Working Paper Series

Frictional Wage Dispersion in Search Models: A Quantitative Assessment

WP 06-07

Andreas Hornstein
Federal Reserve Bank of Richmond

Per Krusell
Princeton University

Giovanni L. Violante
New York University

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Frictional Wage Dispersion in Search Models: A Quantitative Assessment *

Andreas Hornstein† Per Krusell‡ Giovanni L. Violante§

Federal Reserve Bank of Richmond Working Paper 2006-07

Abstract

Standard search and matching models of equilibrium unemployment, once properly calibrated, can generate only a small amount of frictional wage dispersion, i.e., wage differentials among ex-ante similar workers induced purely by search frictions. We derive this result for a specific measure of wage dispersion—the ratio between the average wage and the lowest (reservation) wage paid. We show that in a large class of search and matching models this statistic (the "mean-min ratio") can be obtained in closed form as a function of observable variables (i.e., interest rate, value of leisure, and statistics of labor market turnover). Looking at various independent data sources suggests that, empirically, residual wage dispersion (i.e., inequality among observationally similar workers) exceeds the model's prediction by a factor of 20. We discuss three extensions of the model (risk aversion, volatile wages during employment, and on-the-job search) and find that, in their simplest version, they can improve its performance, but only modestly. We conclude that either frictions account for a tiny fraction of residual wage dispersion, or the standard model needs to be augmented to confront the data.

Keywords: labor market, wage inequality, search frictions, job search
JEL Classification: D83, E24, J31, J41, J63, J64

*We are grateful for comments from Gadi Barlevy, Bjoern Bruegemann, Zvi Eckstein, Chris Flinn, Rasmus Lentz, Dale Mortensen, Giuseppe Moscarini, Fabien Postel-Vinay, Richard Rogerson, Robert Shimer, and Philippe Weil, and from seminar participants at the Cleveland Fed, the Cowles Foundation Workshop, ESSIM 2004 (Cyprus), Georgetown, NYU, Philadelphia Fed, Penn State, Princeton, SED 2006, Southern California, Yale, and York University. Greg Kaplan and Matthias Kredler provided excellent research assistance. The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. Our e-mail addresses are: `andreas.hornstein@rich.frb.org`, `pkrusell@princeton.edu`, and `gianluca.violante@nyu.edu`

†Federal Reserve Bank of Richmond

‡Princeton University, CEPR, and NBER

§New York University and CEPR

1 Introduction

The economic success of individuals is largely determined by their labor market experience. For centuries, economists have been interested in studying the determinants of earnings dispersion among workers. The standard theories of wage differentials in competitive environments are three. Human capital theory suggests that a set of individual characteristics (e.g., individual ability, education, labor market experience, job tenure) are related to wages because they correlate to productive skills, either innate or cumulated in schools, and on the job. The theory of compensating differentials posits that wage dispersion arises because wages compensate for non-pecuniary characteristics of jobs and occupations such as fringe benefits, amenities, location, and risk. Models of discrimination assume that certain demographic groups are discriminated against by employers and, as such, they earn less for similar skill levels.

Mincerian wage regressions based on cross-sectional individual data proxy all these factors through a large range of observable variables, but typically they can explain at most 1/3 of the total wage variation. A vast amount of residual wage variation is left unexplained. In practice, measurement error is large, and the available covariates capture only imperfectly what the theory suggests as determinants of wage differentials. However, even if we could perfectly measure what these competitive theories require, we should not expect to explain all observed wage dispersion.

Theories of frictional labor markets which build on the seminal work of McCall (1970), Mortensen (1970), Lucas and Prescott (1974), Burdett (1978), Pissarides (1985), Mortensen and Pissarides (1994), and Burdett and Mortensen (1998) predict that wages can diverge among ex-ante similar workers looking for jobs in the same labor market (e.g., the market for janitors in Philadelphia) because of informational frictions and luck in the search and matching process. We call this type of wage inequality inherently associated to frictions *frictional wage dispersion.*[1]

The canonical search and matching model provides a natural starting point for thinking about frictional wage dispersion. We begin by asking how much frictional wage dispersion the model can generate, and we arrive at a surprisingly general answer. We show that in

[1]Mortensen (2005), which reviews the theoretical and empirical investigations on the subject, calls it *pure* wage dispersion.

the three standard models (the sequential search model, the island model, and the random matching model) one obtains the same analytical expression for a particular measure of frictional wage dispersion: the ratio between the average wage and the lowest (reservation) wage paid in the economy. We call this measure the mean-min (Mm) ratio. The Mm ratio has the convenient property that it does not depend directly on the shape of the wage-offer distribution, which is hard to observe, but only on a small set of structural parameters that can be readily calibrated to reproduce well known features of the U.S. economy. It must also be noted here that our result, since it merely exploits rational search by workers, is consistent with many views on where the wage-offer distribution comes from, including different sorts of search-matching theory as well as efficiency-wage theory.

A calibration of the model—for plausible values of the job finding rate, separation rate, worker's discount/interest rate, and "flow utility value of being unemployed"—predicts $Mm = 1.036$. I.e., the model only generates a 3.6% differential between the average wage and the lowest wage. The reason is that in the search model "good things come only to those who wait", but the data on unemployment duration show that workers do not wait for very long. Thus, the observed search behavior of workers rationalizes only a tiny amount of dispersion in the wage distribution.

The natural follow-up question is: how large is frictional wage dispersion in actual labor markets? Ideally, one would like to access individual wage observations for ex-ante similar workers searching in the same labor market. These requirements pose several challenges that we can only partially address, given data availability. We exploit three alternative data sources: the November 2000 survey from the Occupational Employment Statistics (OES) program, the 1967-1996 waves of the Panel Study of Income Dynamics (PSID), and the 5% Integrated Public Use Microdata Series (IPUMS) sample of the 1990 U.S. Census. Overall, from the empirical work we gauge that the observed Mm ratios are at least twenty times larger than what the model predicts.

Residual wage dispersion measured in the data includes both wage differentials due to frictions, and wage differentials due to unobservable, possibly time-varying, skills that cannot be fully controlled for, given data constraints. Thus, the large discrepancy between model and data we document can be resolved in two ways. First, one can "blame the data": unobserved heterogeneity would account for the large Mm ratios we document, and the remaining frictional wage dispersion is very small, as predicted by the theory. Second, one can "blame the theory": more elaborate, or different, search models need to be explored in order to account for our recorded Mm ratios. We remain agnostic. We do, however, consider it important to pursue each possibility further. As for further data analysis, this paper contains rather detailed analysis already; in particular, our PSID data does control for fixed (time-invariant) individual effects. However, it would be valuable to try to measure more individual-specific components (directly measured time-invariant variables, such as test scores, and time-varying factors such as events altering the value of leisure, e.g., fertility or health shocks). This work is important, but it is beyond the scope of this paper to pursue it here.

As for developing theory—which one would need if indeed our Mm estimates represent true frictional wage dispersion—that can overcome the spectacular failure of the baseline model, we do begin this pursuit here. Our investigation maintains focus on the Mm ratio. This statistic is naturally obtained from the reservation wage equation, a cornerstone of virtually every search model, and in all the three extensions we study, we are able to derive simple closed-form expressions for the Mm.

The first extension introduces risk aversion: risk-averse workers particularly dislike the low-income state (unemployment) and set a low reservation wage, which allows the model to generate a larger Mm ratio. We find that even when we exclude agents from using any form of insurance, a very high risk-aversion coefficient (around eight) is needed to generate what we see in the data. Moreover, we know that the availability of self-insurance (say, with precautionary saving) would help consumers a lot, thus requiring much higher risk aversion still.

The second extension allows for stochastic wage fluctuations during employment, with endogenous separations. If wages vary over the employment spell, then the wage offer drawn during unemployment is not very informative about the value of a job, so a large dispersion of wages could coexist with a small dispersion of job values which is what drives unemployment duration. Also in this case, though, for reasonable calibration the quantitative improvement of the model is modest: wages need to be virtually i.i.d. during the employment relationship for the model to produce a high Mm ratio; in reality, wages are instead close to a random walk.

The third extension allows for on-the-job search, along the lines of the basic job-ladder model of Burdett (1978).[2] The ability to search on the job for new employment opportunities makes unemployed workers less demanding, which reduces their reservation wage and allows the model to generate a higher Mm ratio. This latter modification is the one that shows most promise but, at least in its simplest form, it still falls short of explaining the data. To generate dispersion, this model needs a high arrival rate of offers on the job. However, the high arrival rate implies separations at a frequency that is almost twice that observed.

In conclusion, we find that for plausible parameterizations a remarkably large class of search models has trouble generating the observed amount of residual wage dispersion. This result is also helpful for understanding why the existing empirical structural search literature (see Eckstein and Van den Berg, 2005, for a recent survey) systematically finds very low (negative and large) estimates of the value of non-market time, extremely high estimates of the interest rate, or very large estimates of measurement error or of unobserved worker heterogeneity. Finally, we note that our findings indicate a link between the struggle in understanding the cross-sectional wage distribution and the recently discussed difficulty of replicating the time series of unemployment and vacancies using search theory (see, e.g., Hall, 2005, and Shimer, 2005b): parameter values that help resolve the former problem make it harder to resolve the latter, and vice versa.

The rest of the paper is organized as follows. Section 2 derives the common expression for the Mm ratio in three canonical search models and quantifies their implications. Section 3 contains the empirical analysis. Section 4 makes a number of first attempts at rescuing the canonical model. Sections 5, 6, and 7 then outline the three significant extensions of the model mentioned above and evaluate them quantitatively. Section 8 discusses the empirical search literature from our perspective. Section 9 concludes the paper. Finally, some of the theoretical propositions in the present paper are proved in a separate Technical Appendix: Hornstein, Krusell, and Violante (2006).

[2]Since none of our derivations depend on the shape of the wage offer distribution, our results hold in the wage-posting equilibrium versions of the job-ladder model (e.g., Burdett and Mortensen, 1998) as well.

2 Frictional wage dispersion in canonical models of equilibrium unemployment

The three canonical models of frictional labor markets are the sequential search model developed by McCall (1970) and Mortensen (1970), the island model of Lucas and Prescott (1974), and the random matching model proposed by Pissarides (1985).

In what follows, we show that all three models lead to the same analytical expression for a particular measure of frictional wage dispersion: the mean-min ratio, i.e., the ratio between the average wage and the lowest wage paid in the labor market to an employed worker. Then, we explore the quantitative implications of this class of models for this particular statistic of frictional wage dispersion.

2.1 The basic search model

We begin with the basic sequential search model formulated in continuous time. Consider an economy populated by ex-ante equal, risk-neutral, infinitely lived individuals who discount the future at rate r. Unemployed agents receive job offers at the instantaneous rate λ_u. Conditionally on receiving an offer, the wage is drawn from a well-behaved distribution function $F(w)$ with upper support $w^{\max}$. Draws are i.i.d. over time and across agents. If a job offer w is accepted, the worker is paid a wage w until the job is exogenously destroyed. Separations occur at rate σ. While unemployed, the worker receives a utility flow b which includes unemployment benefits and a value of leisure and home production, net of search costs. Thus, we have the Bellman equations

$$rW(w) = w - \sigma [W(w) - U] \quad (1)$$

$$rU = b + \lambda_u \int_{w^*}^{w^{\max}} [W(w) - U] \, dF(w), \quad (2)$$

where $rW(w)$ is the flow (per period) value of employment at wage w, and rU is the flow value of unemployment. In writing the latter, we have used the fact that the optimal search behavior of the worker is a reservation-wage strategy: the unemployed worker accepts all wage offers w above $w^* = rU$, at a capital gain $W(w) - U$. Solving equation (1) for $W(w)$ and substituting in (2) yields the reservation wage equation

$$w^* = b + \frac{\lambda_u}{r+\sigma} \int_{w^*}^{w^{\max}} [w - w^*] \, dF(w) .$$

Without loss of generality, let $b = \rho \bar{w}$, where $\bar{w} = E[w | w \geq w^*]$. Then,

$$\begin{aligned} w^* &= \rho \bar{w} + \frac{\lambda_u [1 - F(w^*)]}{r+\sigma} \int_{w^*}^{w^{\max}} [w - w^*] \frac{dF(w)}{1 - F(w^*)} \\ &= \rho \bar{w} + \frac{\lambda_u^*}{r+\sigma} [\bar{w} - w^*] , \end{aligned} \tag{3}$$

where $\lambda_u^* \equiv \lambda_u [1 - F(w^*)]$ is the job-finding rate. Equation (3) relates the lowest wage paid (the reservation wage) to the average wage paid in the economy through a parsimonious set of structural parameters of the model.

If we now define the mean-min wage ratio as $Mm \equiv \bar{w}/w^*$ and rearrange terms in (3), we arrive at

$$Mm = \frac{\frac{\lambda_u^*}{r+\sigma} + 1}{\frac{\lambda_u^*}{r+\sigma} + \rho} . \tag{4}$$

The mean-min ratio Mm is our new measure of frictional wage dispersion, i.e., wage differentials entirely determined by luck in the random meeting process. This measure has one important property: it does not depend *directly* in any way on the shape of the wage distribution F. Put differently, the theory allows predictions on Mm without requiring any information on F. The reason is that all it is relevant to know about F, i.e., its probability mass below w^*, is already contained in the job finding rate λ_u^* that we can measure directly through labor market flows from unemployment to employment and treat as a parameter.

The model's mean-min ratio can thus be written as a function of a four-parameter vector, $(r, \sigma, \rho, \lambda_u^*)$, which we can try to measure independently. Thus, looking at this relation, if we measure the discount rate r to be high (high impatience), for given estimates of σ, ρ, and λ_u^*, an increased Mm must follow. Similarly, a higher measure of the separation rate σ increases Mm (because it reduces job durations and thus decreases the value of waiting for a better job opportunity). A lower estimate of the value of non-market time ρ would also increase Mm (agents are then induced to accept worse matches). Finally, a lower measure of the contact rate λ_u^* pushes Mm up, too (because it makes the option value of search less attractive).

2.2 The search island model

We outline a simple version of the island model as in Rogerson, Shimer, and Wright (2006). Consider an economy with a continuum of islands. Each island is indexed by its productivity level p, distributed as $F(p)$. On each island there is a large number of firms operating a linear technology in labor $y = pn$, where n is the number of workers employed. In every period, there is a perfectly competitive spot market for labor on every island. An employed worker is subject to exogenous separations at rate σ. Upon separation, she enters the unemployment pool. Unemployed workers search for employment and at rate λ_u they run into an island drawn randomly from $F(p)$.

