_{it} are easily computed by using the cross-fit variance estimates introduced in (7). For example, Kline, Saggio, and Sølvsten (2020) considered a second step regression wherein $\boldsymbol{Z}$ included a constant, the share of workers over age 35, firm size, and their interaction. Note that, as in our earlier discussion of cross-fitting, interest centers on the finite population of J firms actually measured in our dataset rather than an abstract "super-population" from which those firms were drawn. Replacing the unknown $\mathbb{V}_{\varepsilon}\left[\hat{\psi}\right]$ with $\hat{\mathbb{V}}_{\varepsilon}\left[\hat{\psi}\right]$ yields an unbiased estimate of the variance of the second step regression coefficient that can be used for inference. Fortunately, computation does not require that the entire $\hat{\mathbb{V}}_{\varepsilon}\left[\hat{\psi}\right]$ matrix be computed or stored.[21]

While it is straightforward for research agencies to release fixed effect estimates to the public and their (squared) standard errors, it is not feasible to release entire variance matrices. In principle, one could conduct inference relying only on the fixed effect standard errors by considering worst case correlation patterns. However, doing so could lead to extremely conservative inferences. An interesting area for future work is understanding what low dimensional features of $\hat{\mathbb{V}}_{\varepsilon}\left[\hat{\psi}\right]$ can be reported that would enable accurate inference on projection coefficients without knowledge of the $\boldsymbol{Z}$ under consideration by the research team.[22]

5.3 Revisiting the firm size wage premium

Figure 4 illustrates the use of these methods by studying how the relationship between firm effects and firm size varies by province in Veneto. Returning to the firm effect estimates studied in Table 2, the matrix $\boldsymbol{Z}$ is chosen to include indicators for the firm size categories utilized by Bloom et al. (2018) interacted with indicators for which of Veneto's seven provinces contains the firm in question. As a normalization, the smallest firm size category of 1-10 employees has been set to zero in each province, so that each of the included estimates represents a within province firm size "premium." To reduce clutter, we have dropped the province of Rovigo which is so small that it lacks any firms in the largest two size categories. By contrast, more than one thousand firms are present in each size category of the pictured provinces.

[21] A computationally efficient approach to estimation of $\mathbb{V}_{\varepsilon}\left[\hat{\theta}\right]$ is automated and detailed in the LeaveOutTwoway package available at https://github.com/rsaggio87/LeaveOutTwoWay.

[22] For example, in a lower dimensional context, Firth and De Menezes (2004) propose reporting "quasi-variances" that can be used for inference on unknown contrasts.

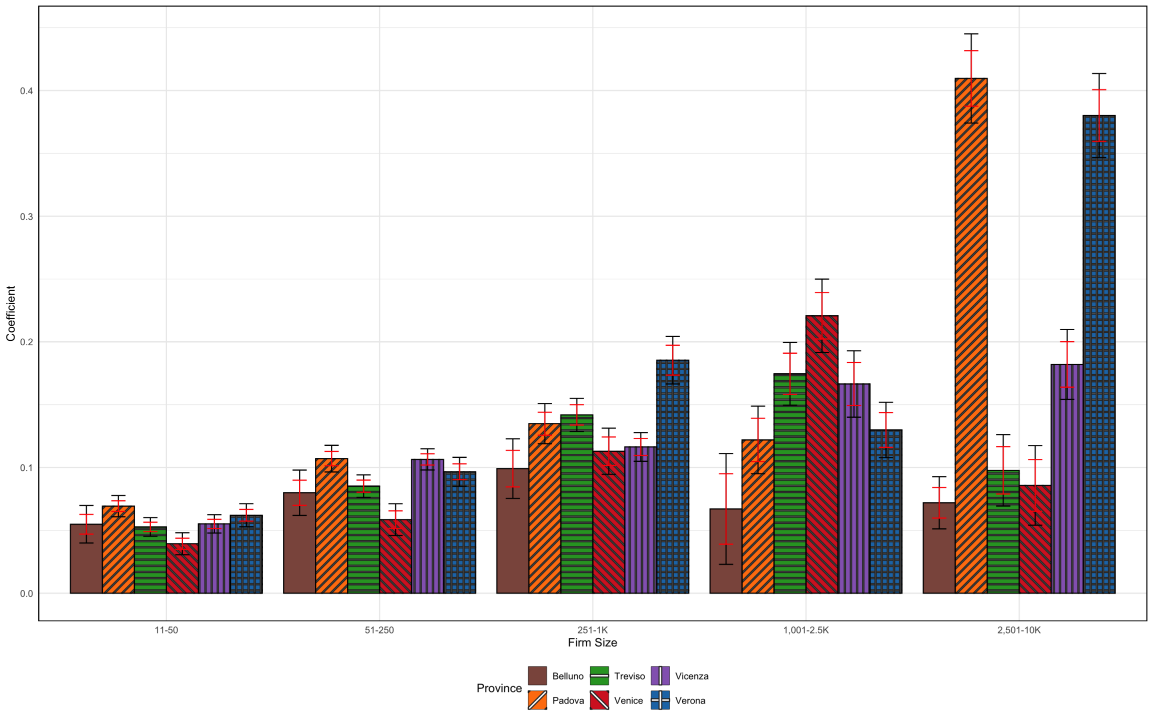

Figure 4: Mean firm effects by firm size and province

Notes: Sample comprised of firms in leave-out connected set described in Table 2. Bar height gives coefficient from second step regression of firm effects onto province indicators plus interactions with indicators for firm size category. Omitted firm size category in each province is 1-10. Both confidence intervals derived by adding ± 1.96 standard errors to the point estimate. Outer confidence interval (depicted in black) relies on $\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\hat{\mathbb{V}}_{\varepsilon}\left[\hat{\psi}\right]\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}$ as estimator of the asymptotic variance. Inner confidence interval (depicted in red) relies on $\frac{J}{J-k}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\left\{\text{diag}\left(\boldsymbol{M}_Z\hat{\psi}\right)\right\}^2\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}$ as estimator of the asymptotic variance, where $\boldsymbol{M}_Z = \boldsymbol{I} - \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'$.

Confidence intervals based on naive heteroscedasticity-robust standard errors computed via a second step regression are shown alongside those based on the cross-fit variance matrix $\hat{\mathbb{V}}_{\varepsilon}\left[\hat{\psi}\right]$. The cross-fit standard errors reflecting uncertainty attributable to ε turn out to be about 75% larger than the naive standard errors on average. As a result the 95% confidence intervals based on cross-fitting turn out to bracket those based on naive standard errors in all cases. Evidently, the downward bias in naive standard errors attributable to neglecting correlation among the estimated firm effects outweighs the upward bias attributable to treating the firm effects as random draws from a broader population.

In all six pictured provinces, firm effects tend to increase with firm size. However, the size profiles differ substantially across provinces and in some cases appear non-monotone. In Verona and Padova the largest firms exhibit fixed effects averaging approximately 40 log points more than the smallest firms, while in Venice the corresponding gap in firm wage effects is only about 9 log points. These orderings reverse, however, in the next largest firm size category. In Venice, for example, firms with 1,000-2,500 employees are estimated to pay roughly 22 log points more than firms with 1-10 employees, while in Verona the premium is only 13 log points.

