Journal of Political Economy

Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence

--Manuscript Draft--

Manuscript Number:	2009599R4
Full Title:	Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence
Article Type:	Regular Article
Keywords:	labor; Macroeconomics; Search; sorting; complementarities; heterogeneity; firm wage differentials
Corresponding Author:	Rafael Lopes de Melo, Ph.D. UNITED STATES
Order of Authors:	Rafael Lopes de Melo, Ph.D.
Abstract:	Why do firms pay different wages? Empirical evidence suggests presence of substantial differences in firm pay controlling for worker skill. Moreover, these differences are uncorrelated with skills, indicating absence of sorting. We show that the face value interpretation is inconsistent with evidence on coworker segregation. We interpret the evidence by applying a sorting model and show that the correlation is biased. We identify non-monotonicities in wages as the reason for this bias and show that a measure of worker-coworker sorting is more accurate. By calibrating the model to US data, we confirm that the model matches many job market characteristics.
Response to Reviewers:	N/A

Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence *

Rafael Lopes de Melo

University of Edinburgh

October 10, 2016

*An earlier version of this paper was part of my dissertation, and had the title "Sorting in the Labor Market: Theory and Measurement". I am indebted to Giuseppe Moscarini for the extensive advice and support. I would like to thank Robert Shimer, Fabian Lange, Fabien Postel-Vinay, Björn Brügemann, Jeremy Lise, Tzuo Hann Law and Iourii Manovskii for their valuable comments and help. Very useful feedback was also received from participants at the seminars at Yale, UCL, University of Chicago, University of Chicago School of Business, Northwestern University, Princeton University, University of Pennsylvania, Boston University, University of Western Ontario, University of Montreal, CREI, SED, LACEA, PUC-Rio, and Ipea. I alone am responsible for any shortcomings in the paper.

Abstract

Why do firms pay different wages? Empirical evidence suggests presence of substantial differences in firm pay controlling for worker skill. Moreover, these differences are uncorrelated with skills, indicating absence of sorting. We show that the face value interpretation is inconsistent with evidence on coworker segregation. We interpret the evidence by applying a sorting model and show that the correlation is biased. We identify non-monotonicities in wages as the reason for this bias and show that a measure of worker-coworker sorting is more accurate. By calibrating the model to US data, we confirm that the model matches many job market characteristics.

1 Introduction

It is well documented that a substantial share of wage dispersion takes place between firms. For example, Davis and Haltiwanger (1991) performed a variance decomposition of US manufacturing wages and found that around 60% of dispersion occurs between, rather than within firms. Similarly, the documented large firm-size wage differentials (e.g. Oi and Idson (1999)) and firm industry differentials (e.g. Krueger and Summers (1988)) have been the focus of a number of studies. These differentials could be attributed to differences in worker skill levels across firms as well as discrepancies in wage policies between firms. The latter could arise due to non-competitive features of labor market. Abowd, Kramarz and co-authors (1999, 2002 and 2004 - hereafter, AKM) attempted to quantify the relative importance of worker versus firm components in determining wages. The authors focused on estimating a wage regression that includes both worker and firm fixed effects, which was achieved through the use of longitudinal matched employer employee datasets. Their methodology was subsequently applied to the datasets of several different countries with consistent results. Authors of extant works based on this approach reported substantial firm fixed effects differentials, which account for a sizable share of firm-size differentials and industry wage differentials. Another robust result arising from this methodology is that the correlation between the two sets of fixed effects is close to zero, or sometimes even negative. Card, Heining and Kline (2013) recently quantified how changes in the dispersion of worker and firm fixed effects, and their correlation (referred to as sorting), explains changes in inequality over time.

Our work aims to provide structural interpretation of the facts obtained by applying the AKM methodology, using explicit models of labor market dynamics. We examine models that include both worker and firm heterogeneity, as well as a job matching process that yields mobility and assortative matching. Our first aim is to make explicit the connection between the AKM model and the widely used equilibrium job search model, which we label "the piece-rate model". The model features a dynamic matching process between heterogeneous workers and firms in a frictional environment, and has been shown to successfully explain

several characteristics pertaining to labor market transitions and wage dynamics. According to its postulates, the frictions in the economy cause more productive firms to pay higher wages to identical workers, resulting in a self-selection process of workers into firms. However, this process is independent of worker skill level, which implies absence of sorting in equilibrium. As a result, this model provides a "face value" interpretation for the facts. More specifically, the worker fixed effects capture differences in skill levels across individual employees, whereas the firm fixed effects capture wage differentials paid by more productive companies due to frictions. Under this interpretation, the high dispersion of firm fixed effects found in the data is indicative of a sizable degree of frictional wage dispersion. Moreover, the fact that the correlation between fixed effects is close to zero is supportive of an equilibrium with no sorting between worker and firm types.

Our first contribution stems from demonstrating that the piece-rate model is inconsistent with new evidence, which we sourced from a Brazilian matched employer-employee dataset. Applying the AKM methodology, we obtain results largely similar to those reported in the extant literature with respect to the facts already described. In addition, we also provide evidence of a high correlation between the fixed effects of workers and their coworkers, which suggests a strong degree of worker segregation among firms.[1] This violates the model's implication that the distribution of workers is the same across firms. Furthermore, we use data on education, occupation, sectors and locations to demonstrate that this correlation is not driven by composition of observables, a possibility that has been allowed for in empirical applications of the model.

After presenting the evidence of coworker sorting, we shift our focus toward a variant of the job search model that actually features assortative matching in equilibrium. A clear challenge stems from the need for the model to account for the zero correlation between fixed effects while permitting sorting in equilibrium. Thus, we build on the frictional matching

[1] Bagger and Lentz (2014) computed our measure on Danish data, and Tzuo Hann Law computed it on German data, both obtaining similar results to ours. We also relate our findings to the ones in Iranzo, Schivardi and Tosetti (2008), who report a measure of segregation of worker skill.

model with assortative matching proposed by Shimer and Smith (2000). The main feature we add to the model is the addition of the on-the-job search (both voluntary and involuntary), which is a pervasive feature of labor markets and is shown to substantially affect the extent of frictional wage dispersion in the model (e.g. Hornstein, Krusell and Violante (2011)).

The proposed model incorporates an equilibrium wage function that does not conform to the AKM specification. In AKM, firms give an extra bonus to all their employees (the fixed effect), whereas our model yields wages that are non-monotone with respect to firm productivity. In the sorting economy, each worker's skills and preferences can be matched with an ideal firm type, which is reflected in wages. If a worker gains a position in a more productive firm than his/her ideal one (because of frictions), he/she will earn less because of the need to compensate the more productive firm for giving up on finding a more skilled worker. These non-monotonicities have two implications with respect to how the model maps into the AKM methodology. First, since they distort the mapping between firm types and firm fixed effects, the correlation between fixed effects no longer captures correctly the extent of sorting in the economy. More specifically, even if the model allows for a large degree of sorting, this empirical measure may still be negligible. A similar finding was reported independently by Lise, Meghir and Robin (2016) and Bagger and Lentz (2014). However, to the best of our knowledge, our work is the first to relate this finding to the non-monotonicities in wages.[2] Second, the same non-monotonicity that helps explain the correlation between fixed effects may cause problems when attempting to elucidate the observed variance of firm fixed effects. More specifically, since each worker has a different ideal firm type, a given firm will increase the wage of some workers (relative to another firm), and decrease the wage of others, which may lead to a small average "firm effect". To our knowledge, we are the first to make this observation.

We calibrate the model by choosing a set of parameters in order to match a number of features of the US economy. We show that the model is able to explain a number of

[2]Eeckhout and Kircher (2011) show that this mechanism does not rely on frictions, and applies to the frictionless matching model as well.

facts about labor market transitions, the distribution of wages and wage dynamics. The model also matches well the correlations between fixed effects in AKM, generating a sizable correlation between workers and coworkers, while generating a correlation between workers and firms that is negligible. However, the model generates less variance in firm effects than can be discerned from the data analysis. This is consistent with the tension induced by the non-monotonicity described earlier. Nonetheless, we perform a second and much less parsimonious calibration, verifying that the results are robust to the more flexible parametrization. This leads to the conclusion that neither the piece-rate model nor this version of the sorting model is fully consistent with the reduced form evidence, suggesting that additional features should be incorporated. We discuss some possibilities for the modification of our model in the conclusion. However, we note a route we pursue in a companion paper (Lopes de Melo (2015)) as a particularly fruitful one, as we extend the Shimer and Smith model to include a second dimension of firm heterogeneity: compensating differentials.

The remainder of the paper is structured as follows. section 2 summarizes the related literature, while the basics of the piece-rate model are presented in section 3, where we also summarize the empirical evidence and describe our own work. section 4 is designated for the detailed description of the theoretical model. In section 5, we present the model calibration, aiming to match its output to a number of labor market facts. Finally, section 6 concludes the paper, along with with a discussion of future research.

2 Related Literature

The work described here relates to several strands of literature. First, the essence of the empirical methodology used throughout the paper is derived from the extensive work of Abowd, Kramarz and co-authors (1999, 2002 and 2004), described earlier in this paper.

Second, our model is grounded in the theoretical assignment literature. Becker (1973) presents a model with two-sided heterogeneity, demonstrating that, in the presence of com-

plementarities in production, the equilibrium exhibits perfect sorting. Shimer and Smith (2000) introduced search frictions (via random search) in the Becker model. This approach adds noise to the equilibrium allocations and thus requires stronger complementarities for the equilibrium to exhibit positive assortative matching. Eeckhout and Kircher (2010) developed a directed search model that can be viewed as an intermediate case between the frictionless model and the economy with search frictions. That framework requires fewer complementarities than the search case, but more than the frictionless economy, to induce positive sorting. This approach is related to the work reported by Shi (2001).

Thirdly, authors of a recent stream of studies applied models with firm and worker heterogeneity and matching to study labor markets. In some of these works, matching models with worker and firm heterogeneity are proposed while failing to capture assortative matching in equilibrium. Notable examples of studies based on this framework are those conducted by Christensen et al. (2005), Barlevy (2008), and Postel-Vinay and Robin (2002), among many others.

Finally, another recent stream of research is based on models that allow for incorporating more elaborate patterns of matching between workers and firms to understand labor markets. We believe that our work should be classified in this category. In one of such studies, Bagger and Lentz (2014) formulate a search model with endogenous search intensity, and structurally estimate it using the Danish matched employer-employee dataset. This strategy requires weaker complementarities in production to achieve positive sorting, characterized by a highly skilled worker searching with a higher intensity. In their work, Lise, Meghir and Robin (2016) build on the model developed by Shimer and Smith (2000), introducing on-the-job search, a wage mechanism akin to that introduced by Cahuc, Postel-Vinay and Robin (2006), and a dynamic process that causes firms to change productivity over time. These authors estimate the model using NLSY data and refer to its output to explain certain features of wage dynamics and labor market transitions. In related work, Eeckhout and Kircher (2011) used variations of the Becker (1973) to argue that, by analyzing the wage

data alone, it is not possible to distinguish a model that features positive sorting from that based on negative sorting. However, they also argue that wages can provide information about the strength of sorting, which is related to the underlying mechanisms of the worker-coworker measure described in our paper. Finally, Hagedorn, Law and Manovskii (2012) provide results pertaining to the non-parametric identification of the model developed by Shimer and Smith (2000) using matched employer-employee data.