It is immediate that one can obtain exactly the same set of equations (1)-(2) for the worker, while for the firm in each island, we can write its flow value of producing as

$$rJ(p) = pn - wn.$$

Competition among firms drives profits to zero, and thus in equilibrium $w = p$. At this point the mapping between the island model and the search model is complete.[3] The search island model yields the same expression for Mm as in (4).

2.3 The basic random matching model

There are three key differences between the search setup described above and the matching model (e.g., Pissarides, 1985). First, there is free entry of vacant firms (or jobs). Second, the flow of contacts m between vacant jobs and unemployed workers is governed by an aggregate matching technology $m(u, v)$. Let the workers' contact rate be $\lambda_u = m/u$ and the firm's contact rate be $\lambda_f = m/v$. Third, workers and firms are ex-ante equal, but upon meeting they jointly draw a value p, distributed according to $F(p)$ with upper support $p^{\max}$, which determines flow output on their potential match. Once p is revealed, they bargain over the match surplus in a Nash fashion and determine the wage $w(p)$. Let β be the bargaining power of the worker; then the Nash rule for the wage establishes that

$$w(p) = \beta p + (1 - \beta) rU, \tag{5}$$

[3] One can also allow firms to operate a constant returns to scale technology in capital and labor, i.e., $y = pk^{\alpha}n^{1-\alpha}$. If capital is perfectly mobile across islands at the exogenous interest rate r, then firms' optimal choice of capital allows us to rewrite the production technology in a linear fashion, and the equivalence across the two models goes through.

where rU is the flow value of unemployment.[4] This equation uses the free entry condition of firms that drives the value of a vacant job to zero.

From the worker's point of view, it is easy to see that equations (1)-(2) hold with a slight modification:

$$\begin{aligned} rW(p) &= w(p) - \sigma[W(p) - U] \\ rU &= b + \lambda_u \int_{p^*}^{p^{\max}} [W(p) - U]\, dF(p), \end{aligned}$$

i.e., the value of employment is expressed in terms of the value p of the match drawn; similarly, the optimal search strategy is expressed in terms of a reservation productivity p^*. Rearranging these two expressions, we arrive at an equation for the reservation productivity

$$p^* = b + \frac{\lambda_u \beta}{r+\sigma} \int_{p^*}^{p^{\max}} [p - p^*]\, dF(p), \tag{6}$$

where we have used the fact that $p^* = rU$. Substituting (5) into (6), we obtain

$$\begin{aligned} w^* &= b + \frac{\lambda_u [1 - F(p^*)]}{r+\sigma} \int_{p^*}^{p^{\max}} [w(p) - w^*] \frac{dF(p)}{1 - F(p^*)} \\ &= b + \frac{\lambda_u^*}{r+\sigma} [\bar{w} - w^*]. \end{aligned}$$

Using the definition $b = \rho\bar{w}$ in the last equation, we again obtain the formula in (4) for the mean-min ratio. Finally, note that nothing in this derivation depends on the shape of the matching function.

2.4 Quantitative implications for the mean-min ratio

How much frictional wage dispersion can these models generate when plausibly calibrated? We set the period to one month.[5] An interest rate of 5% per year implies $r = 0.0041$. Shimer (2005a) reports, for the period 1967-2004, an average monthly separation rate σ (EU flow) of 2% and a monthly job finding probability (UE flow) of 39%. These two numbers imply a mean unemployment duration of 2.56 months, and an average unemployment rate of 4.88%.

[4]See, for example, Pissarides (2000), Section 1.4, for a step-by-step derivation of this wage equation.

[5]The Mm ratio has the desirable property of being invariant to the length of the time interval. A change in the length of the period affects the numerator and denominator of the ratio $\lambda_u^*/(r+\sigma)$ proportionately, leaving the ratio unchanged. The parameter ρ is unaffected by the period length.

The OECD (2004) reports that the net replacement rate of a single unemployed worker in the U.S. in 2002 was 56%. The fraction of labor force eligible to collect unemployment insurance is close to 90% (Blank and Card, 1991) which implies a mean replacement rate of roughly 50%. Of course, unemployment benefit are only one component of b. Others are the value of leisure, the value of home production (both positive), and the search costs (negative). Shimer (2005b), weighting all these factors, sets ρ to 41%. As discussed by Hagedorn and Manovskii (2006), this is likely to be a lower bound. For example, taxes increase the value of ρ since leisure and home production activities are not taxed. Since higher values for b will strengthen our argument, we proceed conservatively and set $\rho = 0.4$, and in Section 4 we perform a sensitivity analysis.

This choice of parameters implies $Mm = 1.036$: the model can only generate a 3.6% differential between the average wage and the lowest wage paid in the labor market.[6] This number appears very small. What explains the inability of the search/matching model to generate quantitatively significant pure wage dispersion? In the model, workers remain unemployed if the option value of search is high. The latter, in turn, is determined by the dispersion of wage opportunities. The short unemployment durations, as in the U.S. data, reveal that agents in the model do not find it worthwhile to wait because frictional wage inequality is tiny. The message of search theory is that "good things come to those who wait", so if the wait is short, it must be that good things are not likely to happen.

We now turn to the obvious next question: how large is frictional wage dispersion in actual labor markets?

3 An attempt to measure frictional wage dispersion

The aim of our analysis is to quantify the empirical counterpart of the model's mean-min ratio Mm. Ideally, one would like to access individual wage observations for ex-ante similar workers searching in the same labor market. This requirement poses three major challenges that we address by exploiting three alternative data sources: the November 2000 OES survey, the 1967-1996 waves of the PSID, and the 5% IPUMS sample of the 1990 U.S. Census.

[6] The reason why this ratio is close to one is that the term $\lambda_u^* / (\sigma + \rho)$ is very large—around 16—compared to both 1 and ρ, the other two terms of the expression in 4.

The first challenge is to define a "labor market". The most natural boundaries across labor markets are, arguably, geographical, sectoral, and occupational, and possibly combinations of all these dimensions. The PSID sample is too small to construct detailed labor markets. OES and Census data, however, allow us to look at the wage distribution in thousands of separate labor markets in the U.S. economy.

Second, differences in annual earnings may reflect differences in hours worked. To avoid this problem, one should focus on hourly wages. However, it is well known that measurement error in hours worked plagues household surveys, and large measurement error will generate an upward bias in estimates of wage dispersion. The OES is an establishment survey where measurement error should be negligible. PSID and Census information on hourly wages, though, may suffer from measurement error bias. In particular, estimates of the reservation wage through the lowest wage observation are especially subject to outliers due to reporting or imputing errors. As an alternative, we estimate the reservation wage from the 1st, 5th, and 10th percentile of the wage distribution. These percentiles are less volatile, even though upward biased, estimators of the reservation wage in the empirical wage distribution. We denote the corresponding mean-min ratios as $Mp1$, $Mp5$, and $Mp10$, respectively.

Third, we would like to eliminate all wage variation due to ex-ante differences across individuals in observable characteristics (e.g., experience, education, race, gender, etc.), as well as all wage variation due to heterogeneity in unobservables (e.g., innate ability, value of leisure, etc.). The OES does not provide any demographic information on workers. In the Census data, we can control for a wide set of observable characteristics. When using the PSID, we can exploit the panel dimension to also purge fixed individual heterogeneity from the wage data.

Overall, none of the three data sets is ideal, but each of them is informative in its own way.

3.1 OES Data

The first data source we use is the OES. Appendix A contains a description of the survey and of the sample selection criteria we adopt.

Table 1: Dispersion measures from the 2000 OES

	Number of labor mkts	Ratio of mean wage to 10th percentile
Occupation	637	1.68
Occ./Industry	6,293	1.60
Occ./Geog Area	106,278	1.48

OES data are collected at the level of the establishment. Each establishment reports the average hourly wage paid within each occupation. To the extent that there are within-establishment differences in wages due to luck or frictions among similar workers, these data underestimate frictional wage dispersion.

We use three different levels of aggregation: nation-wide data by occupation (3-digit), occupation $\times$ metropolitan area data, and occupation (3-digit) $\times$ industry (2-digit) data. The publicly available survey reports the average, the 10th, 25th, 50th, 75th and 90th percentiles of the hourly wage distribution.

The best possible estimate (which clearly is downward-biased) of the mean-min ratio is the ratio between the average wage and the 10th percentile ($Mp10$). We exclude all those cells where hourly wage data are not available, and those where the 90th percentile is top coded (at \$70 per hour)—a sign that wages are heavily censored in that cell.[7] In each cell (i.e., a labor market) we compute the $Mp10$, and then we calculate the median $Mp10$ ratio across labor markets. The median is preferable to the mean because we consistently found that the empirical distributions of mean-min ratios are very skewed to the right, and we are interested in the mean-min ratio of the wage distribution in a typical labor market.

The median $Mp10$ ratio across occupations in the U.S. economy is 1.68. For the classification of labor markets based on 2-digit industry (58 industries) and occupation, the median $Mp10$ ratio is 1.60. Finally, when we define labor markets by metropolitan area (337 areas) and occupation, the median $Mp10$ ratio is estimated to be 1.48. Table 1 summarizes the results.

[7] The first restriction mainly excludes workers in the Education sector (25-000), while the second mainly excludes Healthcare Practitioners (29-000).

As the definition of labor market becomes more refined, wage dispersion falls for two reasons. First, there is less worker heterogeneity within a specific occupation in a given industry, or in a given geographical area than at the country level. Second, as we keep disaggregating, the number of establishments sampled within each cell falls. For example, for cells defined by occupation and metropolitan area, we have on average only 11 establishments per cell. With such a low number of observations, the estimate of the 10th percentile could be severely upward biased, and in turn the mean-min ratio underestimated.

3.2 PSID Data

Our second data source is the Panel Study of Income Dynamics (PSID). In Appendix B, we describe in detail our sample selection. With our final sample in hand, for every year in the period 1967-1996, we run an OLS regression on the cross-section of individual log hourly wages where we control for gender, 3 race dummies (white, non-white, Hispanics), 5 education dummies (high-school dropout, high-school degree, some college, college degree and post-graduate degree), a cubic in potential experience (age minus years of education minus five), a dummy for marital status, 6 regional dummies, 25 two-digit occupation dummies, and interaction between occupation and experience to capture occupation-specific tenure profiles. We face a trade-off in the choice of covariates between (1) the appropriate filtering out of the variation in hourly wages due to observable individual characteristics which are rewarded in the labor market, and (2) the risk of overfitting the data. On average, these year-by-year regressions yield an R^2 between 0.42 and 0.45.

Next, we use the panel dimension of the data and identify individual-specific effects in wages. Let ε_{it} be the residual of the first stage for individual $i = 1, ...I$ in year $t = 1, ..., T$. We limit the sample to those whose number of wage observations in the panel (N_i) is at least ten and estimate $\bar{\varepsilon}_i = \sum_{t=1}^{N_i} \varepsilon_{it}/N_i$ for every individual. The vector $\{\bar{\varepsilon}_i\}_{i=1}^{I}$ captures the variation in fixed unobserved individual factors (e.g., innate ability, preference for leisure) which affect wages. Let $\tilde{w}_{it} = \exp(\varepsilon_{it} - \bar{\varepsilon}_i)$. For each year t, we then calculate our indexes of residual inequality across workers on $\tilde{w}_{it}$.

We report our results for the PSID in Figure 1. For comparison with the other data sources, we comment on the values for the last part of the sample (the 1990-1996). The ratio between mean wage and lowest wage residual from the basic Mincer regressions is $Mm = 4.47$, but the estimate is clearly very noisy. When the reservation wage is estimated from the 1st and the 5th percentile of the wage distribution, the noise is much reduced and we obtain, respectively, $Mp1 = 2.73$ and $Mp5 = 2.08$. The coefficient of variation of the regression residuals is 0.50.

Controlling for individual effects drastically cuts the estimate of the mean-min ratios by more than half. For the period 1990-1996, we estimate $Mm = 3.11$, $Mp1 = 1.90$, and $Mp5 = 1.46$. The coefficient of variation of the residuals net of individual effects falls to 0.25.[8] One should be cautious in interpreting these results, however. The estimated fixed effects confound worker-specific characteristics with match- (or firm-) specific effects. This is especially true for long-lived matches. Removing the estimated fixed effects may therefore eliminate some of the variation in the data we want to explain.

To facilitate the comparison with the OES data, we also report ratios between the mean and the 10th percentile. The $Mp10$ on the residuals of the first-stage regression equals 1.77, and the $Mp10$ on the residuals net of individual-specific means is 1.32. The corresponding statistics for the OES data all lie somewhere in between (see Table 1).

3.3 Census Data

Our third data source is the 5% Integrated Public Use Microdata Series (IPUMS) sample of the 1990 United States Census. In Appendix C, we outline our sample selection. As for the PSID, we run an OLS regression on the log of individual hourly wage to control for the variation in wages due to observable characteristics which are rewarded in the labor market. Among the covariates, we include gender dummies, 3 race dummies, 5 education dummies, and a cubic in potential experience. We weight each observation by its Census sample weight. The regression explains 31% of the total variation in log hourly wages.[9]

Next, we group the (exponent of the) regression residuals by labor markets. Our

[8] In passing, we note that Figure 1 is consistent with the views that residual wage dispersion has risen significantly over the period, and that the rise in prices for unobserved innate characteristics is a key component of this phenomenon.

[9] This R^2 is sizeably lower than the one for the PSID regression, and it is more in line with typical Mincerian regressions. The reason is that in the PSID regression we included occupational dummies and occupation-specific experience profiles, which have strong explanatory power. Moreover, the PSID sample is smaller by a factor of 1,500 compared to the Census sample. See Appendices B and C.

first definition of labor market is the individual occupation. The Census allows us to distinguish between 487 distinct occupations (variable OCC). Our second definition is the combination of occupation and place of work (for the main job). Our indicator for place of work is constructed as follows. Whenever possible, we use the 329 metropolitan areas (PWMETRO). For rural areas we use a variable (PWPUMA) which defines a geographical area by following the boundaries of groups of counties, or census-defined "places" which contain up to 200,000 residents. Overall, we end up with 799 different geographical areas. For each cell identified as a labor market, we calculate the mean-min ratio and we report the median mean-min ratio across labor markets.