While the premiums relative to the base firm size category are precisely estimated in each province, it is not completely obvious which of these premiums differ from one another given that the estimates are all correlated. A useful rule of thumb is that we can conclude that the estimands are different from one another if their confidence intervals do not overlap.[23] Based on this heuristic, we can safely infer that in both Venice and Treviso the firm size premium in the 1,000-2,500 employee category exceeds the premium in the 2,500-10,000 employee category, indicating that the firm size premiums are not monotone in these regions.

We can also infer that size premiums tend to differ by region, though our visual rule of thumb becomes less decisive in smaller firm size categories. To obtain a more accurate assessment of regional differences in the average premium for firms with 10-50 employees, I reparameterize $\boldsymbol{Z}$ to include interactions between province and firm size categories. The resulting standard error estimates reveal that it is possible to reject at the 5% level the null hypothesis that the premiums in the 10-50 employee category are equal in Venice and Vicenza. By contrast, the premiums in Vicenza and Treviso cannot be distinguished from each other even at the 10% level.

6 Hiring origins and state dependence

The basic AKM specification views wage determination as fundamentally static: the expected wage arising from a match between a worker and firm depends only on their underlying (time-invariant) types. Search theoretic models, by contrast, often predict that wages should be influenced by the circumstances surrounding how the match was formed – e.g., whether the worker was "poached" from another firm or hired from unemployment, as unemployed workers typically have worse outside options than their employed counterparts. Consistent with this view, Faberman et al. (2022) provide survey based evidence that job offers received by currently employed workers pay higher wages than those received by unemployed workers with similar characteristics.

An influential framework for modeling such state dependence comes from the class of sequential auction models pioneered by Postel-Vinay and Robin (2002a) and Postel-Vinay and Robin (2002b), where on the job

[23] For any two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, we have $\mathbb{V}\left(\hat{\theta}_1-\hat{\theta}_2\right)=\mathbb{V}\left(\hat{\theta}_1\right)+\mathbb{V}\left(\hat{\theta}_2\right)-2\mathbb{C}\left(\hat{\theta}_1,\hat{\theta}_2\right)\leq\left\{\mathbb{V}\left(\hat{\theta}_1\right)^{1/2}+\mathbb{V}\left(\hat{\theta}_2\right)^{1/2}\right\}^2$, where the upper bound binds with equality when the two estimators are perfectly negatively correlated. Consequently, $\mathbb{V}\left(\hat{\theta}_1\right)^{1/2}+\mathbb{V}\left(\hat{\theta}_2\right)^{1/2}$ provides a conservative standard error on the difference between the estimators. A test that evaluates whether $\left|\hat{\theta}_1-\hat{\theta}_2\right|>c\cdot\left[\mathbb{V}\left(\hat{\theta}_1\right)^{1/2}+\mathbb{V}\left(\hat{\theta}_2\right)^{1/2}\right]$ for some critical value c (e.g., 1.96 as in Figure 4) amounts to evaluating whether $\left[\hat{\theta}_1-c\mathbb{V}\left(\hat{\theta}_1\right)^{1/2},\hat{\theta}_1+c\mathbb{V}\left(\hat{\theta}_1\right)^{1/2}\right]\cap\left[\hat{\theta}_2-c\mathbb{V}\left(\hat{\theta}_2\right)^{1/2},\hat{\theta}_2+c\mathbb{V}\left(\hat{\theta}_2\right)^{1/2}\right]=\emptyset$.

search gives rise to a series of bilateral competitions between firms for workers. These competitions mirror second-price auctions, with firms tailoring their wage bids based upon the willingness to pay of the rival they face. Consequently, the wages offered to new hires differ based on where a worker is hired from and which firm is hiring them. Tailoring of this nature can, in principle, contribute greatly to cross-sectional inequality by amplifying the role of luck: an early job displacement can lower wages throughout a worker's career by persistently degrading their outside options.

Di Addario et al. (2023) study the empirical predictions of sequential auction models for hiring wages using an extension of the AKM model, in which a separate fixed effect is allowed for each possible hiring origin. These hiring origin fixed effects are meant to proxy for the worker's outside option. Letting Y_{im} denote the log hiring wage of the ith worker at their m'th job, they consider a linear model taking the form:

$$Y_{im} = \alpha_i + \underbrace{\psi_{\mathbf{j}(i,m)}}_{\text{destination effect}} + \underbrace{\lambda_{\mathbf{h}(i,m)}}_{\text{origin effect}} + X_{im}'\delta + \varepsilon_{im}, \quad \text{for } i \in [n], m \in [M_i]. \tag{10}$$

Here, the function $\mathbf{j} : [n] \times [M_i] \to [J]$ returns the identity of the firm hiring the worker at their m'th job. The function $\mathbf{h} : [n] \times [M_i] \to [J] \cup \{U\}$ returns the origin of the new hire, which can either be the identity of a prior employer from which the worker was "poached" or unemployment (denoted as "U"). Thus, each firm j has a pair (ψ_j, λ_j) of fixed effects.

Di Addario et al. (2023) term the specification in (10) a "dual wage ladder" (DWL) model because hiring wages depend on two dimensions of firm heterogeneity. As in the AKM model, α_i is a person fixed effect that can be ported from employer to employer, while the vector X_{im} includes time varying covariates, including work experience and indicators for the year that the match was formed. The term $\psi_{\mathbf{j}(i,m)}$ is a destination firm fixed effect that, like the traditional AKM firm effect, must be forfeited upon separating from the employer. The distinctive feature of the DWL specification is the origin firm fixed effect, $\lambda_{\mathbf{h}(i,m)}$, which captures a form of state dependence in wage setting. According to the DWL model, two workers with the same α_i, hired by the same firm from two different origins – e.g., non-employment and the most productive firm in the economy – will be paid different wages.

The error ε_{im} measures omitted factors that vary across matches at the time of hiring. Each of these errors is assumed to have mean zero, which is a version of the traditional exogenous mobility assumption used to justify least squares estimation. An important feature of standard sequential auction models is that bilateral competitions are presumed to be efficient: i.e., the more productive firm always wins the auction. If firm productivity is time invariant, then conditioning on $\mathbf{j}(i,m)$ and $\mathbf{h}(i,m)$ is equivalent to conditioning on the productivity of the origin and destination firm, which given log-linear wage contracts implies the errors

$\{\varepsilon_{im}\}_{i\in[n],m\in[M_i]}$ are strictly exogenous.

6.1 Structural interpretation

Di Addario et al. (2023) show formally that the model of Bagger, Fontaine, et al. (2014), which nests the seminal model of Postel-Vinay and Robin (2002a) when consumption utility is assumed to be logarithmic, yields (10) as the reduced form for hiring wages. It is useful to review this argument both to understand the structural interpretation of the origin and destination fixed effects and the justification for the exogenous mobility assumption on the reduced form errors. The Bagger, Fontaine, et al. (2014) model implies that the log hiring wage offered by a firm of productivity level p, to a worker of productivity type ϵ, with labor market experience $\mathcal{X}$, who is currently employed at a firm of productivity q can be written as the generalized linear function

$$\alpha(\epsilon) + g(\mathcal{X}) + \psi(p) + \lambda(q) + \mathcal{E}.$$

Hires from unemployment follow the same equation with the productivity of the incumbent firm q set equal to the flow value of leisure b, which is assumed to be common for all workers.