3 Wage Dispersion and Sorting at the Firm Level: the Old View

In this section, we present facts related to wage dispersion and sorting at the firm level, as well as the "face value" interpretation of these facts. First, we introduce a matching model with frictions that features no sorting in equilibrium, in line with models that are widely used in the related literature. The model yields a number of testable predictions and its structure is directly related to the wage decomposition of AKM, which we make explicit. We proceed with summarizing the related empirical evidence, as well as presenting our own results obtained using a Brazilian matched employer-employee dataset. We argue that the "face value" interpretation is incongruent with the facts, which motivates the analyses and discussions in the remainder of the paper.

3.1 A Model of Matching Model with Heterogeneity and no Sorting

This model is based on the islands model developed by Lucas and Prescott (1974), whereby we reinterpret islands as firms.The model is set in continuous time and is populated by a unit mass of workers, each with a type x drawn from distribution $L(x)$. At each point in time, the worker is either "in transit" or occupies an island of type y. Workers in transit are unemployed

and earn a flow benefit xb. These workers sample islands at rate λ^0 by taking draws from a distribution $G(y)$. Within islands, there are competitive labor markets, where workers immediately find work and earn their marginal product $w(x,y) = F(x,y) = xy$. Employed workers have the opportunity to move to island y' at rate λ^1, where the new island is drawn from the same distribution $G(y')$. These workers also face a risk of exogenous displacement at rate δ. We can express the value function of unemployed, $U(x)$, and employed workers, $W(x,y)$ as follows:

$$\begin{aligned} rU(x) &= xb + \lambda^0 \int [W(x,y') - U(x)]\, dG(y') \\ (r+\delta) W(x,y) &= xy + \delta U(x) + \lambda^1 \int [W(x,y') - W(x,y)]\, dG(y') \end{aligned}$$

The problem the workers face stems from the need to decide when to accept job opportunities in this economy. Employed workers move from one island to another, if the new island is more productive than the current one, $y' > y$. Unemployed workers accept jobs above a reservation productivity level $R(x)$, which is such that $W(x, R(x)) = U(x)$. It can be easily demonstrated that the solution to this problem has the following properties:

1. $U(x) = x\overline{U}$ and $W(x,y) = x\overline{W}(y)$.

2. $R(x) = \overline{R}$ and is such that $\overline{W}(R) = \overline{U}$.

Since every worker adopts the same reservation strategy, it follows that $x \perp y$ in equilibrium. More specifically, workers self select into better jobs, y; however, that matching process does not depend on the type of the worker, resulting in no assortative matching between firm and worker types.

This framework encompasses a class of models of bargaining and competition with firm heterogeneity, which we label the "piece-rate" model.[3] In those models, y does not represent

[3]This model is very similar to the ones presented in the works of Barlevy (2008), Christensen et al. (2005) and others. Postel-Vinay, Robin and coauthors (2002; 2006; 2014) use variations of this model, allowing wage

the marginal product per unit of skill, but rather the piece-rate, i.e. the amount of flow output that the worker appropriates in negotiations per unit of skill. Mortensen (2003) discusses two wage setting mechanisms that generate a piece-rate offer distribution, one where firms make wage offers under full commitment (i.e. as described by Burdett and Mortensen (1998)), and another where firms and workers negotiate a wage in every period. The details are different but both sets of assumptions have the same implication: piece rates are a strictly increasing function of firm productivity. That assertion implies that those models have the same implications for sorting, and the same log-additive wage equation that we analyze in section 3.2.

3.2 AKM Methodology

The wage equation implied by this model is closely related to the AKM empirical methodology, which we make explicit. We assume that each island in the economy represents a firm, and that log-wages are measured with error: $log(w) = log\,(xy) + \varepsilon$, where ε denotes classical measurement error. The log-wage of a worker i at firm J at time t can be decomposed into two sets of fixed effects, pertinent to the worker and to the firm, respectively, as well as an error term

$$\begin{aligned} log(w_{iJ(i,t)}) &= log\,(x_i) + log\left[y_{J(i,t)}\right] + \varepsilon_{it} \\ &= \theta_i + \psi_{J(i,t)} + \varepsilon_{it}. \end{aligned} \tag{1}$$

where $\theta_i = log\,(x_i)$ and $\psi_{J(i,t)} = log\left[y_{J(i,t)}\right]$.

In order to estimate Equation 1, it is necessary to have access to a matched employer-employee dataset, which follows workers and firms over time. This equation has been estimated using datasets from several countries, including that provided by Abowd, Kramarz and Margolis (1999), who applied the method to French data. The econometric model is

contracts to be renegotiated while on the job, akin to the model presented in section 4, while still featuring the no-sorting implication.

based on a number of identifying assumptions. First, it is assumed that the assignment of workers to firms is uncorrelated with the error term ε_{it}. This condition is satisfied in our baseline model if ε_{it} represents classical measurement error. Second, identification in this model is only assured within groups of workers and firms that are connected. According to Abowd, Creecy and Kramarz (2002), a group of persons and firms is connected when it comprises of all the workers who have ever worked for any of the firms in the group and all the firms at which any of the workers have been employed at some point in time. In typical matched employer-employee datasets, the largest group comprises consists of over 95% of the observations. Thus, restricting attention to the largest group of the sample is a common solution. Under these assumptions, this statistical model can be estimated by ordinary least squares, which is a challenging mathematical problem, since usual samples comprise of data pertaining to millions of workers and hundreds of thousands of firms. Abowd, Creecy and Kramarz (2002) propose the use of an iterative conjugated gradient algorithm, which is used by authors of most extant studies, and which we use adopt in for our own calculations. Finally, it is worth nothing that typical implementations of this equation include an extra term $x_{it}\beta$, which captures the effects of time-varying observables (e.g. the returns to experience and time effects).

The "piece-rate" model allows predictions of assignment patterns, and provides a structural interpretation for the firm fixed effects. Recall that, in the model, worker and firm types are independent in equilibrium. One implication of this relation is that worker and firm fixed effects should be uncorrelated in the cross-section: $Corr\left(\theta_i, \psi_{J(i,t)}\right) = 0$. Another testable implication of the model is that the distribution of worker types is the same across firms. In particular, the model predicts that worker types should be uncorrelated to the average type of their coworkers: $Corr\left(\theta_i, \widetilde{\theta}_{J(i,t)}\right) = 0$, where $\widetilde{\theta}_{J(i,t)}$ is the mean value of θ among the co-workers of worker i.[4] Finally, it is worth noting that the model gives provides

[4]In samples comprising of large firms, this moment is identical to the index of segregation proposed by Kremer and Maskin (1996), using θ as the measure of skill: $\frac{Var(\theta|j)}{Var(\theta)} = Corr\left(\theta_i, \overline{\theta}_{J(i,t)}\right)$, where $\overline{\theta}_J$ denotes the average value of θ in firm J.

a structural interpretation for the fixed effects. More specifically, θ reflects differences in pay due to worker skill level, while ψ reflects frictional wage dispersion due to workers finding better or worse jobs. Consequently, we also highlight the relative variance between worker and firm fixed effects, $\frac{Var(\psi)}{Var(\theta+\psi)}$.

3.3 Evidence across Datasets

This methodology has been applied to datasets from many different countries with consistent results. Here, we summarize the results of studies from five countries (US, France, Germany, Italy and Denmark), alongside our results pertaining to the Brazilian matched employer-employee dataset. We use the labor market census RAIS (Relacao Anual de Informacoes Sociais), an administrative dataset collected annually by the Brazilian labor ministry, which includes all firms in the Brazilian formal sector and provides information for all their workers. We restrict our focus to the country's richest state, Sao Paulo, as well as to workers who have at least a high school diploma. In addition, our work analyses pertain only to workers that have at least a moderate degree of formal labor market attachment, and are included in the sample in at least 5 of the 11 years. These restrictions are intended to minimize the problem of the lack of coverage of the informal sector of the economy. We provide further details pertinent to the data, as well as a description of the sampling selection criteria in Appendix 7.1.

Table 1 summarizes the results from six countries, focusing on the moments described in section 3.1. It should first be noted that the AKM regression has a very high explanatory power, accounting for approximately 90% of wage variation in most studies. It is well known that typical observable terms in mincer regressions (experience, gender, etc.) have a low explanatory power; hence, most of the explained variation arises from the fixed effects terms θ and ψ. Second, across all studies the firm fixed effects also account for a sizable share of wage dispersion, whereby $\frac{Var(\psi)}{Var(\theta+\psi)}$ ranges from 0.19 to 0.32. Third, across all datasets the value of $Corr\,(\theta, \psi)$ is either very close to zero (as in the case of the US, Italy or Brazil)

or it takes a small negative value (as in France or Germany). If we interpret this evidence using the piece-rate model, most of the wage dispersion is driven by worker heterogeneity, while frictions still account for a significant fraction of wage dispersion, as reflected by value of $\frac{Var(\psi)}{Var(\theta+\psi)}$ at around $20-30\%$. In addition, the evidence that $Corr\left(\theta,\psi\right)$ is close to zero in most datasets seems supportive of the model implication of no sorting in labor markets.[5] However, this last conclusion is incongruent with the last piece of evidence, as $Corr\left(\theta,\widetilde{\theta}\right)$ ranges from 0.17 to 0.52 across datasets. This correlation suggests a substantial degree of clustering of workers across firms based on their skill level, which contradicts the implication that the distribution of worker skills is the same across firms.

3.4 Sorting across Sub-markets

In section 3.3, we concluded that the piece-rate model is inconsistent with the high degree of clustering of workers and coworkers reflected in the values of $Corr\left(\theta,\widetilde{\theta}\right)$ across a number of datasets. In order for the piece-rate model to explain those facts, the clustering can be assumed to be driven mechanically by composition. The argument supporting this strategy is as follows. Assume that the data is generated by the model presented in section 3.1, but separately across "groups" of workers and firms. A firm under this interpretation consists of the set of workers that belong to the respective "group" in a given establishment. Using an example of lawyers and law firms and call centers and their workers as two separate sub-markets, it can be assumed that, on average, the lawyers have a higher skill level relative to call center employees. Even though within groups there would be no firm-worker sorting, the law firms are in greater need of highly skilled workers, which would result in clusters of such workers in these firms. Most empirical implementation of piece-rate models allow for this type of sorting patterns by assuming existence of sub-markets divided by education levels or occupation categories.[6] It is worth noting that, in a more general setting with assortative

[5] Christensen et al. (2005) and Postel-Vinay and Robin (2002) use this moment to justify absence of a sorting equilibrium.

[6] For example, Postel-Vinay and Robin (2002) allow for seven occupation classes using French data, and Christensen et al. (2005) allow for six occupation classes using Danish data. In addition, see the discussion

matching, the correlations within groups do not need to be smaller than the unconditional correlation. This can be shown on an example of managers and maintenance staff as the two groups, assuming, once again, that most managers are more skilled than maintenance workers. Assume that firm one hires both better managers and maintenance staff than firm two, but it has a higher ratio of staff to managers. It is easy to construct cases where the average skill level in the two firms is identical, even though segregation within groups exists.

We shed evidence on this hypothesis using our Brazilian sample. To do so, we first divide our data into groups consisting of combinations of education categories and occupations, sectors or location data. We compute the value of the AKM moments for each of the groups, before calculating the average across groups, weighting by the size of each group: $E\left[Corr\left(\theta,\psi|\, G\right)\right]$, $E\left[Corr\left(\theta,\widetilde{\theta}\middle|\, G\right)\right]$ and $E\left[\frac{Var(\psi|G)}{Var(\theta+\psi|G)}\right]$. We restrict the sample to firms with at least two workers of a given group, and groups with at least 1,000 observations and multiple firms. The results are displayed in Table 2. Since the resulting sample is different for each case, we also include the unconditional moments. In all cases, the worker-coworker measure is well above zero, and in the case of education/occupations, it is even greater than the unconditional one, which suggests that our results are not driven by composition. It is also worth pointing out that the relative dispersion in firm fixed effects, $\frac{Var(\psi)}{Var(\theta+\psi)}$, is higher within groups than overall, irrespective of the choice of groups. This is consistent with the results obtained by Bagger et al. (2014), who report values of $\frac{Var(\psi)}{Var(\theta)}$ between 1.13 and 2 using Danish data and three educational groups. This congruence in findings suggests that some forces that are not captured by the model, such as compensating differentials between sectors or differences in local prices/amenities, are not the main component behind the variability in ψ.

preceding footnote 19 in Postel-Vinay and Robin (2002).