In the top panel of Figure 2, we show one example of the wage distribution for Janitors and Cleaners, excluding Maids and House Cleaners (code 453) in the Philadelphia metropolitan area (code 616). These are the wage residuals of a regression that controls for the demographics listed above, restricted to those working full time (35-45 hours per week), full year (48-52 weeks per year) to reduce the role of measurement error. Overall, we have 572 observations. As reported in the figure, the ratio between the mean and the first percentile is 2.24. In the bottom panel, we display the distribution of $Mp1$ ratios, obtained exactly as for the Philadelphia cell, for Janitors and Cleaners across all geographical areas in the U.S. economy for which we have at least 50 individual wage observations (131 areas). There are local labor markets displaying more and markets displaying less residual dispersion than in Philadelphia, but the bottom line seems to be that, even within a very unskilled occupation such as Janitors, and even after selecting the sample to minimize the role of measurement error, wage differentials remain large: the median $Mp1$ ratio for Janitors across the U.S. is 2.20.

Table 2 reports our results in a number of formats. We use various estimators of the mean-min ratio: $Mm, Mp1, Mp5$, and $Mp10$. We condition on cells with at least 50 observations and with at least 200. Larger cells usually display higher Mm ratios both because of higher unobserved heterogeneity that we did not capture in the first-stage regression, and because they permit a less biased estimate of the low percentiles of the wage distribution.

	Min. obs. per cell	Number of labor mkts	Ratio of mean wage to min.	1st pct.	5th pct.	10th pct.	CV
(1) Occupation		487	4.54	2.83	2.13	1.83	0.47
(2) Occ./Geog. Area	(N≥50)	13,246	2.94	2.66	2.04	1.76	0.41
	(N≥200)	2,321	3.85	2.88	2.13	1.82	0.44
(3) Occ./Geog. Area	(N≥50)	7,195	2.74	2.49	1.92	1.66	0.35
Full time/Full year	(N≥200)	1,117	3.58	2.68	1.98	1.71	0.37
(4) Occ./Geog. Area	(N≥50)	2,810	2.64	2.46	1.92	1.68	0.40
Experience ≤ 10	(N≥200)	406	3.33	2.57	1.97	1.73	0.44
(5) Occ./Geog. Area	(N≥50)	1,152	2.51	2.37	1.98	1.77	0.45
Unskilled Occ.	(N≥200)	191	2.95	2.57	2.08	1.83	0.49
(6) Occ./Geog. Area	(N≥50)	13,246	2.61	2.39	1.88	1.64	0.39
Within cell regression	(N≥200)	2,321	3.33	2.66	2.02	1.75	0.42

Table 2: Dispersion measures for hourly wage from the 1990 Census

To get at the measurement error issue, we condition our analysis on full-time, full-year workers who report weekly hours between 35 and 45 and annual weeks worked between 48 and 52. Going from row (2) to row (3) in Table 2, wage dispersion falls with respect to the full sample, but it remains very high.[10]

To eliminate the importance of individual-specific differences in cumulated skills not perfectly correlated with experience (accounted for in the first-stage regression), we condition on workers with less than 10 years of experience, Table 2 row (4), and on a set of very low-skilled occupations, Table 2 row (5), where occupation and firm-specific skills are arguably not very important.[11] Once again, going from row (2) to either row (4) or row (5) the findings are barely affected.

[10] The coefficient of variation falls by 16%. For comparison, Bound and Krueger (Table 6, 1991) compare matched Current Population Survey data to administrative Social Security payroll tax records and find that the measurement error explains between 7% and 19% of the total standard deviation of log earnings. Recall that the standard deviation of the logs has the same scale of the coefficient of variation.

[11] This list includes, inter alia, Launderers and ironers, Crossing guards, Waiters and waitresses, Food counter, fountain and related occupations, Janitors and cleaners, Elevator operators, Pest control occupations, and Baggage porters and bellhops.

Finally, we also run the first-stage regressions within each occupation/area cell, to account for the fact that the role of demographic characteristics in wage determination may be different across occupations. Going from row (2) to row (6), estimates of dispersion fall by less than 10%.

We conclude that, except for estimates of Mm based on the lowest observed wage that are quite volatile, all the other statistics in Table 2 remain very robust to all these controls and strongly support the view that residual wage dispersion is large.

3.4 Summary and interpretation

Our three independent data sources offer a fairly consistent view of the size of residual wage dispersion within narrowly defined labor markets. If we focus our attention on the $Mp5$ estimate of the mean-min ratio, a review of our findings yields $Mp5 = 1.46$ from PSID. The PSID estimate could be upward-biased because of measurement error, but at the same time the individual wage demeaning could filter out too much variation, including variation due to "persistent luck" components that should be included in measures of frictional dispersion. From the Census sample restricted to full-time, full-year workers (where measurement error in hours should be negligible) we have estimated $Mp5 = 1.98$. Given the OES estimate of the $Mp10$, and the fact that the other two data sets suggest that $Mp5$ are roughly 10%-15% larger, we conjecture that the $Mp5$ in the OES data may be around 1.67. An average across the three data sets yields 1.70, which we use as a target in the rest of our analysis.

This appraisal of residual wage dispersion—based not on the minimum wage observed, but on the 5th percentile, and hence quite conservative—is about *20 times larger* than that implied by the textbook models of Section 2. How can we resolve this enormous discrepancy between the size of residual dispersion in the data and the model-implied frictional dispersion?

One reaction to our findings is that the actual wage data hide large differentials due to unobservable skills that we cannot fully control for with our given data. This is possible, though one has to bear in mind that with our PSID data, we have controlled for unobserved heterogeneity that is fixed over time.[12] Thus, for heterogeneity to explain our large

[12] In an unreported set of regressions on PSID data, we also allowed fixed differences in (linear and

Mm ratios, it would have to involve time-varying, unobserved skills, or preferences, which influence remuneration in the labor market. Such heterogeneity cannot be ruled out a priori, and it is important to continue incorporating more detailed worker information to isolate this source in future work.[13]

One can also imagine the presence of firm-specific skills that are not perfectly correlated with experience. Estimates of returns to firm tenure vary widely. Topel (1991) estimates that 10 years of seniority increase log hourly wages by 25%. Altonji and Shakotko (1987) report estimates below 7%. Recently, Altonji and Williams (2004) have reassessed the evidence, concluding that returns to tenure over 10 years could be around 11%, most of them occurring in the first 5 years of the employment spell.[14] If we assume that average tenure is roughly 4 years (see section 7), then this factor may account for at most 1/10th of the wage difference between the average worker (with average tenure) and the lowest paid worker (with zero tenure) in a given occupation/geographical area. Residual wage dispersion remains very high.

In conclusion, we have made efforts along several fronts to isolate frictional wage dispersion as well as possible, and our "residual" measure remains large—much larger than what textbook models predict. If indeed this discrepancy is due to poor data and, after all, unobservable worker characteristics (which vary over time) do account for the bulk of our computed Mm ratios, then one would hope that more careful future work will reveal this. We do not, however, consider it satisfactory to simply stop here. Many hold the prior that "markets are close to frictionless, because any significant wage differences for identical workers would be exploited by profit-maximizing employers" but, after all, this is a belief, and we must strive to update and test our beliefs, especially when preliminary data analysis suggests that such priors may be off. Therefore, we also proceed on a parallel

quadratic) time trends across workers—possibly capturing differences in learning ability—and this did not significantly change our findings either.

[13]Incidentally, a large class of quantitative macroeconomic models of the Bewley-Huggett-Aiyagari style implicitly takes this view, presuming idiosyncratic risks which are modelled as a stochastic process for efficiency units of labor, priced in a frictionless labor market. An example of this approach in the search literature is Ljungqvist and Sargent (1998).

[14]Kambourov and Manovskii (2004) argue that the bulk of returns to specific human capital is occupational-specific: 5 years of occupational tenure increase wages between 12% and 20%. They find that once occupation is taken into account, returns to human capital specific to industry and employer become virtually zero. Hence, their findings would change the nature of wage differentials due to unobserved heterogeneity specific skills, but not their magnitude.

front, which is to examine the consequences of the data at hand actually revealing large true frictional wage dispersion: what, then, is so spectacularly wrong with textbook theory? To find out, we first make some attempts to rescue the baseline model, without changing its key features. Next, we investigate ways to augment the model so as to improve its performance.

4 Five attempts to rescue the baseline model

4.1 Unemployment vs. wage dispersion

In defense of the model, one might argue that it is designed to explain unemployment, not wage dispersion. This argument is flawed: in the search model, there is a tight link between the existence of unemployment and the existence of wage dispersion. Unemployment exists because of the option value of searching for better wage opportunities. Let us reverse our logic and suppose that, given the amount of frictional wage dispersion observed empirically, we want to predict unemployment duration, i.e., use equation (4) and the empirical value of $Mm = 1.7$ to compute the implied value for λ_u^*. We would obtain $\lambda_u^* = 0.011$. In other words, a search model consistent with the amount of wage dispersion in the data predicts an expected unemployment duration of 91 months, 35 times the average duration in U.S. data.[15]

4.2 Alternative parameterizations

To calibrate the pair (λ_u^*, σ), we used the UE and EU flow data. One could argue that we should also incorporate flows in and out of the labor force. Taking this into account, Shimer (2005a) reports the monthly separation rate to be 3.5%, and the monthly job finding rate to be 61%. For the same values of r and ρ used in the baseline calibration, we obtain $Mm = 1.038$.[16]

[15]In one of the most commonly used search setups, Mortensen and Pissarides (1994), unemployment duration is not connected to wage dispersion, because in that model, unemployed workers always receive the maximum wage upon employment: they never consider turning down a job offer. There, however, the separation rate is determined by a reservation-wage strategy since there are wage shocks on the job. Thus, that model instead links wage dispersion to the observed separation rate. We explore this link in Section 6 below.

[16]Since, under the new parameterization, both the job finding rate and the separation rate increase, the effect on the Mm ratio is negligible.

With respect to the interest rate r, we have used a standard value, but it is possible that unemployed workers, especially the long-term unemployed, face a higher effective interest rate if they wanted to borrow. Much less is known about ρ. To assess the robustness of our conclusions to the choice of values for these two parameters, in Figure 3, we plot the pairs (r, ρ) which are consistent with an Mm ratio of 1.7, together with the region of "reasonable pairs" based on our prior. This region covers the area where $\rho \in (0, 1)$ and r is at most 27% per year.

The results are striking and suggest the baseline model cannot be rescued: even for annual interest rates around 40% per year, one would need agents to value one month of time away from the market the equivalent of *minus* three times the average monthly wage. Positive net values of non-market time are consistent with the observed wage dispersion only for interest rates beyond 1,350% per year.[17]

4.3 Ability differences

Wage inequality can, of course, naturally arise from ability differences. A very simple illustration, extending the above search setting, goes as follows: there are two worker types, and type 1 is more productive than type 2 by μ percent in the following sense: $F_2(w) \equiv F(w)$ and $F_1(w) = F(w/(1+\mu))$ for all w. Suppose also that the workers have the same values for ρ, r, σ, and λ_u. Then

$$w_i^* = \rho \bar{w}_i + \frac{\lambda_u(1 - F_i(w_i^*))}{r + \sigma} (\bar{w}_i - w_i^*) \tag{7}$$

for each type i. It is easy to show that if w_2^* and $\bar{w}_2 = \int_{w_2^*} wF(dw)$ solve (7) for $i = 2$, then, using the assumed symmetry, $w_1^* = (1+\mu)w_2^*$ implies $F_1(w_1^*) = F_2(w_2^*)$ and $\bar{w}_1 = \int_{w_1^*} wF_1(dw) = (1+\mu)\bar{w}_2$, and therefore solves (7) for $i = 1$. Thus, in this model the observed wage distribution for the type-1 worker is a μ-percent scaling up of that of type-2 workers, and $\bar{w}_1/w_1^* = \bar{w}_2/w_2^*$, which will be a small number given the above analysis.

[17] A number of authors in the health and social behavioral sciences have argued that unemployment can lead to stress-related illnesses due to financial insecurity or to a loss of self-esteem. This psychological cost would imply an additional negative component in b. Economists have argued that this empirical literature outside economics has not convincingly solved the serious endogeneity problem underlying the relationship between employment and health status and even have, at times, reached the opposite conclusion, i.e., that there is a positive association between time spent in non-market activity and health status (e.g., Ruhm 2003).

However, for the population, $\bar{w} = \alpha\bar{w}_1 + (1-\alpha)\bar{w}_2$, where α is the share in population of type 1. So the population-wide mean-min ratio, which the econometrician observes, will equal

$$Mm = \frac{\alpha\bar{w}_1 + (1-\alpha)\bar{w}_2}{w_2^*} = \left(\alpha\frac{\bar{w}_1}{\bar{w}_2} + 1 - \alpha\right)\frac{\bar{w}_2}{w_2^*} = (1+\alpha\mu)\,\frac{\bar{w}_2}{w_2^*}.$$

Thus, if μ is large and α is not too small, we can obtain large population mean-min values, even though mean-min values within groups are small. In particular, large enough ability difference will generate any desired mean-min ratio for the overall population.

The model just described, however, is not one of frictional wage inequality, but rather one of ability-driven wage differences. Moreover, our PSID-based empirical analysis eliminates individual-specific effects, and we still find very large values for the mean-min ratio. A related model could also be constructed assuming that the types are random; perhaps they follow a Markov chain, depicting the evolution of human capital on (and perhaps also off) the job. Similar settings have been used by Ljungqvist and Sargent (1998) and Kambourov and Manovskii (2004). One can use arguments along the lines of those above to demonstrate that such settings also allow larger wage inequality, but only due to the skill differences between types being large: for a given type, wage inequality is still small, and thus large frictional wage inequality—as we view it—does not result here either.