The term $\alpha(\epsilon)$ is a worker fixed effect capturing general human capital, which is rewarded equally by all employers. Likewise, $g(\mathcal{X})$ captures the returns to experience, while the error term $\mathcal{E}$ captures idiosyncratic innovations to the worker's general human capital. By assumption, neither of these terms influence worker mobility, which depends solely on the firm productivities p and q. The destination firm effect, $\psi(p)$, equals $\beta \ln p + I(p,\beta)$, while the hiring origin effect, $\lambda(q)$, is given by $(1-\beta)\ln q - I(q,\beta)$, where $\beta \in [0,1]$ indexes worker bargaining strength. The function $I(p,\beta)$, which is decreasing in both its arguments and obeys $I(p,1) = 0$, captures the expected utility of the wage growth associated with moving from a firm with productivity p to the most productive firm in the economy. Hence, the difference $I(p,\beta) - I(q,\beta)$ captures the expected utility of the wage growth associated with moving from an incumbent firm with productivity q to a poaching firm with productivity p.

Inspection of these equations reveals that when β is small, the destination effect $\psi(p)$ will be decreasing in p, which can be interpreted as a compensating differential for the anticipated wage growth associated with moving. By contrast, for any value of $\beta < 1$, $\lambda(q)$ will be increasing in q, which reflects that it is more difficult to poach workers from firms that can afford to pay them more. When $\beta = 1$, the term $\lambda(q)$ becomes zero and the model reduces to a version of the AKM model with only destination firm effects. Remarkably, $\psi(p) + \lambda(p) = \ln p$ for any value of β, implying that a firm's productivity can be recovered by summing its origin and destination effects. Since workers view more productive firms as fundamentally more desirable than less productive firms, this sum recovers the ordering of the underlying "job ladder" in expected utility governing worker flows.

6.2 Testable restrictions

In the Bagger, Fontaine, et al. (2014) model, firms are differentiated only by productivity. Consequently, the origin and destination effects are deterministic functions of one another. Di Addario et al. (2023) show that it is possible to exploit this feature of the model to bound the bargaining power of workers using the excess variance of the destination effects over the origin effects. Letting $\mathbb{V}_p$ denote the variance across firms, the following bound on β is obtained by exploiting the fact that $\frac{\partial}{\partial \ln p} I(p, \beta) \in \left[-(1-\beta)^2/\beta, 0\right]$:

$$\beta \geq 1/2 + \frac{\mathbb{V}_p[\psi(p)] - \mathbb{V}_p[\lambda(p)]}{2\mathbb{V}_p[\psi(p) + \lambda(p)]}. \tag{11}$$

As discussed earlier, if β were very close to 1, we should expect the origin effects to be negligible and for destination effects to be large as workers extract from firms the greatest wage they can afford: $\ln p$. This bound formalizes the converse idea that when destination effects are large relative to origin effects, worker bargaining power must be strong. When $\beta > 1/2$, the following lower bound can be shown to hold on the correlation between the two dimensions of firm heterogeneity:

$$\operatorname{corr}(\psi(p), \lambda(p)) \geq \sqrt{\frac{\mathbb{V}_p[\psi(p)]}{\mathbb{V}_p[\psi(p) + \lambda(p)]}} \left(1 - \frac{3}{10}\sqrt{\frac{\mathbb{V}_p[\lambda(p)]}{\mathbb{V}_p[\psi(p) + \lambda(p)]}}\right).$$

Intuitively, when β is large, both the origin and destination effects must be strongly increasing in productivity, yielding a high correlation. However, a large β also yields relatively larger destination effects than origin effects. The correlation bound formalizes this link, effectively providing a test of the presence of a unidimensional firm hierarchy.[24]

6.3 It ain't where you're from, it's where you're at

Di Addario et al. (2023) fit (10) to Italian social security data using the average daily wage of each worker in their first year of employment with a firm as a proxy for their hiring wage. A poaching event is presumed to have taken place whenever a worker resigns from their job as opposed to being laid off or fired for cause. If the worker did not resign from their previous job, they are assumed to have been hired from unemployment. While there are reasons to suspect that stated resignations provide an imperfect proxy of when bilateral competition between firm pairs is taking place (McLaughlin 1991; Postel-Vinay and Turon 2014), Italian

[24] Roussille and Scuderi (2023) reject a unidimensional model of firm valuations in favor of a mixture model with three distinct hierarchies using data from an online job board for software engineers.

workers poached according to this criterion turn out to have much shorter durations of non-employment between jobs than workers involved in other sorts of separations and to exhibit lower chances of experiencing a wage cut when switching jobs.

Di Addario et al. (2023) find a roughly 3.5 log point gap between the estimated value of λ_U (the origin effect associated with unemployment) and the average origin effect of poached workers, $\mathbb{E}_n\left[\lambda_{\mathbf{h}(i,m)} \mid \mathbf{h}(i,m) \neq U\right]$, implying a modest penalty for being hired from unemployment. The bias-corrected variance of origin effects, $\mathbb{V}_n\left[\lambda_{\mathbf{h}(i,m)}\right]$, turns out to be extremely small, accounting for less than 1% of the variance of hiring wages across job movers. By contrast, the variance of destination effects, $\mathbb{V}_n\left[\psi_{\mathbf{j}(i,m)}\right]$, explains 24% of the variance of hiring wages.

As mentioned in section 4, variance shares can be somewhat difficult to interpret given that noise levels vary across samples. The estimated standard deviation of destination effects in their sample of job movers is roughly 0.26, which is only slightly above the typical bias-corrected standard deviation of AKM firm effects reported for the US and Italy in Figure (3). By contrast, the origin effects have a standard deviation among all job movers of 0.04 and a standard deviation of 0.08 among the roughly 1/3 of job transitions that involve poaching a worker from another firm.

While an 8% wage change is not negligible, this standard deviation of origin effects turns out to be far less than would be predicted by the Bagger, Fontaine, et al. (2014) model. The standard deviation across firms of the destination effects, $\mathbb{V}_p\left[\psi(p)\right]^{1/2}$, is 0.26 (the same as was found across workers), while the corresponding standard deviation of origin effects, $\mathbb{V}_p\left[\lambda(p)\right]^{1/2}$, is only 0.07. Applying the formula in (11) implies that $\beta \geq 0.88$. In addition to being intuitively implausible, this value of β would require an extremely high correlation between the origin and destination effects of 0.84. In practice, the bias-corrected correlation is only 0.25, indicating that the model cannot rationalize the covariance structure of the origin and destination effects under *any* distribution of firm productivities.

6.4 Information and conduct

The order of magnitude difference in scale between firm origin and destination effects suggests either that the identity of one's current employer doesn't convey much information about outside options at the time of a poaching attempt or that firms are unable (or unwilling) to tailor offers to those outside options. To assess the former hypothesis, one could collect more granular proxies of outside options. Perhaps interacting the identity of the incumbent firm with detailed job titles or tenure would be more predictive of hiring wages? The second possibility, that firms are not able or willing to tailor wage offers, is more difficult to evaluate. Firms often report having some latitude to tailor wages to worker circumstances and Caldwell, Haegele, and

Heining (2024a) provide evidence that wages are strongly related to previous firm pay among those firms that engage in bargaining. On the other hand, survey evidence suggests offer matching is rare empirically (Faberman et al. 2022; Caldwell, Haegele, and Heining 2024a). Moreover, receiving an outside offer does not seem to be associated with large wage gains on average (Guo 2023).