4 A Sorting Model with Frictions

Having concluded that the sorting patterns in the data are inconsistent with the piece-rate model, we introduce a model that incorporates sorting in equilibrium. In this section, we outline the main features of the model, leaving the details to the Appendix. The model shares the basic characteristics of that proposed by Shimer and Smith (2000), in that it describes economy comprising of heterogeneous workers and firms, as well as search frictions. The main innovation that we introduce into the model stems from the addition of the on-the-job search, which occurs in both voluntary and involuntary manner. This is important because a large fraction of employment separations involve job-to-job transitions (e.g. Fallick and Fleischman (2001)), and because it substantially affects the way frictions shape wage inequality (e.g. Hornstein, Krusell and Violante (2011)).

4.1 The Environment

This is a continuous time economy, with a unit mass of workers and a measure J of jobs. We assign a value $x \in [\underline{x}, \overline{x}]$ to each worker and $y \in \left[\underline{y}, \overline{y}\right]$ to each job, where the worker index is supposed to capture his/her level of human capital, whereas the job index reflects entrepreneurial talent, differences in the stock of capital, etc. It is assumed that these types are observable and the distributions are known to all agents. Workers and jobs are distributed according to density $l(x)$ and $g(y)$, respectively. In this economy, each firm represents a collection of jobs of a certain type without cross-complementarities. In line with the work of Shimer and Smith (2000), the equilibrium characterization is initially described in terms of jobs. However, when the model is applied to the data at the firm level, our notion of firms will be applied. We further assume existence of a measure E of firms, with distribution $\upsilon\left(y\right)$. For each value of y, all jobs are evenly distributed across the firms of such type such that $E\int_{\underline{y}}^{y} n\left(y'\right)\upsilon\left(y'\right)dy' = J\int_{\underline{y}}^{y} g\left(y'\right)dy'$, where $n\left(y\right)$ is the mass of jobs per firm of type y.

We study a stationary environment in which workers and jobs are matched via a random search technology. Workers search for jobs both while employed and unemployed, whereas firms offering jobs only search for suitable candidates while these positions are vacant. In addition, unemployed workers encounter vacancies at a rate λ^U, while employed workers come across vacancies at rate λ^E. Matches are dissolved either exogenously at rate δ or when workers face "reallocation shocks" at rate λ^R, which forces the worker to leave his/her current employer and immediately encounter a new firm. As discussed by Jolivet, Postel-Vinay and Robin (2006) and Lopes de Melo (2007), these shocks help explain the high fraction of job-to-job transitions that involve a wage cut. On the firm side, vacancies are compared with the characteristics of unemployed workers at a rate λ^F, employed workers at rate λ^{FE}, and workers subjected to reallocation shocks at rate λ^{FR}. Even though these rates are exogenous for our purposes, they need to satisfy steady state conditions which we describe in the Appendix.

When a worker of type x is matched with a job of type y, this results in flow output $F(x, y)$, which we assume to be increasing in both arguments. Unemployed workers receive a flow value $b(x)$ and vacant jobs are assigned zero flow value. Workers and jobs discount time at rate r and have linear preferences over income and flow profits, respectively.

The presence of search frictions yields match rents, which requires us to specify a rule to divide the match surplus between the two parties. We follow the approach developed by Cahuc, Postel-Vinay and Robin (2006) and others and assume that wages are determined via a sequential auctions bargaining game. In this game, the surplus of the match is apportioned by applying a generalized Nash Bargaining rule, where the worker receives a share β, and the outside options available to workers depend on their labor market history. We describe more details of the game in the next section as well as in the Appendix.

4.2 Description of the Equilibrium

The equilibrium of this economy consists of value functions and distributions such that workers and firms match optimally, and the distributions satisfy steady state flow conditions. It is useful to introduce the following notation. An employed worker of type x working for a firm of type y with a contracted wage of w has value $W(x, y, w)$. Alternatively, if this worker is unemployed, his/her value is $U(x)$. A job of type y that pays w has value $J(x, y, w)$ if matched with a worker of type x, or $V(y)$ if vacant. The surplus of a match between worker x and job y is given by $S(x, y) \equiv [W(x, y, w) - U(x)] + [J(x, y, w) - V(y)]$. Note that surplus of the match does not depend on wages because workers and firms have linear preferences, in line with our wage setting mechanism. The steady state density of employed matches is $e(x, y)$. The density of unemployed workers for each type is $u\,(x)$, and the density of vacancies is $v\,(y)$. Finally, the unemployment rate, u, and the number of vacancies in the economy, v, are also determined endogenously.

In this frictional economy, workers and jobs are paired at a Poisson rate, whereby they can match or not. Moreover, upon matching, the surplus of the match is split between the two via a Nash Bargaining protocol, whereby a fraction β of the surplus is given to workers and the remainder is received by firms. One important element of the bargaining protocol is the presence of outside options for both workers and firms partaking in negotiations. For firms, this option it is always the value of keeping a vacancy, $V\,(y)$. For workers, it will depend on their labor market status. When coming from unemployment, or after facing a reallocation shock, the outside option is the value of search while unemployed, $U\,(x)$. Then, as employed workers find new matches, they use them as leverage in negotiations, gaining the full surplus of that match as bargaining power. We can explain this using the example of an individual currently working for firm y with second best option y'. In this case, his/her outside option is $U\,(x) + S\,(x, y')$. When this worker is offered a job by a new employer y'', three possibilities emerge. First, if $S\,(x, y'') < S\,(x, y')$, there is no change. Second, if $S\,(x, y') < S\,(x, y'') < S\,(x, y)$, the worker remains with the current employer

but uses the new firm to leverage his/her negotiation position, and the new outside option becomes $U(x) + S(x, y'')$. Finally, if $S(x, y'') > S(x, y)$, then the worker leaves the current employer, whereby his/her alternative option becomes $U(x) + S(x, y)$. Note that workers and firms only match if they generate positive surplus $S(x, y) > 0$, and when workers are offered employment in other firms, they choose to work at the firm that generates the highest surplus.

Our assumptions imply that the expected value of unemployed workers is equal to:

$$rU(x) = b(x) + \beta\lambda^U \int_{\underline{y}}^{\overline{y}} [S(x, y')]^+ v(y')\, dy', \tag{2}$$

where $[A]^+ = AI[A > 0]$. In the Appendix, we show the analogous equation for $V(y)$, which is similar to this one but accounts for the fact that vacancies are also considered by employed workers. We also show in the Appendix that the surplus of the match can be computed as

$$\left(r + \delta + \lambda^R\right) S(x, y) = F(x, y) - rU(x) - rV(y) + \beta\lambda^R \int_{\underline{y}}^{\overline{y}} [S(x, y')]^+ v(y')\, dy' \tag{3}$$
$$+\beta\lambda^E \int_{\underline{y}}^{\overline{y}} [S(x, y') - S(x, y)]^+ v(y')\, dy'.$$

The first three terms on the RHS of the equation are standard and are presented in the work of Shimer and Smith (2000). In addition, although matches subject to reallocation shocks are eliminated, the worker can appropriate a share β of the surplus of a new match if he/she finds a suitable one. Finally, workers who find better jobs are able to extract a share β of the excess surplus of the new job versus the old one.

In addition to introducing the asset value equations, the equilibrium of this model includes the endogenous distributions $e(x, y)$, $v(y)$ and $u(x)$. In steady state, those are fixed by equating the inbound and outbound flows of matches of a particular types, as well as market clearing conditions. These are described in the Appendix.

One very useful property of this framework is that $S(x, y)$, $U(x)$ and $V(y)$ can be

computed jointly with the endogenous distributions without the need to compute wages. This facilitates the computation of the equilibrium, since it is not necessary to keep track of the second best job as a state variable. With knowledge of the value functions and equilibrium distributions, the implied wage function can be retrieved by using the surplus sharing conditions imposed by Nash Bargaining. The wage of worker x, working for firm y, with second best option q satisfies the condition

$$W(x, y, w(x, y, q)) - [U(x) + S(x, q)] = \beta [S(x, y) - S(x, q)], \tag{4}$$

where $W(x, y, w(x, y, y'))$ is computed via an asset value equation described in the Appendix. Having the wage function allows the flow profits of a job to be computed as

$$\pi(x, y, q) = F(x, y) - w(x, y, q).$$

4.3 Sorting, Wages and Intuition

Next, we illustrate some properties of equilibrium. As emphasized by Becker (1973) and Shimer and Smith (2000) complementarities in production are the main driver behind assortative matching: positive (negative) complementarities induce positive (negative) assortative matching, PAM (NAM). We consider both types of complementarity and illustrate the case of an economy without on-the-job search, $\lambda^E = \lambda^R = 0$, which helps in developing intuition. In this case, there are no poaching offers, and the wage function depends on the worker and firm types as follows:

$$w^{NOJ}(x, y) = \beta [F(x, y) - rV(y)] + (1 - \beta) U(x) \tag{5}$$

It is easy to see that $U'(x) > 0$, and $V'(y) > 0$, which implies that wages are increasing with worker skill level, whereas they may be non-monotone with respect to firm productivity. We can rearrange Equation 5 and obtain $w^{NOJ}(x, y) = U(x) + \beta S(x, y)$. This implies that wages

are at their minimum at the edges of the matching set (where $S(x,y)=0$), which ensures non-monotone wages for all workers with interior matching sets. These non-monotonicities are a natural reflection of job scarcity and optimal assignment. Indeed, in the matching economy, each worker has an ideal job (the one that generates the highest surplus), and workers choose them due to wages. Thus, if the ideal match of a worker with mid-level skills is a mid-level productivity firm, then that is the firm that will pay him/her the most. Highly productive firms would pay that worker less because hiring him/her entails giving up on the opportunity to hire a more suitable worker, which is taken into account in negotiations. It is worth noting that similar non-monotonicities also emerge in the profit function, $\pi(x,y,q)$, this time with respect to x.

We show two examples of symmetric economies where $\beta = 0.5$, $J = 1$, $b(h) = 0$ and $l = g$, one with supermodular technology and another with submodular technology.[7] Figure 1 depicts plots of log-output and log-wages as a function of job productivity, where each line represents a different percentile of worker skill. For each worker type, the support of the plot only includes the firms that belong to his/her matching set, which provides a visual representation of these sets. We can see that highly skilled workers match with highly productive firms in the case of PAM, and firms characterized by low productivity in the case of NAM, while the converse is true for low skilled workers. It is also evident that in both NAM and PAM cases wages are increasing in x, whereas they are non-monotone with respect to y, as described above.