4.4 Targeting European data

In Section 2.4 we have indicated that the short duration of unemployment in the U.S. reveals that frictional wage dispersion must be small. It is well known that in Europe unemployment spells last much longer, on average. For example, Machin and Manning (1999, Table 1) document that in 1995 the proportion of workers unemployed for more than 12 months was less than 10% in the U.S., but over 40% in France, Germany, Greece, Italy, Portugal, Spain, and the United Kingdom. Does this observation give hope to the model to be more successful in explaining European wage dispersion? To answer this question, recall that in a stationary equilibrium, unemployment is $u = \sigma/\left(\sigma + \lambda_u^*\right)$. Using this formula in expression (4) allows one to rewrite the Mm ratio as

$$Mm = \frac{\frac{\sigma}{r+\sigma}\frac{1-u}{u} + 1}{\frac{\sigma}{r+\sigma}\frac{1-u}{u} + \rho} \approx \frac{\frac{1-u}{u} + 1}{\frac{1-u}{u} + \rho},$$

where the "approximately equal" sign is obtained by setting $r = 0$, a step justified by the fact that r is second order compared to the other parameters in that expression. Setting the unemployment rate to 10%, with $\rho = 0.4$, one obtains $Mm = 1.076$. The reason for the small improvement is that in European data both unemployment duration and employment spells are much longer than in the U.S. labor market. While the first fact is consistent with larger equilibrium wage dispersion, the second implies that unemployed workers are more selective and set their reservation wage high, which reduces frictional wage dispersion.

For the argument to be fully convincing, one would need to document the extent of residual wage dispersion, as well as the magnitude of the value of non market time, in European countries. A systematic investigation goes beyond the scope of this paper, but conventional wisdom would suggest that while inequality is lower in Europe, social benefits for the unemployed are much more generous.[18]

4.5 Implications for other dispersion measures

Admittedly, the mean-min ratio is not a common index of dispersion. One may argue that even though the model fares poorly in terms of this statistic, its performance along more common measures of dispersion, such as the coefficient of variation (cv), could be satisfactory. To answer this question, we need to make further assumptions about the equilibrium wage distribution. Given a parametric specification for this distribution, we can map predicted mean-min ratios into cv's, i.e., we can determine the value of the cv corresponding to a certain value for the mean-min ratio.

The Gamma distribution (see Mood et al., 1974, for a standard reference) is a convenient choice because it is a flexible parametric family and has certain properties that are useful in our application. Let wages w be distributed according to the density

$$g(w; w^*, \alpha, \gamma) = \frac{\left(\frac{w-w^*}{\alpha}\right)^{\gamma-1} \exp\left(-\frac{w-w^*}{\alpha}\right)}{\alpha \Gamma(\gamma)}, \tag{8}$$

with $\gamma, \alpha > 0$, and with $\Gamma(\gamma)$ denoting the Gamma function. The value for w^* is the location parameter and determines the lowest wage observation, α is the scale parameter

[18]For example, Hansen (1998) calculates benefits replacement ratios with respect to the average wage up to 75% in some European countries.

determining how spread out the density is on its domain, whereas γ is the parameter that determines the shape of the function (e.g., exponentially declining, bell-shaped, etc.).[19]

The mean and standard deviation of a random variable distributed with $g(w; w^*, \alpha, \gamma)$ are given by, respectively, $\bar{m} = w^* + \alpha\gamma$ and $sd = \alpha\sqrt{\gamma}$. Recalling that $Mm = \bar{m}/w^*$, it is easy to obtain a relation between the coefficient of variation cv and the mean-min ratio, and this relation only depends on the shape parameter γ:

$$cv = \frac{1}{\sqrt{\gamma}} \left[\frac{Mm - 1}{Mm} \right]. \tag{9}$$

The empirical analysis of section (3) suggest that $cv = 0.30$ and $Mm = 1.7$ are reasonably conservative estimates for the coefficient of variation and the mean-min ratio of the wage distribution within labor markets.[20] From equation (9), this implies $\gamma = 1.88$.

A search model generating a mean-min ratio of 1.036, under the Gamma wage offer distribution assumption, would generate a coefficient of variation for hourly wages of 0.025, i.e., 1/12th of the coefficient of variation in the wage data. We conclude that the failure of the model generalizes to more common measures of dispersion.

4.6 Non-pecuniary job attributes

In many jobs, wages are only one component of total compensation. In a search model where a job offer is a bundle of a monetary component and a non-pecuniary component, short unemployment duration can coexist with large wage dispersion, as long as non-pecuniary job attributes are *negatively correlated* with wages so that the dispersion of total job values is indeed small.

This hypothesis, which combines the theory of compensating differentials with search theory, does not show too much promise. First, it is well known that certain key non-monetary benefits such as health insurance tend to be positively correlated with the wage, e.g., through firm size.[21] Second, illness or injury risks are very occupation-specific and our measures of frictional wage dispersion are within-occupation indexes. Third, differences

[19] The Gamma family is very flexible: it includes the Weibull (hence, the exponential) distribution for $\gamma = 1$ and the lognormal, in the limit as $\gamma \to \infty$.

[20] This value for cv is an average between the PSID estimate (0.25) and the estimate on Census sample of full-time, full-year workers (0.35).

[21] For example, the mean wage in jobs offering health insurance coverage is 15%-20% higher than in those not offering it; see Dey and Flinn (2006).

in work shifts and part-time penalties are quantitatively small. Kostiuk (1990) shows that genuine compensating differentials between day and night shifts can explain at most 9% of wage gaps. Manning and Petrongolo (2005) calculate that part time penalties for observationally similar workers are around 3%.

We now examine three extensions of the baseline search model that go, qualitatively, in the direction of producing more frictional wage dispersion: risk aversion, stochastic wages during the employment relationship, and on-the-job search.

5 Risk aversion

Risk-averse workers particularly dislike states with low consumption, like unemployment. Compared to risk-neutral workers, ceteris paribus, they lower their reservation wage in order to exit unemployment rapidly, thus allowing Mm to increase.

Let $u(c)$ be the utility of consumption, with $u' > 0$, and $u'' < 0$. To make progress analytically, we assume workers have no access to storage, i.e., $c = w$ when employed, and $c = b$ when unemployed. It is clear that this model will give an *upper bound* for the role of risk aversion: with any access to storage, self-insurance or borrowing, agents can better smooth consumption, thus becoming effectively less risk-averse.

To obtain the reservation wage equation with risk aversion, observe that in the Bellman equations for the value of employment and unemployment, the monetary flow values of work and leisure are simply replaced by their corresponding utility values. The reservation wage equation (3) then becomes

$$u(w^*) = u(\rho\bar{w}) + \frac{\lambda_u^*}{r+\sigma}\left[E(u(w)) - u(w^*)\right]. \tag{10}$$

A second-order Taylor expansion of $u(w)$ around the conditional mean $\bar{w}$ yields

$$u(w) \simeq u(\bar{w}) + u'(\bar{w})(w-\bar{w}) + \frac{1}{2}u''(\bar{w})(w-\bar{w})^2.$$

Take the conditional expectation of both sides of the above equation and arrive at

$$E(u(w)) \simeq u(\bar{w}) + \frac{1}{2}u''(\bar{w})\, var(w). \tag{11}$$

Let $u(w)$ belong to the CRRA family, with θ representing the coefficient of relative risk aversion. Then, using (11) in (10), and rearranging, we obtain

$$Mm \equiv \frac{\bar{w}}{w^*} \simeq \left[\frac{\frac{\lambda_u^*}{r+\sigma}\left(1+\frac{1}{2}\left(\theta-1\right)\theta cv^2\left(w\right)\right)+\rho^{1-\theta}}{\frac{\lambda_u^*}{r+\sigma}+1} \right]^{\frac{1}{\theta-1}}. \tag{12}$$

It is immediate to see that, for $\theta = 0$, the risk-neutrality case, the expression above equals that in equation (4).

To assess the quantitative role of risk aversion, we use the same parameterization of the risk-neutral case, and based on the evidence provided in section 3, we set $cv(w) = 0.30$. Figure 4 plots the pairs of (ρ, θ) consistent with $Mm = 1.70$. For $\theta = 8$, the model can match the data. Recall, however, the upper-bound nature of our experiment: in fact, plausibly calibrated models of risk-averse individuals who have access to a risk-free bond for saving and borrowing are much closer to full insurance than to autarky (see, e.g., Aiyagari, 1994). For example, it is well known that as $r \to 0$, the bond economy converges to complete markets (Levine and Zame, 2002).

6 Wage shocks during employment

We now extend the basic search model by allowing wages to fluctuate stochastically along the employment spell.[22] Unemployed workers draw wage offers from the distribution $F(w)$ at rate λ_u, but now these wage offers are not permanent. At rate δ, the wage changes, and the worker draws again from $F(w)$. Draws are i.i.d. over time. Separations are now endogenous and will occur at rate $\sigma^* \equiv \delta F(w^*)$, where w^* is the reservation wage.

The reason why this generalization can potentially generate a larger Mm ratio is that the particular value drawn from $F(w)$ by an unemployed worker is not a good predictor of the continuation value of employment, if the wage is very volatile. Unemployed workers will therefore be more willing to accept initially low wage offers, which reduces w^* and increases dispersion.

The Bellman equations for employment and unemployment are, respectively,

$$\begin{aligned} rW(w) &= w + \delta \int_{w^*}^{w^{\max}} \left[W(z) - W(w)\right] dF(z) - \delta F(w^*) \left[W(w) - U\right] \\ rU &= b + \lambda_u \int_{w^*}^{w^{\max}} \left[W(z) - U\right] dF(z). \end{aligned}$$

[22] The Technical Appendix, Hornstein et al. (2006), contains the details of all the derivations in this and the next sections.

With respect to equation (1), the value of employment is modified in two ways. First, the endogenous separation rate is now $\delta F(w^*)$. Second, there is a surplus value from accepting a job at wage w which is given by the second term on the right-hand side. Exploiting the definition of reservation wage $W(w^*) = U$, integrating by parts, and using $W'(w) = 1/(r+\delta)$, we arrive at

$$w^* = b + \frac{\lambda_u - \delta}{r + \delta} \int_{w^*}^{w^{\max}} [1 - F(z)]\, dz.$$

Now, using the definition of the conditional mean wage $\bar{w}$, we obtain

$$w^* = b + \frac{(\lambda_u - \delta)[1 - F(w^*)]}{r + \delta} (\bar{w} - w^*).$$

Therefore, imposing $b = \rho \bar{w}$, and rearranging, we can write the Mm ratio in this model as

$$Mm = \frac{\bar{w}}{w^*} = \frac{\frac{\lambda_u^* - \delta + \sigma^*}{r+\delta} + 1}{\frac{\lambda_u^* - \delta + \sigma^*}{r+\delta} + \rho}. \tag{13}$$

As $\delta \to 0$, the Mm ratio converges to equation (4) with $\sigma^* = 0$, since without any shock during employment every job lasts forever. As $\delta \to \lambda_u$, unemployed workers accept every offer above b since being on the job has an option value equal to being unemployed.

The parameter δ maps into the degree of persistence of the wage process during employment. In particular, in a discrete time model where $\delta \in (0, 1)$ the autocorrelation coefficient of the wage process is $1 - \delta$.[23] Individual panel data suggest that residual wages are very persistent, indeed near a random walk, so plausible values of δ are close to zero.

We repeat the exercise in Section 4 on the (r, ρ) pair. Given values for $(\lambda_u^*, \sigma^*, r)$ one can search for the values of (δ, ρ) that generate Mm ratios of the observed magnitude. Figure 5 reports the results. Once again, the model is very far from the data for reasonable values of ρ and of the degree of wage persistence of wage shocks. For example, for $\delta = 0.1$ (corresponding to an annual autocorrelation coefficient of 0.9) the model requires $\rho = -13$. Only for virtually i.i.d. wage shocks ($\delta \approx 1$) would the model succeed.

The setup of Mortensen and Pissarides (1994) is similar to that described here, with one difference: upon employment, all workers start with the highest wage, $w^{\max}$, and thus

[23] It is easily seen that a discrete time version of this model leads exactly to equation (13).

they only sample from $F(w)$ while employed. This model, however, does not offer higher mean-min ratios: the resulting Mm ratio can be shown to be strictly bounded above by that in equation (13).

7 On-the-job search

Allowing on-the-job search goes, qualitatively, in the right direction for reasons similar to the model with stochastic wages. If the arrival rate of offers on the job is high, workers are willing to leave the ranks of unemployment quickly since they do not entirely forego the option value of search. This property breaks the link between duration of unemployment and wage dispersion that dooms the baseline model. However, for a proper test, we now need to explore the implied labor-market flows, which now include employment-to-employment transitions.

We generalize the model of Section 2 and turn it into the canonical job ladder model outlined by Burdett (1978). A worker employed with wage $\hat{w}$ encounters new job opportunities w at rate λ_w. These opportunities are drawn from the wage offer distribution $F(w)$ and they are accepted if $w > \hat{w}$.

A large class of equilibrium wage posting models, starting from the seminal work by Burdett and Mortensen (1998), derives the optimal wage policy of the firms and the implied equilibrium wage offer distribution as a function of structural parameters. It is not necessary, at any point in our derivations, to specify what $F(w)$ looks like. Our expression for Mm will hold in any equilibrium wage posting model that satisfies the following two assumptions. First, employed workers accept any wage offer above their current wage; second, every worker (employed or unemployed) faces the same wage offer distribution. Moreover, without loss of generality, to simplify the algebra, we posit that no firm would offer a wage below the reservation wage w^*; thus, $F(w^*) = 0$.

The flow values of employment and unemployment are

$$\begin{aligned} rW(w) &= w + \lambda_w \int_w^{w^{\max}} [W(z) - W(w)]\, dF(z) - \sigma [W(w) - U] \\ rU &= b + \lambda_u \int_{w^*}^{w^{\max}} [W(z) - U]\, dF(z), \end{aligned}$$

and the reservation wage equation becomes

$$w^* = b + (\lambda_u - \lambda_w) \int_{w^*}^{w^{\max}} \frac{1 - F(z)}{r + \sigma + \lambda_w [1 - F(z)]} dz. \tag{14}$$

It is easy to see that, in steady state, the cross-sectional wage distribution among employed workers is

$$G(w) = \frac{\sigma F(w)}{\sigma + \lambda_w [1 - F(w)]}. \tag{15}$$

Using this relation between $G(w)$ and $F(w)$ in the reservation wage equation (14), and exploiting the fact that the average wage is

$$\bar{w} = w^* + \int_{w^*}^{w^{\max}} [1 - G(z)] \, dz, \tag{16}$$

we arrive at the new expression for the Mm ratio,

$$Mm \approx \frac{\frac{\lambda_u - \lambda_w}{r + \sigma + \lambda_w} + 1}{\frac{\lambda_u - \lambda_w}{r + \sigma + \lambda_w} + \rho}, \tag{17}$$

in the model with on-the-job search.[24]

Note that, if the search technology is the same in both employment states and $\lambda_w = \lambda_u$, the reservation wage will be equal to b, since searching when unemployed gives no advantage in terms of arrival rate of new job offers. Indeed, for $\lambda_w > \lambda_u$, unemployed workers optimally accept jobs below the flow value of non-market time b in order to access the better search technology available during employment.

The crucial new parameter of this model is the arrival rate of offers on the job λ_w. To pin down λ_w, note that average job tenure in the model is given by

$$\tau = \int_{w^*}^{w^{\max}} \frac{dG(w)}{\sigma + \lambda_w [1 - F(w)]} = \frac{\sigma + \lambda_w / 2}{\sigma (\sigma + \lambda_w)} \in \left(\frac{1}{2\sigma}, \frac{1}{\sigma} \right). \tag{18}$$

Since we set the monthly separation rate σ to 0.02, the model can only generate average tenures between 25 and 50 months. Based on the CPS Tenure Supplement, the BLS

[24] The "approximately equal" sign originates from one step of the derivation where we have set

$$\frac{r + \sigma}{r + \sigma + \lambda_w [1 - F(z)]} \approx \frac{\sigma}{\sigma + \lambda_w [1 - F(z)]},$$

a valid approximation since, for plausible calibrations, r is negligible compared to σ.

(2006, Table 1) reports that *median* job tenure (with current employer) for workers 16 years old and over, from 1983-2004, was 3.64 years, or 43.7 months.[25]

An alternative way to restrict the choice of λ_w is to compute the average separation rate χ implied by the model, which is given by

$$\chi = \sigma + \lambda_w \int_{w^*} [1 - F(w)]\, dG(w) = \frac{\sigma(\lambda_w + \sigma)\log\left(\frac{\sigma+\lambda_w}{\sigma}\right)}{\lambda_w}, \tag{19}$$

where the first equality states that the separation rate equals the EU flow rate (σ) plus the EE flow rate (the integral). Recent studies (Fallick and Fleischman, 2001; Nagypal, 2005) estimate the data counterpart of χ to be around 4%.[26]

In what follows, we ask whether the on-the-job search model can generate the observed Mm ratio while, at the same time, being consistent with tenure lengths higher than 43.7 months or, alternatively, a monthly average separation rate of 4%. Since the Mm ratio is increasing in λ_w, it should be clear that the model will have a good chance to generate a large Mm ratio for low job tenures/high separation rates (which imply high values of λ_w).

Figure 6 illustrates that, once λ_w is chosen so that the model can generate transitions as in the U.S. data, the model will produce high Mm ratios only for negative values of ρ. Although the model is still far from a success, it is clear that the introduction of on-the-job search represents a significant quantitative improvement. Compare Figure 3 with Figure 6. In the former graph, one needs $\rho = -5$ to reproduce $Mm = 1.70$, whereas in the latter setting $\rho = -0.5$ allows the model to roughly match both the separation and the wage dispersion facts.[27] However, the independent evidence on average job tenure seems to put more strain on the model: wage dispersion and average tenure are jointly replicated only for values of ρ around -4.

This result contains an important lesson. One may be tempted to think that, given the relationship between $F(w)$ and $G(w)$ in equation (15), it suffices to assume enough dispersion in firms' productivities, and hence in the wage offer distribution $F(w)$, to

[25]For two reasons, this number is a lower bound as the empirical counterpart of τ. First, the implied average job tenure would be longer, since tenure distributions are notoriously skewed to the right. Second, the BLS data refer to the truncated tenure distribution. A higher estimate for average tenure strengthens our argument.

[26]There is no inconsistency between average tenure exceeding 44 months and an average separation rate of 4%, since the employment hazard is not constant.

[27]For $\rho = 0.4$, the model calibrated to match the aggregate separation rate gives rise to $Mm = 1.10$.

generate large wage differentials among employed workers. Equation (14), instead, tells us that a high variation in $F(w)$ is consistent with a low reservation wage, and thus a high Mm ratio, only for negative values of b, or high values of r.

Finally, there exist versions of on-the-job search models which have much weaker implications for workers' flows, because it is possible that many outside offers to workers do not result in job-to-job flows, but rather in a matching of the outside offer by the current employer.[28] While this is common in the economics profession, however, it is arguably (as emphasized and discussed in some detail in Mortensen, 2005) more rare in the broader labor market.

Arguably, models with on-the-job search today represent the most vibrant research area within search theory. Next, we analyze two recent theoretical developments of this class of models and evaluate their potential to generate frictional wage dispersion.

7.1 Endogenous search effort

Recently, Christensen et al. (2005, CLMNW therafter) have generalized Burdett's on-the-job search model by introducing an endogenously chosen level of search intensity that determines the contact rate λ.[29] The model assumes full symmetry between employment states in the sense that the search effort cost $c(\lambda)$—a convex function of the offer arrival rate chosen by the worker—is the same for unemployed and employed workers. Given that this cost function is independent of the income earned, whereas the return to search is clearly declining in such income, the optimal policy $\lambda(w)$ is decreasing in the wage w for employed workers.

In this model, the reservation wage w^* equals the flow value of leisure b, and thus $Mm = 1/\rho$. The model, therefore, has the same potential to generate sizeable frictional wage dispersion as the standard on-the-job search model with $\lambda_u = \lambda_w$. The advantage is that, while this latter parameterization induces implausibly high workers' flows, CLMNW show that their model, estimated on employer-employee matched Danish data, can replicate the empirical separation rate as a function of wage, $\chi(w) = \sigma + \lambda(w)[1 - F(w)]$, as well.

[28] For a model of this sort, see, e.g., Postel-Vinay and Robin (2002).

[29] See also the recent work by Lise (2006).

This extension of the basic job ladder model appears, at first sight, a quick and reasonable fix to the fundamental shortcoming we discuss throughout the paper. However, a more careful look reveals that this is not the case, for two reasons.

First, under this symmetric specification for the disutility of search, the model has the implication that unemployed workers and workers employed at the reservation wage make the same optimal search effort choice, i.e., the hazard rate from unemployment must equal the job-to-job flow rate at w^*. CLMNW (Table 2) estimate the latter to be 0.07 at the monthly frequency, with an extremely tight standard error due to their large sample size. They do not, however, use data on worker flows out of unemployment and do not test this implication. Rosholm and Svarer (2004) document that, for Denmark, the monthly exit rate from unemployment is 0.11, so one would strongly reject the hypothesis that the job finding rate equals the job-to-job transition rate at the lowest wage—a key prediction of this model.[30] In particular, for the model to be consistent with this aspect of the data, the symmetry assumption must be abandoned in favor of a specification where unemployed workers (i) face extra costs of being unemployed (so they have the incentive to exit unemployment more quickly), which is exactly the type of problem we encountered in all the models studied so far, or (ii) they search more cheaply than employed workers.[31] In the latter case, though, the reservation wage becomes higher than b, again making it harder for the model to generate wage dispersion.

The second reason why the CLMNW model does not provide an entirely satisfactory solution is related to one of the key observations of our paper: to generate sizeable wage dispersion, the search model needs a large and implausible disutility of unemployment. Recall that in the CLMNW model, the disutility flow from being jobless is $b - c(\lambda^*)$. Therefore, one should investigate how large the search cost component $c(\lambda^*)$ is once the model is calibrated.

Through a number of manipulations of the FONC for search effort, it is possible to derive a bound on the marginal search cost as a discounted value of the expected gain

[30] This rejection is for the Danish data. It remains to be studied whether U.S. labor market data, on which we based all our empirical analysis, would be more favorable to the model. At the moment, data constraints prevent one from performing such exercise.

[31] For example, one could assume that searching for a job takes time out of work or leisure; then the effective search cost is lower for the unemployed, and is increasing in the wage for employed workers.

from search:

$$c'(\lambda^*) \geq \frac{\bar{w} - w^*}{r + \sigma + \lambda^*}.$$

Assuming an iso-elastic specification for the cost function, as done in CLMNW, i.e., $c(\lambda) = \lambda^\gamma$, and using the fact that in the model $Mm = 1/\rho$ we have that

$$\frac{c(\lambda^*)}{\bar{w}} \geq \frac{1}{\gamma} \frac{\lambda^* (1 - \rho)}{r + \sigma + \lambda^*},$$

which establishes a lower bound for the search cost of the unemployed in terms of the average wage. Using the baseline parameterization of the paper ($\lambda^* = 0.39, r = 0.004, \sigma = 0.02, \rho = 0.4$) together with $\gamma = 2$ based on the elasticity estimated by CLMNW, one obtains that the model requires a search cost for the unemployed that is at least 28% of the average wage paid in the economy. This magnitude, once again, seems very large: it implies a net utility flow from unemployment $b - c(\lambda^*)$, of *at most* 12% of the average wage.

7.2 Reallocation shocks

Micro data indicate that a non-negligible fraction of job-to-job movers receive a wage cut. In line with this observation, some recent contributions (e.g., Jolivet et al., 2006) advocate that on-the-job search models, to be successful, must introduce "reallocation shocks", i.e., a situation where employed workers receive a job offer with an associated wage drawn from $F(w)$ that cannot be rejected. In other words, under the scenario of a reallocation shock the outside option of a worker is not keeping the current job (and the current wage), but becoming unemployed which is always dominated by accepting any new job offered. This category of employment to employment (EE) transitions may include, for example, search activity during a notice period, or a geographical move for non-pecuniary motives.

If we let ϕ be the arrival rate of such an event, it is easy to see that the mean-min ratio in the model becomes

$$Mm \approx \frac{\frac{\lambda_u - \lambda_w - \phi}{r + \sigma + \lambda_w + \phi} + 1}{\frac{\lambda_u - \lambda_w - \phi}{r + \sigma + \lambda_w + \phi} + \rho}, \tag{20}$$

a magnitude that is increasing in ϕ.[32] Ceteris paribus, the reallocation shock shortens the life of a job and makes unemployed workers less picky, and thus they reduce their

[32] The approximately equal sign holds here for the same reason as in equation (17).

reservation wage, which helps raising frictional wage dispersion. Moreover, it is straightforward to derive that the average separation rate in this model is exactly as in (19) with σ replaced by $\sigma + \phi$.

Even though this extension goes in the right direction, quantitatively it falls short of succeeding because, as in the standard on-the-job search model, the model with forced job-to-job mobility cannot reconcile the observed wage dispersion with labor turnover data. To replicate $Mm = 1.70$ with positive values of ρ, the separation rate would have to be implausibly high.

A variant of the reallocation shock story which shows more promise is suggested by Nagypal (2006). In her model, employed workers move within the wage distribution in two ways: through job-to-job movements following on-the-job search with contact rate λ_w, and through shocks occurring at rate ϕ that change the wage during an employment relationship without inducing a separation: at rate ϕ, workers make another wage draw from $F(w)$ that is uncorrelated to their current wage and that has to be accepted, or else separation takes place. Effectively, this model is a combination of the canonical on-the-job search model and the model with stochastic wages we presented in Section 6.

The mean-min ratio in this model has exactly the same form as in (20), but now raising ϕ does translate into a wage change without increasing the separation rate. To gauge the quantitative implications of this version of the on-the-job search model, we perform the following exercise: keeping the values for (σ, ρ) unchanged relative to our previous analysis, we set the pair (λ_w, ϕ) to match the aggregate separation rate (4%) and $Mm = 1.70$. Then, we calculate the average arrival rate κ of a wage cut among employed workers in the model. It is given by the formula

$$\kappa = \phi \int F(w)\, dG(w) = \phi \left[1 + \frac{\sigma + \phi}{\lambda_w} - \frac{(\sigma + \phi)(\lambda_w + \sigma + \phi) \log\left(\frac{\sigma+\phi+\lambda_w}{\sigma+\phi}\right)}{\lambda_w^2} \right].$$

The model implies that 9.5% of the workforce is subject to a wage cut every month, i.e., 70% of the stock of employees experience at least a salary cut from month to month, in any given year.[33] This extreme implication is similar to our finding of Section 6, where

[33]The calibration yields a value for the Poisson arrival rate κ of 10% which implies a monthly probability of receiving a wage cut of $1 - \exp(-\kappa)$=9.5%.

we concluded that the model with wage shocks (but without on-the-job search) generates individual wage profiles that are too volatile.

8 The literature on structural estimation of search models

Since the pioneering effort of Flinn and Heckman (1983), a rather vast literature on structural estimation of search models has developed (see Eckstein and van den Berg, 2005, for a recent survey). We view these contributions as having generated many valuable insights. From our perspective here, it is important to comment on how the literature has dealt with the baseline model's apparent inability to generate frictional wage dispersion. We summarize our reading of the literature as follows: it has either (1) simply "accepted" implausible parameter estimates for the value of non-market time and the interest rate, or (2) introduced unobserved skill heterogeneity and measurement error that soak up the large wage residuals in the data. We now proceed to discussing a number of selected papers in more detail.

In one of the first attempts at a full structural estimation, Eckstein and Wolpin (1990) estimate the Albrecht and Axell (1984) search model with worker heterogeneity in the value of non-market productivity and conclude that their model cannot generate any significant wage dispersion, and that almost all of the observed wage dispersion is explained through measurement error. Eckstein and Wolpin (1995) reach a far better match between model and data, by introducing a five-point distribution of unobserved worker heterogeneity within each race/education group (8 groups in total). In spite of such heterogeneity, however, for many of the groups the estimates of b remain extremely small or negative (see their Table 7, page 284). In this work, thus, wage dispersion is for the most part accounted for by heterogeneity in observable and unobservable characteristics. In our view, this procedure, which is quite frequent in this literature, can perhaps be categorized more as part of the human-capital theory of wages: it does not deliver large frictional wage dispersion.[34]

[34] There is also a theoretical argument against models of frictional wage dispersion based on heterogeneity. Gaumont et al. (2006) demonstrate that wage dispersion in an Albrecht and Axell (1984) model with worker heterogeneity in the value of leisure is fragile. As soon as an arbitrarily small search cost is introduced, the equilibrium unravels and we are back to the "Diamond paradox", i.e., a unique wage

Negative estimates of the net value of non market time are quite common. The survey paper by Bunzel et al. (2001) estimates several models with on the job search on Danish data. When firms are assumed to be homogeneous, the point estimate for ρ is -2. With heterogeneity in firms' productivity it increases just about to $\rho = 0$. Only the model with measurement error produces a large and positive estimate of ρ.[35] Flinn (2006) estimates a Pissarides-style matching model of the labor market, without on-the-job search, to evaluate the impact of the minimum wage on employment and welfare. In his model, as is typical in estimation exercises, the pair (ρ, r) is not separately identified. Setting r to 5% annually in his model implies roughly $\rho = -4$.[36]

Another example of extreme parameter estimates that can be found in Postel-Vinay and Robin (2002). Under risk neutrality, their estimates of the discount rate r always exceed 30% per year in every occupational group, reaching 55% for unskilled workers, where they find no role for unobserved heterogeneity. Risk aversion (in the form of log utility, without storage) does not help: the estimates of r decline by just 3 percentage points on average. Recall, from Figure 3, that a negative value for ρ and a high value for r are two sides of the same coin.

Whenever authors have restricted (r, ρ) to plausible values ex-ante, not surprisingly, they end up with finding that frictions play a minor role. For example, Van den Berg and Ridder (1998) estimate the Burdett-Mortensen model on Dutch data allowing for measurement error and observed workers' heterogeneity (58 groups defined by education, age and occupation). They set r to zero and b to equal the average unemployment benefit for each group, i.e., roughly 60% of the average wage. They conclude that observed heterogeneity and measurement error account for over 80% of the empirical wage variation. Moscarini (2003) develops an equilibrium search model where workers learn about their match values, based on Jovanovic (1979). When the model is calibrated, r is set to 5% annually and ρ to 0.6. His model generates a Mm ratio of just 1.16 (Moscarini, 2005, Table 2).

One may note that a number of papers in the literature do claim that the (on-the-

arising in equilibrium.

[35]These values for ρ are obtained from Bunzel et al. (2001) by dividing the estimates of b for the entire sample, in Tables II and V, by the average wage from Table I.

[36]Calculations are available upon request.

job search) model is successful in simultaneously matching the wage distribution and labor market transition data (see, e.g., Bontemps et al., 2000, and Jolivet et al., 2005). These claims of success, clearly, need to be properly reinterpreted in light of our findings. The typical strategy in these papers is, first, to estimate the wage distribution $G(w)$ non-parametrically without using the search model. Next, the model is used to predict the wage-offer distribution $F(w)$ through a steady-state relationship like (15), where the structural parameters of the relationship $(\sigma, \lambda_u, \lambda_w)$ are estimated by matching transition data. Success is then expressed as a good fit (in some specific metric). However, the exercise is not a full success because it neglects the implications of the joint estimates of $F(w)$ and of the transition parameters for the relative value of leisure ρ (or, similarly, for the interest rate r). The key additional "test" that we are advocating would thus entail using the estimated $F(w)$ in the reservation-wage equation (14) and, given an estimate of w^* (for example, the bottom-percentile wage observed), backing out the implied value for ρ. In light of our results, we maintain that ρ would be negative, or in any case unreasonably low.

In conclusion, while we are definitely of the view that there is an important amount of progress in this empirical literature, the success is really only partial: the literature has not yet managed to match the data with plausible parameter values. In short, important parameters such as b and r (the value of leisure and the discount rate, respectively) are considered free parameters, and estimates that are far from what we view as reasonable are thus "accepted". Alternatively, unobserved heterogeneity or measurement error must be introduced, also with amounts that are free parameters, in order to match the data. Our contribution in this context is thus to point to a specific prediction—which is quite sharp in most search models and which has not been noted before—that arguably is somewhat of an Achilles heel for this class of models: the mean-min wage ratio is too low. Marked improvements in model performance may thus benefit from further analysis of the determinants of this measure of wage dispersion.

9 Conclusions

Search theory maintains that similar workers looking for jobs in the same labor market may end up earning different wages according to their luck in the matching process. This paper has proposed a simple strategy for evaluating the quantitative ability of search models to generate frictional wage dispersion. The strategy is based on a particular measure of wage differentials, the mean-min ratio, that arises very naturally, in closed form, from the reservation wage equation, the cornerstone of a vast class of search models.

We have demonstrated that, when plausibly calibrated, the textbook search and matching model implies that frictions play virtually no role in the labor market: the mean-min ratio is less than 4%. We have made several attempts to save the model. The most promising extension is the one with on-the-job search. However, within the simplest version of the job ladder model, it is hard to produce large wage dispersion while, at the same time, being consistent with the observed labor market transition data. Other attempts seem far less promising. Risk aversion can be successful only if one believes that self insurance is unimportant, but a decade of quantitative investigations of Bewley-type models speaks against this possibility. Volatility in wages during employment can be successful only if one believes that wages are as volatile during an employment spell as in the cross-section.

The data we have analyzed tell a striking story: residual wage dispersion among observationally similar workers in narrowly defined labor markets (by occupation and geographical area) is twenty times larger than what predicted by the model. This leaves two possible, but radically different, conclusions on the table. First, residual wage inequality in the data is attributable to unobserved (and probably time-varying) skills that are remunerated in a near frictionless labor market; put differently, the role of frictions is actually small in the data as well. This conclusion is viewed as plausible by many, but it does call for more careful testing. In particular, one must look for more individual-specific information (both time-invariant, like test scores or attitudes towards leisure and work, and time-varying, like significant events altering the value of leisure, e.g., need for child or health care in individuals' lives) in order to reduce the current residual dispersion. One may also be able to look more deeply at implications for on-the-job wage variability, which should be quite large if indeed individual-specific and time-varying skill heterogeneity accounts for the wage dispersion. Alternatively, the second possible conclusion which puts more faith in the measures of frictional wage dispersion reported here, is that our basic search theory needs further development. Distinguishing between these two conclusions should be a central task in macroeconomics and labor economics: the relative roles of skills and luck in the labor market lie at the heart of policy design.

Finally, an implicit but interesting implication of the present work is closely connected to the recent debate on whether the matching model is able to generate enough time-series fluctuations in aggregate unemployment and vacancies; see Shimer (2005b), Hagedorn and Manovskii (2006), and Hall (2005). There, it is pointed out that the matching model, at least if one is to avoid the incorporation of significant real-wage rigidity, requires a very high benefit of non-market time (denoted ρ above, expressing this benefit as a *fraction* of the average wage) in order to produce sharp movements in vacancy and unemployment rates. But, as we showed here, the higher is the value of ρ, the more difficult it is to explain wages cross-sectionally. More specifically, the time-series facts necessitate a value of ρ close to 1 to explain the data, and our cross-sectional facts demand a value significantly below 0. We believe that it is important in future work to keep both "puzzles" in mind while developing and using search-matching theories of the labor market.

10 Appendix

A: THE OCCUPATIONAL EMPLOYMENT STATISTICS PROGRAM

The Occupational Employment Statistics (OES) program collects data on employees in approximately 200,000 non-farm establishments to produce employment and wage estimates for 821 occupations classified based on the Standard Occupational Classification (SOC).

Since November 2003, the program samples 200,000 establishments semi-annually. Before then, it sampled 400,000 once a year. The OES survey is designed to produce occupational wage and employment estimates using six panels (3 years) of data. The BLS Employment Cost Index is used to adjust survey data from prior panels before combining them with the current panel's data. The full six-panel sample of 1.2 million establishments allows the construction of occupational estimates at detailed levels of geography and industry. Estimates based on geographic areas are available at the National, State, and Metropolitan Area levels. Industry classifications correspond to 3, 4, and 5-digit North American Industry Classification System (NAICS) industry groups.

The OES survey form sent to establishments defines wages as straight-time, gross pay, exclusive of premium pay. Base rate, cost-of-living allowances, guaranteed pay, hazardous-duty pay, incentive pay including commissions and production bonuses, tips, and on-call pay are included. Excluded are back pay, jury duty pay, overtime pay, severance pay, shift differentials, non-production bonuses, employer cost for supplementary benefits, and tuition reimbursements.

The OES survey groups wages in 12 discrete intervals. In November 2004, the lowest interval was "Under \$6.75" and the highest was "\$70 and over". Mean hourly wage rate for an occupation equals total wages that all workers in the occupation earn in an hour divided by the total employment of the occupation. The same concept applies to more disaggregated levels such as occupations within metropolitan areas or industries. The mean and percentiles for an occupation are calculated by uniformly distributing the workers inside each wage interval, ranking the workers from lowest paid to highest paid.

B: THE PANEL STUDY OF INCOME DYNAMICS

The Panel Study of Income Dynamics (PSID) began collecting information on a sample of approximately 5,000 households in 1968. Of these, about 3,000 are representative of the U.S. population as a whole (core sample), while the rest are low-income families (SEO sample). Since then, both the original households and their split-offs (members of the original household forming a family of their own) have been followed over time. Questions on labor income and hours are retrospective: information collected in the year t wave refers to the calendar year $t-1$.

Our initial sample comprises every head and spouse between 20-60 years old in the 1968-1997 waves of the PSID core sample. We then exclude individuals currently in school, self-employed, or disabled, and those with annual hours below 520 and above 5096 to reduce the role for measurement error in hours, which leaves 80,979 individual/year records in the sample.[37] Next, we exclude all individuals whose earnings are top-coded or whose hourly wage (computed as annual earnings divided by annual hours) is below the federal minimum wage, which eliminates

[37]French (2005) uses the PSID Validation Study to assess the size of measurement error in hourly wages. He concludes that it accounts for 24% of the standard deviation of log wages. By trimming the hours distribution below 520 and above 5096, we eliminate many outliers that are due to reporting errors.

around 4,141 individual/years observations. At the end of this selection, we are left with 76,848 individual/year observations in our final sample, i.e., roughly 2,500 per year.

The estimation of the individual fixed effect described in the main text is based on workers with at least ten reported wage observations which reduces the sample size to 49,010 individual/year observations, i.e., 1,633 per year on average.

C: THE 5% IPUMS SAMPLE OF THE 1990 CENSUS

The 1990 Census of the U.S. population uses a single long-form questionnaire for sample questions completed by one half of persons in locations with a population under 2,500, one sixth of persons in other tracts and block numbering areas with fewer than 2,000 housing units, and one eighth of persons in all other areas. Overall, about one sixth of all housing units complete a long form. Within each state, the Bureau divides the sample questionnaires into an appropriate number of 1-percent samples. For example, if 20 percent of the population of a state completed long forms, the sample questionnaires for that state are divided into twenty subsamples of equal size. The 5-percent files are then selected at random from the 1-percent subsamples for each state. Weights are attached to each case representing the number of individuals in the general population represented by any particular case in the sample.

The original data set contains over 12,500,000 person-level observations. To create our sample, first we exclude every person below 20 and above 60 years old, as well as every individual currently in school, self-employed, or disabled, which leaves 4,636,759 individual records in the sample. Next, we exclude all individuals who report zero wage income or zero weeks worked over the year, and individuals whose annual earnings are top-coded, i.e. higher than $140,000 (19,890 cases). We also remove altogether occupational groups where more than 3% of all workers are top coded. Eleven occupations are excluded based on this criterion, e.g., Airplane pilots, Athletes, Dentists, Physicians, Podiatrists, and Judges. Finally, we eliminate individuals for whom estimated hourly wages are below the federal minimum wage ($3.35 in 1989), i.e., roughly 400,000 observations.[38] We are then left with 3,923,744 individual records in our final sample.

We construct hourly wage as annual wage and salary income (variable INCWAGE) divided by the product between the number of weeks worked last year (WKSWRK1) and usual weekly hours worked (UHRSWORK).[39]

[38] Through validation with CPS data, Baum-Snow and Neal (2006) find that a significant fraction of workers report usually working 8 hours per week on the census' long form when they actually usually worked 40 hours per week; thus, they respond as if the question meant to report "usual hours per day". However, this type of measurement error which plagues the 1980 Census is much less frequent in the 1990 census. Moreover, these respondents are excluded from our selection criteria on annual hours.

[39] See Ruggles et al. (2000) for a detailed explanation of the IPUMS Census data.

References

[1] Aiyagari, R. (1994). "Uninsured Idiosyncratic Risk and Aggregate Saving," Quarterly Journal of Economics, vol. 109(3), pages 659-684.

[2] Albrecht J., and B. Axell (1984). "An Equilibrium Model of Search Unemployment," Journal of Political Economy, vol. 92, pages 824-840.

[3] Altonji, J., and R. Shakotko (1987). "Do Wages Rise With Job Seniority?" Review of Economic Studies, vol. 65(3), pages 437-460.

[4] Altonji, J., and N. Williams (2004). "Do Wages Rise with Job Seniority? A Reassessment," forthcoming in the Industrial and Labor Relations Review.

[5] Baum-Snow, N., and D. Neal (2006). "Data Problems and the Measurement of Racial and Gender Wage Gaps Over Time," mimeo, University of Chicago.

[6] Blank, R., and D. Card (1991). "Recent Trends in Insured and Uninsured Unemployment: Is There an Explanation?" Quarterly Journal of Economics, vol. 106(4), pages 1157-1189.

[7] Bontemps, C., J.-M. Robin, and G. J. Van den Berg (2000). "Equilibrium Search with Continuous Productivity Dispersion Theory and Non Parametric Estimation," International Economic Review, vol. 41(2), pages 305-358.

[8] Bound, J., and A. Krueger (1991). "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" Journal of Labor Economics, vol. 9(1), pages 1-24.

[9] Burdett, K. (1978). "A Theory of Employee Job Search and Quit Rates," American Economic Review, vol. 68, pages 212-220.

[10] Burdett, K., and D. T. Mortensen (1998). "Wage Differentials, Employer Size, and Unemployment," International Economic Review, vol. 39(2), pages 257-273.

[11] Bureau of Labor Statistics (2006). "Employee Tenure Summary," http://www.bls.gov.

[12] Bunzel, H., B. J. Christensen, P. Jensen, N. M. Kiefer, L. Korsholm, L. Muus, G. R. Neumann, and M. Rosholm (2001). "Specification and Estimation of Equilibrium Search Models," Review of Economic Dynamics, vol. 4, pages 90-126.

[13] Christensen, B. J., R. Lentz, D. T. Mortensen, G. R. Neumann, and A. Wervatz (2005). "On the Job Search and the Wage Distribution," Journal of Labor Economics, vol. 23(1), pages 31-58.

[14] Dey, M. and C. Flinn (2006). "Household Search and Health Insurance Coverage," mimeo, NYU.

[15] Eckstein, Z., and G. Van den Berg (2005). "Empirical Labor Search: A Survey," IZA Discussion paper.

[16] Eckstein, Z., and K. Wolpin (1990). "Estimating a Market Equilibrium Search Model from Panel Data on Individuals," Econometrica, vol. 4, pages 783-808.

[17] Eckstein, Z., and K. Wolpin (1995). "Duration to First Job and the Return to Schooling: Estimates from a Search-Matching Model," Review of Economic Studies, vol. 62, pages 263-286.

[18] Fallick, B. and C. A. Fleischman (2001). "The importance of Employer-to-Employer Flows in the U.S. Labor Market," Finance and Economics Discussion Series 2001-18, Board of Governors of the Federal Reserve System.

[19] Flinn, C. (2006). "Minimum Wage Effects on Labor Market Outcomes Under Search, Bargaining, and Endogenous Contact Rates," Econometrica, vol. 74, pages 1013-1062.

[20] Flinn, C. and J. Heckman (1983). "Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States?" Journal of Labor Economics, vol. 1(1), pages 28-42.

[21] French, E. (2005). "The Labor Supply Response to Mismeasured (but Predictable) Wage Changes," forthcoming in the Review of Economics and Statistics.

[22] Gaumont, D., M. Schindler, and R. Wright (2006). "Alternative Theories of Wage Dispersion," European Economic Review, vol. 50(4), pages 831-848.

[23] Hagedorn, M., and I. Manovskii (2006). "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited," mimeo, University of Pennsylvania.

[24] Hall, R. E. (2005). "Employment Fluctuations with Equilibrium Wage Stickiness," American Economic Review, vol. 95(1), pages 50-65.

[25] Hansen, H. (1998). "Transition from Unemployment Benefits to Social Assistance in Seven European OECD Countries," Empirical Economics, vol. 23(1), pages 5-30.

[26] Hornstein, A., P. Krusell and G.L. Violante (2006). "Technical Appendix for 'Frictional Wage Dispersion in Search Models: A Quantitative Assessment'," Federal Reserve Bank of Richmond Working Paper 2006-07, available at http://www.richmondfed.org/publications/economic_research/working_papers/index.cfm.

[27] Kambourov, G. and I. Manovskii (2004). "Occupational Mobility and Wage Inequality," mimeo.

[28] Jolivet, G., F. Postel-Vinay and J.-M. Robin (2006). "The Empirical Content of the Job Search Model: Labor Mobility and Wage Dispersion in Europe and the U.S.,"European Economic Review, vol. 50(4), pages 877-907

[29] Jovanovic, B. (1979). "Job Matching and the Theory of Turnover," Journal of Political Economy, vol. 87(5), pages 972-990.

[30] Kostiuk, P. (1990). "Compensating Differentials for Shift Work," Journal of Political Economy, vol. 98, pages 1054-1075.

[31] Levine, D. and W. Zame (2002). "Does Market Incompleteness Matter?" Econometrica, vol. 70(5), pages 1805-1839.

[32] Lise, J. (2006). "On-the-Job Search and Precautionary Savings: Theory and Empirics of Earnings and Wealth Inequality," mimeo, Queens University.

[33] Ljungqvist, L. and T.J. Sargent (1998). "The European Unemployment Dilemma," Journal of Political Economy, vol. 106, pages 514-550.

[34] Lucas, R. E., and E. Prescott (1974). "Equilibrium Search and Unemployment," Journal of Economic Theory, vol. 7(2), pages 188-209.

[35] Machin, S., and A. Manning (1999). "The Causes and Consequences of Long-Term Unemployment in Europe," in Handbook of Labor Economics, (edited by Ashenfelter, O., and D. Card), North-Holland.

[36] Manning, A., and B. Petrongolo (2005). "The Part-Time Pay Penalty," CEP Discussion Paper No. 679.

[37] McCall, J. (1970). "Economics of Information and Job Search," Quarterly Journal of Economics, vol. 84(1), pages 113-126.

[38] Mood, A., F. Graybill, and D. Boes (1974). Introduction to the Theory of Statistics, McGraw-Hill

[39] Mortensen, D. T. (1970). "Job Search, the Duration of Unemployment, and the Phillips Curve," American Economic Review, vol. 60(5), pages 847-62.

[40] Mortensen, D. T. (2005). Wage Dispersion: Why Are Similar Workers Paid Differently?, MIT Press.

[41] Mortensen, D. T., and C. Pissarides (1994). "Job Creation and Job Destruction in the Theory of Unemployment," Review of Economic Studies, vol. 61(3), pages 397-415.

[42] Moscarini, G. (2003). "Skill and Luck in the Theory of Turnover," mimeo, Yale University.

[43] Nagypal, E. (2005). "Worker Reallocation Over the Business Cycle: The Importance of Job-to-Job Transitions," mimeo, Northwestern University.

[44] Nagypal, E. (2006). "On the Extent of Job to Job Transitions," mimeo, Northwestern University.

[45] OECD (2004). Taxes and Benefits in the United States, OECD.

[46] Pissarides, C. (1985). "Short-run Equilibrium Dynamics of Unemployment Vacancies, and Real Wages," American Economic Review, vol. 75(4), pages 676-90.

[47] Pissarides, C. (2000). Equilibrium Unemployment Theory, MIT Press.

[48] Postel-Vinay, F. and J.-M. Robin (2002). "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," Econometrica, vol. 70(6), pages 2295-2350.

[49] Rogerson, R., R. Shimer, and R. Wright (2006). "Search Theoretic Models of the Labor Market," Journal of Economic Literature, vol. 43 (4), pages 959-988.

[50] Rosholm, M. and M. Svarer (2005). "Endogenous Wage Dispersion in a Search-Matching Model," Labor Economics, vol. 11(5), pages 623-645.

[51] Ruggles, S., M. Sobek, A. Trent, C. A. Fitch, R. Goeken, P. K. Hall, M. King, and C. Ronnander (2004). "Integrated Public Use Microdata Series: Version 3.0," Minneapolis, MN: Minnesota Population Center. http://www.ipums.org.

[52] Ruhm, C. (2003). "Good Times Makes You Sick," Journal of Health Economics, vol. 22(4), pages 637-658.

[53] Shimer, R. (2005a). "Reassessing the Ins and Outs of Unemployment", mimeo, University of Chicago.

[54] Shimer, R. (2005b). "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," American Economic Review, vol. 95(1), pages 25-49.

[55] Topel, R. (1991). "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," Journal of Political Economy, vol. 99(1), pages 145-76.

[56] Van den Berg, G., and G. Ridder (1998). "An Empirical Equilibrium Search Model of the Labor Market," Econometrica, vol. 66(5), pages 1183-1221.

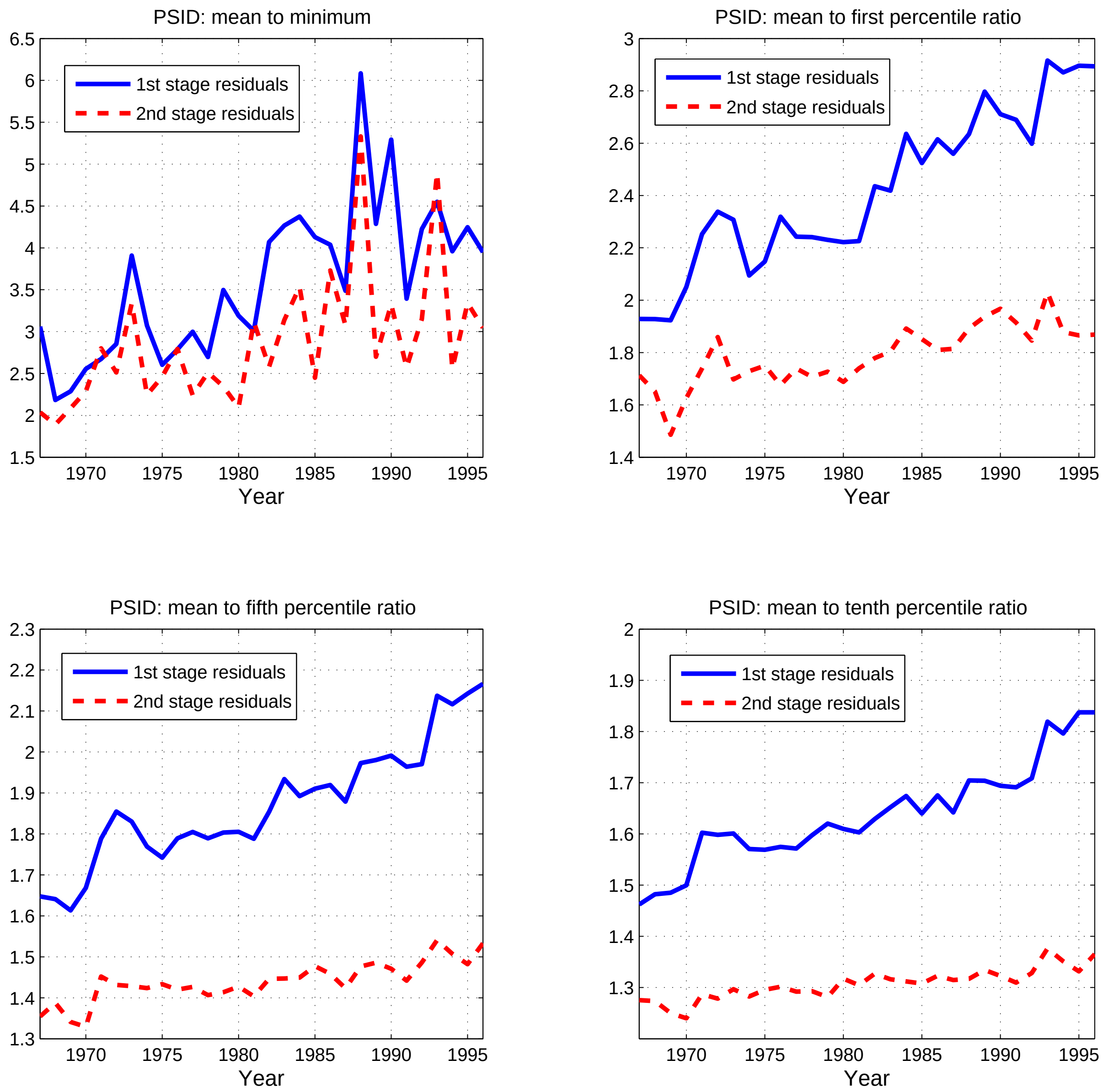
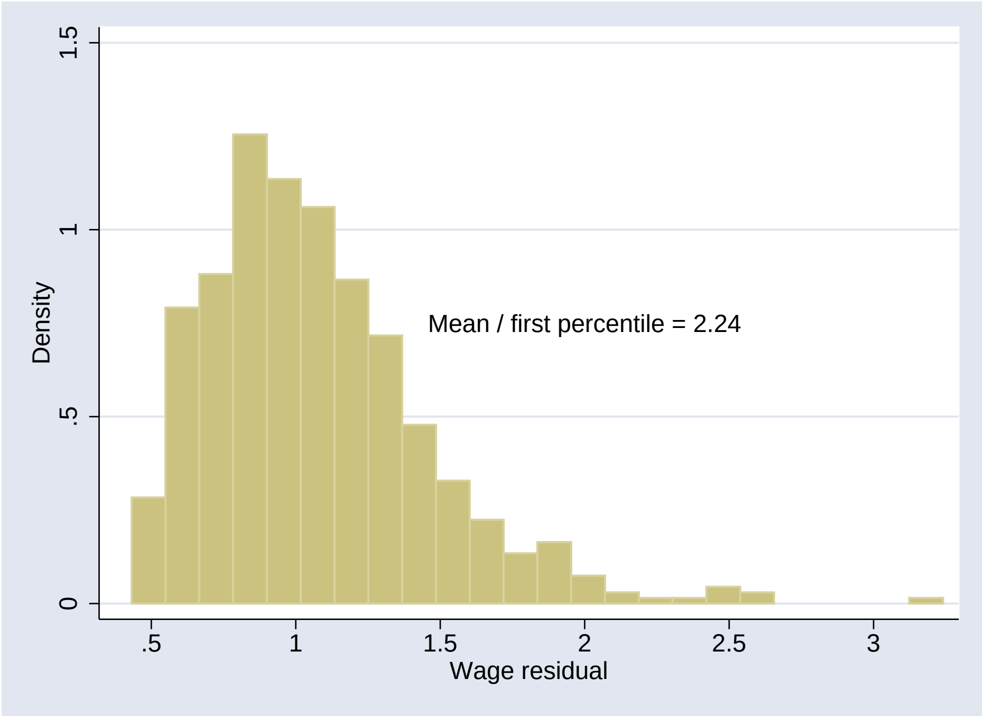


Figure 1: Empirical analysis on PSID data. The first stage residuals refer to the regression on observable covariates. The second stage residuals are the first stage residuals demeaned individually.



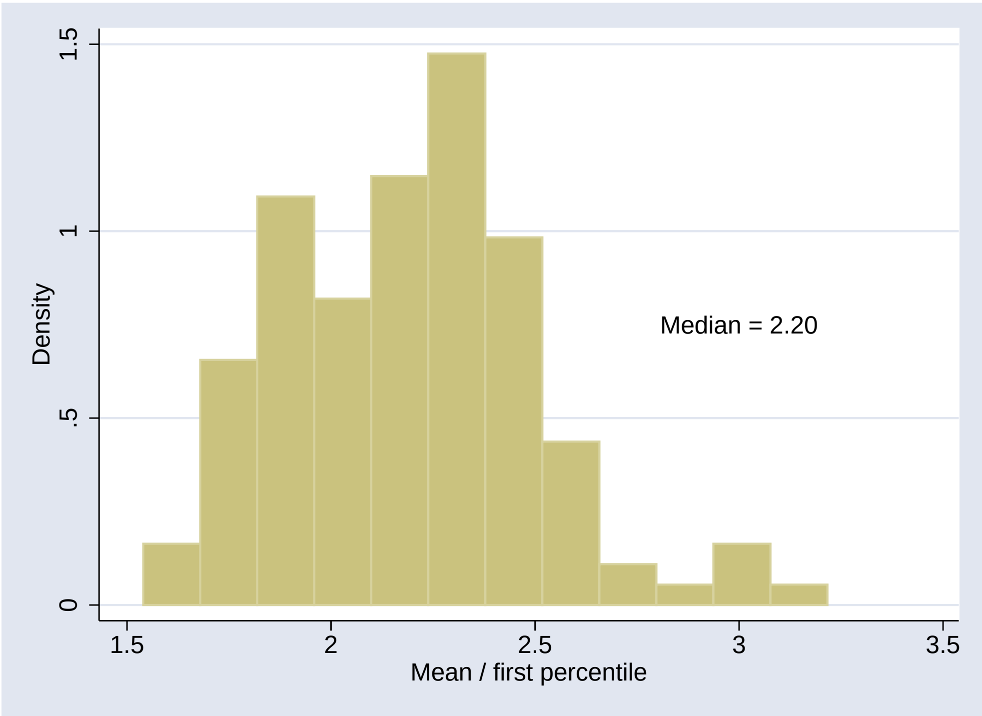


Figure 2: Top panel: Residual wage distribution for full-time, full-year janitors and cleaners in the Philadelphia area. Bottom panel: Distribution of mean-min ratios for full-time, full-year janitors and cleaners across U.S. geographical areas.

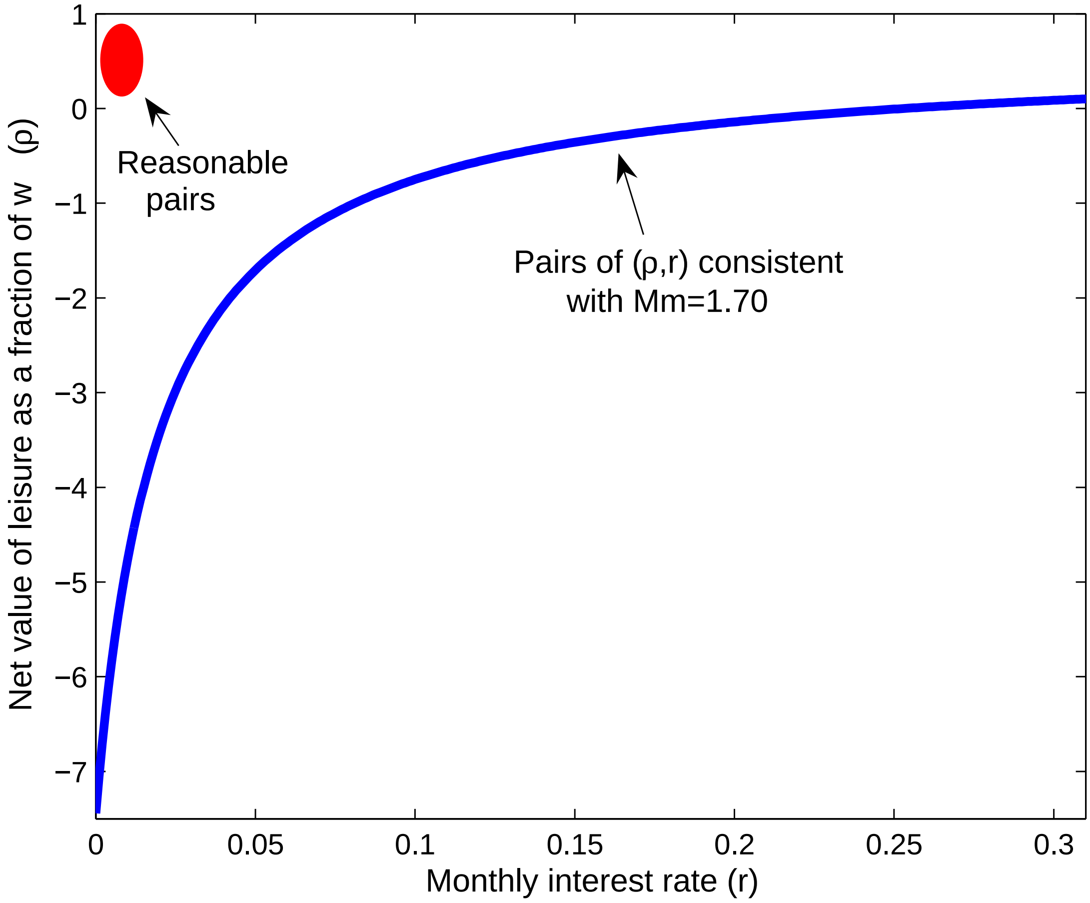


Figure 3: Pairs of the value of non-market time and the interest rate that can generate Mm=1.70

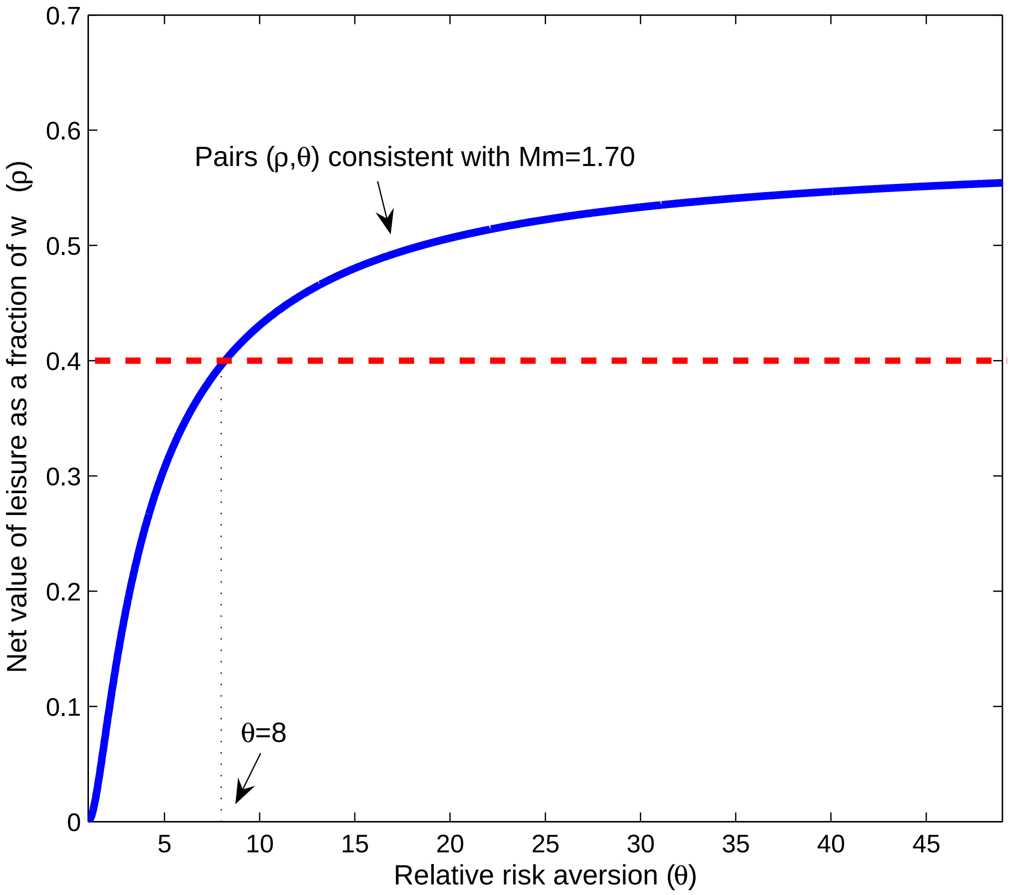


Figure 4: Pairs of the value of non-market time and risk aversion that can generate Mm=1.70

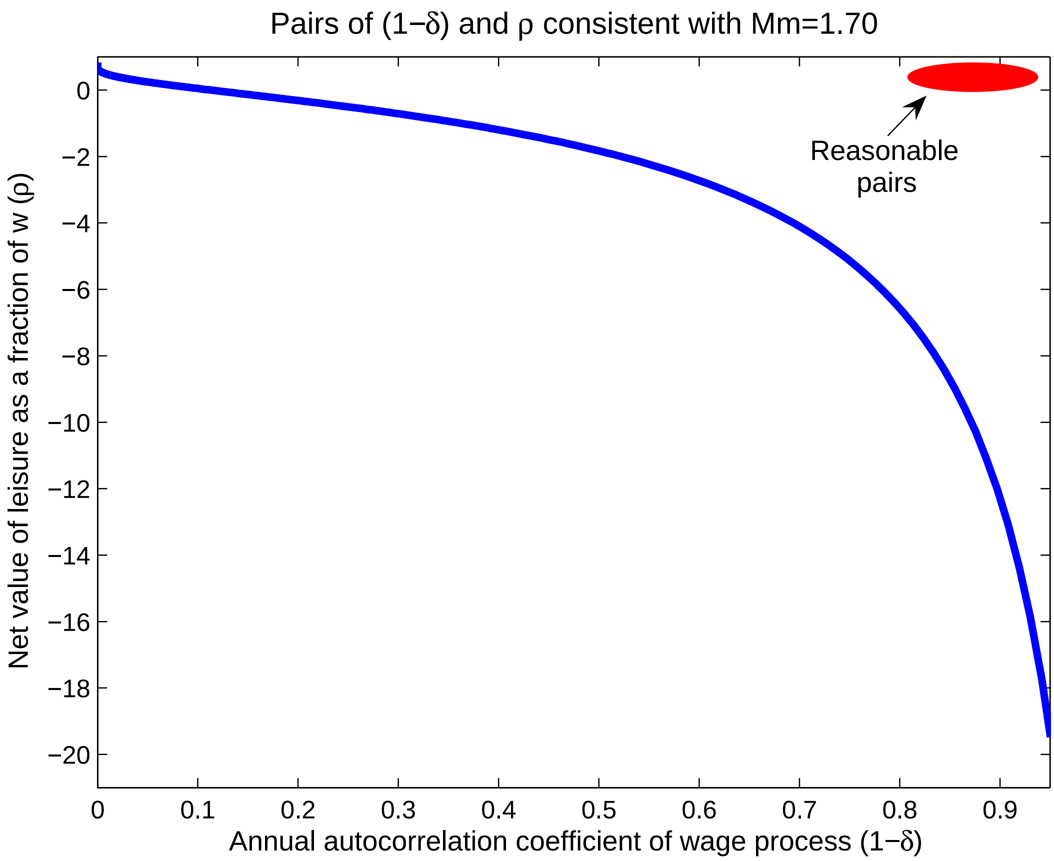


Figure 5: Pairs of the value of non-market time and the wage autoregression coefficient that can generate Mm=1.70

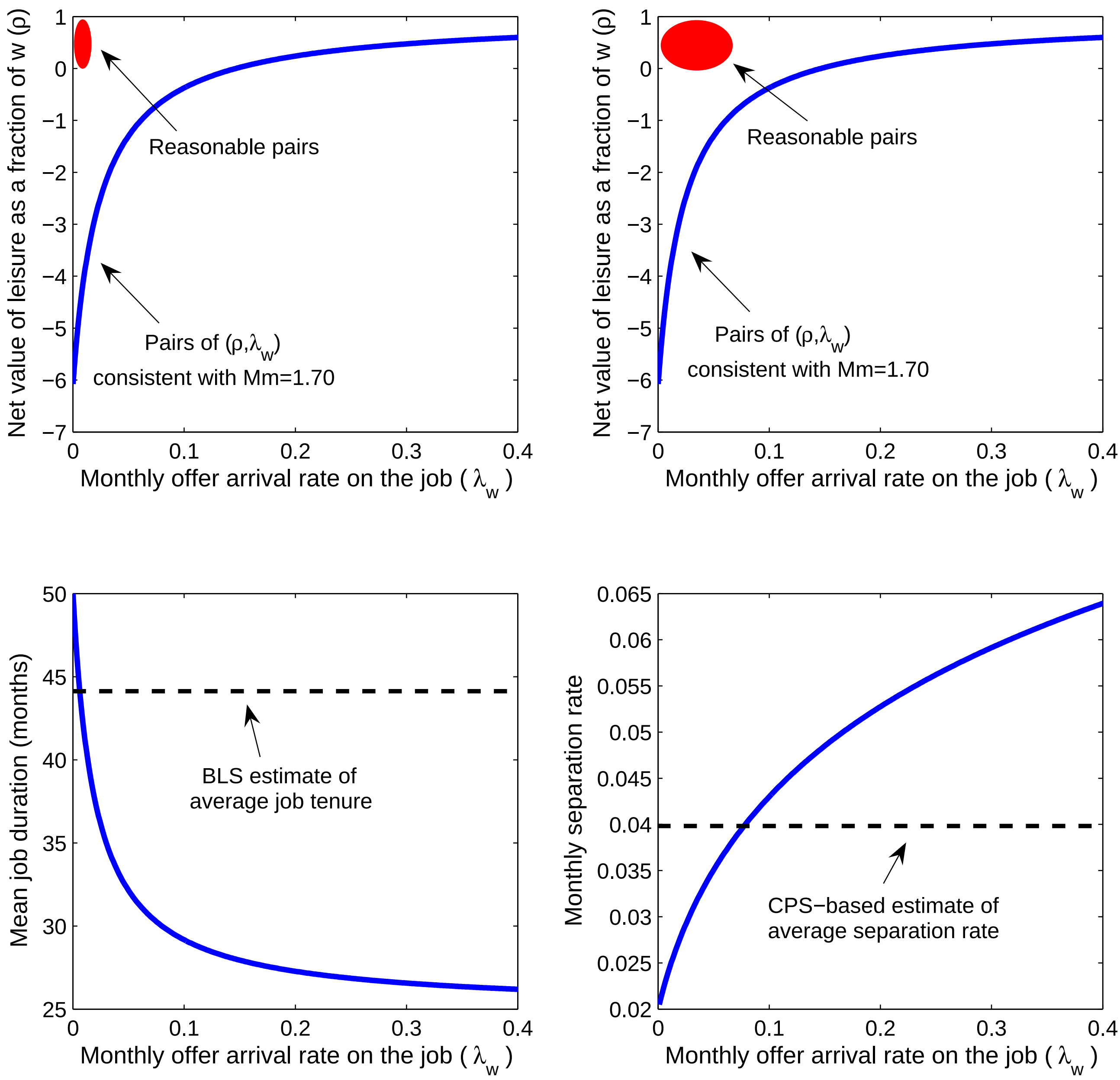


Figure 6: Top panels: pairs of the value of non-market time and the job offer rate during employment that can generate Mm=1.70. Bottom panels: mapping between the arrival rate of offers on the job and average job tenure (left panel), and separation rate (right panel).