Even if firms typically do have the ability to tailor wages, the informational requirements of tying wage offers to best predictors of outside options are formidable. Sequential auction models are predicated on a perfect information benchmark where each firm knows the willingness to pay of the rival firm for the worker in question, leading them to offer a rival dependent wage.[25] By contrast, the famous Burdett and Mortensen (1998) model effectively assumes that firms know nothing about workers' outside options, which is why they offer the same wages to unemployed workers and workers searching on the job. As Postel-Vinay and Robin (2002b) acknowledge "reality lies somewhere in between our complete information story and Burdett's and Mortensen's incomplete information assumption."

How to think about this middle ground between wage posting and sequential auction models remains a frontier area of research. One approach is to view the economy as comprised of a mixture of wage posting firms ala Burdett and Mortensen (1998) and tailoring firms ala Postel-Vinay and Robin (2002a). While coherent models of this nature have been proposed (Postel-Vinay and Robin 2004; Flinn and Mullins 2017), empirical evidence on how wage setting conduct varies across employers remains in its infancy. A recurrent finding from estimation of these models is that counter offers and negotiation are more common among higher skilled workers (Caldwell and Harmon 2019; Flinn and Mullins 2021). This finding likely resonates among academic economists, many of whom have experienced the majority of their salary growth by receiving outside offers. Indeed, the sequential auction paradigm of bilateral competition appears to be a good one for academia, which is a hierarchical industry where employers have good information about the ability of rival institutions to compete for talent. It is unclear how many other labor markets are characterized by this sort of competition.

Breaking their variance decompositions down by industry, Di Addario et al. (2023) find that destination effects are orders of magnitude more variable than origin effects in most sectors of the Italian economy. The key exceptions are finance/banking and the legal sector, where origin and destination effects exhibit comparable variability. Both of these sectors are hierarchical and plausibly exhibit more information regarding the ability of firms to pay to retain workers than other sectors. The finance/banking industry is the only sector where the correlation bound is satisfied, suggesting perhaps that it too exhibits the sort of unidimensional competition described in sequential auction models. In less skilled sectors, by contrast, employers are likely

[25] Workers are also assumed to be fully informed about the match surplus available at the two rival firms. Jäger et al. (2024) provide evidence suggesting that workers at low wage firms tend to underestimate their outside options.

more difficult to rank. As a result, less information may be conveyed by the identity of one's previous employer. In these settings, worker outside options seem more likely to be private information, an idea that is central to the idea of monopsonistic models of wage setting.

7 Conclusion

While much has been learned about which firms pay high wages and their contribution to wage inequality, plenty of work remains. Some questions this review has touched upon that appear particularly ripe for exploration include:

1. **Dispersion and Development**: Why are firm wage effects more dispersed in less developed countries? One possibility is that labor market frictions are more pronounced in these economies, leading to greater misallocation. Another is that measurement differences, especially the prevalence of informal work, play a confounding role.

2. **Accounting for Cycles**: What accounts for the cyclic component of edge effects? Cycles could reflect either economic shocks shared by closely connected firms or differences in the sorts of workers moving along different parts of the mobility network. The former view has difficulty explaining the documented stability of firm effects. The latter interpretation suggests important dimensions of heterogeneity may have been missed by existing models of non-separable wages, estimates of which typically exhibit small departures from linearity.

3. **Intransitive Firms:** To what extent do firm rankings, in both wages and desirability, vary with worker and job characteristics? Does accounting for this heterogeneity amplify or mute the total contribution of firms to inequality?

4. **Hiring Origins and Conduct**: When and where do hiring origins matter for wage determination? Do markets where the dispersion of origin effects is larger exhibit greater wage effects of receiving outside offers? How does the reason for separation (e.g., ostensible quits vs layoffs) influence the degree of state dependence in wages?

5. **Worker Mobility Post-Layoff**: Why do mass layoffs sometimes lead workers to move to higher-wage firms? Does the prevalence of this behavior vary with labor market institutions?

6. **Understanding Network Structure**: What network formation models produce realistic mobility patterns? How effective are these models at predicting the next firm that will employ a worker? How do network-based definitions of labor markets align with workers' perceptions as measured in surveys?

7. **Reproducibility**: How can fixed effect estimates be shared most effectively? Transparency and replicability are crucial components of the data science revolution (Donoho 2024). Future work could en-

able wider access not only to point estimates but also to measures of uncertainty, lowering the barriers to downstream inference and prediction.

References

Abowd, J. M. and D. Card (1989). "On the Covariance Structure of Earnings and Hours Changes". In: *Econometrica* 57.2, pp. 411–445.

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002). *Computing person and firm effects using linked longitudinal employer-employee data*. Tech. rep. Center for Economic Studies, US Census Bureau.

Abowd, J. M., F. Kramarz, P. Lengermann, K. L. McKinney, and S. Roux (2012). "Persistent inter-industry wage differences: rent sharing and opportunity costs". In: *IZA Journal of Labor Economics* 1, pp. 1–25.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). "High wage workers and high wage firms". In: *Econometrica* 67.2, pp. 251–333.

Abowd, J. M., P. Lengermann, and K. L. McKinney (2003). *The measurement of human capital in the US economy*. Tech. rep. Citeseer.

Abowd, J. M. and K. L. McKinney (2023). "Mixed-Effects Methods for Search and Matching Research". In: *arXiv preprint arXiv:2308.15445*.

Abowd, J. M., K. L. McKinney, and I. M. Schmutte (2019). "Modeling endogenous mobility in earnings determination". In: *Journal of Business & Economic Statistics* 37.3, pp. 405–418.

Akerlof, G. A. and J. L. Yellen (1990). "The fair wage-effort hypothesis and unemployment". In: *The Quarterly Journal of Economics* 105.2, pp. 255–283.

Amemiya, T. (1978). "A note on a random coefficients model". In: *International Economic Review*, pp. 793–796.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171.3, pp. 673–697.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2012). "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias". In: *Economics Letters* 117.3, pp. 824–827.

Angrist, J. D. and W. K. Newey (1991). "Over-identification tests in earnings functions with fixed effects". In: *Journal of Business & Economic Statistics* 9.3, pp. 317–323.

Armstrong, T. B. and M. Kolesár (2018). "Optimal inference in a class of regression models". In: *Econometrica* 86.2, pp. 655–683.

Armstrong, T. B. and M. Kolesár (2021). "Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness". In: *Econometrica* 89.3, pp. 1141–1177.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014). "Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics". In: *American Economic Review* 104.6, pp. 1551–96.

Bagger, J. and R. Lentz (2019). "An empirical model of wage dispersion with sorting". In: *The Review of Economic Studies* 86.1, pp. 153–190.

Bassier, I. (2023). "Firms and inequality when unemployment is high". In: *Journal of Development Economics* 161, p. 103029.