This non-monotonicity of wages with respect to y suggests that the firm fixed effects in a wage regression , ψ, are unlikely to capture the underlying level of firm productivity well. This has two implications with respect to mapping the primitives of the model to the empirical fixed effects methodology. First, the correlation $Corr(\theta,\psi)$ is distorted, and thus no longer accurately reflects the degree of sorting between the model primitives x and y. Sec-

[7] We use a CES production function $F(x,y) = [0.5x^{\eta} + 0.5y^{\eta}]^{\frac{1}{\eta}}$, where $\eta \longrightarrow 0$ (Cobb Douglas) on the supermodular example and $\eta = 3$ on the submodular example .We also assume that the distributions are Log-Normal with variance 1, $\rho = 0.005$, $\delta = 0.02$ and $\lambda^U = 0.3$.

ond, the same non-monotonicity tends to suppress the amount of dispersion in ψ generated by the model. As noted previously, each worker has his/her ideal firm type. Thus, when the difference in wages for a worker in two firms is compared, $w^{NOJ}(x,y) - w^{NOJ}(x,y')$, the outcome may be positive for some workers and negative for others, introducing ambiguity into the average "firm effect" on wages. This second point is well illustrated by an example described by Eeckhout and Kircher (2011), using a simplified version of the frictional matching model with uniform distribution of types. In that example, when applying the AKM methods to the equilibrium wage function of the model, the implied firms fixed effects do not vary with their type regardless of the degree of frictions, implying that $Var(\psi) = 0$.[8] Recall that $Var(\psi) \gg 0$ on the data, which may be a challenge for the model to match. In general, the variability of ψ will depend on the joint distribution of worker skill and productivity in equilibrium. Consequently, when parametrizing the model, in our approach, we allow for very flexible shapes for the distributions of worker skill and job productivity.

The same graph suggests that wage data can still be useful for inferring the strength of sorting for two reasons. First, in our model, the highly skilled workers work for the high-productivity firms in the case of PAM (or the low-productivity in the case of NAM), and thus have highly skilled coworkers. Second, wages are monotonic in x, suggesting that the relationship between worker and coworker wages reflects that of the primitives. Thus, the moment $Corr\left(\theta, \widetilde{\theta}\right)$ is a promising way to measure the intensity of sorting in the economy. One limitation of the measure is that it cannot distinguish the sign of sorting (PAM or NAM), just its intensity.

In sum, to ensure that this model is consistent with the empirical evidence presented in section 3, the economy should exhibit a substantial degree of sorting, as reflected by $Corr\left(\theta, \widetilde{\theta}\right)$, whereas $Corr(\theta, \psi)$ would be biased downward because of non-monotonicity in wages. However, this non-monotonicity cannot be too strong, such that the model still yields $Var(\psi) \gg 0$. Whether the model can meet those targets simultaneously is a quantitative

[8] See Equation 18 and the discussion on pages 15 and 16. However, the authors do not discuss the implications for the variance of those fixed effects, only the implications for sorting.

matter and is investigated in section 5.

On-the-job search and sequential auctions make this relationship more complex because the wage function includes extra terms that reflect the potential gains to the worker from on-the-job search that are included in the negotiations and depress workers' wages. This mechanism can invalidate the monotonicity in x described above. To explain this argument, we can consider the example of the wage function of workers hired from unemployment, $q = \emptyset$, in an economy where workers have low bargaining power, $\beta = 0$, and $b(x) = b$. Equations 2, 7 and 4 imply that

$$w^{\beta=0}(x, y, \emptyset) = b - \lambda^E \int_{\underline{y}}^{\overline{y}} [S(x, y') > 0] \, min \{S(x, y'), S(x, y)\} \, v(y') \, dy'.$$

Now, let us consider any firm y^0 with an interior matching set, and the least skilled employee of that firm, $\underline{x}(y^0)$. Note that by construction $S(\underline{x}(y^0), y^0) = 0$, which implies that this worker will have the highest wage among the workers just hired from unemployment to take positions in that firm. Higher values of β, an increasing $b(x)$ and the arrival of competing offers, $q \neq \emptyset$, all counteract this force and restore monotonicity in worker skill level.

5 Calibration

In this section, we select a parametrization of the model and set the parameters in order to match a number of labor market facts emphasized by the empirical job-search literature (e.g. Jolivet, Postel-Vinay and Robin (2006)). We perform two calibrations. First, we select a "parsimonious" parametrization and do not include the moments described in section 3.3 as part of the target moments. We show that, while the model succeeds in matching the targeted moments, it fails to explain some of the facts highlighted in section 3.3. In particular the model generates less variation in firm fixed effects ψ than observed in the data. For this reason, in the method presented in Appendix 7.5, we adopt a more flexible parametrization, and include the aforementioned facts as target moments. This approach improves the fit of

the model, but still fails to reproduce some aspects of the data.

5.1 Calibration

In this section, we describe model calibration, performed by selecting parameters that guarantee that the model matches a number of features of the US data. In order to do so, we need to make certain parametric assumptions. First, we assume that the production function takes a CES form $F(x,y) = [\varphi x^{\eta} + (1-\varphi) y^{\eta}]^{\frac{1}{\eta}}$, which allows for a wide degree of complementarities. If $\eta < 1$, the production function is supermodular, whereas if $\eta > 1$, it is submodular. Second, we assume that the flow value of leisure for the worker takes the following form: $b(x) = bF(x, \phi(x))$, where $b \in [0,1]$, and $\phi(x)$ denotes the optimal job assigned to worker x in the frictionless economy.[9] We posit that this is a parsimonious way to allow more productive workers to be more productive in non-market activities as well. We assume that worker and firm types follow a Lognormal distribution with zero mean and respectively σ^W and σ^F standard deviations. It is worth noting that the mean of these distributions cannot be distinguished from φ, which is why they are set to 0. [10]

Following the discussion in section 4.3 we restrict $\eta < 1$ (positive complementarities), as our empirical measure of sorting only helps identify the strength of sorting, not its sign. We set the discount rate to match a 5% annual discount rate. We also set $J = 1$ to ensure that, if frictions are removed from the economy, every worker and job would find a match. Finally, we also set $\varphi = 0.5$. Thus, nine parameters remain, which are selected in order to match a series of facts about labor markets emphasized in the job-search literature. Although our model performs in continuous time, in practice, we approximate the economy with the discrete-time correspondent, setting the periodicity to one week. We simulate the path of 10,000 workers that are matched to 1,000 firms over a 7-year period. These numbers are chosen to match

[9] $\phi(x)$ is such that $\int_x^1 l(x')\,dx' = J \int_{\phi(x)}^1 g(y')\,dy'$.

[10] This holds because we can always re-write $F(x,y) = [\varphi(ax)^{\eta} + (1-\varphi)(by)^{\eta}]^{\frac{1}{\eta}} \propto \left[\frac{\varphi a^{\eta}}{\varphi a^{\eta}+(1-\varphi)b^{\eta}} x^{\eta} + \frac{(1-\varphi)b^{\eta}}{\varphi a^{\eta}+(1-\varphi)b^{\eta}} y^{\eta}\right]^{\frac{1}{\eta}} = [\varphi' x^{\eta} + (1-\varphi') y^{\eta}]^{\frac{1}{\eta}}$.

the length and the ratio of workers per firm of the US matched employer-employee dataset, as shown by Woodcock (2015). The number of workers is smaller than that found in typical matched employer-employee datasets. However, we found that increasing this number has little effect on the results, while it increases the computational time substantially. We present results using a much larger sample (300K workers) for our calibrated parameter values. We compute the moments on a yearly basis before calculating the average across years. The conditional moments are computed (both data and model), restricting attention to workers who remain employed throughout the year and experience at least one job transition.

In order to initialize our sample, we begin with a sample of unemployed workers drawn from distribution $l(x)$ discretized to a 100 points grid and simulate their paths for a 10-year period, in order to start the economy in a steady state. From that point, we collect 7 years of data, which is the basis of our calculations. On the firm side, we discretize the firm types grid to 500 points and allocate the firms to those grid points according to distribution $\upsilon(y)$, which we assume is such that firms have an equal number of total jobs (vacant and filled).[11] Whenever the worker finds a firm he/she takes a draw from the (endogenous) distribution of vacancies and is allocated randomly to one of the firms at the corresponding type. Note that the choice of $\upsilon(y)$ does not affect the wage function, matching decisions, the distribution of vacancies or the joint densities of worker and job types. As a result, we have found that this distribution has little impact on the moments under consideration.

We perform two sets of calibrations, one for workers with high school education, and another one for workers with at least a college degree. It is well documented that there are substantial differences in unemployment inflow rates across workers with different levels of education (e.g. Elsby, Hobijn and Şahin (2010)), and this model is not designed to explain those endogenously. We select a number of data moments as targets for our calibration. Next, we discuss identification informally, thus providing some level of intuition as to why that moment should be directly affected by that parameter.

[11] We have more gridpoints on the support of firm types to prevent ties on the worker job decisions, including job-to-job transitions. We use interpolation methods to obtain values on the expanded gridpoints.

First, we have the transition rates λ^U, δ and λ^E, which have a clear association with the observed unemployment inflow rates, unemployment outflow rates and job-to-job transition rates. These can be computed in our model as follows.

$$
\begin{aligned}
UErate &= \lambda^U \int_{\underline{y}}^{\overline{y}} \int_{\underline{x}}^{\overline{x}} [S(x', y') > 0]\, u(x')\, v(y')\, dx' dy' \\
EUrate &= \delta + \lambda^R \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} \int_{\underline{x}}^{\overline{x}} [S(x', y'') \leq 0]\, e(x', y')\, v(y'')\, dx' dy' dy'' \\
EErate &= \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} \int_{\underline{x}}^{\overline{x}} \left[\lambda^R [S(x', y'') > 0] + \lambda^E [S(x', y'') > S(x', y')]\right] e(x', y')\, v(y'')\, dx' dy' dy''
\end{aligned}
$$

We use the CPS microdata for the period $1996 - 2003$ and compute $UErate$ corrected for time aggregation for both educational groups following the methodology proposed by Shimer (2012). We focus on this period to match the time span used to compute our $EErates$ target. Next, we use unemployment rates by education provided by the BLS (4.3% for high school graduates, and 2.2% for college graduates) and adopt the standard steady state approximation $u \approx \frac{EUrate}{EUrate+EUrate}$ to compute the $EUrate$ of each group. Finally, we make use of the results reported by Nagypál (2008) who provides $EErates$ by education level for the same period.[12] The values are shown in Table 3.

Next, we use a series of moments related to the distribution of wages and wage dynamics to calibrate the parameters σ^W, σ^F, λ^R and β. We employ data from the SIPP 1996 panel, which tracks workers at a weekly frequency and is well suited for measuring job transitions. First, increasing σ^W increases the overall variability of log wages $Var(w)$, which we use as a target moment. In this environment, characterized by frictions and firm heterogeneity, the same worker can be paid differently in different firms, and this is affected by the extent of firm heterogeneity, σ^F. Consequently, we use as a target moment the variance of wage changes for workers who experience a job-to-job transition, $Var(\Delta w | JJT)$. As documented by Lopes de Melo (2007), Jolivet, Postel-Vinay and Robin (2006) and Postel-Vinay and

[12] We adopt her results using SIPP data, because it pertains to time aggregation. To obtain the implied $EErates$, we use her reported shares of job-to-job transitions as a fraction of all separations (49.7% for high school and 58.5% for college).

Robin (2002), a substantial fraction of job-to-job transitions involves a wage cut, which is hard to match in a model in which workers are only permitted to switch jobs for monetary gains. Thus, we set the rate of involuntary reallocation shocks, λ^R, to match the fraction of wage cuts that involve job-to-job transitions, $Pr\left(\Delta w < 0 | JJT\right)$. In addition, recall that in this model, there is a certain extent of backloading in wages as workers start jobs at low wages and negotiate increases as they consider roles in other firms. The extent of backloading is affected by the parameter β, whereby higher β corresponds to the more immediate wage gains. In fact, as shown by Postel-Vinay and Robin (2002) if β is very low, when switching jobs, some workers may accept a wage cut in exchange for future wage growth. Thus, we set β to match the mean wage gain in job-to-job transitions, $E\left(\Delta w | JJT\right)$.

At this point, only b and η remain, whereby the latter controls the degree of complementarity in the production function. Without frictions, sorting is perfect and positive, as long as $\eta < 1$. In the presence of frictions, workers and firms tolerate mismatch, the extent of which will depend on the degree of complementarities. As discussed in section 5 the moment $Corr\left(\theta, \widetilde{\theta}\right)$ captures well the intensity of sorting among workers. Thus, the value of this moment must be set as a means of fixing η. We choose a target value of 0.4 for $Corr\left(\theta, \widetilde{\theta}\right)$, which corresponds to the Denmark case and tends towards the top of the range found across datasets (0.17 in Italy, and ~ 0.5 in Brazil and Germany). In our extended calibration presented in section 7.5, we also test an alternative value of 0.2 for each education group. Finally, we need to set the parameter value of home production, b. We follow the approach used by Shimer (2005) and assume that b is such that the average value of $b(x)$ is 40% of the average wage, a number derived from the average replacement rate from the US unemployment insurance system. However, since the pool of unemployed workers may differ from the overall pool of workers we compute $\frac{b(x)}{E(w|x)}$ for each worker type, and then average over the distribution of unemployed workers, obtaining our target $\int_{\underline{x}}^{\overline{x}} \frac{b(x')}{E(w|x')} u\left(x'\right) dx'$.

We choose the combination of parameters that minimize the distance between the target

moments and the model moments.[13] The calibration results are shown in Table 3, panels (A) and (B). As can be seen from Table 3, the model is successful in matching the proposed moments, especially for the sample of college educated workers (there are too many wage cuts in job switches in the sample comprising of workers with high school education). When considering the estimated parameters, it is worth mentioning that there is a substantial difference in the estimated degree of complementarities pertaining to the high school versus the college sample. While both η values (-0.99 and 0.59, respectively) reflect supermodularity, the degree of compementarities in the market populated by workers with high school diploma is stronger. Our intuition for that result is that there are "more frictions" in the market serving high school graduates ($\frac{\lambda^U}{\delta} \approx 30$ in HS, and $\frac{\lambda^U}{\delta} \approx 50$ in college) which leads to less sorting. However, since we are imposing the same degree of sorting on both markets ($Corr\left(\theta, \widetilde{\theta}\right) = 0.4$)), we need stronger complementarities to achieve that.

5.2 AKM Regression

Given that our model has been calibrated to match several features of the US economy, its ability to reproduce the AKM empirical moments needs to be verified. Table 4 summarizes the results. First, as expected, the worker-coworker correlation $Corr\left(\theta, \widetilde{\theta}\right)$ (a targeted moment) captures the "true" degree of sorting (the value of $Corr\left(x, \widetilde{x}\right)$) in the economy quite well. Second, while the $Corr\left(\theta, \psi\right)$ value is much lower than its theoretical counterpart, this value is still well above that measured in any of the datasets described in section 3. Third, and perhaps more significant, the dispersion in firm fixed effects, $\frac{Var(\psi)}{Var(\theta+\psi)}$, is substantially smaller than that observed in the data. Our intuition for these results derives from the non-monotonicities explained in section 4.3. Since they distort the mapping between true firm types, y, and firm fixed effects, ψ, the model is consistent with a low value for $Corr\left(\theta, \psi\right)$ despite incorporating positive sorting. However, the same non-monotonicities reduce the

[13] We use the global optimizer "covariance matrix adaptation evolution strategy" (CMA-ES), which combines evolutionary algorithms with traditional gradient methods, and is well suited for non-smooth problems with potentially many local optima (see Hansen and Ostermeier (2001) for details). We use the code provided by the authors, with the default parameters, and at least three (possibly manual) restarts.

dispersion in ψ, as firms determine the wages of different workers in different ways. In addition, we included results of a simulation based on a sample of $300K$ workers (instead of $10K$) to establish whether sample size affects the results. As can be seen, the results based on the larger sample are very similar to our baseline results.

6 Conclusion and Future Directions

Our objective in this paper was to provide a structural interpretation for the facts pertinent to firm-wage differentials and worker-firm sorting obtained by adopting the AKM methodology. First, we reviewed reduced form evidence from the AKM regressions, including our own, focusing on three moments: a significant variance of firm fixed effects, worker and firms fixed effects that are uncorrelated, and a significant correlation between the fixed effects of workers and their coworkers. We demonstrated that these three moments are inconsistent with the "face value" interpretation provided by the piece-rate model. This is the case because the degree of coworker segregation that we documented violates the implication that the worker skill distribution is independent across firms. We extended our analyses further and verified that this fact is not driven by composition using Brazilian data.

In the next step of our investigation, we used a natural alternative to explain those facts: a frictional model that allows for assortative matching. A clear challenge for the model was to generate a zero correlation between fixed effects, while allowing for sorting patterns. We showed that the model incorporates a mechanism that allows for that: the wage function is non-monotone in firm's productivity, which distorts the mapping between firm types and fixed effects, biasing the correlation. This non-monotonicity does not affect the mapping between worker types and their wages; hence, the worker-coworker measure is not contaminated and accurately captures the degree of sorting in the economy. However, the same mechanism can generate problems in matching with respect to the first fact. As we have shown, because wages are non-monotone, a firm will pay some workers more and other

workers less, which depresses the dispersion in firm fixed effects. Thus, it is a quantitative matter if the model can succeed or not in matching the data. To address that issue, we calibrated the model to US data, confirming that the model can explain a number of labor market transition characteristics, wage dynamics and the wage distribution. The model also performs well in explaining the correlations of the AKM regression. However, the model generates too little dispersion in firm fixed effects relative to the data, which can be explained by the aforementioned non-monotonicities. The result remains valid after robustness checks with a much less parsimonious calibration.

These findings lead to the conclusion that neither model fully succeeds in matching the data, indicating that features need to be changed or added. We discuss a few possibilities. One possible extension would be to allow for types of workers or jobs to be stochastic. It is clear from the frictionless matching model that this has strong implications for the variance of the fixed effects. The argument is as follows. The model described in section 4, with $J = 1$, no frictions and a production function such that $F_{hp} > 0$ can be applied. This is a version of the Becker (1973) marriage model, which features perfect assortative matching (top worker with top firm, and vice versa) under such a parametrization. Now, the same economy can be replicated from one period to the next, allowing workers to switch types while keeping the distribution unchanged. In such a scenario, firms will hire new workers, but the new hires will have the same level of skill and will be paid the same wage. Therefore, a set of firm dummies would explain 100% of wage variation (over both periods), whereas the same is not true for worker dummies. Naturally, we can reverse the argument and fix worker types while allowing firm types to change. In that case, worker dummies would explain 100% of wage variation. It is not well known how these conclusions change when both types change simultaneously, and how these interact with frictions.[14] Second, since our co-worker measure points to sorting between workers, another extension would be to consider explicitly models with team production and cross-complementarities in production. This is a challenging and

[14]Lise, Meghir and Robin (2016) allow for stochastic job types, but only consider the implications for the correlation between fixed effects, $Corr\,(\theta, \psi)$.

promising area of research.

Another promising direction for future investigations is the one we pursued in the companion paper Lopes de Melo (2015). In that work, we extend the frictional sorting model to include firm attributes that affect workers' preferences, in line with the theory of compensating differentials (e.g. Rosen (1986)). That makes the model analysis more complex because it requires identifying two distinct dimensions of firm heterogeneity (firm productivity and the amenities). In that model, we still have complementarities in production between worker skill level and firm productivity, which induces assortative matching. Moreover, we have the compensating differential that shifts the wages of all workers in the same direction, which resembles the firm fixed effects in the AKM methodology. Consequently, the model can explain the correlations in AKM, as well as generate a fair amount of dispersion in firm fixed effects.

7 Appendix

7.1 Data Description and Sample Selection

We use the labor market census RAIS (Relacao Anual de Informacoes Sociais), an administrative dataset collected annually by the Brazilian Labor Ministry, which includes all firms in the Brazilian formal sector and provides information for all their workers. The ministry collects demographic information of all workers, such as age, education and sex, along with some information about establishments, such as sector and location. In addition, it provides information about the job, such as the average wage earned during that year (the measure we use), the wage in December, the average number of hours worked, occupation, dates of admission and resignation, type of contract, and causes for the termination of employment. [15]

We use data pertaining to a 11-year period, ranging from 1995 to 2005. Although data from previous years was available, we opted for this time frame because, until 1994, Brazil suffered from extremely high inflation, which caused serious measurement problems in variables such as wages, and also had structural implications for the macroeconomy. We restrict the sample to the state of Sao Paulo, which is the richest state of the country, contributing to the GDP and industrial production by over 13% and 30%, respectively. Sao Paulo is characterized by a much smaller level of informal employment relative to all other regions of the country. Work informality is an important feature in Brazil, just as it is in many developing countries, and is excluded in our analysis. Furthermore, RAIS is an enormous dataset, and reducing the number of states makes the data employed in analyses more manageable.

We start with a sample of 25,856,195 observations pertaining to workers aged between 20 and 60, with at least a high school diploma, employed in the State of Sao Paulo, in the 95-05 period, for whom wage observations are available and who worked for a particular firm during the full year. We then apply a series of inclusion criteria on this sample. First, we

[15] The remaining variables are race, nationality, a measure of disability and the juridic nature of the firm.

select individuals that worked at least an average of 25 hours per week. Then, we collapse observations to individual, establishment and year level. If data for a particular worker are given for the same establishment multiple times, we select the observation with the highest wage. Next, we restrict our focus to workers with at least a moderate degree of formal labor market attachment, whose data is included in the sample for at least 5 of the 11 years, and we eliminate observations pertaining to individuals who have worked in more than three firms per year. We also restrict the sample to firms with at least two workers in order to compute coworker measures. These restrictions yield a final sample comprising of 16,253,899 observations. Finally, we use the algorithm described by Abowd, Creecy and Kramarz (2002) for computing the connected groups in the sample, and select the largest group. As discussed previously, applying the AKM methodology allows for identifying only worker and firm fixed effects within groups of connected workers. The largest group contains 16,027,426 observations, or 98.6% of the sample. Table 5 provides some statistics related to this sample.