Bassier, I., A. Dube, and S. Naidu (2022). "Monopsony in movers: The elasticity of labor supply to firm wage policies". In: *Journal of Human Resources* 57.S, S50–s86.

Bellmann, L., B. Lochner, S. Seth, and S. Wolter (2020). *AKM effects for German labour market data*. Tech. rep. Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for ...

Bertheau, A., E. M. Acabbi, C. Barcelo, A. Gulyas, S. Lombardi, and R. Saggio (2022). *The Unequal Cost of Job Loss across Countries*. Tech. rep. IZA DP No. 15033.

Bertheau, A., E. M. Acabbi, C. Barceló, A. Gulyas, S. Lombardi, and R. Saggio (2023). "The unequal consequences of job loss across countries". In: *American Economic Review: Insights* 5.3, pp. 393–408.

Bickel, P. J. and D. A. Freedman (1981). "Some asymptotic theory for the bootstrap". In: *The Annals of Statistics*, pp. 1196–1217.

Bloom, N., F. Guvenen, B. S. Smith, J. Song, and T. von Wachter (2018). "The disappearing large-firm wage premium". In: *AEA Papers and Proceedings*. Vol. 108, pp. 317–22.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2023). "How much should we trust estimates of firm effects and worker sorting?" In: *Journal of Labor Economics* 41.2, pp. 291–322.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). "A distributional framework for matched employer employee data". In: *Econometrica* 87.3, pp. 699–739.

Boza, I. and B. Reizer (2024). "The Role of Flexible Wage Components in Gender Wage Difference". In.

Bozzo, E. (2013). "The Moore–Penrose inverse of the normalized graph Laplacian". In: *Linear Algebra and its Applications* 439.10, pp. 3038–3043.

Brown, C., J. Hamilton, and J. L. Medoff (1990). *Employers large and small*. Harvard University Press.

Brown, C. and J. Medoff (1989). "The employer size-wage effect". In: *Journal of political Economy* 97.5, pp. 1027–1059.

Bruns, B. (2019). "Changes in workplace heterogeneity and how they widen the gender wage gap". In: *American Economic Journal: Applied Economics* 11.2, pp. 74–113.

Burdett, K. and D. T. Mortensen (1998). "Wage differentials, employer size, and unemployment". In: *International Economic Review*, pp. 257–273.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006). "Wage bargaining with on-the-job search: Theory and evidence". In: *Econometrica* 74.2, pp. 323–364.

Caldwell, S., I. Haegele, and J. Heining (2024a). "Bargaining in the Labor Market". In.

— (2024b). "Firm Pay and Worker Search". In.

Caldwell, S. and N. Harmon (2019). "Outside options, bargaining, and wages: Evidence from coworker networks". In: *Unpublished manuscript, Univ. Copenhagen*.

Cappelli, P. and K. Chauvin (1991). "An interplant test of the efficiency wage hypothesis". In: *The Quarterly Journal of Economics* 106.3, pp. 769–787.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). "Firms and labor market inequality: Evidence and some theory". In: *Journal of Labor Economics* 36.S1, S13–S70.

Card, D., A. R. Cardoso, and P. Kline (2016). "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women". In: *The Quarterly journal of economics* 131.2, pp. 633–686.

Card, D., J. Heining, and P. Kline (2013). "Workplace heterogeneity and the rise of West German wage inequality". In: *The Quarterly journal of economics* 128.3, pp. 967–1015.

Card, D., J. Rothstein, and M. Yi (2023). *Industry wage differentials: A firm-based approach*. Tech. rep. National Bureau of Economic Research.

— (2024). "Industry wage differentials: A firm-based approach". In: *Journal of Labor Economics* 42.S1, S11–S59.

Casarico, A. and S. Lattanzio (2024). "What firms do: Gender inequality in linked employer-employee data". In: *Journal of Labor Economics* 42.2, pp. 325–355.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018). "Inference in linear regression models with many covariates and heteroscedasticity". In: *Journal of the American Statistical Association* 113.523, pp. 1350–1361.

Chamberlain, G. (1982). "Multivariate regression models for panel data". In: *Journal of econometrics* 18.1, pp. 5–46.

— (1984). "Panel data". In: *Handbook of econometrics* 2, pp. 1247–1318.

Chen, M., I. Fernández-Val, and M. Weidner (2021). "Nonlinear factor models for network and panel data". In: *Journal of Econometrics* 220.2, pp. 296–324.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). "How does your kindergarten classroom affect your earnings? Evidence from Project STAR". In: *The Quarterly Journal of Economics* 126.4, pp. 1593–1660.

Chetty, R. and N. Hendren (2018). "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects". In: *The Quarterly Journal of Economics* 133.3, pp. 1107–1162.

Condorcet, M. d. (1785). "Essay on the Application of Analysis to the Probability of Majority Decisions". In: *Paris: Imprimerie Royale*.

Coudin, E., S. Maillard, and M. Tô (2018). *Family, firms and the gender wage gap in France*. Tech. rep. IFS Working Papers.

Crane, L. D., H. R. Hyatt, and S. M. Murray (2023). "Cyclical labor market sorting". In: *Journal of Econometrics* 233.2, pp. 524–543.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009). "Dealing with limited overlap in estimation of average treatment effects". In: *Biometrika* 96.1, pp. 187–199.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the royal statistical society: series B (methodological)* 39.1, pp. 1–22.

Di Addario, S., P. Kline, R. Saggio, and M. Sølvsten (2023). "It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages". In: *Journal of Econometrics* 233.2, pp. 340–374.

Donoho, D. (2024). "Data science at the singularity". In: *Harvard Data Science Review* 6.1.

Dostie, B., J. Li, D. Card, and D. Parent (2023). "Employer policies and the immigrant–native earnings gap". In: *Journal of Econometrics* 233.2, pp. 544–567.

Drenik, A., S. Jäger, P. Plotkin, and B. Schoefer (2023). "Paying outsourced labor: Direct evidence from linked temp agency-worker-client data". In: *Review of Economics and Statistics* 105.1, pp. 206–216.

Dustmann, C., B. Fitzenberger, U. Schönberg, and A. Spitz-Oener (2014). "From sick man of Europe to economic superstar: Germany's resurgent economy". In: *Journal of economic perspectives* 28.1, pp. 167–188.

Dustmann, C., A. Lindner, U. Schönberg, M. Umkehrer, and P. Vom Berge (2022). "Reallocation effects of the minimum wage". In: *The Quarterly Journal of Economics* 137.1, pp. 267–328.

Eeckhout, J. and P. Kircher (2011). "Identifying sorting—in theory". In: *The Review of Economic Studies* 78.3, pp. 872–906.

El Karoui, N. and E. Purdom (2018). "Can we trust the bootstrap in high-dimensions? the case of linear models". In: *The Journal of Machine Learning Research* 19.1, pp. 170–235.

Engbom, N. and C. Moser (2022). "Earnings inequality and the minimum wage: Evidence from Brazil". In: *American Economic Review* 112.12, pp. 3803–3847.

Engbom, N., C. Moser, and J. Sauermann (2023). "Firm pay dynamics". In: *Journal of Econometrics* 233.2, pp. 396–423.