7.2 Derivation of the Value Functions and Surplus

Vacancies can meet three kinds of workers: unemployed, employed from reallocation shocks and unemployed from job-to-job transitions. The first two kinds of workers have as their outside option unemployment, so the firm is able to extract a share $(1-\beta)$ of the surplus of the newly formed match. The latter kind has as outside option the total surplus generated on her current match, which leads the firm to gain a share $1-\beta$ of the excess surplus over the previous match. This implies the following value equation.

$$rV(p) = \lambda^{FE}(1-\beta)\int_{\underline{y}}^{\overline{y}}\int_{\underline{x}}^{\overline{x}} \alpha^{E}(x',y',y)\left[S(x',y)-S(x',y')\right]e(x',y')\,dx'dy' + \tag{6}$$
$$\lambda^{F}(1-\beta)\int_{\underline{x}}^{\overline{x}} \alpha^{U}(x',y)\,S(x',y)\,u(x')\,dx' + \lambda^{FR}(1-\beta)\int_{\underline{y}}^{\overline{y}}\int_{\underline{x}}^{\overline{x}} \alpha^{U}(x',y)\,S(x',y)\,e(x',y')\,dx'dy'.$$

Next we compute the net present discounted value of a worker employed at wage $w(x,y,q)$.[16]

$$\left(r+\delta+\lambda^R+\lambda^E\int_{\underline{y}}^{\overline{y}}\left[S(x,y')>S(x,q)\right]v(y')\,dy'\right)\left[W(x,y,w(x,y,q))-U(x)\right]= \tag{7}$$

$$w(x,y,q)-rU(x)+\lambda^R\beta\int_{\underline{y}}^{\overline{y}}\left[S(x,y')>0\right]S(x,y')\,v(y')\,dy'$$

$$+\lambda^E\int_{\underline{y}}^{\overline{y}}\left[S(x,y')>S(x,y)\right]\left[\beta max\left\{S(x,y'),S(x,y)\right\}+(1-\beta)\,min\left\{S(x,y'),S(x,y)\right\}\right]v(y')\,dy'.$$

The firm analogous of this Equation is given by

$$\left(r+\delta+\lambda^R+\lambda^E\int_{\underline{y}}^{\overline{y}}\left[S(x,y')>S(x,q)\right]v(y')\,dy'\right)\left[J(x,y,w(x,y,q))-V(y)\right]= \tag{8}$$

$$F(x,y)-w(x,y,q)-rV(y)+\lambda^E\int_{\underline{y}}^{\overline{y}}\left[S(x,y')>S(x,q)\right]\left[(1-\beta)\left[S(x,y)-min\left\{S(x,y'),S(x,y)\right\}\right]\right]v(y')\,dy'.$$

If we add Equations 7 and 8 and do some algebra and we obtain Equation 3. Note that this argument applies to matches that pay equilibrium wages. Following footnote 16 one can easily adapt these equations to compute values at arbitrary wages, which yields the same surplus function.

7.3 Steady State Flows

We now describe the equilibrium equations that jointly determine the stationary distributions $e(x,y)$, $u(x)$ and $v(y)$, the idleness rates u and v, and the necessary restrictions on λ^F, λ^{FE} and λ^{FR}. The first equilibrium equation that we describe is between the flows in and out of employed matches of type (x,y). If $S(x,y)\leq 0$ then $e(x,y)=0$. Otherwise, $e(x,y)$ is

[16] One can use a very similar equation to compute the value for the worker x of matching with firm y at an arbitrary wage w, $W(x,y,w)$. The only modification would be that job-to-job transitions affect the worker value whenever the surplus of the new match exceeds the value of the current contract for the worker: $S(x,y')>W(x,y,w)-U(x)$.

determined by equating the inflows to the outflows,

$$\left[\delta+\lambda^{R}+\lambda^{E}\int_{\underline{y}}^{\overline{y}}\left[S\left(x,y'\right)>S\left(x,y\right)\right]v\left(y'\right)dy'\right]\left(1-u\right)e(x,y)= \tag{9}$$
$$u\lambda^{U}u\left(x\right)v\left(y\right)+\lambda^{R}\left(1-u\right)v\left(y\right)\int_{\underline{y}}^{\overline{y}}e(x,y')dy'+\lambda^{E}\left(1-u\right)v\left(y\right)\int_{\underline{y}}^{\overline{y}}\left[S\left(x,y\right)>S\left(x,y'\right)\right]e(x,y')dy'.$$

On the LHS of the equation, matches of type (x, y) can be ended in three ways: exogenous destruction shocks, reallocation shocks and successful job-to-job transitions. On the other hand, the inflows into those matches come from workers hired from unemployment, and employed workers of type x that switch to jobs of type y either via reallocation shocks, or job-to-job transitions.

In addition, we use two consistency conditions to determine the densities of unemployed workers and vacancies, $u\left(x\right)$ and $v\left(y\right)$. The mass of employed workers of type x has to equal the total mass of workers of that type minus the unemployed ones.

$$1-u\left(x\right)u=\left(1-u\right)\int_{\underline{y}}^{\overline{y}}e\left(x,y'\right)dy'. \tag{10}$$

A similar condition needs to hold for jobs of type y

$$J-v\left(y\right)v=\left(J-v\right)\int_{\underline{x}}^{\overline{x}}e\left(x',y\right)dx'. \tag{11}$$

Next, the unemployment rate is determined by integrating (9) over the full support of x and y.

$$\left(1-u\right)\left[\delta+\lambda^{R}\int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}}\int_{\underline{x}}^{\overline{x}}\left[S\left(x',y''\right)\leq 0\right]v\left(y''\right)e(x',y')dx'dy'dy''\right] \tag{12}$$
$$=u\lambda^{U}\int_{\underline{y}}^{\overline{y}}\int_{\underline{x}}^{\overline{x}}\left[S(x',y')>0\right]u\left(x'\right)v\left(y'\right)dx'dy'.$$

Another equilibrium requirement for equilibrium is that the total number of employed workers has to equal the number of filled jobs.

$$1 - u = J - v. \tag{13}$$

Finally, the contact rates of workers and jobs must such the total number of contacts is the same in equilibrium:

$$\begin{aligned} u\lambda^U &= v\lambda^F \\ (1-u)\,\lambda^E &= v\lambda^{FE} \\ (1-u)\,\lambda^R &= v\lambda^{FR} \end{aligned} \tag{14}$$

7.4 Computation Algorithm

We use the following method to solve the model. First, we select N grid points on the supports of x and y where we compute our endogenous objects.[17] We begin with initial guesses for $e(x,y)$, $S(x,y)$ and u on the selected node points. We then adopt the following interactive procedure:

1. Given the old value of γ^O, S^O and u^O we update these using equilibrium Equations 2 through 14, thus obtaining γ^N, S^N and u^N. When computing integrals we use numerical integration using as node points the N grid points previously selected.

2. We compute the distance between the old value and the update. We first compute the maximum absolute value of $\frac{X^N - X^O}{1+X^O}$ on the grid for each of the three objects, and then take the maximum value between the three.

3. Repeat Steps 1 and 2 until the norm is sufficiently small.

[17] We select evenly spaced grid points (the midpoints of N evenly spaced intervals) instead of Chebychev nodes because the endogenous matching sets imply that the regions of integration of the program are worker and firm specific.

Because we do not have results showing that these mappings are contractions, we cannot anticipate that this algorithm works. Adding on-the-job search increases the computational burden of the problem substantially, as it requires us to compute double integrals. Because of that we use the solution to the problem without on-the-job search as the initial condition for the model with on-the-job search. We set N to 30 when solving the model with on-the-job search, which we found to be a good compromise between speed and accuracy of the approximation. Also, because of discretization it may happen sometimes that the algorithm does not converge because matching decisions change discontinuously with minimal changes to the surplus. In order to smooth that we assume that the decision rule takes a logistic form $\frac{1}{1+e^{-100S}}$, instead of indicator functions.[18] Furthermore, we have found that "slowing" the updates $\gamma^O = \alpha\gamma^N + (1-\alpha)\,\gamma^O$ improves the performance of the code.

7.5 Expanded Calibration

On section 5 we showed that our baseline calibration fails to match jointly all the moments from the AKM regression. One possibility is that this was due to some of the parameter restrictions that we imposed in our calibration. In order to address that we perform a second calibration, this time with more flexibility and many more free parameters in order to see if the model can match the AKM moments jointly with the moments that we had previously targeted for some configuration of parameters.

We make the following modifications to the exercise. First, we allow the discount rate, ρ, the measure of jobs in the economy J and the weight of the CES function, φ to vary as part of the calibration. We cap the discount rate at 20% a year, and J at 1.1, which would imply a vacancy rate around 14.5%, and a monthly vacancy filling rate around 0.25 for the high school market.[19] This rate is well below the values documented in Davis, Faberman and Haltiwanger (2010) using JOLTS data (reducing J increases the job-filling

[18] This only influences very small surplus values, below 0.1 in modulus. Surplus values can take the value of hundreds or thousands in typical calibrations.

[19] The vacancy filling rate is $\frac{uUErate+(1-u)EErate}{J-1-u}$.

rate). Second, we use Beta distributions instead of Log-Normal, which allows for very flexible shapes. This adds four shape parameters (in addition to the two variance parameters) to the calibration, leaving us with 16 parameters on the new formulation. Finally, we allow for negative complementarities in production, by allowing for $\eta > 1$.

We choose to match the same target moments as before, as well as the additional moments from the AKM regression, $Corr\left(\theta, \psi\right)$, $\frac{Var(\psi)}{Var(\theta+\psi)}$ and R^2_{AKM}, and the skewness of wages to assure that the wage distribution has a reasonable shape. Another modification is that we also consider a value of 0.2 for $Corr\left(\theta, \widetilde{\theta}\right)$, in line with the lower values for that statistic presented in Table 1. Finally, we drop $\int_0^1 \frac{b(x')}{E(w|x')} u\left(x'\right) dx'$ as one of the target moments from the calibration because there is a debate in the related literature about the appropriate value for the flow value of leisure (e.g. see Hall and Milgrom (2008)). This leaves us with an overparametrized specification with 12 target moments and 16 parameters.

The main results are displayed on Tables 6 and 7. As we can see the expanded calibration allows the model to match the AKM moments better, especially $\frac{Var(\psi)}{Var(\theta+\psi)}$. However, the model still fails short in some dimensions, such as a too high value value of $Corr\left(\theta, \psi\right)$ relative to the data. We see this as consistent with the tension to match $\frac{Var(\psi)}{Var(\theta+\psi)}$ and $Corr\left(\theta, \psi\right)$ jointly due to the non-monotonicities explained in section 4.3. Using the lower value for $Corr\left(\theta, \widetilde{\theta}\right)$ of 0.2 we obtain our closest fit of the AKM moments in the high school sample. However, that calibration falls short with respect to other moments (e.g. too little dispersion in wages, and a too low $UErate$), and features a discount rate of 17.5% a year.

References

Abowd, J., Creecy, R. and Kramarz, F. (2002), 'Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data', *unpublished manuscript* .

Abowd, J., Kramarz, F., Lengerman, P. and Perez-Duarte, S. (2004), 'Are Good Workers Employed by Good Firms? A Test of a simple assortative matching model for France and the United States', *unpublished manuscript* .

Abowd, J., Kramarz, F. and Margolis, D. (1999), 'High Wage Workers and High Wage Firms', *Econometrica* **67**(2), 251–333.

Andrews, M., Gill, L., Schank, T. and Upward, R. (2008), 'High wage workers and low wage firms: negative assortative matching or limited mobility bias?', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **171**(3), 673–697.

Bagger, J., Fontaine, F., Postel-Vinay, F. and Robin, J.-M. (2014), 'Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics', *The American Economic Review* **104**(6), 1551–1596.

Bagger, J. and Lentz, R. (2014), 'An empirical model of wage dispersion with sorting', *NBER working paper* .

Barlevy, G. (2008), 'Identification of search models using record statistics', *The Review of Economic Studies* **75**(1), 29–64.

Becker, G. (1973), 'A Theory of Marriage: Part I', *The Journal of Political Economy* **81**(4), 813–846.

Burdett, K. and Mortensen, D. (1998), 'Wage Differentials, Employer Size, and Unemployment', *International Economic Review* **39**(2), 257–273.

Cahuc, P., Postel-Vinay, F. and Robin, J. (2006), 'Wage Bargaining With On-The-Job Search: Theory and Evidence', *Econometrica* **74**(2), 323–364.

Card, D., Heining, J. and Kline, P. (2013), 'Workplace heterogeneity and the rise of west german wage inequality*', *The Quarterly Journal of Economics* .

Christensen, B., Lentz, R., Mortensen, D., Neumann, G. and Werwatz, A. (2005), 'On the job search and wage dispersion', *Journal of Labor Economics* **23**, 31–58.

Davis, S. J., Faberman, R. J. and Haltiwanger, J. C. (2010), The establishment-level behavior of vacancies and hiring, Technical report, National Bureau of Economic Research.

Davis, S. J. and Haltiwanger (1991), 'Wage dispersion between and within us manufacturing plants, 1963-86', *Brookings Papers on Economic Activity. Microeconomics* **1991**, 115–200.

Eeckhout, J. and Kircher, P. (2010), 'Sorting and decentralized price competition', *Econometrica* **78**(2), 539–574.

Eeckhout, J. and Kircher, P. (2011), 'Identifying Sorting-In Theory', *The Review of Economic Studies* **78**(3), 872–906.

Elsby, M. W., Hobijn, B. and Şahin, A. (2010), 'The labor market in the great recession', *Brookings Papers on Economic Activity* pp. 1–69.

Fallick, B. and Fleischman, C. (2001), 'The Importance of Employer-to-employer Flows in the US Labor Market', *unpublished manuscript* .

Hagedorn, M., Law, T. H. and Manovskii, I. (2012), 'Identifying equilibrium models of labor market sorting', *unpublished manuscript* .

Hall, R. E. and Milgrom, P. R. (2008), 'The limited influence of unemployment on the wage bargain', *The American Economic Review* **98**(4), 1653–1674.

Hansen, N. and Ostermeier, A. (2001), 'Completely derandomized self-adaptation in evolution strategies', *Evolutionary computation* **9**(2), 159–195.

Hornstein, A., Krusell, P. and Violante, G. L. (2011), 'Frictional wage dispersion in search models: A quantitative assessment', *American Economic Review* **101**(7), 2873–98.

Iranzo, S., Schivardi, F. and Tosetti, E. (2008), 'Skill dispersion and firm productivity: An analysis with employer-employee matched data', *Journal of Labor Economics* **26**(2), 247–285.

Jolivet, G., Postel-Vinay, F. and Robin, J.-M. (2006), 'The empirical content of the job search model: Labor mobility and wage distributions in europe and the us', *European Economic Review* **50**(4), 877–907.

Kremer, M. and Maskin, E. (1996), 'Wage inequality and segregation by skill', *NBER working paper* .

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

Lise, J., Meghir, C. and Robin, J.-M. (2016), 'Matching, sorting and wages', *Review of Economic Dynamics* **19**, 63–87.

Lopes de Melo, R. (2007), 'The Implications of Search Models for Wage Dynamics: an Empirical Assessment', *unpublished manuscript* .

Lopes de Melo, R. (2015), 'How firms affect wages: a structural decomposition', *unpublished manuscript* .

Lucas, R. and Prescott, E. (1974), 'Equilibrium search and unemployment', *Journal of Economic Theory* **7**(2), 188–209.

Mortensen, D. (2003), *Wage Dispersion: Why are Similar Workers Paid Differently*, MIT Press.

Nagypál, É. (2008), 'Worker reallocation over the business cycle: The importance of employer-to-employer transitions', *unpublished manuscript* .

Oi, W. Y. and Idson, T. L. (1999), 'Firm size and wages', *Handbook of labor economics* **3**, 2165–2214.

Postel-Vinay, F. and Robin, J. (2002), 'Equilibrium Wage Dispersion with Worker and Employer Heterogeneity', *Econometrica* **70**(6), 2295–2350.

Rosen, S. (1986), 'The theory of equalizing differences', *Handbook of labor economics* **1**, 641–692.

Shi, S. (2001), 'Frictional assignment. i. efficiency', *Journal of Economic Theory* **98**(2), 232–260.

Shimer, R. (2005), 'The cyclical behavior of equilibrium unemployment and vacancies', *American economic review* pp. 25–49.

Shimer, R. (2012), 'Reassessing the ins and outs of unemployment', *Review of Economic Dynamics* **15**(2), 127–148.

Shimer, R. and Smith, L. (2000), 'Assortative Matching and Search', *Econometrica* **68**(2), 343–369.

Woodcock, S. D. (2015), 'Match effects', *Research in Economics* **69**(1), 100–121.

Table 1: Summary of evidence based on AKM regressions

Country	US 1[a]	US 2	FR	DE	IT	DK	BR
$Var\left(x\beta\right)$	0.03	0.14	0.02	—	0.01	—	0.02
$Var\left(\theta\right)$	0.29	0.23	0.21	0.05	0.05	0.07	0.40
$Var\left(\psi\right)$	0.08	0.053	0.08	0.013	0.01	0.014	0.18
$\frac{Var(\psi)}{Var(\theta+\psi)}$	0.22	0.19	0.32	0.22	0.23	0.16	0.31
$Corr\left(\theta,\psi\right)$	-0.01	-0.03	-0.28	-0.10	0.04	—	0.04
$Corr\left(\theta,\widetilde{\theta}\right)^b$	—	—	—	0.51	0.17	0.40	0.52
R^2	0.89	0.9	0.84	0.79	—	—	0.93
Sample Statistics							
Years	90-99	84-93	76-87	93-97	81-97	85-03	95-05
Nobs	$37.7M$	$4.3M$	$5.3M$	$4.8M$	—	$600K$	$16.0M$
Nworkers	$5.2M$	$293K$	$1.2M$	$1.8M$	$1.7M$	$66K$	$2.0M$
Nfirms	$476K$	$80K$	$500K$	1821	$421K$	$25K$	$137K$
% 1st Group[c]	—	99.1%	88.3%	94.9%	99.5%	—	98.6%

[a] "US1" is sourced from Woodcock (2015), which covers two non-identified states, and includes all workers who were employed in 1997. "US2" and "Fr" are derived from Abowd et al. (2002). The US data covers 1/10 of workers in the state of Washington, whereas the French data pertains to 1/25 of all workers. "DE" is obtained from Andrews et al. (2008) and is based on data from around 2000 establishments in West Germany. "IT" is sourced from Iranzo et al. (2008), which covers 1200 plants with at least 50 workers. "DK" is sourced from Bagger et al. (2014) and includes all workers with 15-20 years of education. "BR" refers to our own calculations.

[b] The coworker correlation for Germany was provided by Tzuo Hann Law, using a different sample from the one employed by Andrews et al. (2008). For Italy, Iranzo et al. (2008) compute the index of segregation proposed by Kremer and Maskin (1996), using θ as their measure of skill. When firms are large (as in their sample) that measure is very similar to our worker co-worker measure. However, these authors use Pearson correlations instead of rank correlations. The coworker correlation for Denmark was provided by Bagger and Lentz (2014).

[e] This denotes the fraction of the sample in the largest connected group.

Table 2: Correlation within groups

			Unconditional			Within Groups		
Group	Ngroups[a]	Nobs	$(A)^b$	(B)	(C)	(A)	(B)	(C)
Education	2	$15.8M$	0.05	0.52	0.33	0.07	0.48	0.45
Ed./Occupations[c]	351	$9.5M$	0.03	0.55	0.32	0.00	0.57	0.41
Ed./Sectors	788	$15.6M$	0.05	0.52	0.33	0.02	0.40	0.36
Ed./Location	510	$15.5M$	0.05	0.52	0.33	0.06	0.44	0.43

[a] In our sample, two levels of education (high school diploma and college degree) are included. Occupation data is given at a 3-digit level, while the sectors is presented using 5 digits. Location data is at the municipality level.

[b] $(A) = Corr\,(\theta, \psi)$, $(B) = Corr\left(\theta, \widetilde{\theta}\right)$ and $(C) = \frac{Var(\psi)}{Var(\theta+\psi)}$.

[c] Occupation classification changed after 2002, so we only use this data for the period 95-02.

Table 3: Calibration Targets, Model Fit and Parameters

(A) Calibration Targets					(B) Calibrated Parameter		
	High School		College			High School	College
Target	Data	Model	Data	Model			
$UErate$	0.412	0.412	0.373	0.367	λ^U	0.495	0.404
$EUrate$	0.019	0.017	0.008	0.008	δ	0.016	0.008
$EErate$	0.018	0.020	0.012	0.013	λ^E	0.031	0.034
$Var\left(w\right)$	0.22	0.224	0.303	0.302	σ^W	0.151	0.163
$Corr\left(\theta,\widetilde{\theta}\right)$	0.4	0.394	0.4	0.400	η	-0.986	0.587
$\int_{\underline{x}}^{\overline{x}} \frac{b(x\prime)}{E(w\|x\prime)} u\left(x\prime\right) dx\prime$	0.4	0.405	0.4	0.400	b	0.048	0.043
$Var\left(\triangle w\|JJT\right)^a$	0.096	0.092	0.090	0.086	σ^F	0.296	0.252
$Pr\left(\triangle w<0\|JJT\right)$	0.27	0.375	0.260	0.289	λ^R	0.010	0.004
$E\left(\triangle w\|JJT\right)$	0.039	0.042	0.047	0.050	β	0.147	0.167

[a] The $\triangle$ refers to yearly change. JJT refers to workers who were continuously employed during that year but switched jobs at least one time.

Table 4: AKM Moments on Calibrated Economy

AKM moments	Data	Model		Model (large sample)[a]	
		High School	College	High School	College
$Corr\left(\theta,\psi\right)$	-0.01	0.175	0.338	0.173	0.335
$Corr\left(\theta,\widetilde{\theta}\right)$	0.40	0.394	0.400	0.397	0.395
$\frac{Var(\psi)}{Var(\theta+\psi)}$	0.22	0.023	0.042	0.023	0.044
R^2	0.89	0.892	0.943	0.890	0.942
$Corr\left(x,y\right)$	—	0.486	0.473	0.488	0.478
$Corr\left(x,\widetilde{y}\right)$	—	0.415	0.411	0.418	0.409
$\frac{E[Var(w\|x)]}{Var(w)}$	—	0.198	0.122	0.200	0.121

[a] The larger simulated sample has $300K$ workers.

Table 5: Sample Statistics

Whole Sample					
		Firms		Workers	
Nobs	$25,9M$	Number	$579K$	Number	$6,1M$
% college	36.2%	avg Nobs	44.59	avg Nobs	4.18
% female	49%	avg Fsize	11.7	avg Nfirms	1.36
avg. age	35.5	avg Nyears	3.81	avg Nyears	4.01
Selected Sample					
		Firms		Workers	
Nobs	$16,0M$	Number	$137K$	Number	$2.0M$
% college	38%	avg Nobs	116.27	avg Nobs	7.97
% female	49%	avg Nworkers	22.65	avg Nfirms	1.72
avg. age	36.7	avg Nyears	5.13	avg Nyears	7.71

Table 6: Expanded Calibration

	$Corr\left(\theta,\widetilde{\theta}\right)=0.4$				$Corr\left(\theta,\widetilde{\theta}\right)=0.2$			
	High School		College		High School		College	
Target	Data	Model	Data	Model	Data	Model	Data	Model
$UErate$	0.412	0.392	0.373	0.371	0.412	0.337	0.373	0.363
$EUrate$	0.019	0.015	0.008	0.007	0.019	0.016	0.008	0.007
$EErate$	0.018	0.022	0.012	0.015	0.018	0.023	0.012	0.015
$Var\left(w\right)$	0.220	0.170	0.303	0.240	0.220	0.133	0.303	0.257
$Corr\left(\theta,\widetilde{\theta}\right)$	0.400	0.303	0.400	0.347	0.200	0.195	0.200	0.192
$Var\left(\triangle w\vert JJT\right)^{(a)}$	0.096	0.099	0.090	0.106	0.096	0.110	0.090	0.099
$Pr\left(\triangle w<0\vert JJT\right)$	0.270	0.347	0.260	0.290	0.270	0.368	0.260	0.285
$E\left(\triangle w\vert JJT\right)$	0.039	0.046	0.047	0.053	0.039	0.044	0.047	0.053
$skew\left(w\right)$	-0.028	-0.028	-0.487	-0.492	-0.028	-0.027	-0.487	-0.491
$Corr\left(\theta,\psi\right)$	-0.013	0.171	-0.013	0.154	-0.013	-0.047	-0.013	0.097
$\frac{Var(\psi)}{Var(\theta+\psi)}$	0.217	0.165	0.217	0.156	0.217	0.160	0.217	0.166
R^2_{AKM}	0.889	0.879	0.889	0.922	0.889	0.794	0.889	0.944

Table 7: Extended Calibration - Parameter Values

	$Corr\left(\theta, \widetilde{\theta}\right) = 0.4$		$Corr\left(\theta, \widetilde{\theta}\right) = 0.2$	
	High School	College	High School	College
λ^U	0.535	0.468	0.388	0.379
δ	0.013	0.006	0.015	0.007
λ^E	0.050	0.052	0.042	0.047
λ^R	0.012	0.005	0.012	0.005
σ^W	2.94	3.62	7.04	1.458
σ^F	52.01	19.66	9.78	63.83
η	0.448	0.565	2.16	0.635
β	0.244	0.234	0.172	0.276
b	0.000	0.000	0.138	0.000
ρ	0.009	0.015	0.016	0.016
J	1.050	1.016	1.072	1.055
φ	0.224	0.234	0.455	0.312
α_1^W	48.94	51.57	58.71	12.88
α_2^W	35.03	7.18	43.26	13.75
α_1^F	57.16	3.52	50.46	12.65
α_2^F	3.41	0.50	16.86	0.54

Figure 1: Equilibrium Output and Wages with PAM and NAM, without On-the-Job Search.