Faberman, R. J., A. I. Mueller, A. Şahin, and G. Topa (2022). "Job search behavior among the employed and non-employed". In: *Econometrica* 90.4, pp. 1743–1779.

Fenizia, A. (2022). "Managers and productivity in the public sector". In: *Econometrica* 90.3, pp. 1063–1084.

Finkelstein, A., M. Gentzkow, and H. Williams (2016). "Sources of geographic variation in health care: Evidence from patient migration". In: *The Quarterly Journal of Economics* 131.4, pp. 1681–1726.

Firth, D. and R. X. De Menezes (2004). "Quasi-variances". In: *Biometrika* 91.1, pp. 65–80.

Flinn, C. and J. Mullins (2017). *Firms' Choices of Wage-Setting Protocols in the Presence of Minimum Wages*. Tech. rep. Discussion paper, New York University.

— (2021). *Firms' choices of wage-setting protocols*. Tech. rep. Discussion paper, New York University.

Forgy, E. W. (1965). "Cluster analysis of multivariate data: efficiency versus interpretability of classifications". In: *biometrics* 21, pp. 768–769.

Garcia-Louzao, J. and A. Ruggieri (2023). "Labor market competition and inequality". In.

Gerard, F., L. Lagos, E. Severnini, and D. Card (2021). "Assortative matching or exclusionary hiring? the impact of employment and pay policies on racial wage differences in brazil". In: *American Economic Review* 111.10, pp. 3418–3457.

Gibbons, R. and L. Katz (1992). "Does unmeasured ability explain inter-industry wage differentials?" In: *The Review of Economic Studies* 59.3, pp. 515–535.

Gibbons, R., L. F. Katz, T. Lemieux, and D. Parent (2005). "Comparative advantage, learning, and sectoral wage determination". In: *Journal of labor economics* 23.4, pp. 681–724.

Goldschmidt, D. and J. F. Schmieder (2017). "The rise of domestic outsourcing and the evolution of the German wage structure". In: *The Quarterly Journal of Economics*, qjx008.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2022). *Contamination bias in linear regressions*. Tech. rep. National Bureau of Economic Research.

Gottfries, A. and G. Jarosch (2023). "Dynamic monopsony with large firms and an application to non-competes". In.

Groshen, E. L. (1991). "Sources of intra-industry wage dispersion: How much do employers matter?" In: *The Quarterly Journal of Economics* 106.3, pp. 869–884.

Guo, J. (2023). "The Response of Wages to Rejected Offers". In.

Guo, N., L. Zhang, R. Zhang, and B. Zou (2024). "Migration Restrictions and the Migrant-Native Wage Gap: The Role of Wage Setting and Sorting". In.

Haanwinckel, D. (2023). *Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution*. Tech. rep. National Bureau of Economic Research.

Haltiwanger, J., H. R. Hyatt, and J. R. Spletzer (2024). "Rising Top, Falling Bottom: Industries and Rising Wage Inequality". In: *American Economic Review* 114.10, pp. 3250–3283.

Hanushek, E. A. (1974). "Efficient estimators for regressing regression coefficients". In: *The American Statistician* 28.2, pp. 66–67.

Hoaglin, D. C. and R. E. Welsch (1978). "The hat matrix in regression and ANOVA". In: *The American Statistician* 32.1, pp. 17–22.

Holzer, H. J., L. F. Katz, and A. B. Krueger (1991). "Job Queues and Wages". In: *Quarterly Journal of Economics* 106.3, pp. 739–768.

Hsieh, C.-T. and P. J. Klenow (2009). "Misallocation and manufacturing TFP in China and India". In: *The Quarterly journal of economics* 124.4, pp. 1403–1448.

Huitfeldt, I., A. R. Kostøl, J. Nimczik, and A. Weber (2023). "Internal labor markets: A worker flow approach". In: *Journal of Econometrics* 233.2, pp. 661–688.

Imbens, G. W. (2010). "Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)". In: *Journal of Economic literature* 48.2, pp. 399–423.

Jäger, S., C. Roth, N. Roussille, and B. Schoefer (2024). "Worker beliefs about outside options". In: *The Quarterly Journal of Economics*, qjae001.

Jarosch, G., J. S. Nimczik, and I. Sorkin (2024). "Granular search, market structure, and wages". In: *Review of Economic Studies*, rdae004.

Jochmans, K. and M. Weidner (2019). "Fixed-effect regressions on network data". In: *Econometrica* 87.5, pp. 1543–1560.

Katz, L. F., L. H. Summers, R. E. Hall, C. L. Schultze, and R. H. Topel (1989). "Industry rents: Evidence and implications". In: *Brookings Papers on Economic Activity. Microeconomics* 1989, pp. 209–290.

Kline, P., R. Saggio, and M. Sølvsten (2020). "Leave-out estimation of variance components". In: *Econometrica* 88.5, pp. 1859–1898.

Koutis, I., G. L. Miller, and D. Tolliver (2011). "Combinatorial preconditioners and multilevel solvers for problems in computer vision and image processing". In: *Computer Vision and Image Understanding* 115.12, pp. 1638–1646.

Krueger, A. B. and L. H. Summers (1988). "Efficiency wages and the inter-industry wage structure". In: *Econometrica: Journal of the Econometric Society*, pp. 259–293.

Lachowska, M., A. Mas, R. Saggio, and S. A. Woodbury (2023). "Do firm effects drift? Evidence from Washington administrative data". In: *Journal of Econometrics* 233.2, pp. 375–395.

Lachowska, M., A. Mas, and S. A. Woodbury (2020). “Sources of displaced workers’ long-term earnings losses”. In: *American Economic Review* 110.10, pp. 3231–3266.

Lagos, L. (2024). “Union Bargaining Power and the Amenity-Wage Tradeoff”. In: *IZA Working Paper No. 17034*.

Lamadon, T., M. Mogstad, and B. Setzler (2022). “Imperfect competition, compensating differentials, and rent sharing in the US labor market”. In: *American Economic Review* 112.1, pp. 169–212.

Lehmann, T. (2023). “Non-wage job values and implications for inequality”. In: *Available at SSRN 4373816*.

Lei, L. and B. Ross (2023). “Estimating Counterfactual Matrix Means with Short Panel Data”. In: *arXiv preprint arXiv:2312.07520*.

Lentz, R., S. Piyapromdee, and J.-M. Robin (2022). “The Anatomy of Sorting-Evidence from Danish Data”. In.

Li, J., B. Dostie, and G. Simard-Duplain (2023). “Firm pay policies and the gender earnings gap: the mediating role of marital and family status”. In: *ILR Review* 76.1, pp. 160–188.

Lloyd, S. (1982). “Least squares quantization in PCM”. In: *IEEE transactions on information theory* 28.2, pp. 129–137.

MacKinnon, J. G. and H. White (1985). “Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties”. In: *Journal of econometrics* 29.3, pp. 305–325.

MaCurdy, T. E. (1982). “The use of time series processes to model the error structure of earnings in a longitudinal data analysis”. In: *Journal of econometrics* 18.1, pp. 83–114.

Maestas, N., K. J. Mullen, D. Powell, T. Von Wachter, and J. B. Wenger (2023). “The value of working conditions in the United States and implications for the structure of wages”. In: *American Economic Review* 113.7, pp. 2007–2047.

Mammen, E. (1993). “Bootstrap and wild bootstrap for high dimensional linear models”. In: *The annals of statistics*, pp. 255–285.

Mas, A. (2024). “Compensating differentials”. In: *Handbook of Labor Economics*.

McLaughlin, K. J. (1991). “A theory of quits and layoffs with efficient turnover”. In: *Journal of Political Economy* 99.1, pp. 1–29.

Meghir, C. and L. Pistaferri (2004). “Income variance dynamics and heterogeneity”. In: *Econometrica* 72.1, pp. 1–32.

Murphy, K. M. and R. H. Topel (1990). “Efficiency wages reconsidered: Theory and evidence”. In: *Advances in the Theory and Measurement of Unemployment*. Springer, pp. 204–240.

Newey, W. K. and J. R. Robins (2018). “Cross-fitting and fast remainder rates for semiparametric estimation”. In: *arXiv preprint arXiv:1801.09138*.

Nimczik, J. S. (2017). "Job mobility networks and endogenous labor markets". In.

Oi, W. Y. and T. L. Idson (1999). "Firm size and wages". In: *Handbook of labor economics* 3, pp. 2165–2214.

Page, L., S. Brin, R. Motwani, T. Winograd, et al. (1999). "The pagerank citation ranking: Bringing order to the web". In.

Pérez Pérez, J. and J. G. Nuño-Ledesma (2022). *Workers, workplaces, sorting, and wage dispersion in Mexico*. Tech. rep. Working Papers.

Postel-Vinay, F. and J.-M. Robin (2002a). "Equilibrium wage dispersion with worker and employer heterogeneity". In: *Econometrica* 70.6, pp. 2295–2350.

— (2002b). "The distribution of earnings in an equilibrium search model with state-dependent offers and counteroffers". In: *International Economic Review* 43.4, pp. 989–1016.

— (2004). "To match or not to match?: Optimal wage policy with endogenous worker search intensity". In: *Review of Economic Dynamics* 7.2, pp. 297–330.

Postel-Vinay, F. and H. Turon (2014). "The Impact of Firing Restrictions on Labour Market Equilibrium in the Presence of On-the-job Search". In: *The Economic Journal* 124.575, pp. 31–61.

Rambachan, A. and J. Roth (2023). "A more credible approach to parallel trends". In: *Review of Economic Studies* 90.5, pp. 2555–2591.

Roussille, N. and B. Scuderi (2023). "Bidding for Talent: A Test of Conduct in a High-Wage Labor Market". In.

Roy, A. D. (1951). "Some thoughts on the distribution of earnings". In: *Oxford economic papers* 3.2, pp. 135–146.

Schmieder, J. F., T. Von Wachter, and J. Heining (2023). "The costs of job displacement over the business cycle and its sources: evidence from Germany". In: *American Economic Review* 113.5, pp. 1208–1254.

Shapiro, C. and J. E. Stiglitz (1984). "Equilibrium unemployment as a worker discipline device". In: *The American economic review* 74.3, pp. 433–444.

Slichter, S. H. (1950). "Notes on the Structure of Wages". In: *The Review of Economics and Statistics* 32.1, pp. 80–91.

Sockin, J. (2022). "Show Me the Amenity: Are Higher-Paying Firms Better All Around?" In.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019). "Firming up inequality". In: *The Quarterly journal of economics* 134.1, pp. 1–50.

Sorkin, I. (2018). "Ranking firms using revealed preference". In: *The quarterly journal of economics* 133.3, pp. 1331–1393.

Sorkin, I. and M. Wallskog (2023). "The slow diffusion of earnings inequality". In: *Journal of Labor Economics* 41.S1, S95–S127.

Spielman, D. (2019). *Spectral and algebraic graph theory.* URL: http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf.

Stigler, G. J. (1962). "Information in the labor market". In: *Journal of political economy* 70.5, Part 2, pp. 94–105.

Vafa, K., E. Palikot, T. Du, A. Kanodia, S. Athey, and D. M. Blei (2022). "CAREER: A Foundation Model for Labor Sequence Data". In: *arXiv preprint arXiv:2202.08370*.

Walters, C. R. (2024). "Empirical Bayes Methods in Labor Economics". In: *Handbook of Labor Economics*.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". In: *Econometrica: journal of the Econometric Society*, pp. 817–838.

Yeh, C., C. Macaluso, and B. Hershbein (2022). "Monopsony in the US labor market". In: *American Economic Review* 112.7, pp. 2099–2138.

Young, P. (1995). "Optimal voting rules". In: *Journal of Economic Perspectives* 9.1, pp. 51–64.

Appendix: Covariance between person and firm effects

Here, I detail how to construct an unbiased estimator of $\mathbb{C}_\varepsilon\left[\hat{\alpha}_i, \hat{\psi}_j\right]$ for each (i, j) pair in $[N] \times [J-1]$ that can be used to bias correct the covariance. I then discuss how bounds can be formed on the covariance.

From (6), we can write the OLS estimators

$$\hat{\alpha} = \alpha + \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \tilde{\boldsymbol{D}}_m\right)^{-1} \sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \varepsilon_m,$$

$$\hat{\psi} = \psi + \left(\sum_{m\in[M]} \tilde{\boldsymbol{F}}'_m \tilde{\boldsymbol{F}}_m\right)^{-1} \sum_{m\in[M]} \tilde{\boldsymbol{F}}'_m \varepsilon_m,$$

where $\tilde{\boldsymbol{D}}_m$ is the matrix of worker indicators after having partialled out the matrix of firm indicators. Hence,

$$\begin{aligned}
(\hat{\alpha} - \alpha)\left(\hat{\psi} - \psi\right)' &= \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \tilde{\boldsymbol{D}}_m\right)^{-1} \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \varepsilon_m\right) \left(\sum_{m\in[M]} \tilde{\boldsymbol{F}}'_m \varepsilon_m\right)' \left(\sum_{m\in[M]} \tilde{\boldsymbol{F}}'_m \tilde{\boldsymbol{F}}_m\right)^{-1} \\
&= \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \tilde{\boldsymbol{D}}_m\right)^{-1} \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \varepsilon_m \varepsilon'_m \tilde{\boldsymbol{F}}_m\right) \left(\sum_{m\in[M]} \tilde{\boldsymbol{F}}'_m \tilde{\boldsymbol{F}}_m\right)^{-1} \\
&+ \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \tilde{\boldsymbol{D}}_m\right)^{-1} \left(\sum_{m\in[M]} \sum_{l\neq m} \tilde{\boldsymbol{D}}'_m \varepsilon_m \varepsilon'_l \tilde{\boldsymbol{F}}_l\right) \left(\sum_{m\in[M]} \tilde{\boldsymbol{F}}'_m \tilde{\boldsymbol{F}}_m\right)^{-1}.
\end{aligned}$$

Independence across matches implies that the final line has expectation zero, allowing us to write

$$\mathbb{E}_\varepsilon\left[(\hat{\alpha} - \alpha)\left(\hat{\psi} - \psi\right)'\right] = \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \tilde{\boldsymbol{D}}_m\right)^{-1} \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \Omega_m \tilde{\boldsymbol{F}}_m\right) \left(\sum_{m\in[M]} \tilde{\boldsymbol{F}}'_m \tilde{\boldsymbol{F}}_m\right)^{-1}.$$

We can estimate this covariance matrix with

$$\hat{\mathbb{E}}_\varepsilon\left[(\hat{\alpha} - \alpha)\left(\hat{\psi} - \psi\right)'\right] = \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \tilde{\boldsymbol{D}}_m\right)^{-1} \left(\sum_{m\in[M]} \tilde{\boldsymbol{D}}'_m \hat{\Omega}_m \tilde{\boldsymbol{F}}_m\right) \left(\sum_{m\in[M]} \tilde{\boldsymbol{F}}'_m \tilde{\boldsymbol{F}}_m\right)^{-1}.$$

The lower triangle of this estimated matrix gives the relevant unbiased estimators $\hat{\mathbb{C}}_\varepsilon\left[\hat{\alpha}_i, \hat{\psi}_j\right]$ of $\mathbb{C}_\varepsilon\left[\hat{\alpha}_i, \hat{\psi}_j\right]$. The debiased estimator of covariance between person and firm effects is:

$$\hat{\theta}_{\alpha,\psi} = \mathbb{E}_n\left[\hat{\alpha}_i \hat{\psi}_{\mathbf{j}(i,t)} - \hat{\mathbb{C}}_\varepsilon\left[\hat{\alpha}_i, \hat{\psi}_{\mathbf{j}(i,t)}\right]\right].$$

To bound this covariance in the broader connected sample that is not leave-out connected, we can again apply the bound $0 \leq \mathbf{\Omega}_m \leq \mathbb{E}\left[Y_m Y_m'\right]$. Upwardly and downwardly biased estimators of the relevant covariances $\mathbb{E}_\varepsilon\left[(\hat{\alpha}-\alpha)\left(\hat{\psi}-\psi\right)'\right]$ are obtained by replacing $\hat{\mathbf{\Omega}}_m$ with $Y_m Y_m'$ or zero, respectively, in just-connected matches contributing to $\hat{\mathbb{E}}_\varepsilon\left[(\hat{\alpha}-\alpha)\left(\hat{\psi}-\psi\right)'\right]$. One then applies the bias correction formula above replacing $\hat{\mathbb{C}}_\varepsilon\left[\hat{\alpha}_i, \hat{\psi}_{\mathbf{j}(i,t)}\right]$ with either its upwardly or downwardly biased estimate.

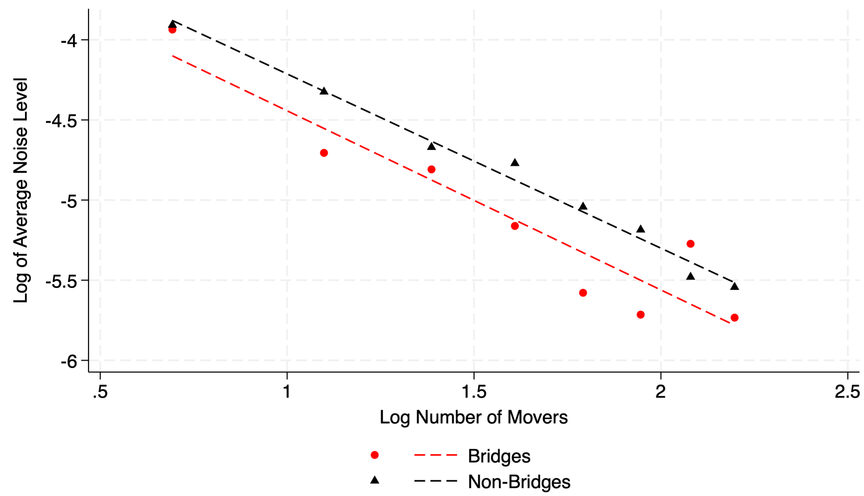

Figure A.1: Noise level by number of movers per edge

Notes: The vertical axis depicts the natural logarithm of the average of $\hat{\mathbb{V}}\left[\hat{\Delta}_\ell\right]$ among edges with a given number of movers. Number of movers considered ranges from 2 to 10. The line of best fit (depicted above) has intercept -3.325 and slope -1.118 when fit on edges that are bridges and an intercept of -3.126 and slope -1.087 when fit on edges that are not bridges. Using these lines of best fit to impute the variance of singleton edges, the imputed noise level of singleton bridges is $\exp(-3.325) \approx 0.036$ and the imputed noise level for singleton edges that are not bridges is $\exp(-3.12) \approx 0.044$.

Study	Source	Country / Region	Years	Total Var	Worker Var	Firm Var	2*Cov
Kline, Saggio, and Sølvsten 2020	Table 2	Veneto, Italy	1999, 2001	0.184	0.112	0.024	0.029
Lachowska, Mas, Saggio, et al. 2023	Table 3	Washington State	2002-2014	0.407	0.250	0.043	0.075
	Table 2	Washington State	2002-2003	0.360	0.285	0.032	0.024
	Table 2	Washington State	2013-2014	0.426	0.333	0.025	0.052
Engbom and Moser 2022	Table 2	Brazil	1994-1998	0.709	0.176	0.187	0.120
		Brazil	2014-2018	0.444	0.154	0.072	0.070
Haanwinckel 2023	Table 1	Brazil	1998-2001	0.688	0.419	0.116	0.098
		Brazil	2011-2013	0.577	0.384	0.056	0.097
Engbom, Moser, and Sauermann 2023	Table 3	Sweden	1985-2016	0.268	0.079	0.032	0.005
Bassier 2023	Table 2	South Africa	2011-2016	1.320	0.488	0.370	0.106
Casarico and Lattanzio 2024	Table D.2	Italy, Men	1995-2015	0.176	0.070	0.029	0.043
		Italy, Women	1995-2015	0.160	0.047	0.041	0.024
Gerard et al. 2021	Table 2	Brazil, white men	2002-2014	0.449	0.163	0.073	0.102
		Brazil, non-white men	2002-2014	0.332	0.100	0.054	0.058
		Brazil, white women	2002-2014	0.498	0.219	0.075	0.123
		Brazil, non-white women	2002-2014	0.324	0.144	0.045	0.060
Di Addario et al. 2023	Table 3 / A.3	Italy	2005-2015	0.279	0.083	0.066	0.047
Garcia-Louzao and Ruggieri 2023	Table 2	Lithuania	2000-2020	0.595	0.156	0.171	0.053
Pérez Pérez and Nuño-Ledesma 2022	Table A.4	Mexico	2004-2008	0.596	0.252	0.193	0.099
		Mexico	2009-2013	0.627	0.234	0.226	0.111
		Mexico	2014-2018	0.628	0.220	0.234	0.121
Abowd and McKinney 2023	Table 4	LEHD	2012-2014	3.330	0.086	0.037	0.023
		LEHD	1994-2017	3.423	0.017	0.038	0.018
Guo et al. 2024	Table 2	China, Native	2006-2014	0.530	0.219	0.125	0.054
		China, Migrant	2006-2014	0.572	0.145	0.204	0.081

Table A.1: Studies included in Figure 3