(a) PAM

(b) NAM

Note: Each line refers to a different percentile of worker skill, and for each worker the line only includes the firms that belong to his/her matching set.

Figure 1 a

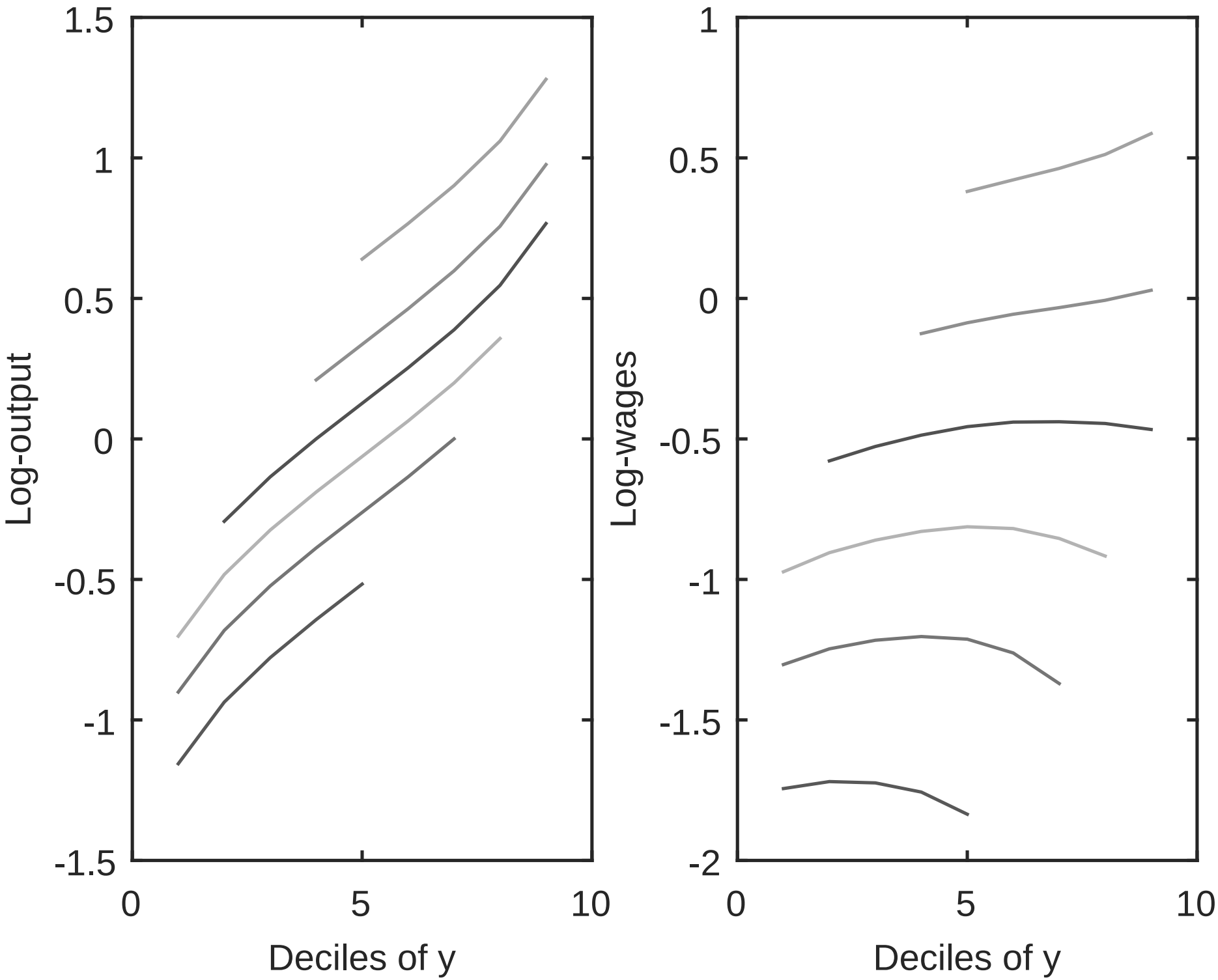
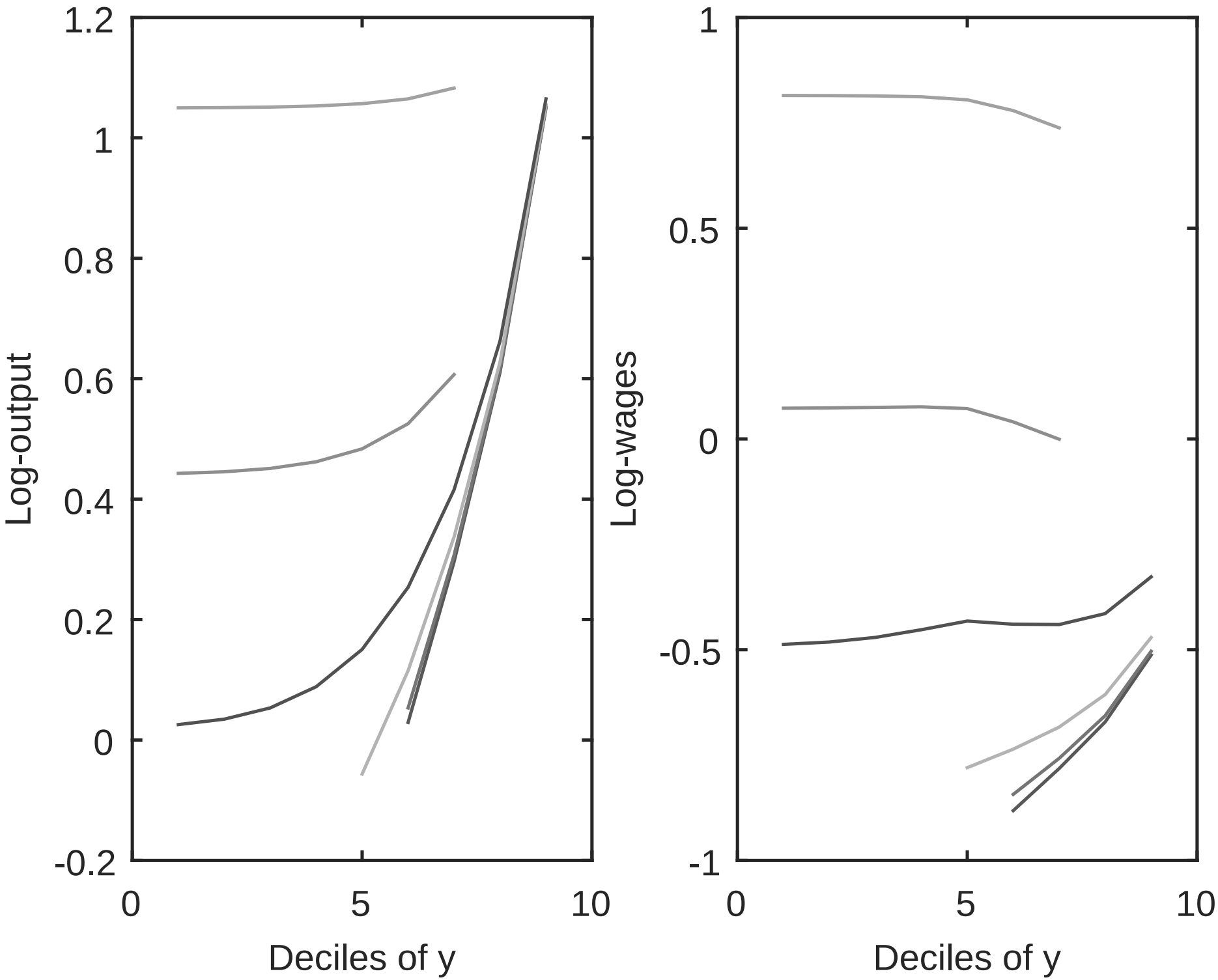

1.5
1
0.5
0
-0.5
-1
-1.5
Log-output
0
5
10
Deciles of y
1
0.5
0
-0.5
-1
-1.5
-2
Log-wages
0
5
10
Deciles of y

Figure 1 b


1.2
1
0.8
0.6
0.4
0.2
0
-0.2
Log-output
0
5
10
Deciles of y
1
0.5
0
-0.5
-1
Log-wages
0
5
10
Deciles of y

Full title of article: Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence

Suggested short title: Firm-Wages and Sorting

Full author names and affiliations, as they should appear in print:

Rafael Lopes de Melo, University of Edinburgh

Abstract, as it should appear in print (please adhere to strict 100-word limit):

Why do firms pay different wages? Empirical evidence suggests presence of substantial differences in firm pay controlling for worker skill. Moreover, these differences are uncorrelated with skills, indicating absence of sorting. We show that the face value interpretation is inconsistent with evidence on coworker segregation. We interpret the evidence by applying a sorting model and show that the correlation is biased. We identify non-monotonicities in wages as the reason for this bias and show that a measure of worker-coworker sorting is more accurate. By calibrating the model to US data, we confirm that the model matches many job market characteristics.

Does your paper include any special formatting or characters (e.g., any unusual math or symbols)? If yes, please explain.

No

Are there any appendixes? yes

If so, please label them clearly for print or online-only publication. *Print appendixes* will be edited and typeset with the main text. *Online-only appendixes* should be provided as PDF files that are ready for as-is posting.

How many tables does your article contain? 7

(Please exclude tables in online-only appendixes from total.)

How many figures does your article contain? 1

(Please exclude figures in online-only appendixes from total.)

Do you wish for any of your figures to appear in color? No

(There is a $300 per page charge for color printing.)

Is your paper subject to the JPE data policy? (http://www.press.uchicago.edu/journals/jpe/datapolicy.html?journal=jpe)

If not, briefly explain why:

It is subject to the policy, but because I use administrative data I followed the instructions of what to do in those cases.

If your paper is subject to the data policy, please send your data and program files in a .zip folder to the managing editor at jpe@press.uchicago.edu. If your files are too large for email transmittal, please provide a link to an institutional server, Dropbox, or other data storage system from which the files can be retrieved.

Special Notes: