TI 2011-010/3
Tinbergen Institute Discussion Paper

Sorting and the Output Loss due to Search Frictions

Pieter A. Gautier[1,3,4]
Coen N. Teulings[2,3,4,5]

[1] *Faculty of Economics and Business Administration, VU University Amsterdam;*
[2] *Faculty of Economics and Business, University of Amsterdam;*
[3] *Tinbergen Institute;*
[4] *CEPR, CESifo, IZA;*
[5] *Netherlands Bureau of Economic Policy Analysis CPB, The Hague.*

Sorting and the output loss due to search frictions*

Pieter A. Gautier[†]and Coen N. Teulings [‡]

April, 2010

Abstract

We analyze a general search model with on-the-job search and sorting of heterogeneous workers into heterogeneous jobs. This model yields a simple relationship between (i) the unemployment rate, (ii) the value of non-market time, and (iii) the max-mean wage differential. The latter measure of wage dispersion is more robust than measures based on the reservation wage, due to the long left tail of the wage distribution. We estimate this wage differential using data on match quality and allow for measurement error. The estimated wage dispersion for the US is consistent with an unemployment rate of 4-6%. We find that without search frictions, output would be between 7.5% and 18.5% higher, depending on whether or not firms can ex ante commit to wage payments.

1 Introduction

Relative to a competitive economy, an economy with search frictions generates less output because (i) there are idle resources like unemployed workers, (ii) resources are spent on recruitment activities, and (iii) the assignment of workers to jobs is sub-optimal. Heterogeneity is crucial when assessing the importance of search frictions. If all unemployed workers and jobs were alike, it would be hard to imagine why it would take workers

*We thank seminar participants at MIT, the 2009 Sandjberg conference on search models of the labor market, and the Sciences Po conference on sorting for useful comments and discussions.

[†]VU University Amsterdam, Tinbergen Institute, CEPR, Ces-Ifo, IZA, email: pgautier@feweb.vu.nl

[‡]Netherlands Bureau of Economic Policy Analysis, University of Amsterdam, Tinbergen Institute, CEPR,.CeS-Ifo, IZA, email: C.N.Teulings@cpb.nl

months to find a good match. The more important the quality of the match, the costlier are search frictions.

This paper analyses a class of search models with on-the-job search (OJS) and heterogeneity among workers and jobs. As a starting point, we take the framework of Gautier, Teulings, and Van Vuuren (2010) where the productivity of a match is decreasing in the distance between worker and job type and where employed workers continue moving towards the most productive jobs. We use a production technology that can be interpreted as a second order Taylor approximation of a more general production function. Within this framework, various wage mechanisms can be analyzed like wage posting with full commitment as in Burdett and Mortensen (1998) and wage mechanisms without commitment as in Coles (2001) and Shimer (2006). The key difference between wage setting with and without commitment is that in the first case, firms pay both hiring and no quit premia to hire/ keep workers whereas in the latter case, the only reason for firms to pay workers above their reservation wage is to prevent them from quitting. The equilibrium is characterized by a relationship between just three statistics: (i) the unemployment rate, (ii) the value of non-market time, and (iii) a summary statistic for wage dispersion between identical workers, the *max-mean* wage differential. We show that this statistic is more informative and more robust than alternatives like the ratio of the mean wage to the reservation wage (i.e. the *mean-min* ratio of Hornstein et.al., 2010). This relation hardly depends on any of the model's parameters, except for the relative efficiency of on- versus off-the-job search, ψ. For the calculation of the total output loss due to search frictions for a given unemployment rate, even the effect of ψ is a higher order phenomenon.

The combination of two-sided heterogeneity and search frictions relates our model to the literature on hedonic pricing in the spirit of Rosen (1974), Sattinger (1975) and Teulings (1995,2005). These models are hierarchical, in the sense that better skilled workers have a comparative advantage in more complex jobs. Hence, there is a least and a best skilled worker, and there is a least and most complex job type. In a Walrasian equilibrium, there is perfect sorting. With search frictions, this perfect correlation between worker and job types breaks down, since workers cannot afford to wait for ever till the optimal match comes along. Shimer and Smith (2000) and Teulings and Gautier (2004) are early examples of assignment models with search frictions. Hierarchical models are difficult to solve because matching sets in the corners of the type space do not have interior boundaries. We therefore first transform the hierarchical model into a circular model in the spirit of Marimon and Zilibotti (1999) and Gautier, Teulings, and Van Vuuren (2008). The idea is that profits are decreasing in the distance between worker and job types. This makes it possible to derive a closed form solution of the equilibrium. When turning to the empirical analysis of data on individual wages, and on worker and job characteristics, we reintroduce the hierarchical aspect of the model.

Ultimately, the usefulness of our model depends on how well it can simultaneously describe the observed wage dispersion due to mismatch, the unemployment rate, and the ratio of job-to-job versus unemployment-to-job worker flows. We show that the equilibrium unemployment rate in our model that is consistent with the observed amount of wage dispersion is between 4% and 6% which seems reasonable. Given that our model performs well, we can calculate the total output loss due to search frictions which is between 8% and 12% if firms can commit to wage payments, and between 14.5% and 18.5% if they cannot. The majority of the output loss is due to recruitment activities and mismatch.

Hornstein et.al. (2010) also derive a simple relationship between the unemployment rate and wage dispersion that holds for a large class of search models. They argue that (most) search models without OJS cannot explain the coexistence of low unemployment rates and substantial wage dispersion because the first suggests low frictions and the latter suggests high frictions. Gautier and Teulings (2006) made the related point that without OJS, estimates of output losses due to search frictions based on the unemployment rate are substantially lower than estimates based on wage dispersion. We show that allowing for OJS and unobserved heterogeneity can resolve this issue. OJS lowers the reservation wage which increases wage dispersion for a given unemployment rate.

Hornstein et.al. (2010) propose a wage dispersion measure based on the ratio of the reservation wage to the mean wage (the *mean-min* ratio). Similarly, Eeckhout and Kircher (2010) construct a measure based on the distance between the lowest and highest wage. One disadvantage of relating wage dispersion to the lowest observed wage is that the wage distribution for a given skill type has a long left tail. This long tail is consistent with OJS for the reasons spelled out in Burdett and Mortensen (1998): (i) OJS reduces the lowest acceptable job type because less option value of continued search has to be given up when accepting a job, and (ii) less workers quit from good matches and more workers accept good matches. Empirically, it matters a lot whether one takes the 1^{th} or the 2^{nd} percentile of the wage distribution as a proxy for the reservation wage. Therefore, the difference between the highest wage at the optimal assignment and the mean wage (the max-mean differential) is a more robust measure for wage dispersion. Moreover, the max-mean differential is less sensitive to the relative efficiency of on- versus off-the-job search, ψ. The reason for this is that there are two offsetting effects of an increase in the efficiency of OJS on the mean wage while the maximum wage is hardly affected. First, more job offers for employed workers decrease the lowest acceptable wage which has a small negative effect on the mean wage, and secondly, more job offers imply that workers move faster towards the optimal assignment and this increases competition between firms for workers, resulting in a small positive effect on both the maximum and the mean wage.

When on- and off-the-job search are equally efficient and when the well known congestion effects of opening vacancies are switched off (i.e. by a quadratic contact technology), the equilibrium where firms are able to commit to their posted wages is constrained efficient. In the equilibrium where firms are unable to commit, quasi-rents per worker are higher than in the social optimum due to a business-stealing externality. Under free entry, these quasi-rents are spent on (excess) vacancy creation, see Gautier, Teulings and van Vuuren (2010). When on and off- the-job search are equally efficient, we estimate the output loss due to this business externality to be up to 6% (if no firm commits).

The empirical estimate of the max-mean wage differential for identical workers in heterogeneous jobs is an extension of the framework of Gautier and Teulings (2006) with OJS. Our estimate is based on the intercept of a simple quadratic wage regression with appropriately normalized measures for worker and job characteristics. This type of inference is highly sensitive to measurement error in both characteristics because an observed sub-optimal matched worker can either reflect true mismatch, or simply imply measurement error. Our estimation procedure accounts for this problem. Gautier and Teulings (2006) use second order terms in worker and job characteristics to capture the concavity of wages around the optimal assignment that is implied by search models with sorting. However, there is a crucial difference between a model with and without OJS. In a model without OJS, wages are a linear transformation of the match surplus. Since the match surplus is a differentiable function of the match quality indicator, so is the wage function. This simple relation no longer applies with OJS, see Shimer (2006). The wage function turns out to be non-differentiable at the optimal assignment. At that point, firms are prepared to pay the highest premiums to raise hiring and to reduce quitting. In our empirical application, we take this into account. Allowing for OJS is important since Fallick and Fleischman (2004) and Nagypal (2005) show that job-to-job flows are twice as large as the job-to-unemployment flow.

Lise, Meghir and Robin (2008) and Lopes de Melo (2008) also look at sorting in models with OJS. Their focus is on interpreting the correlations between worker and firm fixed effects and how this relates to complementarities between worker and job types in the production technology. Finally, Eeckhout and Kircher (2009) consider a simple model based on Atakan (2006) where workers and jobs are randomly matched and have the option to dissolve and at some cost move to a competitive sector with perfect sorting. They derive a similar hump shaped relation where productivity is highest at the optimal assignment and decreases in the distance from this optimal assignment. This framework is however less suitable to bring to the data.

The structure of the rest of this paper is as follows. Section 2 presents the basic framework. Section 3 discusses the basic steps in our argument. In this section, we also reinstall the hierarchical features of the model and derive the wage function that comes with it. We also show how we can normalize worker and job skills such that we can relate the constant in a simple wage regression to the max-mean wage differential. Finally, section 4 concludes.

2 The model

2.1 Why a circular model?

Shimer and Smith (2000) and Teulings and Gautier (2004) analyze an assignment model with search frictions and without OJS. Though the idea is straightforward, the analysis is complicated. Figure 1 provides an intuition for why this is the case. The figure shows the space of potential matches between skill types, s, and worker types, c. The Walrasian equilibrium assignment is depicted by the main diagonal. Comparative advantage of skilled workers in complex jobs implies that it is upward sloping. Perfect sorting implies that it is a one-to-one correspondence. Search frictions and sufficient complementarities between worker and job types imply that the equilibrium assignment is a set rather than a point where x measures the distance of worker type s to her optimal assignment. Away from the corners, the optimal match is in the middle of the matching set while close to the corners, the optimal assignment is close to the boundary. Teulings and Gautier (2004) who do not allow for OJS use Taylor expansions for the middle part and show that for small search frictions and if worker and job types are normally distributed, the corner

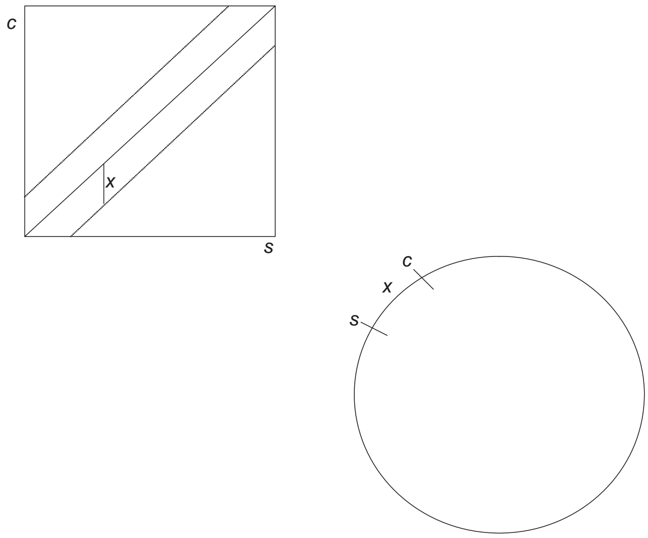

Figure 1: The hierarchical versus the circular model

problem can be ignored. With OJS this approach does not work well. Gautier, Teulings, and Van Vuuren (2005) show that by taking out the hierarchical aspect of the model, the south-west and the north-east corner of the matching space can be "glued" together. Then, a circular model in the spirit of Marimon and Zilibotti (1999) can be used, see the lower part of Figure 1. The distance x to the optimal assignment is now measured along the circumference of the circle. All conclusions from the analysis of the hierarchical model without OJS in Teulings and Gautier (2004) based on Taylor expansions carry over to the circular model. The intuition for this is that if there is relatively little mass around the corners, all that matters for productivity is the distance to the optimal job. We follow this idea in this paper, but now for a model with OJS. First, we take out the hierarchical feature and provide a closed form analysis of a search equilibrium with sorting in the context of the circular model. Then, we reintroduce the hierarchical aspect for the empirical analysis of wage differentials, using data with information on worker and job characteristics, which are hierarchical by nature.

The interpretation of the circular search model as an approximation of the hierarchical model has an important implication for the production structure. Figure 2 depicts Rosen's hedonic equilibrium in the wage-skill space. The upper panel is the Walrasian case. The

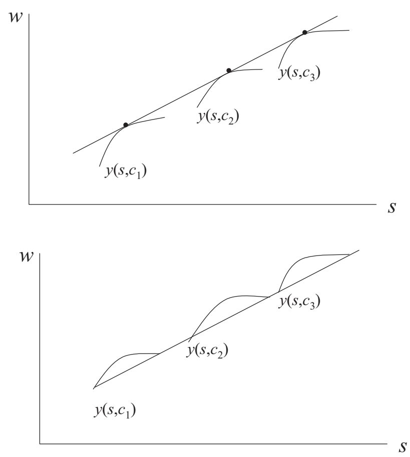

Figure 2: Rosen's offer curves and the shape of matching sets

upward sloping line is the market wage for a worker with skill type s, the offer curves represent the value of output in jobs with a particular level of complexity c when occupied by worker types with different skill types. The point of tangency is the optimal assignment, the only assignment that is relevant in the Walrasian case. The lower panel is the case with search frictions. The upward sloping line is now the reservation wage of the worker. All values of s enveloped by this reservation wage and the offer curve are now part of the matching set of that job type. Taking out the hierarchical aspect of the model implies that the (reservation) wage function becomes horizontal. What is crucial is that these functions keep their shape by this operation. Hence, these functions are differentiable in their maximum. We impose this feature in the theoretical model.

2.2 Assumptions

Production

There is a continuum of worker types s and job types c; s and c are locations on a circle. Workers can only produce output when matched to a job. The productivity of a match of worker type s to job type c depends on the "distance" $|x|$ between s and c along the circumference of the circle, where x is defined as $x \equiv s - c$. $Y(x)$ has an interior maximum at $x = 0$; it is symmetric around this maximum, which is normalized to unity: $Y(0) = 1$;

$Y(x)$ is twice differentiable and strictly concave. We consider the simplest functional form that meets these criteria:

$$Y(x) = 1 - \frac{1}{2}\gamma x^2. \tag{1}$$

x is the *mismatch* indicator. The parameter γ determines the cost of sub-optimal assignment. $Y(x)$ can be interpreted as a second order Taylor approximation around the optimal assignment of a more general production technology. Since, the first derivative of a continuous production function equals zero in the optimal assignment, $Y'(0) = 0$, the first order term drops out. We are interested in equilibria where unemployed job seekers do not accept all job offers, which imposes a minimum constraint on γ.[1]

Labor supply and the value of non market time

Labor supply per s-type is uniformly distributed over the circumference of the circle. Total labor supply in period t equals $L(t)$. Unemployed workers receive the value of non market time B. Employed workers supply a fixed amount of labor (normalized to unity) and their payoff is equal to the wage they receive. Workers live for ever. They maximize the discounted value of their expected lifetime payoffs.

Labor demand

There is free entry of vacancies for all c-types. The cost of maintaining a vacancy is equal to K per period. After a vacancy is filled, the firm's only cost is the worker's wage. The supply of vacancies is determined by a zero profit condition: firms open new vacancies up till the point where the discounted value of expected profits is equal to the cost of keeping that vacancy open. Vacancies are uniformly distributed over the circumference of the circle; $v(c) = v$ is the measure of vacancies per unit of c.

Job search technology

We use a reduced form specification of the job search technology:

$$\lambda = \lambda(u, v), \lambda_v > 0,$$

and assume that the rate at which unemployed workers meet jobs is λ and the rate at which employed workers meet jobs is $\psi\lambda$. The exact functional form of this relation is irrelevant for the analysis in this paper. E.g. our analysis applies irrespective of the

[1] A sufficient condition for this is that $Y(x) < 0$ for at least some x. Let C be the circumference of the circle, then $0 \leq x \leq \frac{1}{2}C$. Hence $\gamma > 8C$.

returns to scale. The parameter $\psi, 0 \leq \psi \leq 1$, measures the relative efficiency of on- relative to off-the-job search; $\psi = 0$ is the case without OJS, as analyzed in the stochastic matching model of Pissarides (2000), Marimon and Zilibotti (1999) and Teulings and Gautier (2004); for $\psi = 1$, on- and off-the-job search are equally efficient. When a worker quits her old job to accept a new job, the old job disappears. Hence, she cannot return to her previous job.

Job destruction

Matches between workers and jobs are destroyed at an exogenous rate $\delta > 0$.

Golden-growth path

We study the economy while it is on a golden-growth path, where the discount rate $\rho > 0$ is equal to the growth rate of the labor force. We normalize the labor force at $t = 0$ to one. Hence, the size of the labor force is $L(t) = \exp(\rho t)$. The assumption of a golden-growth path buys us a lot in terms of transparency and tractability. The implications of the golden growth assumptions are equivalent to those that follow from the assumption that the discount rate ρ converges to zero, an assumption that is often applied in the wage posting literature, see for example Burdett and Mortensen (1998) because discounting reduces future output while population growth increases it. New workers enter the labor force as unemployed.

Labor supply per worker type, the productivity in the optimal assignment $Y(0)$ are all normalized to one. Hence, in the absence of search frictions, the output of this economy equals one.

Wage setting

Wages, denoted by $W(x)$, are set unilaterally by the firm, conditional on the mismatch indicator x in the current job. We analyze wage setting under two different assumptions on the ability of firms to commit to future wage payments. Under the first assumption, firms can commit to a future wage payment contingent on x. Then, firms pay both no-quit and hiring premiums, that is, they account for the positive effect of a higher wage offer on reduced quitting and increased hiring. Under the second assumption, firms are unable to commit to future wage payments. In this case, hiring premiums are non-credible because immediately after the worker has accepted the job, the firm has no incentive to continue paying a hiring premium, since the worker cannot return to her previous job. Workers anticipate this, and will therefore not respond to this premium in the first place, and hence, firms will not offer it. No-quit premiums are credible even without commitment because it is in the firm's interest to pay them as soon as the worker has accepted the job. Starting to pay a no-quit premium only when the worker gets an offer is non-credible again, because the firm has no incentive to continue paying the no-quit premium after the worker has declined the offer. Hence, the only way for the firm to gain credibility in paying no-quit premiums is to pay these premiums right from the start.[2]

2.3 The asset values of (un)employment and vacancies

The golden-growth assumption is particularly useful for the derivation of the asset values of employment, unemployment, and vacancies.

Asset value of unemployment

In the appendix, we show that the asset value of unemployment, denoted by V^U is a weighted average of the worker's payoff while unemployed, B, and the expected wage when employed, $\mathrm{E}_x W$, the weights being the unemployment and the employment rate, respectively:

$$\rho V^U = uB + (1-u)\,\mathrm{E}_x W. \tag{2}$$

Why does this relation take such a simple form? The reason is that the growth rate of the workforce is equal to the worker's discount rate. Therefore, the expected payoff of a worker with one year of experience is equal to the average payoff of the cohort of workers that entered the labor force one year ago. Likewise, the expected payoff of a worker with two years of experience is equal to the average payoff of the cohort of workers that entered the labor market two years ago, etc. The asset value of unemployment is equal to the weighted sum of expected payoffs for each level of experience, future payoffs being discounted at a rate ρ per year. This weighted sum is exactly equal to the sum of payoffs for the current workforce. The fact that older cohorts are smaller than younger cohorts due to the growth of the labor force at a rate ρ exactly offsets the effect of discounting future payoffs for the calculation of the asset value of unemployment. The term $(1-u)\mathrm{E}_x W$ can be interpreted as the option value of finding a job. Alternatively, when $\rho \to 0$, workers spend a fraction u of their life as unemployed and the rest of the time they are employed.

Asset value of employment in the marginal job

[2] See also Bontemps van den Berg and Robin (2000) for wage setting with and Coles (2001) and Shimer (2006) for wage setting without commitment.

Let $V^E(x)$be the asset value of holding a job with mismatch indicator x. At $\bar{x}$, an unemployed job seeker is indifferent between accepting the job or waiting for a better offer: $V^E(\bar{x}) = V^U$. Again, the asset value for this job is a weighted average of $W(\bar{x})$ and $\mathrm{E}_x W$:

$$\rho V^E(\bar{x}) = \frac{uW(\bar{x}) + \psi(1-u)\mathrm{E}_x W}{u + \psi(1-u)}, \tag{3}$$

see Appendix A.1 for the derivation. The factor $u + \psi(1-u)$ is the effective supply of job seekers, namely u unemployed job seekers and $(1-u)$ employed job seekers, which are discounted by a factor ψ due to their lower search effectiveness. Hence, the weights in equation (3) are the shares of unemployed and employed respectively in the effective supply. The intuition for this equation is that the option value of finding a better job is the same as for an unemployed job seeker, since both an unemployed and a worker employed in the marginal job type $x = \bar{x}$ accept any job: $0 \leq |x| < \bar{x}$. However, the option value of an employed worker is only a fraction ψ of the option value of an unemployed due to their lower contact rate. At first sight, one would expect that the value of non market time B would show up in the equation, because at a rate δ the worker is fired and receives the asset value of unemployment ρV^U, see equation (2). However, since $V^E(\bar{x}) = V^U$, we can substitute $V^E(\bar{x})$ for V^U, thereby eliminating B from the equation. For $0 < \psi < 1$, unemployed job seekers give up part of the option value of search by accepting a job. An unemployed job seeker accepts a job offer if and only if she is compensated for this loss in option value, implying that $W(\bar{x}) > B$. For $\psi = 1$, on- and off-the-job search are equally efficient, so an unemployed job seeker does not give up any option value by accepting a job. Hence: $W(\bar{x}) = B$, and equation (3) simplifies to the same expression as equation (2):

$$\rho V^E(\bar{x}) = uB + (1-u)\mathrm{E}_x W.$$

Asset value of vacancies

Adding up the zero profit condition for all vacancies implies that the total cost of maintaining vacancies must be equal to the aggregate rents that firms make in currently filled jobs. In Appendix A.1 we show that

$$vK = (1-u)(\mathrm{E}_x Y - \mathrm{E}_x W). \tag{4}$$

The left hand side of this equation is the total cost of vacancies at $t = 0$. The right hand side is employment $1-u$ times the quasi-rents per worker, which is equal to expected output $E_x Y$ minus expected wages $E_x W$. One would expect that quasi-rents should be discounted, since firms must first post a vacancy before they can make a profit. However, as for the asset value of employment, the right-hand side is again a weighted average of cohorts of vacancies and discounting is offset by the growth of the economy.

The reservation value of the mismatch indicators

The definition of $\bar{x}$ as the mismatch indicator of a job which is just acceptable to an unemployed job seeker implies:

$$W(\bar{x}) = Y(\bar{x}) = 1 - \frac{1}{2}\gamma\bar{x}^2, \tag{5}$$

To understand this condition, note that if $W(\bar{x}) < Y(\bar{x})$, there would be a job type $|x| > \bar{x}$ for which an employer could offer a wage W, $W(\bar{x}) < W < Y(x)$, which would be attractive to both the worker and the firm, which is inconsistent with profit maximization, while if $W(\bar{x}) > Y(\bar{x})$, the firm would be better off by not hiring the worker at all, which is again inconsistent with profit maximization. Substitution of equation (2) and (4) in the condition $V^E(\bar{x}) = V^U$ yields,

$$W(\bar{x}) = [u + \psi(1-u)]B + (1-\psi)(1-u)E_x W. \tag{6}$$

When on and off the job search are equally efficient, $\psi = 1$, equation (6) simplifies to:

$$W(\bar{x}) = B = 1 - \frac{1}{2}\gamma\bar{x}^2. \tag{7}$$

where the last step follows from (5). Hence, the relation between γ and $\bar{x}$ does not depend on expected wages in this case, and consequently neither on whether or not firms can commit on paying hiring premiums.

The output loss due to search frictions

The definition of the output loss due to search frictions is given by:

$$X = (1-u)(1-E_x Y) + u(1-B) + vK = 1 - (1-u)E_x Y - uB + vK. \tag{8}$$

The output loss is equal to employment, $1-u$, times the difference between productivity in the optimal assignment, $Y(0) = 1$, and the expected productivity in the actual assignment, $E_x Y$, plus unemployment, u, times the difference between the productivity in the optimal assignment and the value of non market time, $1-B$, plus the cost of keeping vacancies open, vK. Substitution of equation (2) and (4) in (8) yields:

$$X = 1 - \rho V^U = u(1-B) + (1-u)(1-E_x W). \tag{9}$$

The last step follows from the fact that by the zero profit condition, the cost of maintaining vacancies is equal to the surplus of expected productivity over expected wages, see equation (4). The first equality tells us that the output loss is equal to the output in the optimal assignment ($Y(0) = 1$) minus the asset value of unemployment. As frictions become smaller workers find jobs close to the optimal assignment quickly and X becomes smaller. The second equality in (9) tells us that the output loss is equal to the sum of the output loss for unemployed and for employed workers. The former is equal to the lost output in the optimal assignment minus the value of non market time, while the latter is equal to the foregone wage income. Under free entry, the difference between wages and productivity is spent on vacancies.

2.4 Equilibrium flow conditions

Under both assumptions for wage setting, commitment and no-commitment, wages are a decreasing function of x for $x \geq 0$, $W_x(x) > 0$, implying that workers accept any job-offer with a lower mismatch indicator $|x|$ than their current job. Hence, we can analyze job-to-job flows independent of the exact form of the wage policy $W(x)$. Let $G(x), x \geq 0$, be the fraction of workers employed in jobs at smaller distance from their optimal job than $|x|$. This implies that $G(0) = 0$ and $G(\bar{x}) = 1$, since $\bar{x}$ is the largest acceptable value of $|x|$. The golden growth assumption requires that the number of workers employed in jobs with a mismatch indicator lower than x grows at a rate ρ:

$$2\lambda x \left\{u + \psi(1-u)\left[1 - G(x)\right]\right\} - \delta(1-u)G(x) = \rho(1-u)G(x)\,. \tag{10}$$

The first term on the left-hand side is the number of people that find a job with mismatch indicator lower than x, either from unemployment (the first term in braces), or by mobility from jobs with a larger mismatch indicator (the second term in brackets). The number of better jobs is given by $2x$, since the worker can accept jobs both to the left and to the right of her favorite job type $x = 0$. The second term in brackets is weighted by the factor ψ, reflecting the efficiency of on- relative to off-the-job search. The final term on the left-hand side is the outflow of workers due to job-destruction. The right-hand side reflects that at the balanced growth path, employment grows at a rate ρ at all levels including the class of workers with a mismatch indicator smaller than x, $G(x)$. Mobility within this class is irrelevant because the disappearance of the old match and the emergence of the new one cancel. Evaluating (10) at $\bar{x}$ yields:

$$\delta(1-u)+\rho-2\lambda\bar{x}u=\rho u.$$

Solving for u yields:

$$u=\frac{1}{1+\kappa\bar{x}}, \tag{11}$$
$$\text{where: } \kappa\equiv\frac{2\lambda}{\rho+\delta}.$$

Substitution of (11) into condition (10) yields:

$$G(x)=1-\frac{\bar{x}-x}{(1+\psi\kappa x)\,\bar{x}}, \tag{12}$$
$$g(x)=\frac{1+\psi\kappa\bar{x}}{\bar{x}\,(1+\psi\kappa x)^2},$$

where $g(x)$ is the density function of x among employed workers.

2.5 Wage formation

The wage formation processes are the same as in Gautier, Teulings, and Van Vuuren (2010). We present their main results here. Since the model is symmetric around $x=0$, we can focus on the analysis of $W(x)$ for $x\geq 0$.

Wage setting with commitment

When firms can commit on future wage payments, the optimal wage policy of the firm maximizes the expected value of a vacancy,

$$W(x)=\arg\max_{W}\left(\left[u+\psi(1-u)\,\widehat{G}(W)\right]\frac{Y(x)-W}{\rho+\delta+\psi\lambda F(W)}\right), \tag{13}$$

where $\widehat{G}[W(x)]\equiv 1-G(x)$ is the distribution of wages among employed workers and where $F[W(x)]\equiv 1-\frac{x}{\bar{x}}$ is the wage offer distribution, using the fact that the distribution of x is uniform by assumption. The effect of $\widehat{G}(W)$ on the optimal wage offer is the hiring premium, the effect of $F(W)$ is the no-quit premium. The first order condition of this problem reads:

$$W_x(x)=-2\frac{\psi\kappa}{1+\psi\kappa x}\left[Y(x)-W(x)\right]. \tag{14}$$

This differential equation can be solved analytically for $W(x)$, using equation (5) as an initial condition, see Appendix A.4.

Wages setting without commitment

When firms cannot commit on future wage payments, hiring premiums are non-credible and. Hence, the term $\widehat{G}(W)$ in equation (13) is replaced by $1 - G(x)$ reflecting that the first term in brackets does not depend on the wage and that the wage maximizes the value of a filled job rather than the value of a vacancy. Then, the first order condition reads,

$$W_x(x) = -\frac{\psi\kappa}{1+\psi\kappa x}\left[Y(x) - W(x)\right]. \tag{15}$$

The only difference with equation (14) is a factor two, reflecting the fact that firms pay hiring and no quit premia in the case of commitment, while they pay only a no quit premium in the case without commitment. Again, this differential equation can be solved analytically for $W(x)$, using equation (5) as an initial condition.

3 Measuring wage dispersion, mismatch and the output loss

This section shows that our model yields robust predictions on the relation between wage dispersion, unemployment, the ratio of job-to-job and unemployment-employment flows, and the output loss due to search frictions and that this depends on only a few easily observable statistics. We also offer a methodology to test those predictions empirically. The argument requires a number of steps. As a map for the reader, we first provide an overview of these steps.

1. We show that we can normalize $\kappa = 1$ without loss of generality, and that this normalization implies that the equilibrium is fully characterized by just three parameters: the equilibrium rate of unemployment u, the value of non market time B, and the relative efficiency of on versus off the job search ψ.

2. We derive a simple and robust measure of wage dispersion and provide an analytical characterization of the equilibrium. The measure we propose is the max-mean wage differential, $W(0) - \mathrm{E}_x W$. This measure is less sensitive for the extreme long right tail of the mismatch indicator $|x|$ for identical workers and to the precise value of ψ than dispersion measures based on the lowest wage. Since the analysis becomes much simpler for the special case: $\psi = 1$, we proceed by focussing on this case. In step 6 we look how sensitive our results are to introducing lower values of ψ.

3. In an empirical application, data on x will be only partially observed. This causes special problems due to the non-differentiability of $\ln W(x)$ at $x = 0$. Define

$$x = r + q,$$

where r is the observed and q is the unobserved component of x, with $\mathrm{Cov}[r, q] = 0$. We show that values of r close to zero provide only limited information on the effect of x on $W(x)$. We simulate data on x, q, and $W(x)$ and use these data to estimate the quadratic equation

$$\ln W - \mathrm{E} \ln W = \omega_0 + \omega_2 r^2 + \varepsilon,$$

where ω_0 and ω_2 are parameters and where ε is an error term. We show that ω_0 is informative on the max-mean wage differential and that it depends on the signal-noise ratio.

4. We derive empirical proxies of s and c, and consequently of $x \equiv s - c$, using a methodology outlined in Gautier and Teulings (2006). We provide some additional robustness checks to this method and use these proxies to estimate ω_0.

5. We show how our model relates to the hedonic/assignment models of Rosen (1974), Sattinger (1975), Teulings (1995). These models have direct implications for the elasticity of substitution between high and low skilled workers, as estimated by Katz and Murphy (1992). In particular, there is a one-to-one correspondence between this elasticity of substitution and the second derivative, γ, of the production function $Y(x)$. This relation enables us to establish the units of measurement of x.

6. We investigate the impact of ψ for the expected level of unemployment and the decomposition of the output loss into unemployment, recruitment cost, and the productivity loss due to suboptimal assignment, conditional on the estimated value of ω_0.

These steps will be discussed in the next 6 subsections.

3.1 Normalizing $\kappa = 1$

The first step is to show that κ can be normalized to one without loss of generality. Define a linear transformation of x, $\widetilde{x} \equiv \kappa x$, and the parameters $\widehat{x} \equiv \kappa\bar{x}$ and $\widetilde{\gamma} \equiv \kappa^{-2}\gamma$. Then

$$u = \frac{1}{1+\kappa\bar{x}} = \frac{1}{1+\widehat{x}} \tag{16}$$

$$Y(x) = 1 - \frac{1}{2}\gamma x^2 = 1 - \frac{1}{2}\widetilde{\gamma}\widetilde{x}^2. \tag{17}$$

It can be easily checked that this transformation leaves all other primitives of the model unaffected. Hence, by redefining x as $\widetilde{x}$ and γ as $\widetilde{\gamma}$, we can normalize κ to one. The implication of this normalization is that the equilibrium of this sorting model with search frictions is a function of only three statistics, the rate of unemployment u, the value of non market time B, and the relative efficiency of on- versus off-the-job search, ψ. It does not depend on the curvature of the production function γ. The unemployment rate is obviously not a primitive of the model, but an outcome. However, since we have an idea about the value of the natural rate of unemployment, it is useful to have a characterization of the equilibrium in terms of the value of u. The reason why the parameter γ does not enter the characterization of the equilibrium can most easily be understood for the case $\psi = 1$. (16) and (17) reveal that the normalization of κ to unity implies that the value of the mismatch indicator in the worst acceptable job offer $\widehat{x}$ is a simple function of the unemployment rate. Then, equation (7) (and replacing $\overline{x}$ by $\widehat{x}$) shows that $\widetilde{\gamma}$ can be written as a simple function of the unemployment rate and the value of non market time only.

Figure 3 depicts the density and distribution function of the mismatch indicator $|\widetilde{x}|$ conditional on employment, for the case $u = 5\%$ and $\psi = 1$ (we use these values in all subsequent plots, unless stated otherwise). For a given unemployment rate, $G(\widetilde{x})$ is identical with or without commitment since workers climb the job ladder equally fast in both cases (because wages are in both cases strictly decreasing in x). The main message from Figure 3 is that the distribution of $|\widetilde{x}|$ has a large probability mass close to zero (the optimal assignment) and a long right tail of bad matches. The median value of $\widetilde{x}$ is equal to $(1-u)/(1+u) < 1$, far smaller than $\widehat{x} = (1-u)/u = 19$ (the mismatch indicator in the worst match). The reason for this pattern is that workers who are matched badly quit their jobs fast. The reverse holds for good matches, so their density is high. The skewness of the distribution of $|\widetilde{x}|$ has a number of counter intuitive implications for the wage distribution that are spelled out in greater detail below.

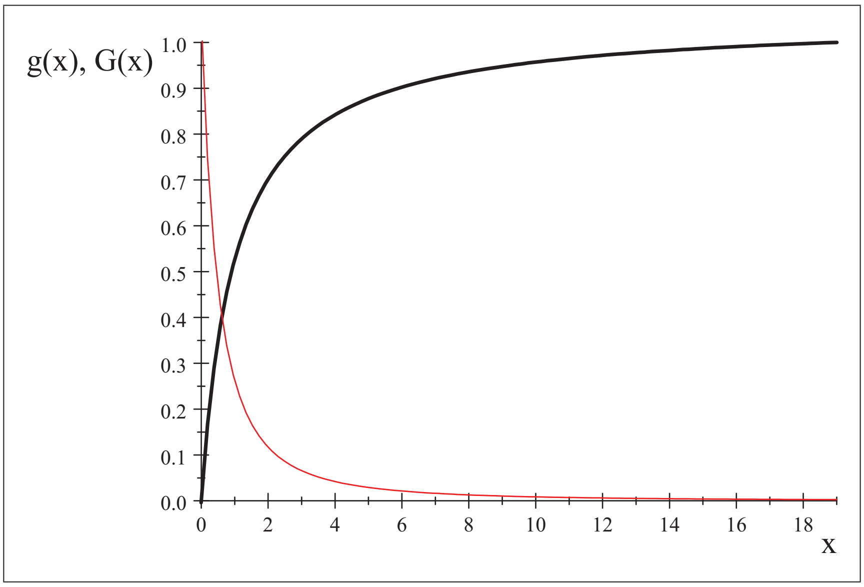

Figure 3: The distribution (thick black) and density function (thin red) of x conditional on employment

The analytical expressions for $W(\widetilde{x}), W_x(\widetilde{x})$, and $\text{E}_x W$ are complicated and highly non-linear, so we placed them in Appendix A.4. Figure 4 depicts $Y(\widetilde{x})$ and $W(\widetilde{x})$ both for the case with and without commitment, setting the value of non market time at $B = 0.4$ (which we do in all subsequent plots). Some authors, i.e. Hagedorn and Manovskii (2008) and Hall (2009) who want to explain the cyclical behavior of unemployment use larger values for B. For these studies, the value of non-market time of the *marginal worker* is relevant whereas here we are interested in the value of non-market time for the *average worker* so a lower value is justified. Contrary to $Y(\widetilde{x})$, $W(\widetilde{x})$ is non-differentiable at $\widetilde{x} = 0$. This is due to the hiring and no-quit premiums that firms pay. Since the density of employment is highest for low values of $|\widetilde{x}|$, the elasticity of labor supply is high for these types of job. A slight variation in wages has large effects both on the probability that workers accept an outside job offer and on the number of workers who are prepared to accept the wage offer (the latter being relevant in the case with commitment only). Hence, firms will bid up wages aggressively for these types of jobs. Since $\psi = 1$, workers accept any job offer that pays more than the value of non market time.

Figure 4 shows that the wage in the optimal assignment is higher when firms can commit than when they cannot, since the ability to commit increases competition between

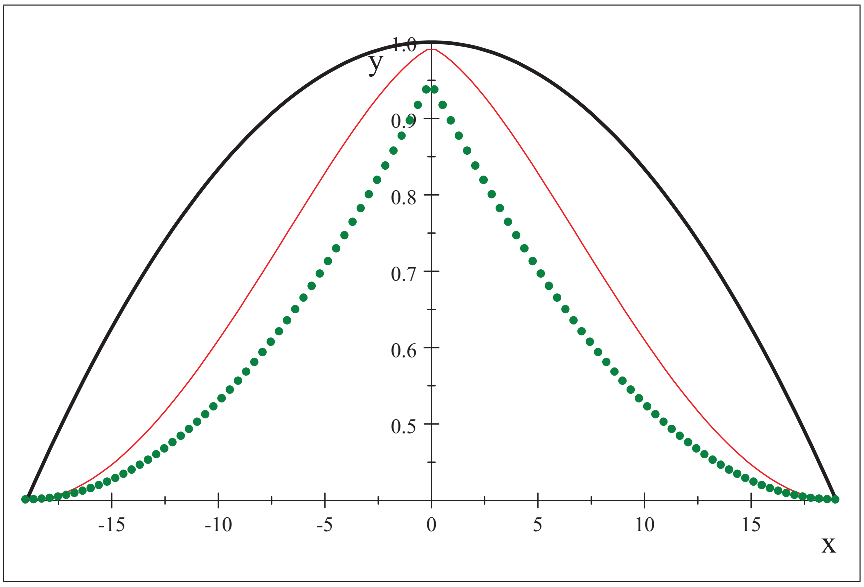

Figure 4: Productivity $Y(x)$ (thick) and wages $W(x)$ with (thin) and without (dotted) commitment

firms for workers. Given the wage setting policy of other firms, the ability to commit allows a firm to make more profits, because the hiring premium enables it to fill its vacancies faster. However, since all other firms do the same, profits are lower compared to the case without commitment. Figure 4 also reveals that for $\widetilde{x} = 0$ the slope of the wage function is smaller (in absolute value) for the case with than without commitment. This is remarkable, since the only difference between the expressions for the slope is a factor 2 in the differential equations for wages for the case with commitment, compare equation (14) and (15). From this point of view, one would expect a steeper wage function for the case with commitment. However, the slope is proportional to the flow profits per worker, i.e. the difference between the productivity and wages, $Y(\widetilde{x}) - W(\widetilde{x})$. This difference is more than twice as large in the case without commitment, yielding a steeper wage function in that case. A comparison with figure 3 reveals that despite the zero derivative of $W(\widetilde{x})$ at $\widetilde{x} = \widehat{x}$, the wage distribution has a fat right tail. Hence, in empirical applications, the reservation wage (the lowest observed wage) is highly sensitive to the exact definition of the lowest wage. The 1^{st} percentile is very different from the 2^{nd} percentile while the $98^{th}$and 99^{th} percentile are very similar. The mean-min wage differential as proposed by Hornstein et.al. (2010) is therefore a less robust statistic for measuring wage dispersion of identical worker types than the max-mean differential.

3.2 Key statistics are largely independent of ψ

The expressions for the wage function $W(\widetilde{x})$ and the expected wage $\mathrm{E}_x W$ can be used to calculate the max-mean and the mean-min wage differentials. These expressions are depicted in Figure 5 for the case with commitment together with the output loss due to search frictions. First, note that the max-mean differential is largely independent of ψ, while the mean-min differential is very sensitive to the precise value of ψ. For the case without commitment (not in the Figure) the max-mean differential varies even less with ψ. It is important to realize that this is not a comparative statics exercise in ψ, for then it would not make sense to keep u constant. The question addressed here is what wage differential and output loss are consistent with a particular value of ψ and an unemployment rate of 5%. Implicitly, the value of γ adjusts to keep the unemployment rate at this level. The reason why ψ does not matter for the max-mean wage differential is that lowering the value of ψ while keeping u constant has two offsetting effects on wage differentials near the optimal assignment. The density of $\widetilde{x}$ at the optimal assignment is equal to $g(0) = \psi + u/(1-u)$, see equation (12). Hence, a lower value of ψ implies that there are fewer workers close to the optimal assignment since search by employed workers is less efficient and since employed job seekers are a particularly relevant source of labour supply for an $\widetilde{x} = 0$ job since all employed job seekers are employed at less optimal jobs. This reduces the mean wage. Hence, holding $W(\widetilde{x})$ constant, a lower value of ψ yields a larger max-mean wage differential. However, holding $\widetilde{x}$ constant, unemployed job seekers become more choosy since they give up a share $1-\psi$ of the option value of search by accepting a job. Therefore, the lowest wage $W(\widehat{x})$ goes up. This reduces wage differentials. For the mean-min wage differential only the latter factor is relevant. Hence, contrary to the max-mean differential, the mean-min differential is particularly sensitive to the value of ψ.

Second, the max-mean differential on the one hand and the mean-min and min-max wage differentials on the other hand tell opposite stories about the effect of commitment on wage differentials. Commitment makes firms compete more fiercely for workers, driving up the maximum wage. Since the minimum wage is the same for the case with and without commitment, this would imply that the max-min differential is larger under commitment. However, since the slope of the wage function close to the optimal assignment is smaller

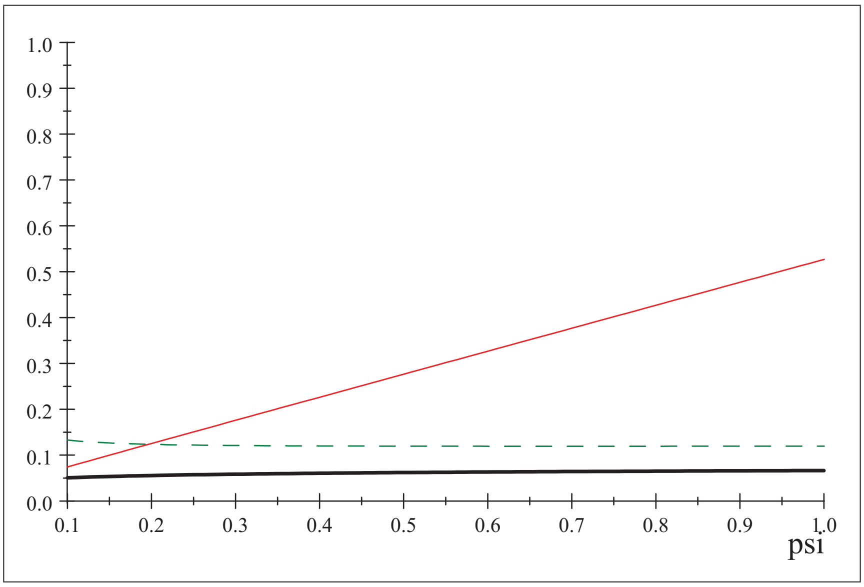

Figure 5: Wage differentials and output loss (with commitment), max-mean differential (thick), mean-min (thin), welfare loss (dashed)

under commitment, see Figure 4, the max-mean differential is actually smaller; 6.7% of the wage in the optimal assignment with commitment versus 9.9% without commitment.

Third, the estimated output loss, conditional on u, is largely independent of ψ. It is much smaller with than without commitment, 10% versus 17% respectively. The reason that the output loss is lower in the commitment case is that in that case the business-stealing externality is internalized, see Gautier et al. (2010). In equation (9), the term $1-\mathrm{E}_x W$, can be written as $[1 - W(0)] + [W(0) - \mathrm{E}_x W]$; since $W(0)$ is close to one, this expression is dominated by the max-mean differential in the second term. To conclude, the precise value of ψ is not very important for the max-mean wage differential and the output loss, but it makes a large difference for the mean-min differential. Setting $\psi = 1$ simplifies the formulas substantially, see the Appendix. All wage differentials and the output loss are proportional to $1 - B$, the difference between output in the optimal assignment and the value of non market time The relation between the max-mean differential, the output loss and the unemployment rate is plotted in Figures 6 and 7 for the commitment and the no-commitment case respectively. The non-linearity is hardly visible in the case of commitment.

Finally, the model generates a sharp prediction on the relation between employment-

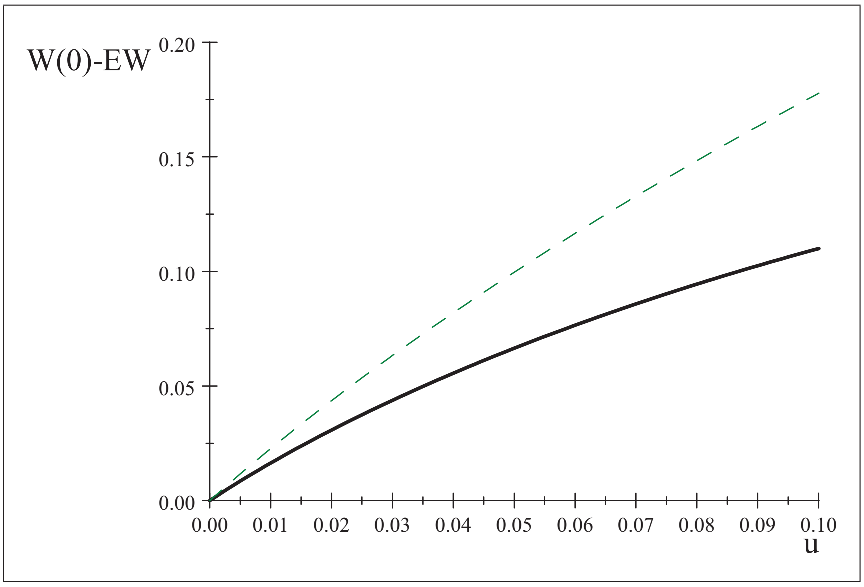

Figure 6: The max-mean wage differential (solid) and the output loss (dashed) (with commitment)

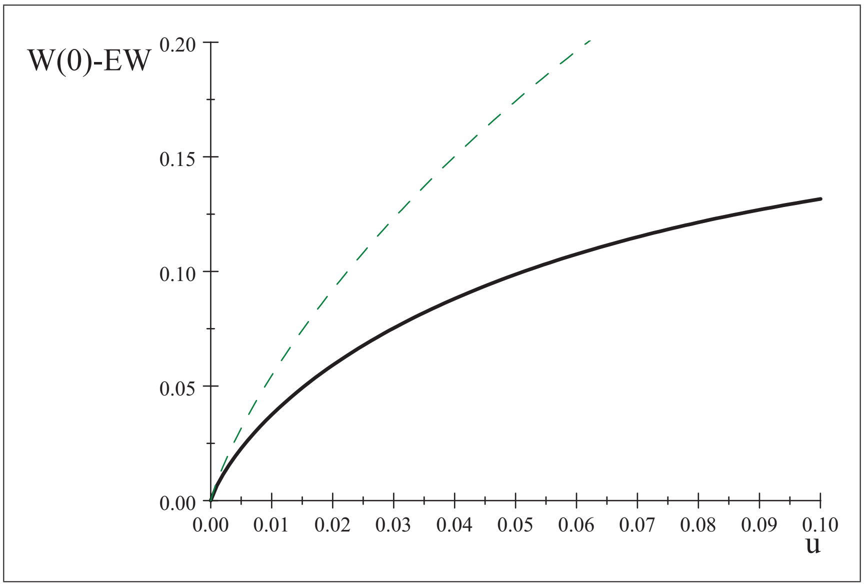

Figure 7: The max-mean wage differential (solid) and the output loss (dashed) (without commitment)

employment, $\{e \to e\}$ flows and unemployment-employment flows, $\{u \to e\}$.

$$\frac{\{e \to e\}}{\{u \to e\}} = -\frac{1}{1-u} \ln u - 1,$$

which is 2.15, for our benchmark parameters, see the Appendix for the derivation for the general case where $\psi \leq 1$. In section 3.6 we discuss the empirical evidence on this value and show how this ratio varies with ψ. For now, we just mention that this value is consistent with the empirical evidence presented by Nagypal (2008).

3.3 Empirical inference on wages, allowing for unobserved heterogeneity in x

Our goal is to confront this model with data on wages, worker-skills s, and job-complexity levels c, (which we use to calculate the mismatch indicator $x \equiv s - c$). Since we have no idea about the units of measurement of the normalized mismatch indicator $\widetilde{x}$, we first create a non-normalized measure for x. Then, in the next section we discuss the appropriate normalization procedure.

We must first deal with the problem that data on s and c provide only partial information on the mismatch indicator. Some variation is unobserved by the researcher. Due to the non-differentiability of $\ln W(x)$ at $x = 0$, this unobserved heterogeneity has large implications for the estimation of the wage equation. Let r and q be the observed and the unobserved component in x respectively:

$$x = r + q,$$

where $\mathrm{Cov}[r, q] = 0$. Let $w \equiv \ln W - \mathrm{E}[\ln W]$.

Since the function $W(x)$ is symmetric and non-differentiable around $x = 0$, a simple approach is to estimate,

$$w = \overline{\omega}_0 - \overline{\omega}_1 |x| + \varepsilon, \tag{18}$$

$\overline{\omega}_0 > 0$ and $\overline{\omega}_1 > 0$ are parameters to be estimated and where $\varepsilon \equiv \ln W(x) - \overline{\omega}_1 |x|$, is an error term with zero mean. We propose to estimate this model in logs rather than levels for two reasons. The main reason is that we normalized output in the optimal assignment to unity, $Y(0) = 1$. Since $W(0) \lesssim Y(0) = 1$, $W_x(0)$ has the interpretation of the relative decline in wages around the optimal assignment. This interpretation carries over to $\overline{\omega}_1$ when we estimate the model in logs. Secondly, we apply the model in the context of an hierarchical model, where the wage in the optimal assignment depends on the worker's skill level s. In that context, we cannot simultaneously normalize the wage in the optimal assignment for all skill levels. Estimating the model in logs resolves this problem.

Suppose we estimate equation (18) by simply replacing x by its observed component r. Equation (18) can be viewed of as a first order Taylor expansion of the exact relation between log wages and the mismatch indicator x around $x = 0$. The parameter $\overline{\omega}_1$ could then be interpreted as an estimate of $d \ln W(x)/dx_{|x=0} = W_x(0)/W(0)$. However, even if this Taylor expansion were a perfect approximation of $\ln W(x)$, $\overline{\omega}_1$ will be a biased estimator of $W_x(0)/W(0)$ due to the convexity of the wage function at $x = 0$. We proceed by approximating the distributions of the observed and unobserved components by normal distributions and then check with numerical simulations how good the approximations are. The subsequent lemma is helpful.

Lemma 1 *Assume:* $r \sim N(0, \sigma_r^2)$ *and* $q \sim N(0, \sigma_q^2)$ *and define:*

$$\sigma_x^2 \equiv \sigma_r^2 + \sigma_q^2.$$

Then the following equalities hold:

$$\begin{aligned} E[|r+q|\,|r] &= \sqrt{2\pi}^{-1}\left(2\sigma_q + \frac{r^2}{\sigma_q}\right) + O(r^4), \\ E[|r+q|] &= \sqrt{2\pi}^{-1} 2\sigma_x, \\ |r+q| &= \sqrt{2\pi}^{-1}\sigma_x^{-1}\left(\sigma_x^2 + \sigma_q^2 + r^2\right) + \upsilon, \end{aligned}$$

where υ *in the final line is interpreted as a residual term and where the coefficients in the second line are set to minimize the residual sum of squares:*

$$\int\int \upsilon^2 \psi(r,q)\, dqdr = \int\int \left[|r+q| - \sqrt{2\pi}^{-1}\sigma_x^{-1}\left(\sigma_x^2 + \sigma_q^2 + r^2\right)\right]^2 \psi(r,q)\, dqdr,$$

and $\psi(r,q)$ *is the joint density function of* r *and* q*. By construction:* $E[\upsilon] = 0$.

The derivation of these relations is in the Appendix. The first equation in lemma 1 gives a fourth order Taylor expansion for the conditional expectation of the absolute value of the mismatch indicator $x = r + q$, conditional on its observed component r. The first and the third derivatives of $E[|x|\,|r]$ with respect to r are zero, so that the first and the third order term in the Taylor expansion drop out. The final line gives a least squares approximation of $|x|$ regressed on an intercept and the observed component squared r^2.

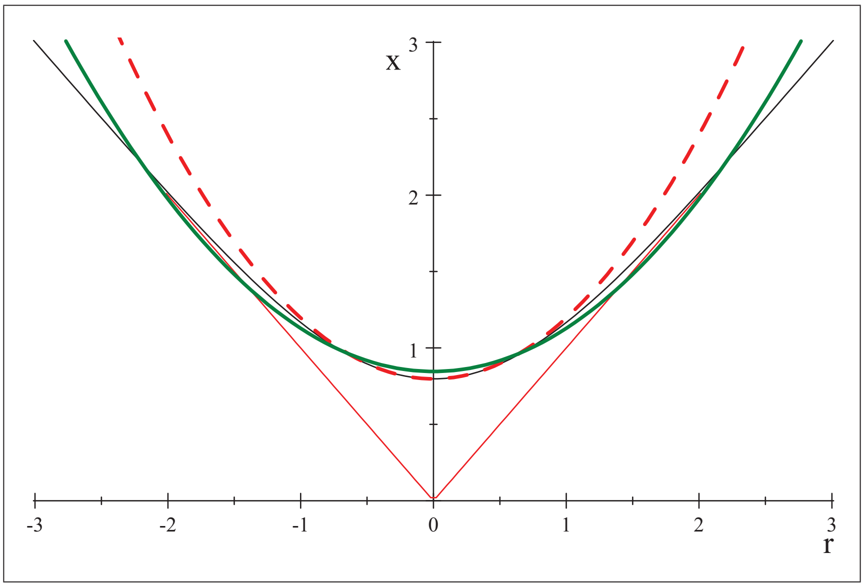

Figure 8: Smoothing of an absolute value function by random mixing for $\sigma_q^2 = \sigma_r^2 = 1$: $|x|$ (thin), $\mathrm{E}[|x|\,|r]$ (thin), Taylor expansion (dashed thick), least squares estimate (thick)

Figure 8 shows the true function $\mathrm{E}[|x|\,|r]$, its Taylor expansion for $\sigma_q^2 = 1$, and its least squares approximation for $\sigma_r^2 = \sigma_q^2 = 1$. The observed component r is not very informative about the value $|x|$ for $r \cong 0$: for $|r|$ varying between 0 and 1, $\mathrm{E}[|x|\,|r]$ varies between 0.80 and 1. Variation of r in the tails of the distribution is much more informative on the actual value of the mismatch indicator; $\lim_{r\to\infty} d\mathrm{E}[|x|\,|r]/dr = 1$ and mutatis mutandis the same for $\lim_{r\to-\infty}$. The Taylor expansion is fairly precise for $|r| < 1$. The precision of the least square approximation is excellent for a much wider range of $|r| < 2.5$.

The results in Lemma 1 and Figure 8 justify the idea of approximating the underlying model (18) by a regression model of w with a quadratic term in r:

$$w = \omega_0 - \omega_2 r^2 + \varepsilon. \tag{19}$$

Under the assumption of joint normality of r and q, the relation between the structural parameter $\overline{\omega}_1$ and the estimated parameters ω_0 and ω_2 can be derived using Lemma 1:

$$\omega_0 = \omega_2 \sigma_r^2 = \overline{\omega}_1 \sqrt{2\pi}^{-1} \sigma_x^{-1} \sigma_r^2, \tag{20}$$

$$\overline{\omega}_1 \mathrm{E}\left[|r+q|\right] = \overline{\omega}_1 \sqrt{2\pi}^{-1} 2\sigma_x = 2\omega_0 \frac{\sigma_x^2}{\sigma_r^2}.$$

The first line establishes a relation between the parameters ω_0 and ω_2 of the regression model (19) and the parameter $\overline{\omega}_1$ of the underlying model (18). The first equality follows from taking expectations in equation (19), using $\mathrm{E}[w] = 0$ and $\mathrm{E}[r^2] = \sigma_r^2$. The second equality is an application of the least square result of Lemma 1 that a least square regression of $|x|$ on an intercept and r^2 yields a coefficient for r^2 equal to $\sqrt{2\pi}^{-1}\sigma_x^{-1}$; a regression on $-\overline{\omega}_1 |x|$ yields therefore a coefficient of $-\omega_2 = -\overline{\omega}_1\sqrt{2\pi}^{-1}\sigma_x^{-1}$. Conditional on the true variance of the mismatch indicator, the coefficient ω_0 is proportional to the variance of x that is observed, σ_r^2/σ_x^2. The second line assesses the expected loss in log wages due to the term $\overline{\omega}_1 |x|$ in the 'true' model in equation (18). The first equality uses the second line in Lemma 1. The second equality uses the first line of (20). The parameter ω_0 underestimates the expected loss by a factor $2\sigma_x^2/\sigma_r^2$. If the mismatch indicator would be perfectly observed, $(\sigma_r^2 = \sigma_x^2)$, ω_0 would still underestimate the expected loss by a factor 2. The reason for this underestimation is that the quadratic approximation of $\mathrm{E}[|r+q|\,|r]$ is not perfect, in particularly not when the variance of q is small, since then r converges to x and since the parabola x^2 is a bad approximation of the absolute value function $|x|$.

The following relations hold:

$$W(0) - \mathrm{E}_x W \lesssim \ln W(0) - \ln \mathrm{E}_x W \lesssim w(0) - \mathrm{E}[w] = \max[w] \gtrsim 2\omega_0. \tag{21}$$

The first inequality follows from the fact that $1 > W(0) > \mathrm{E}_x W$. The second inequality is due to Jensen's inequality, $\ln \mathrm{E}_x W > \mathrm{E}_x \ln W$. The intuition why both inequalities hold approximately with equality is simple. The smaller the wage differentials are (either because u is small or B is close to $Y(0)$), the closer $W(x)$ is to one for all x, so that the approximation $\ln W \lesssim W - 1$ applies. Substituting this approximation in yields both approximate equalities. The third equality follows from the fact that $w(0) = \max[w]$ and $\mathrm{E}[w] = 0$. The final approximation follows from equation (20); the equality holds when the mismatch indicator is fully observed, $\sigma_x^2 = \sigma_r^2$. We propose to use $2\omega_0$ as an estimate of the max-mean wage differential, $W(0) - \mathrm{E}_x W$. Note that the estimate of ω_0 is insensitive to a proportional transformation of r. Suppose that we would use $\widetilde{r}^2 = \zeta r, \zeta > 0$ instead of r^2 as a regressor. That would affect the coefficient ω_2, but not the intercept ω_0. Hence, the proposed measure of wage loss due to search frictions ω_0 is insensitive to the unit of measurement of the observed component r of the mismatch indicator. It is only sensitive to the share of the variance of the mismatch indicator that is actually observed by the researcher.

The above conclusions apply when $\ln W(0) - \ln \mathrm{E}_x W$ is well approximated by $\overline{\omega}_0 - \overline{\omega}_1 |x|$ and when the normal distribution is a good approximation of the true distribution of r and q. However, both conditions do typically not hold. Figure 8 shows that $\overline{\omega}_0 - \overline{\omega}_1 |x|$ is an imperfect approximation of $\ln W(x)$ and Figure 3 shows that the distribution of x is far from normal, and hence, the distributions of its components r and q cannot be jointly normal. We therefore use numerical simulations to evaluate the impact of those violations. We use the benchmark values for the efficiency of on- versus off-the-job search that we used before: $\psi = 1$. We present two results for three values of the unemployment rate u and for four values of the value of non market time B. Table 1 contains some statistics based on the analytical relations derived for the theoretical model, both for the case with and without commitment in wage setting. The first column, $\frac{1}{2}\gamma\sigma_x^2$, is the expected productivity loss due to mismatch, since $\mathrm{E}[Y(0) - Y(x)] = \mathrm{E}\left[\frac{1}{2}\gamma x^2\right] = \frac{1}{2}\gamma\sigma_x^2$. As shown by equation (16), this statistic is insensitive to the units of measurement of x. Moreover, it is the same for the case with and without commitment (keeping constant the unemployment rate), because the pattern of job mobility and the value of γ depend only on the rate of unemployment. Since all wage differentials are proportional to $1 - B$, so is the max-mean differential. The variance of log wages becomes very large for $B = 0$, much larger than is observed empirically even for countries with extreme income inequality and even without controlling for the share of wage dispersion that is accounted for by observable differences in human capital. Hence, just a first look at the data rejects the version of the model where $B \to 0$ and $\psi \to 1$. The reason for this is that for $\psi = 1$ and $B = 0$, the lowest wage $W(\overline{x})$ is equal to zero, and hence the mean-min ratio is $-\infty$. Therefore,, a small probability mass at the bottom of the wage distribution has a disproportional effect on the variance. Stated differently: though wage differentials are proportional to $1 - B$, and hence, the variance of the wage is proportional to $(1 - B)^2$, this proportionality does not carry over to log wage differentials and their variance. Finally, $\max[w]$ is a reasonable estimate for the max-mean wage differential, in particular for the commitment case, but not for $B \leq 0.2$ and $u = 7.5\%$. The final column presents the wage dispersion statistic proposed by Hornstein et.al. (2010), the mean-min wage ratio. Besides that this statistic depends a lot on ψ, it also depends strongly on the value of non market time, B.

Tables 2 and 3 provide information on the final approximation in equation (21). For this analysis, we draw 10,000 values of r and q.[3] We do this for different values of B

[3] Technically we do this by drawing 10,000 values of x from $g(x)$ and by adding to it some measurement error $e \sim \mathrm{N}\left(0, \sigma_x^2\sigma_q^2\sigma_r^{-2}\right)$ to it. We treat $x + e$ as (a monotonic transformation of) the observed

		$\frac{1}{2}\gamma\sigma_x^2$	Var$[w]$		$W(0) - \mathrm{E}_x W$		$\max[w]$		$\frac{\mathrm{E}_x W}{W(\overline{x})}$		X	
commitment		both	yes	no	yes	no	yes	no	yes	no	yes	no
$u(\%)$	B											
2.5	0.0	0.022	0.173	0.238	0.062	0.113	0.103	0.182	∞	∞	0.089	0.182
	0.2	0.018	0.034	0.057	0.050	0.090	0.063	0.122	4.740	4.357	0.071	0.144
	0.4	0.013	0.016	0.021	0.037	0.068	0.046	0.081	2.399	2.261	0.054	0.108
	0.8	0.004	0.001	0.002	0.013	0.023	0.013	0.024	1.234	1.210	0.018	0.036
5.0	0.0	0.041	0.387	0.514	0.111	0.165	0.200	0.326	∞	∞	0.166	0.291
	0.2	0.033	0.066	0.082	0.089	0.132	0.119	0.184	4.510	3.991	0.133	0.232
	0.4	0.024	0.025	0.030	0.067	0.099	0.079	0.123	2.318	2.121	0.100	0.174
	0.8	0.008	0.002	0.002	0.022	0.033	0.024	0.035	1.219	1.186	0.033	0.058
7.5	0.0	0.057	0.608	0.626	0.150	0.197	0.295	0.403	∞	∞	0.234	0.376
	0.2	0.046	0.089	0.104	0.120	0.158	0.168	0.233	4.299	3.381	0.188	0.301
	0.4	0.034	0.034	0.036	0.090	0.118	0.110	0.151	2.242	2.016	0.141	0.226
	0.8	0.012	0.002	0.002	0.030	0.039	0.032	0.042	1.207	1.169	0.047	0.075

Table 1: Numerical values for wage differentials and the variance of log wages

and the signal-noise ratio σ_r^2/σ_q^2 and estimate (20) for both the case with and without commitment. The coefficient ω_0 attains a maximum when the signal to noise ratio is greater than 0.5 (for $u = 2.5\%$ the maximum is at a ratio of 0.5 and when $u = 7.5\%$, the maximum is at a ratio of 2 for commitment and 1 for no commitment). For high values of the signal to noise ratio, the data follow the kink in the wage function tightly, so that the second order Taylor expansion provides a bad approximation of the true function. For low values of the signal to noise ratio, most of the search frictions are not picked up by its observed component. Both effects have a negative impact on the value of ω_0. Table 2 shows that indeed for $\sigma_r^2/\sigma_q^2 = 1/8$, the parameter ω_0 underestimates the max-mean wage differential except for $B = 0$. We took $\sigma_r^2/\sigma_q^2 = 1/8$ because this value is consistent with our estimated ω_0. From Table 3 we see that the log approximations are pretty bad except if the signal-noise ratio is low. In the latter case, downward bias due to the low signal to noise ratio offsets the upward bias due to taking logs. For σ_r^2/σ_q^2 and the more realistic values, $B \geq 0.4$, the estimated value of ω_0 is underestimated by a factor 2 with and a factor 3 without commitment. In general, our approximation works better for the

component r. Since the estimate of ω_0 does not depend on the unit of measurement of r, we do not have to scale $x + e$ back.

case with than without commitment. Our methodology does not allow us to distinguish between the case with and without commitment, since the estimated value of ω_0 is hardly sensitive to this difference (except if $B = 0$ and frictions are large, $u = 7.5\%$).

commitment $u(\%)$	B	yes $W(0) - \mathrm{E}_x W$	ω_0	no $W(0) - \mathrm{E}_x W$	ω_0
2.5	0.0	0.0624	0.0797	0.1128	0.0976
	0.4	0.0374	0.0229	0.0677	0.0239
	0.8	0.0125	0.0059	0.0226	0.0064
5.0	0.0	0.1108	0.0853	0.1645	0.1129
	0.4	0.0665	0.0245	0.0987	0.0252
	0.8	0.0222	0.0066	0.0329	0.0053
7.5	0	0.1503	0.0862	0.1972	0.1242
	0.4	0.0902	0.0271	0.1183	0.0237
	0.8	0.0301	0.0059	0.0394	0.0057

Table 2: Simulated estimates of ω_0 for signal-noise-ratio of 1/8

$u(\%)$	2.5		5.0		7.5	
Commitment?	yes	no	yes	no	yes	no
$W(0) - \mathrm{E}_x W$	0.0374	0.0677	0.0665	0.0987	0.0902	0.1183
$\sigma_r^2/\sigma_q^2 = 1/16$	0.0127	0.0120	0.0140	0.0132	0.0101	0.0126
$\sigma_r^2/\sigma_q^2 = 1/8$	0.0229	0.0239	0.0245	0.0252	0.0271	0.0237
$\sigma_r^2/\sigma_q^2 = 1/4$	0.0328	0.0387	0.0406	0.0381	0.0431	0.0393
$\sigma_r^2/\sigma_q^2 = 1/2$	0.0460	0.0470	0.0551	0.0585	0.0582	0.0555
$\sigma_r^2/\sigma_q^2 = 1$	0.0454	0.0420	0.0610	0.0610	0.0716	0.1284
$\sigma_r^2/\sigma_q^2 = 2$	0.0382	0.0405	0.0592	0.0631	0.0784	0.0758

Table 3: Simulated estimates of ω_0 for B=0.4

3.4 The derivation of the proxies for s and c

The next step in the argument is to construct a proxy for the mismatch indicator r. Our approach is to come up with empirical estimates of workers' skill level s and the level of job complexity c, denoted by $\overline{s}$ and $\overline{c}$ respectively, and to calculate the proxy as the difference between the two, $r \equiv \overline{s} - \overline{c}$. For that purpose, we use a methodology spelled out in Gautier and Teulings (2006) and apply that to data for the United States taken from the March supplements of the CPS 1989-1992, see Gautier and Teulings (2006), for details. First, we must leave the hypothetical circular framework and enter the hierarchical one. Let $Y(s,c)$ denote the productivity of an s-type worker in a c-type job. Let this productivity satisfy the following relation:

$$\ln Y(s,c) = s - \frac{1}{2}\gamma (s-c)^2 . \tag{22}$$

Now, a worker of skill type s has both a comparative advantage in a job-type of the same complexity c as her skill level s and an absolute advantage over other workers with a lower skill, so that he receives a higher wage than these workers when employed in his optimal assignment. The first term captures the absolute advantage of better skilled workers (the hierarchical aspect of the model). The second term captures the match quality. Its specification is equivalent to the circular model in equation (1), in the sense that: $Y''(0)/Y(0) = Y_{cc}(c,c)/Y(c,c) = \gamma$; this parameter is comparable to γ in equation (1). The log supermodularity of $Y(s,c)$ is sufficient for positive assortative matching in a Walrasian equilibrium, see Teulings (1995). The optimal assignment $c(s)$ of worker type s maximizes her output. The first order condition for this problem, $Y_c(s,c) = 0$, implies $c(s) = s$ or $x = 0$. At first sight the linearity of equation (22) in s seems to be a serious limitation to its generality. However, Gautier and Teulings (2006) show that it is not. Since we have not yet defined the units of measurement of s yet, the linearity of the first term is just a matter of proper scaling of the skill index. Hence, the restrictive nature of equation (22) is not in the first but in the second term, namely that the coefficient of the second term, γ, does not vary with s.[4] By a similar argument, the fact that equation (22) is constructed such that the equilibrium assignment is characterized by the simple identity $c(s) = s$ instead of a more general function, is not a restriction to the model, but just a matter of proper scaling of the complexity index c.

The equivalent of the log wage equation (18) that goes with this production function reads:

$$\ln W(s,c) = \ln W(0,0) + s - \omega \cdot |s-c| . \tag{23}$$

Since $c(s) = s$ in the Walrasian equilibrium, c follows the same distribution as s. The mismatch term in equation (23) vanishes for the optimal assignment $c(s) = s$, so that the

[4] Teulings (2005) refers to this case as the constant complexity dispersion equilibrium.

worker's log wage is equal to his skill level: $\ln W(s,c) = \ln W(0,0) + s$. Therefore, we can construct an index of the worker's skill level by running a regression of log wages on individual characteristics. The coefficients of these characteristics measure their contribution to the skill index s. Does the same conclusion apply to the complexity index? At first sight, this is not the case, because the partial derivative of log wages with respect to c is zero in the Walrasian equilibrium, since $c(s) = s$. Hence, s and c are perfectly correlated in that case. However, a similar regression with job instead of worker characteristics does yield an estimate of the contribution of these characteristics to the complexity index c. We construct indices of s and c along these lines by the following procedure:

Normalization: The expectations of w conditional on s and conditional on c among employed workers satisfy:

$$\begin{aligned} \mathrm{E}[w|s] &= s, \\ \mathrm{E}[w|c] &= c. \end{aligned} \tag{24}$$

This normalization is just a convenient way of scaling s and c. It has no empirical content, but it is consistent with equation (22) for the Walrasian equilibrium: (i) $c(s) = s$ and (ii) because there are no search frictions in the Walrasian equilibrium, the zero profit condition implies $Y(s,s) = W(s,s)$; hence: $w = s$. Since s and c are perfectly correlated in the Walrasian equilibrium, it is futile to use the expectation operator in these expressions for this equilibrium, since the distribution of s conditional on c is degenerate, and vice versa. However, in the presence of search frictions, this correlation is no longer perfect and we therefore have to take expectations over the mismatch indicator $x = s - c$. The normalization above implies that we can construct a measure of s and c by running the following regression:

$$\begin{aligned} w &= \vec{x}'\vec{\beta} + \varepsilon_s + \varepsilon_w, \\ w &= \vec{z}'\vec{\alpha} + \varepsilon_c + \varepsilon_w, \end{aligned} \tag{25}$$

where $\vec{x}$ and $\vec{z}$ are vectors of observed worker and job characteristics respectively,[5] where ε_w captures measurement error in log wages, and where ε_s and ε_c capture (i) unobserved

[5] We apply the following personal characteristics: gender, total years of schooling, a third-order polynomial in experience, highest completed education, being married, having a full- or part-time contract as well as various cross terms of experience, education, and being married. As job characteristics, 520 occupation and 242 industry dummies are applied.

characteristics of workers and jobs respectively and (ii) the effect non-optimal assignment on wages. It is convenient to normalize our data on $\vec{x}$ and $\vec{z}$ such that they have zero mean. Since we defined w to have a zero mean too, it does not make sense to include a constant in this regression. The estimated parameter vector can then be used to construct indices for the observed worker and job characteristics,

$$\begin{aligned}\overline{s} &= \vec{x}'\vec{\beta}, \\ \overline{c} &= \vec{z}'\vec{\alpha}.\end{aligned}$$

Again, both indices have zero mean by construction.[6] So the skill measure is the predicted wage conditional on worker characteristics and the job complexity level is the predicted wage conditional on job characteristics. Next, we use these indices and estimate

$$\begin{aligned} w &= \omega_0 + \omega_s \overline{s} + \omega_c \overline{c} - \omega_{ss} \overline{s}^2 + 2\omega_{sc} \overline{sc} - \omega_{cc} \overline{c}^2 + \varepsilon, \qquad (26) \\ w &= \omega_0 + \omega_s \overline{s} + \omega_c \overline{c} - \omega_2 \left(\overline{s} - \overline{c} \right)^2 + \varepsilon. \end{aligned}$$

The second regression imposes two restrictions $\omega_{ss} = \omega_{cc} = \omega_{sc}$. At first sight, this equation seems inadequate to capture model (23). The model includes first order terms for both worker and job characteristics, $\overline{s}$ and $\overline{c}$, where model (23) has only a first order term for the worker's skill s. However, since worker and job characteristics are correlated and since worker characteristics are only partially observed, observed job characteristics will serve as a proxy for unobserved worker characteristics, so that we can expect both ω_s and ω_c to be positive. The problem of establishing the "structural" value of ω_s and ω_c has occupied labor economists ever since the publication of Krueger and Summers' (1998) seminal paper on efficiency wages, see e.g. Bound and Katz (1988) or Abowd, Kramarz, and Margolis (1998). The issue is whether ω_c truly measures the effect of job characteristics, or whether it is merely a proxy for unobserved worker characteristics, see Gautier and Teulings (2006) and Eeckhout and Kircher (2011) for a more elaborate discussion. For now, we adhere to equation (23), which does not allow for a 'true' first order effect of job characteristics. The reason that (26) does not have a term $|x| = |s - c|$ but second order terms for worker and job characteristics, $\overline{s}^2, \overline{sc}$, and $\overline{c}^2$ stems from the fact that the unobserved component of x smooths the non-differentiability of $|x|$ at $x = 0$,

[6] We also normalize $\overline{s}$ and $\overline{c}$ such that in a regression: $w = \beta_1 \overline{s} + \beta_2 \overline{s}^{\,2} + \varepsilon_w$, $\beta_2 = 0$ and the same for $\overline{c}$, see Gautier and Teulings (2006) for details.

as discussed in relation to Lemma 1. Estimating (26) gives,

$$w = \underset{(8.86)}{0.0125} + \underset{(182.4)}{0.61\overline{s}} + \underset{(207.7)}{0.66\overline{c}} - \underset{(21.2)}{0.17\overline{s}^2} - \underset{(21.6)}{0.17\overline{c}^2} + \underset{(36.6)}{0.43\overline{sc}}$$
$$w = \underset{(14.66)}{0.0241} + \underset{(182.2)}{0.61\overline{s}} + \underset{(207.5)}{0.66\overline{c}} - \underset{(35.11)}{0.2007r^2}$$

where as defined before, $r = \overline{s} - \overline{c}$ and $\sigma_r^2 = 0.0306$. The sign restrictions hold for all three second order terms. The F-test[7] rejects the restrictions $\omega_{ss} = \omega_{cc} = \omega_{sc}$ due to the large number of observations that we use but the difference in R^2 between the restricted and the non-restricted model is only 0.0003.[8]

What legitimizes us to interpret the first order term ω_c as capturing unobserved heterogeneity in workers' skills and why can we not interpret the second order terms in the same manner? Why would second order terms really capture the concavity of the wage function in the mismatch indicator x and not unobserved components in either s or c? Gautier and Teulings (2006) argue that while the argument of unobserved heterogeneity applies for ω_c, it is much less likely to apply to the second order terms. They provide three arguments. First, when observed and unobserved worker and job characteristics are distributed jointly normal, it is impossible for second order terms to be a proxy for the unobserved component of a first order term, because the correlation of a second order term in $\overline{s}$ and/or $\overline{c}$ with the unobserved skill index is a third moment, and third moments of a normal distribution are equal to zero. A simple empirical test for this assertion is that including the second order terms should not affect the first order terms, which turns out to be the case. Second, the interpretation of these coefficients as capturing the concavity of the wage function implies sign restrictions, $\omega_{sc} > 0, \omega_{ss} < 0$, and $\omega_{cc} < 0$, which are met for all three coefficients for all six countries for which the model is estimated in Gautier and Teulings (2006). There is no reason why these restriction should hold when the second order terms are to be rationalized from unobserved worker heterogeneity. Moreover $\omega_{sc} = -2\omega_{ss} = -2\omega_{cc}$, which holds approximately. Here, we add a third argument.[9] If the significance of the second order terms is indeed driven by the concavity of the wage function in the mismatch indicator, then their sign would depend on the vectors $\vec{x}$ and $\vec{z}$ which capture worker and job characteristics respectively. If we would compose both

[7] $\frac{\left(R_u^2 - R_{restr}^2\right)/2}{(1-R_u^2)/(222179-6)} = \frac{(0.4479-0.4476)/2}{(1-0.4479)/(222179-6)} = 60.4.$

[8] Actually, the size restrictions are not exactly correct. The exact restrictions account for differences in the degree of observability in s and c, see Gautier and Teulings (2006) for details.

[9] We thank Jean Marc Robin for the idea of this test.

vectors out of mixtures of job and worker characteristics, e.g. experience and occupation dummies in $\vec{x}$ and education and industry dummies in $\vec{z}$, then the concavity result should not come out. Table 4 reports the results for this test. For both alternative combinations, putting education and occupation in s or putting education and industry in s (and the remaining variables in c), the concavity result becomes either much weaker or even breaks down. Hence, the concavity result only survives when worker and job characteristics are separated in two variables and it is not a statistical artifact.

$\overline{s}$ includes:	$\overline{s}$	$\overline{c}$	$\overline{s}^2$	$\overline{c}$	$\overline{sc}$
educ, occ	0.517	0.650	-0.009	-0.037	0.094
(t)	(132.6)	(175.24)	(-0.86)	(-4.10)	6.17)
educ, ind	0.324	0.805	-0.050	0.010	0.053
(t)	(79.9)	(233.2)	(-4.44)	(1.10)	(3.46)

Table 4: Test of concave relation between wages and s and c

3.5 Deriving the share of the variance of x that is observed

The regression (26) implies $\omega_0 = 0.0241$. The results presented in table 3 imply that conditional on the assumption that the value of non market time is 0.4, the signal-to-noise ratio is about 1/8. Let ρ_s^2 be the ratio of the variance of $\overline{s}$ to the variance of $\overline{s} + \varepsilon_s$, and mutatis mutandis the same for ρ_c^2, and let σ^2 be the variance of $\overline{s} + \varepsilon_s$, which is by construction equal to the variance of $\overline{c} + \varepsilon_c$. For the CPS data, we find $\rho_s^2 = 0.395, \rho_c^2 = 0.427$, and $\sigma^2 = 0.402$.[10] As discussed before, the perfect correlation between s and c breaks down in the presence of search frictions. Let ρ be the correlation between s and c. Before we can calculate $\text{Var}[\overline{s} - \overline{c}]$, we must first make the following assumption on their covariance.[11]

[10] We use Bound and Krueger's (1991) finding that the signal noise ratio for wages is about 0.85. The R^2 for the regressions (25) are 0.336 and 0.363 respectively and $\text{Var}[w]$ is 0.342. Dividing those numbers by 0.85 yields the results for ρ_s^2 and ρ_c^2 anf σ^2.

[11] This assumption is equivalent to the following assumption on the covariance matrix:

$$\text{Var}\left[\begin{array}{cccc} s & c & \overline{s} & \overline{c} \end{array}\right] = \sigma^2 \begin{pmatrix} 1 & \rho & \rho_s^2 & \rho\rho_c^2 \\ & 1 & \rho\rho_s^2 & \rho_c^2 \\ & & \rho_s^2 & \rho\rho_s^2\rho_c^2 \\ & & & \rho_c^2 \end{pmatrix}$$

Assumption: $\text{Cov}[\overline{s}, \overline{c}] = \rho\rho_s^2\rho_c^2$.

This assumption implies that the covariance between the observed part of worker characteristics $\overline{s}$ and c (which is $\rho\rho_s^2$) is distributed proportionally between the observed job characteristics $\overline{c}$ and the unobserved component, ε_s. Since the former accounts for a fraction ρ_c^2 in the variance of c, this yields the expression in the assumption. Therefore,

$$\sigma_x^2 \equiv \sigma_{(s-c)}^2 = \sigma^2 + \sigma^2 - 2\rho\sigma^2 = 2(1-\rho)\sigma^2 \tag{27}$$

$$\sigma_{(\overline{s}-\overline{c})}^2 \equiv \sigma_r^2 = \rho_s^2\sigma^2 + \rho_c^2\sigma^2 - 2\rho\rho_s\rho_c\sigma^2 = \left(\rho_s^2 + \rho_c^2 - 2\rho\rho_s^2\rho_c^2\right)\sigma^2$$

In the limit close to the Walrasian equilibrium, $\rho \to 1$, $\sigma_x^2 \to 0$, while $\sigma_r^2 \to (\rho_s^2 + \rho_c^2 - 2\rho_s^2\rho_c^2)\,\sigma^2 > 0$. The reason is that even if s and c are perfectly correlated, as is the case in the Walrasian equilibrium, their observed parts, $\overline{s}$ and $\overline{c}$, are not perfectly correlated because both are contaminated by measurement error. This problem seriously complicates inference on mismatch. Do we observe a worker not assigned to her optimal job type because (i) we mismeasure her skills, (ii) we mismeasure the complexity of the job, or because (iii) true mismatch? The signal-noise ratio can be approximated by:[12]

$$\frac{\sigma_x^2}{\sigma_r^2 - \sigma_x^2} = \frac{2(1-\rho)}{\rho_s^2 + \rho_c^2 - 2\rho\rho_s^2\rho_c^2 - 2(1-\rho)}. \tag{28}$$

Substituting in a signal-to-noise ratio of 1/8 and the values of ρ_s^2 and ρ_c^2 yields $\rho = 0.973$, so[13]

$$\sigma_x^2 = 2 \times 0.027 \times 0.342 = 0.018.$$

An expression for $\frac{1}{2}\gamma\sigma_x^2$ follows directly from the model, see Table 1. The information that is missing is a value for γ that goes with the dimension of x used above. This is the final step in our procedure.

3.6 Deriving γ

For the derivation of a value for γ, we use the fact that the interpretation of the circular model as a simplified representation of an hierarchical assignment model allows us to

[12] The denominator derives the variation of the noise as the difference between the true signal x and its observed value r. Strictly speaking, application of the model signal-measurement error requires that $s-c$ is independent of $\varepsilon_s - \varepsilon_c$, which is not the case due to regression to the mean: the mismatch indicator $x \equiv s - c$ cannot be simultaneously independent of s and c, as it is in the ciruclar model. This problem is closely related to the corner problem discussed in Teulings and Gautier (2004). As long as $\sigma_x^2 \ll \sigma^2$, this problem can be ignored.

[13] $\frac{1}{8} = \frac{2(1-\rho)}{0.395+0.427-2(\rho\times0.395\times0.427+1-\rho)}$. Since ρ is close to unity, indeed: $\sigma_x^2 \ll \sigma^2$, see the previous footnote.

draw an analogy to the theory of imperfect substitution between low and high skilled workers, see Teulings (2005) and Teulings and Van Rens (2008). In fact, γ measures the curvature of Rosen's "kissing" offer and utility curves, see Figure 2. We can establish a relation between our estimation results and the elasticity of substitution between low and high skilled workers as estimated by Katz and Murphy (1992). In the specification of the model considered so far, the distribution of worker-skills and job-complexities was kept constant, and we normalized the definition of the skill index s and complexity index c such that the optimal assignment was given by $c(s) = s$. In this section, the optimal assignment depends on the supply and demand for skill types. We follow the derivation in Teulings and Van Rens (2008). First, we characterize the Walrasian equilibrium of this assignment problem. In that equilibrium, every s-type worker is employed in her optimal assignment $c(s)$. Next, we do comparative statics to analyze the effect on relative wages of a shift in the mean of the skill distribution.

Due to the assumption of comparative advantage of skilled workers in complex jobs, the optimal assignment is an increasing function, $c'(s) > 0$. Let Y^* be aggregate output per worker. We assume that this output is produced by a Leontieff technology, requiring the input of all c-type jobs in fixed proportions. Let $h(c)$ be the density of the input of a c-type job required to produce one unit of aggregate output. Equilibrium on the commodity market for job type $c(s)$ requires the equality of supply and demand for each s-type:

$$Y^* h[c(s)] = g(s) Y[s, c(s)] / c'(s), \tag{29}$$

where $g(s)$ is the skill density function. The left hand side is the demand for the output of job type $c(s)$; it is equal to aggregate output Y^* times the density of job type $c(s)$ required per unit of aggregate output, $h[c(s)]$. The right hand side is the supply of output of job type $c(s)$; it is equal to the density of worker type s, times its productivity in job type $c(s)$, $Y[s, c(s)]$ times the Jacobian $ds/dc = 1/c'(s)$. We assume s and c to be distributed normally with mean μ_s and μ_c respectively and identical standard deviations $\sigma_s = \sigma_c = \sigma$. Without loss of generality, we normalize $\mu_c = \sigma^2$; the only thing that matters in this model turns out to be the difference between μ_s and μ_c. Taking logs in equation (29) and using the density function for $h(\cdot)$ and $g(\cdot)$ yields,

$$\ln Y^* - \frac{1}{2}\left(\frac{c(s) - \sigma^2}{\sigma}\right)^2 = -\frac{1}{2}\left(\frac{s - \mu_s}{\sigma}\right)^2 + s - \frac{1}{2}\gamma [s - c(s)]^2 - \ln c'(s). \tag{30}$$

This equation should hold identically for all s.

Let $\ln W(s)$ be the log wage for worker type s in equilibrium. The zero profit condition implies $W(s) = Y[s, c(s)]$. Firms offering jobs of type c choose their preferred worker type s as to maximize profits, or equivalently, to minimize the log of the cost of production per unit of output, $\ln W(s) - \ln Y(s, c)$. Hence, the equilibrium wage function $W(s)$ satisfies the first order condition

$$\frac{d \ln W(s)}{ds} = \frac{d \ln Y(s, c)}{ds}|_{c=c(s)} = 1 - \gamma [s - c(s)], \tag{31}$$

where we use equation (22) in the second equality. The system of equations (30) and (31) is solved by the following expressions for $c(s)$ and $W(s)$:

$$c(s) = s - \mu_s,$$
$$\frac{d \ln W(s)}{ds} = 1 - \gamma \mu_s.$$

The wage function $\ln W(s)$ is linear in s. $d \ln W(s)/ds = 1 - \gamma \mu_s$ is the return to the human capital index s. This return depends on the supply of human capital, that is, on the mean of the skill distribution. The equilibrium assignment of section 3.4, $c(s) = s$, implies that $\mu_s = 0$. In that case $d \ln W(s)/ds = 1$, as is implied by equation (22). A percentage point upward shift in the mean of the skill distribution, μ_s, reduces the return to human capital by γ % point. Hence, γ is related to the inverse of the elasticity of substitution between high and low skilled workers as estimated by Katz and Murphy's (1992). That relationship can be analyzed more formally. Katz and Murphy split labour into two skill groups, low and high, and consider the effect on relative wages of a shift in labour supply from the one to the other. Let s^* be the cut off level. All worker types with $s > s^*$ are classified as high skilled; all other workers as low skilled. Hence, $\Phi[\sigma^{-1}(s^* - \mu_s)]$ is the share of low skilled workers and $\Phi[-\sigma^{-1}(s^* - \mu_s)]$ is the share of high skilled workers, where $\Phi(\cdot)$ denotes the standard normal distribution function. Katz and Murphy estimate the elasticity η, which is the ratio of the change in relative supply of high and low skilled workers to the change in relative wages.

$$\eta \equiv -\frac{d\left(\ln \Phi[\sigma^{-1}(s^* - \mu_s)] - \ln \Phi[-\sigma^{-1}(s^* - \mu_s)]\right)/d\mu_s}{d\left(\mathrm{E}[\ln W(s) | s > s^*] - \mathrm{E}[\ln W(s) | s < s^*]\right)/d\mu_s} = \frac{1}{\sigma^2 \gamma}. \tag{32}$$

The derivation of the final step is in the Appendix. Katz and Murphy estimate η to be 1.4. Hence:

$$\gamma = \frac{1}{\mathrm{Var}[w]\, \eta_{\text{low-high}}} \cong \frac{1}{0.36 \times 1.4} \cong 2,$$

where we use σ^2 =Var$[w] \cong 0.36$. Teulings and Van Rens (2008) estimate directly the relation between mean years of education and the return to human capital using panel data for some 100 countries during the postwar period. They find a similar value for the compression elasticity. Using $\gamma = 2$ we find:

$$\frac{1}{2}\gamma\sigma_x^2 = 0.019.$$

For $B = 0.4$, this value squares with a value of the unemployment rate of 3.8%, see Table 1. This is in line with the observed value of the unemployment rate and corresponds to an output loss X of 7.8% if firms can commit to wage payments and 14.5% if firms cannot. If firms cannot commit, there is too little competition for workers, which leads to too high quasi rents and consequently to excessive vacancy creation. In other words, an unemployment rate of 3.8% with many vacancies implies a larger output loss than the same unemployment rate with few vacancies. Without commitment, this unemployment rate corresponds to a max-mean wage differential of 8.6% and a mean-min ratio of 2.18, and with commitment, it corresponds to to a max-mean differential of 5.3% and a mean-min ratio of 2.36. These mean-min ratio's are in line with the estimates of Hornstein et.al. (2010). Our model can therefore simultaneously explain the data on unemployment and on wage dispersion among workers with the same human capital.

3.7 The impact of variation in ψ

What is the impact of variations in ψ on our results? First, if we change ψ, we must recalibrate the model. Specifically, the (approximate) value of the signal-noise ratio σ_r^2/σ_q^2 that is consistent with our estimate of $\omega_0 = 0.024$ depends on ψ. For the sake of comparability, we restrict this value of the signal-noise ratio to be the same for wage setting with and without commitment. This only implies that the third decimal of ω_0 cannot always be matched. Next, we calculate the value of ρ that corresponds to this signal-noise ratio, using equation (28), and given this ρ, we calculate the expected productivity loss due to suboptimal assignment, $\frac{1}{2}\gamma\sigma_x^2$, using equation (27). Table 5 shows the results. We can use the values of u that are implied by the observed wage dispersion and mismatch as a test for the performance of our model. A low ψ makes workers more choosy by increasing their reservation wage and therefore unemployment will be higher. It also implies that workers move slower to their optimal job type making the cross-sectional mass of workers around the optimal job type smaller.

ψ	ω_0		σ_r^2/σ_q^2	ρ	$\frac{1}{2}\gamma\sigma_x^2$	$u(\%)$
commitment	yes	no	both	both	both	both
1/2	0.027	0.024	3/4	0.888	0.077	5.8
3/4	0.025	0.022	1/4	0.950	0.034	4.7
1	0.025	0.025	1/8	0.973	0.019	3.8

Table 5: Frictional unemployment for different values of ψ for B=0.4

ψ	$\frac{\{e\to e\}}{\{u\to e\}}$	$W(0) - \mathrm{E}_x W$		$\frac{\mathrm{E}_x W}{W(\overline{x})}$		X	
commitment	both	yes	no	yes	no	yes	no
1/2	1.483	0.076	0.089	1.49	2.10	0.122	0.185
3/4	1.969	0.064	0.089	1.82	2.15	0.097	0.163
1	2.399	0.053	0.086	2.36	2.18	0.078	0.145

Table 6: Key statistics for different values of ψ for B=0.4

Table 6 shows the ratio of job-to-job versus unemployment-to-job worker flows, wage dispersion (max-mean differential and mean-min ratio) and the output loss for various values of ψ. Nagypal (2008) estimates the ratio of job-to-job versus unemployment-to-job worker flows to be $2.2/0.91 = 2.4$ in the US.[14] However, Hornstein et.al. (2010) report a larger value of the unemployment-to-job flow based on Shimer (2005) and get a value of the ratio of both flows close to one.[15] Using tenure data to calculate the job-to-job flow also leads to lower estimates (29% lower according to Nagypal, 2008, implying a ratio of 1.7). Koning, van den Berg and Ridder (1995) structurally estimate ψ with Dutch data and find a value close to one. One way to reconcile those facts is to acknowledge that the distinction between layoffs and quits is blurry. Hence, the available evidence does not allow us to pin down the value of ψ exactly. Moreover, all values of ψ between 0.5 and 1 imply reasonable values of the unemployment rate and the amount of wage dispersion. Our estimates of X based on the benchmark case with $\psi = 1$ is therefore a lower bound. If we accept ψ to be between 0.5 and 1, then if firms can commit to wages, the output loss is between 7.8% and 12.2%, while if firms cannot, it is between 14.5% and 18.5%. The difference in output loss between both wage-determination processes is due to a business stealing externality. For $\psi = 1$, wage setting with commitment yields

[14] Actually, she uses the job-to-job versus the job-to-unemployment ratio, but in steady state both are approximately equal.

[15] They report, $(\{u \to e\}/u)/(\{e \to e\}/e) = 0.43/0.027$. If unemployment is 6% then $\{u \to e\}/\{e \to e\} = 1.02$ and $\psi \simeq 0.2$.

constrained efficiency, see Gautier et.al. (2010). The idea is that without commitment, when opening a vacancy, individual firms do not internalize the future output loss of the firm they will poach a worker from. Although the transitions of workers to better matches are always efficient, the expected productivity gains are too small to justify the entry cost of the marginal firm from a social point of view. For $\psi = 1$, the magnitude of this business-stealing externality is equal to the difference in X with and without commitment.[16] Table 6 shows this difference to be about 6%.

Table 7 shows that the lower is ψ, the larger is the fraction of X that is due to mismatch. Again, there are two effects of decreasing ψ, (i) the reservation wage goes down and (ii) workers move slower to their optimal job type. The first effect reduces the expected mismatch, while the second increases it. It turns out that the second effect dominates. The cost of unemployment makes up for only a small part of the total output loss, in particular if firms cannot ex ante commit to theit posted wages.

ψ	1/2		3/4		1	
Commitment	yes	no	yes	no	yes	no
$u(1-B)$	3.5	3.5	2.8	2.8	2.3	2.3
vK	1.4	7.7	3.7	10.3	2.6	9.3
$(1-u)\frac{1}{2}\gamma\sigma_x^2$	7.3	7.3	3.2	3.2	2.9	2.9
X	12.2	18.5	9.7	16.3	7.8	14.5

Table 7: Decomposition of output loss due to frictions for B=0.4

3.8 Negative or positive assortative matching?

A recent paper by Eeckhout and Kircher (2010) argues that wage data allow us to identify the strength of sorting but not whether there is positive or negative assortative matching (PAM or NAM). Does the methodology outlined in this paper allow us to discriminate empirically between NAM and PAM? As a first observation, equation (22) imposes positive assortative matching by the second derivative of s being negative and the cross partial of s and c being positive. An optimizing worker chooses the job type that satisfies $Y_c(s, c) = 0$. The sign restrictions on the second derivatives impose positive assortative matching. However, by just replacing c by $-c$, one obtains negative assortative matching. Since the

[16] Elliot (2006) discusses other wage mechanisms in a network framework that also internalize the business-stealing externality.

metric of c is undefined, this statement does not have any empirical content. The reason is that by allowing the worker to choose the job type that maximizes her output, the partial derivative with respect to c is zero in the equilibrium assignment. Then, the distinction between high and low productive jobs is meaningless, since for a marginal change in c, net output will be the same, while for larger variations, output will be lower, irrespective of whether it is a more or a less complex job than the observed optimal assignment. The crucial issue is that free entry of firms of any c type implies that if sufficiently many workers prefer to be employed in a particular c type job, those vacancies will be opened. This conclusion can also be inverted: if a model allows for free entry for types on either the worker or the firm side of the market, the issue of positive versus negative assortative matching looses its economic relevance. In absolute-advantage models like Shimer and Smith (2000) and Eeckhout and Kircher (2011), the reason that there exists an interior optimal job type is that there are two opposite forces at work. More complex job types generate more output but they also have a better outside option and therefore require a larger share of the output. The only way to support free entry in that framework is to have higher entry cost for more complex jobs. In that case, we still cannot rank jobs in terms of their *net* profitability (output minus entry cost).

4 Conclusion

This paper showed that within a general search model with on-the-job search and sorting there exists a simple relation between three statistics: the unemployment rate, the value of non market time, and the max-mean wage differential. This relationship does not depend on any parameter, except for the efficiency of on- relative to off-the-job search. However, we show that this dependency is a higher order phenomenon and that the exact value hardly matters for the relation between unemployment and wage dispersion. In addition, we derive the ratio of job-to-job and unemployment-to-employment flows that is implied by the model. Given that the model can jointly explain an unemployment rate of around 5%, the amount of wage-dispersion we observe in the data and the observed labor-market flows, we feel confident to use it to estimate the output loss due to search frictions.

Search frictions directly generate output losses due to the fact that resources are allocated sub optimally and indirectly because decentralized wage mechanisms potentially come with distortions. Allowing for two-sided heterogeneity is extremely important because it is the interaction between the search frictions, the type distributions and the production technology that determines how important these frictions are. If workers are identical and firms are identical then all contacts result in a match. Under two-sided heterogeneity, the production technology matters because it specifies how much output is lost when a given job type is occupied by a sub-optimal worker type. Search frictions generate a lot of output loss if a precize match is very important while if worker types are almost perfect substitutes, the output loss will be will be modest. By combining information on wage dispersion and the substitutability of worker types we can learn about the actual amount of frictions and the importance of a precize match. We then use our model to quantify and decompose this total output loss. As a test for the performance of the model, we calculate the level of unemployment that is consistent with the frictions that are implied by the observed wage dispersion. Depending on the relative efficiency of on-the-job search, we find that the rate of unemployment that our model implies is between 3.5% and 5.5% which is a reasonable range. For an unemployment rate of 5.5%, the total output loss is 12.2% if firms can commit to their posted wages and 18.5% if they cannot. In the latter case, there is excess vacancy creation due to a busines-stealing externality. If firms can commit to wages, 60% of the output loss is due to sub-optimal assignment, 11% is due to vacancy creation and 29% is due to unemployment. Traditionally, most of the macro labor literature focussed on unemployment but those results imply that mismatch is at least as important from an efficiency point of view.

Other contributions of our paper are that we show that the max-mean wage differential is a more robust measure for wage dispersion than measures based on the reservation wage because workers move towards the best jobs so the density around the highest wage is a lot higher than around the lowest wage. We also discuss a simple and tractable method for estimating the size of wage differentials allowing for measurement error. Finally, we apply the theory of imperfect substitution between high and low-skilled workers. This yields a relation between Katz and Murphy's (1992) elasticity of substitution between high and low-skilled workers and the second derivative of the production function of our model. This relation allows us to calculate the total output loss due to mismatch and the corresponding unemployment rate that our model implies, for a given value of non market time.

References

[1] Abowd, J.M., F. Kramarz, and D.N. Margolis (1998), High wage workers and high wage firms, *Econometrica*, 67, 251-333.

[2] Atakan, A. (2006), Assortative Matching with Explicit Search Costs, *Econometrica* 74, 667–680.

[3] Becker, G. S. (1973), A Theory of Marriage: Part I, *Journal of Political Economy* 81(4), 813-46.

[4] Bontemps, C. J.M. Robin and G.J. Van den Berg (2000), Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation, *International Economic Review*, 41, 305-58.

[5] 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*, 9, 1-24.

[6] Burdett K. and D. Mortensen (1998), Equilibrium wage differentials and employer size, *International Economic Review*, 39, 257-274.

[7] Coles M. (2001), Equilibrium Wage Dispersion, Firm Size, and Growth, *Review of Economic Dynamics*, 4-1,159-187.

[8] Dickens, W. and L.F. Katz (1987). Inter-Industry wage differences and theories of wage determination, *NBER* working paper No. 2271.

[9] Eeckhout, J. and P. Kircher (2011). Identifying sorting, in theory, *Review of Economic Studies.* Forthcoming.

[10] Eeckhout, J. and P. Kircher (2010). Sorting and decentralized price competition, *Econometrica 78-2, 539-574.*

[11] Elliott, M. (2010). Inefficiencies in networked markets, mimeo Stanford.

[12] Fallick, B. and C. Fleischman (2004), The importance of employer-to-employer flows in the U.S. labor market, working paper, Federal Reserve Bank Board of Governors, Washington DC.

[13] Gautier, P.A and C.N. Teulings (2006), How large are search frictions? *Journal of the European Economic Association*, 4-6, 1193-1225.

[14] Gautier, P.A, C.N. Teulings and A.P. Van Vuuren (2005), Labor market search with two sided heterogeneity: hierarchical versus circular models, *in:* H. Bunzel, B.J. Christensen, G.R. Neumann, and J.M. Robin, eds., *Structural Models of Wage and Employment Dynamics, conference volume in honor of Dale Mortensen*, Elsevier, Amsterdam.

[15] Gautier, P.A, C.N. Teulings and A.P. Van Vuuren (2008), On the job search and sorting, Tinbergen Institute discussion paper 070/3.

[16] Hall, R.E. (2009), Reconciling Cyclical Movements in the Marginal Value of Time and the Marginal Product of Labor, *Journal of Political Economy*, 117, 281-322.

[17] Hagedorn J. and I. Manovskii (2008), The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited, *American Economic Review*, 98, 1692-1706.

[18] Hornstein, A., P. Krusell and G.L. Violante (2010), Frictional wage dispersion in search models: a quantitative assessment, mimeo, Federal Bank of Richmond.

[19] Katz, L.F. and K.M. Murphy (1992), Changes in relative wages, 1963-1987: supply and demand factors, *Quarterly Journal of Economics*, 107, 35-78.

[20] Koning, P., G. Ridder, and G. J. Van den Berg (1995), "Structural and frictional unemployment in an equilibrium search model with heterogeneous agents." *Journal of Applied Econometrics* 10, 133-151.

[21] Lise, J., C. Meghir, and J.M. Robin (2008), “Matching, Sorting and Wages ”, UCL mimeo, 2008.

[22] Lopes de Melo (2008), “Sorting in the Labor Market: Theory and Measurement”.

[23] Marimon R. and F. Zilibotti (1999), Unemployment vs. mismatch of talents: reconsidering unemployment benefits, *Economic Journal*, 109, 266-291.

[24] Mortensen, D.T. (2000), Equilibrium unemployment with wage posting: "Burdett-Mortensen meets Pissarides", in B.J. Christensen, P. Jensen, N. Kiefer and

D.T. Mortensen, eds., *Panel data and structural labor market models*, Amsterdam, Elsevier science.

[25] Mortensen, D.T. and C. Pissarides (1999), New developments in models of search in the labor market, *in:* O.C. Ashenfelter and D. Card, *Handbook of Labor Economics* 3b, North-Holland, Amsterdam.

[26] Nagypal E. (2005), On the extent of job-to-job transitions, working paper, Northwestern University, Evanston.

[27] Pissarides, C.A. (1994), Search unemployment with on-the-job search, *Review of Economic Studies*, 61, 457-475.

[28] Pissarides, C.A. (2000), *Equilibrium unemployment theory*, 2nd edition, MIT Press, Cambridge.

[29] Postel-Vinay, F. and J.M. Robin (2002), Equilibrium wage dispersion with worker and employer heterogeneity, *Econometrica*, 70, 2295–2350.

[30] Rosen, S. (1974), Hedonic prices and implicit markets: product differentiation in pure competition", *Journal of Political Economy* , 82, 34-55.

[31] Sattinger, M. (1975), Comparative advantage and the distribution of earnings and abilities, *Econometrica*, 43, 455-68.

[32] Shimer, R. (2001), The impact of young workers on the aggregate labor market, *Quarterly Journal of Economics*, 116, 969-1007.

[33] Shimer, R. (2006), On-the-job Search and Strategic Bargaining, *in:* H. Bunzel, B.J. Christensen, G.R. Neumann, and J.M. Robin, eds., *Structural Models of Wage and Employment Dynamics, conference volume in honor of Dale Mortensen*, Elsevier, Amsterdam.

[34] Shimer, R. and L. Smith (2000), Assortative matching and search", *Econometrica*, 68, 343-369.

[35] Teulings, C.N. (1995), The wage distribution in a model of the assignment of skills to jobs, *Journal of Political Economy*, 103, 280-315.

[36] Teulings, C.N. (2005), Comparative advantage, relative wages, and the accumulation of human capital, *Journal of Political Economy*, 113, 425-461.

[37] Teulings, C.N. and P.A. Gautier (2004), The right man for the job, *Review of Economic Studies*, 71, 553-580.

[38] Teulings, C.N. and T. Van Rens (2008), Education, growth, and income inequality, *Review of Economics and Statistics*, 90, 89-104.

[39] Van den Berg, G. and G. Ridder (1998), An empirical equilibrium search model of the labor market, *Econometrica*, 66, 1183-1221.

Appendix

A Derivations and proofs

A.1 Derivation of the asset values

The Bellman equation for the asset value of employment reads

$$\rho V^E(x) = W(x) + 2\psi\lambda \int_0^x \left[V^E(z) - V^E(x)\right] dz - \delta \left[V^E(x) - V^U\right]. \tag{33}$$

Totally differentiating (33) yields

$$V_x^E(x) = \frac{W_x(x)}{\rho + \delta + 2\psi\lambda x}. \tag{34}$$

The solution to this differential equation is

$$V^E(x) = \int_0^x \frac{W_x(z)}{\rho + \delta + 2\psi\lambda z} dz + C_0.$$

Integrating by parts yields

$$V^E(x) = \frac{W(x)}{\rho + \delta + 2\psi\lambda x} - \frac{W(0)}{\rho + \delta} + 2\lambda\psi \int_0^x \frac{W(z)}{(\rho + \delta + 2\psi\lambda z)^2} dz + C_0. \tag{35}$$

Evaluating (35) at $x = 0$ gives an initial condition that can be used to solve for C_0

$$C_0 = V^E(0) = \frac{W(0)}{\rho + \delta} + \frac{\delta}{\rho + \delta} V^U.$$

Substitution of this equation into (35) yields the desired expression. Let $\mathrm{E}_x W \equiv \int_0^z g(x) W(x)\, dx$ be the expected wage of a filled job. Evaluate (3) at z and use the definition of g in (12) to get

$$\rho V^E(z) = \rho V^U = \frac{W(z) + \psi\kappa z \mathrm{E}_x W}{1+\psi\kappa z} = \frac{uW(z) + \psi(1-u)\mathrm{E}_x W}{u + \psi(1-u)}. \tag{36}$$

Next, note that the right-hand side of (33) and (36) are equal which can be used to get an expression for $\int_0^z \left[V^E(x) - V^U\right] dx$. Substitution of this expression into (2) gives ρV^U as a function of $W(z)$, which can be eliminated by solving (36) for $W(z)$. This gives

$$\rho V^U = \frac{B + \kappa\widehat{x}\mathrm{E}_x W}{1+\kappa\widehat{x}} = uB + (1-u)\mathrm{E}_x W, \tag{37}$$

where the final step uses (11).

The free entry condition implies that the option value of a vacancy of type c must be equal to K. Hence, by defining $\mathrm{E}_G Y \equiv \int_0^{\overline{x}} g(x) Y(x)\, dx$, we obtain:

$$\begin{aligned} K &= 2\lambda \int_0^{\overline{x}} \{u + \psi(1-u)[1-G(x)]\} \frac{Y(x) - W(x)}{\rho + \delta + 2\psi\lambda v x} dx \\ &= \frac{1-u}{v}\left(\mathrm{E}_G Y - \mathrm{E}_G W\right). \end{aligned}$$

The first term in the integrand is the effective labor supply, $u + \psi(1-u)[1-G(x)]$ for a vacancy of type x. It is equal to the number of unemployed, u plus the number of workers employed in jobs with a mismatch indicator that exceeds x, $(1-u)[1-G(x)]$. The second factor is the discounted value of a filled vacancy. Just as in the wage equation, we discount current revenue $Y(x) - W(x)$ by the discount rate ρ plus the separation rate δ plus the quit rate $2\psi\lambda v x$. The second line follows from substituting (11) and (12) in.

A.2 Variance of $\widetilde{x}$

$$\begin{aligned} \mathrm{Var}[\widehat{x}] &= \int_0^{\widehat{x}} \widetilde{x}^2 g(\widetilde{x})\, d\widetilde{x} = \frac{1+\psi\widehat{x}}{\widehat{x}} \int_0^{\widehat{x}} \frac{\widetilde{x}^2}{(1+\psi\widetilde{x})^2} d\widetilde{x} = \frac{1+\psi\widehat{x}}{\psi^3\widehat{x}} \int_0^{\widehat{x}} \frac{y^2}{(1+y)^2} dy \\ &= \frac{1+\psi\widehat{x}}{\psi^3\widehat{x}} \left(\widehat{x}\frac{\widehat{x}+2}{\widehat{x}+1} - 2\ln(\widehat{x}+1)\right). \end{aligned}$$

A.3 Job flows

For the sake of convenience, we apply the version where κ is normalized to unity.

$$
\begin{aligned}
\{u \to e\} &= 2\lambda (1-u), \\
\{e \to e\} &= 2\int_0^{\widehat{x}} \psi\lambda (1-u)\, g(\widetilde{x})\, \widetilde{x} d\widetilde{x} = 2\psi\lambda (1-u) \frac{(1+\psi\widehat{x})}{\psi^2 \widehat{x}} \int_0^{\psi\widehat{x}} \frac{q}{(1+q)^2} dq \\
&= 2\lambda (1-u) \left(\frac{1+\psi\widehat{x}}{\psi\widehat{x}} \ln(\psi\widehat{x}+1) - 1 \right), \\
\frac{\{e \to e\}}{\{u \to e\}} &= \frac{1+\psi\widehat{x}}{\psi\widehat{x}} \ln(\psi\widehat{x}+1) - 1.
\end{aligned}
$$

A.4 Wages and expected wages

commitment:

$$
\begin{aligned}
W(\widetilde{x}) &= 1 - \widetilde{\gamma} \left[-\left(\frac{1+\psi\widetilde{x}}{\psi} \right)^2 \log \left(\frac{1+\psi\widetilde{x}}{1+\psi\widehat{x}} \right) - \frac{\widehat{x}}{\psi} \frac{(1+\psi\widetilde{x})^2}{1+\psi\widehat{x}} + \frac{\widetilde{x}}{\psi} + \frac{3}{2}\widetilde{x}^2 \right], \\
W_x(\widetilde{x}) &= -2\widetilde{\gamma} \frac{\psi}{1+\psi\widetilde{x}} \left[-\left(\frac{1+\psi\widetilde{x}}{\psi} \right)^2 \ln \left(\frac{1+\psi\widetilde{x}}{1+\psi\widehat{x}} \right) - \frac{\widehat{x}}{\psi} \frac{(1+\psi\widetilde{x})^2}{1+\psi\widehat{x}} + \frac{\widetilde{x}}{\psi} + \widetilde{x}^2 \right], \\
\mathrm{E}_x W &= \int_0^{\widehat{x}} g(\widetilde{x})\, W(\widetilde{x})\, d\widetilde{x} \\
&= 1 - \int_0^{\widehat{x}} \frac{1+\psi\widehat{x}}{\widehat{x}(1+\psi x)^2} \widetilde{\gamma} \left[-\left(\frac{1+\psi\widetilde{x}}{\psi} \right)^2 \log \left(\frac{1+\psi\widetilde{x}}{1+\psi\widehat{x}} \right) - \frac{\widehat{x}}{\psi} \frac{(1+\psi\widetilde{x})^2}{1+\psi\widehat{x}} + \frac{\widetilde{x}}{\psi} + \frac{3}{2}\widetilde{x}^2 \right] d\widetilde{x} \\
&= 1 - \widetilde{\gamma} \frac{1+\psi\widehat{x}}{\psi^3\widehat{x}} \int_0^{\psi\widehat{x}} \frac{1}{(1+q)^2} \left[-(1+q)^2 \log \left(\frac{1+q}{1+\psi\widehat{x}} \right) - \psi\widehat{x} \frac{(1+q)^2}{1+\psi\widehat{x}} + q + \frac{3}{2} q^2 \right] dq \\
&= 1 - \widetilde{\gamma} \frac{3}{\psi^3\widehat{x}} \left[\psi\widehat{x} + \frac{1}{2}\psi^2\widehat{x}^2 - (1+\psi\widehat{x}) \ln(1+\psi\widehat{x}) \right], \\
\widetilde{\gamma} &= (1-B) \frac{u + \psi(1-u)}{\frac{1}{2}\widehat{x}^2 - (1-\psi)(1-u) \frac{3}{\psi^3\widehat{x}} \left[\psi\widehat{x} + \frac{1}{2}\psi^2\widehat{x}^2 - (1+\psi\widehat{x}) \ln(1+\psi\widehat{x}) \right]}.
\end{aligned}
$$

where the last equation follows from (6).

no commitment

$$
\begin{aligned}
W(\widetilde{x}) &= 1-\widetilde{\gamma}\left[\frac{1+\psi\widetilde{x}}{\psi^2}\ln\left(\frac{1+\psi\widetilde{x}}{1+\psi\widehat{x}}\right)-\frac{\widetilde{x}-\widehat{x}}{\psi}-\frac{1}{2}\widetilde{x}\left(\widetilde{x}-2\widehat{x}\right)\right],\\
\mathrm{E}_xW &= \int_0^{\widehat{x}} g\left(\widetilde{x}\right)W\left(\widetilde{x}\right)d\widetilde{x}\\
&= 1-\int_0^{\widehat{x}}\frac{1+\psi\widehat{x}}{\widehat{x}\left(1+\psi\widetilde{x}\right)^2}\widetilde{\gamma}\left[\frac{1+\psi\widetilde{x}}{\psi^2}\ln\left(\frac{1+\psi\widetilde{x}}{1+\psi\widehat{x}}\right)-\frac{\widetilde{x}-\widehat{x}}{\psi}-\frac{1}{2}\widetilde{x}\left(\widetilde{x}-2\widehat{x}\right)\right]d\widetilde{x}\\
&= 1-\widetilde{\gamma}\frac{1+\psi\widehat{x}}{\psi^3\widehat{x}}\int_0^{\psi\widehat{x}}\frac{1}{\left(1+q\right)^2}\left[\left(1+q\right)\ln\left(\frac{1+q}{1+\psi\widehat{x}}\right)-q+\psi\widehat{x}-\frac{1}{2}q\left(q-2\psi\widehat{x}\right)\right]dq\\
&= 1-\widetilde{\gamma}\frac{1+\psi\widehat{x}}{\psi^3\widehat{x}}\left[-\frac{1}{2}\ln^2\left(1+\psi\widehat{x}\right)+\psi\widehat{x}\ln\left(1+\psi\widehat{x}\right)-\frac{1}{2}\frac{\psi^2\widehat{x}^2}{1+\psi\widehat{x}}\right],\\
\widetilde{\gamma} &= \left(1-B\right)\frac{u+\psi\left(1-u\right)}{\frac{1}{2}\widehat{x}^2-\left(1-\psi\right)\left(1-u\right)\frac{1+\psi\widehat{x}}{\psi^3\widehat{x}}\left(-\frac{1}{2}\ln^2\left(1+\psi\widehat{x}\right)+\psi\widehat{x}\ln\left(1+\psi\widehat{x}\right)-\frac{1}{2}\frac{\psi^2\widehat{x}^2}{1+\psi\widehat{x}}\right)}.
\end{aligned}
$$

where the last equation follows again from (6). In all relations presented above, $\widehat{x}$ can be eliminated using

$$\widehat{x}=\frac{1-u}{u},$$

compare equation (11), while $\widetilde{\gamma}$ can be eliminated using its expressions.

A.5 Wage differentials and the output loss due to search for $\psi=1$

commitment

$$
\begin{aligned}
W\left(0\right)-\mathrm{E}_xW &= 2\left(1-B\right)\left(\frac{u}{1-u}\right)^2\left(\frac{5}{2}-u+\frac{3}{2}u^{-1}+\frac{4-u}{1-u}\ln u\right),\\
\mathrm{E}_xW-W\left(\widehat{x}\right) &= 2\left(1-B\right)\left(\frac{u}{1-u}\right)^2\left(\frac{1}{2}\left(\frac{1-u}{u}\right)^2-\frac{3}{2}\frac{1+u}{u}-3\frac{1}{1-u}\ln u\right),\\
X &= 6\left(1-B\right)\left(\frac{u}{1-u}\right)^2\left(1-u+\frac{1}{2}u^{-1}\left(1-u\right)^2+\ln u\right).
\end{aligned}
$$

no commitment

$$W(0) - \mathrm{E}_x W = 2(1-B)\left(\frac{u}{1-u}\right)^2 \left(-\frac{1}{2}\frac{1}{1-u}(\ln u)^2 - \frac{1+u}{u}\ln u - \frac{3}{2}\frac{1-u}{u}\right),$$

$$\mathrm{E}_x W - W(\widehat{x}) = -2(1-B)\left(\frac{u}{1-u}\right)^2 \left(\frac{1}{2}\left(\frac{1-u}{u}\right)^2 + \frac{1}{2}\frac{1}{1-u}(\ln u)^2 + \frac{1}{u}\ln u + \frac{1}{2}\frac{1-u}{u}\right),$$

$$X = -2(1-B)\left(\frac{u}{1-u}\right)^2 \left(\frac{1}{2}(\ln u)^2 + \frac{1-u}{u}\ln u + \frac{1}{2}\frac{1-u}{u}\right).$$

In all these equations, we use equation (7) to eliminate $\widetilde{\gamma}$.

A.6 The derivation of lemma 1

The conditional expectation of $(r+q)$ is given by,

$$\begin{aligned}
\mathrm{E}\left[|r+q|\,|r\right] &= -\int_{-\infty}^{-r} \frac{r+q}{\sigma_q}\phi\left(\frac{q}{\sigma_q}\right)dq + \int_{-r}^{\infty} \frac{r+q}{\sigma_q}\phi\left(\frac{q}{\sigma_q}\right)dq \\
&= -r\left[1-\Phi\left(\frac{r}{\sigma_q}\right)\right] + \sigma_q\phi\left(\frac{r}{\sigma_q}\right) + r\Phi\left(\frac{r}{\sigma_q}\right) + \sigma_q\phi\left(\frac{r}{\sigma_q}\right) \\
&= r\left[2\Phi\left(\frac{r}{\sigma_q}\right) - 1\right] + 2\sigma_q\phi\left(\frac{r}{\sigma_q}\right).
\end{aligned}$$

The Taylor approximation follows from evaluating the second derivative at $\mu = 0$. The sum of least squares reads:

$$\Sigma = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sigma_r\sigma_q}\left(|r+q| - \omega_0 - \omega_2 r^2\right)\phi\left(\frac{q}{\sigma_q}\right)dq\,\phi\left(\frac{r}{\sigma_r}\right)dr.$$

Define:

$$\sigma \equiv \sigma_x^{-1}\sigma_r\sigma_q,$$
$$\sigma_x \equiv \sqrt{\sigma_r^2 + \sigma_q^2}.$$

The first order conditions for ω_0 implies:

$$
\begin{aligned}
0 &= \int_{-\infty}^{\infty} \frac{1}{\sigma_r \sigma_q} \left[\begin{array}{c} -\int_{-\infty}^{-r} (r+q)\,\phi\left(\frac{q}{\sigma_q}\right) dq + \int_{-r}^{\infty} (r+q)\,\phi\left(\frac{q}{\sigma_q}\right) dq \\ -\int_{-\infty}^{\infty} (\omega_0 + \omega_2 r^2)\,\phi\left(\frac{q}{\sigma_q}\right) dq \end{array} \right] \phi\left(\frac{r}{\sigma_r}\right) dr \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma_r} \left[r \left[2\Phi\left(\frac{r}{\sigma_q}\right) - 1 \right] + 2\sigma_q \phi\left(\frac{r}{\sigma_q}\right) - \left(\omega_0 + \omega_2 r^2\right) \right] \phi\left(\frac{r}{\sigma_r}\right) dr \\
&= \int_{-\infty}^{\infty} \frac{2}{\sigma_r} r \Phi\left(\frac{r}{\sigma_q}\right) \phi\left(\frac{r}{\sigma_r}\right) dr + \int_{-\infty}^{\infty} \sqrt{\frac{2}{\pi}} \frac{\sigma_q}{\sigma_r} \phi\left(\frac{r}{\sigma}\right) dr - \left(\omega_0 + \omega_2 \sigma_r^2\right) \\
&= \int_{-\infty}^{\infty} 2\frac{\sigma_r}{\sigma_q} \phi\left(\frac{r}{\sigma_q}\right) \phi\left(\frac{r}{\sigma_r}\right) dr + \sqrt{\frac{2}{\pi}} \frac{\sigma_q}{\sigma_r} \sigma - \left(\omega_0 + \omega_2 \sigma_r^2\right) = \sqrt{\frac{2}{\pi}} \sigma_x - \left(\omega_0 + \omega_2 \sigma_r^2\right),
\end{aligned}
$$

where in the third line we apply integration by parts. For ω_2 we get,

$$
\begin{aligned}
0 &= \int_{-\infty}^{\infty} \frac{r^2}{\sigma_r \sigma_q} \left[\begin{array}{c} -\int_{-\infty}^{-r} (r+q)\,\phi\left(\frac{q}{\sigma_q}\right) dq + \int_{-r}^{\infty} (r+q)\,\phi\left(\frac{q}{\sigma_q}\right) dq \\ -\int_{-\infty}^{\infty} (\omega_0 + \omega_2 r^2)\,\phi\left(\frac{q}{\sigma_q}\right) dq \end{array} \right] \phi\left(\frac{r}{\sigma_r}\right) dr \\
&= \int_{-\infty}^{\infty} \frac{2}{\sigma_r} r^3 \Phi\left(\frac{r}{\sigma_q}\right) \phi\left(\frac{r}{\sigma_r}\right) dr + \int_{-\infty}^{\infty} \sqrt{\frac{2}{\pi}} \frac{\sigma_q}{\sigma_r} r^2 \phi\left(\frac{r}{\sigma}\right) dr - \left(\omega_0 \sigma_r^2 + 3\omega_2 \sigma_r^4\right) \\
&= \int_{-\infty}^{\infty} \frac{2}{\sigma_r} \left(r^3 - 2\sigma_r^2 r\right) \Phi\left(\frac{r}{\sigma_q}\right) \phi\left(\frac{r}{\sigma_r}\right) dr + \int_{-\infty}^{\infty} 4\sigma_r r \Phi\left(\frac{r}{\sigma_q}\right) \phi\left(\frac{r}{\sigma_r}\right) dr \\
&+ \sqrt{\frac{2}{\pi}} \sigma^3 - \left(\omega_0 \sigma_r^2 + 3\omega_2 \sigma_r^4\right) \\
&= \int_{-\infty}^{\infty} \sqrt{\frac{2}{\pi}} \frac{\sigma_r}{\sigma_q} r^2 \phi\left(\frac{r}{\sigma}\right) dr + \int_{-\infty}^{\infty} 2\sqrt{\frac{2}{\pi}} \frac{\sigma_r^3}{\sigma_q} \phi\left(\frac{r}{\sigma}\right) dr + \sqrt{\frac{2}{\pi}} \frac{\sigma_q \sigma^3}{\sigma_r} - \left(\omega_0 \sigma_r^2 + 3\omega_2 \sigma_r^4\right) \\
&= \sqrt{\frac{2}{\pi}} \frac{\sigma_r^2 \sigma_q^2}{\sigma_x} + 2\sqrt{\frac{2}{\pi}} \frac{\sigma_r^4}{\sigma_x} - \left(\omega_0 \sigma_r^2 + 3\omega_2 \sigma_r^4\right),
\end{aligned}
$$

where in the third line we repeatedly apply integration by parts. Solving these first order conditions for ω_0 and ω_2 yields

$$
\begin{aligned}
\omega_0 &= \sqrt{2\pi}^{-1} \left(\sigma_x + \frac{\sigma_q^2}{\sigma_x}\right), \\
\omega_2 &= \sqrt{2\pi}^{-1} \sigma_x^{-1}.
\end{aligned}
$$

A.7 The derivation of equation (32)

Let $z \equiv \frac{s-\mu_s}{\sigma}$, $z^* \equiv \frac{s^*-\mu_s}{\sigma}$. The effect of an increase in μ_s on the mean log wage of low and high skilled workers respectively reads:

$$\begin{aligned}
\frac{d}{d\mu_s}\mathrm{E}\left[\ln W\left(s\right)|s<s^*\right] &= \frac{d}{d\mu_s}\mathrm{E}\left[\ln W\left(0\right)+(1-\gamma\mu_s)s|s<s^*\right] \\
&= \frac{d}{d\mu_s}(1-\gamma\mu_s)s\sigma\mathrm{E}\left[z|z<z^*\right]-\frac{d}{d\mu_s}(1-\gamma\mu_s)s\mu_s \\
&= \gamma\sigma\frac{\phi\left(z^*\right)}{\Phi\left(z^*\right)}-\frac{1}{2}\gamma\mu_s, \\
\frac{d}{d\mu_s}\mathrm{E}\left[\ln W\left(s\right)|s>s^*\right] &= -\gamma\sigma\frac{\phi\left(z^*\right)}{1-\Phi\left(z^*\right)}-\frac{1}{2}\gamma\mu_s,
\end{aligned}$$

Hence:

$$\frac{d}{d\mu_s}\left(\mathrm{E}\left[\ln W\left(s\right)|s>s^*\right]-\mathrm{E}\left[\ln W\left(s\right)|s<s^*\right]\right)=-\gamma\sigma\frac{\phi\left(z^*\right)}{\Phi\left(z^*\right)\left[1-\Phi\left(z^*\right)\right]}$$

The effect on the number of low and high skilled workers reads:

$$\begin{aligned}
\frac{d\ln\Phi\left[\sigma^{-1}\left(s^*-\mu_s\right)\right]}{d\mu_s} &= \frac{\phi\left(z^*\right)}{\sigma\Phi\left(z^*\right)}, \\
\frac{d\ln\left[1-\Phi\left[\sigma^{-1}\left(s^*-\mu_s\right)\right]\right]}{d\mu_s} &= \frac{\phi\left(z^*\right)}{\sigma\left[1-\Phi\left(z^*\right)\right]}.
\end{aligned}$$

Hence:

$$\begin{aligned}
\frac{d}{d\mu_s}\left(\ln\Phi\left[\sigma^{-1}\left(s^*-\mu_s\right)\right]-\ln\left[1-\Phi\left[\sigma^{-1}\left(s^*-\mu_s\right)\right]\right]\right) &= \frac{1}{\sigma}\frac{\phi\left(z^*\right)}{\Phi\left(z^*\right)\left[1-\Phi\left(z^*\right)\right]}, \\
-\frac{d\left(\ln\Phi\left[\sigma^{-1}\left(s^*-\mu_s\right)\right]-\ln\left[1-\Phi\left[\sigma^{-1}\left(s^*-\mu_s\right)\right]\right]\right)/d\mu_s}{d\left(\mathrm{E}\left[\ln W\left(s\right)|s>s^*\right]-\mathrm{E}\left[\ln W\left(s\right)|s<s^*\right]\right)/d\mu_s} &= \frac{1}{\sigma^2\gamma}.
\end{aligned}$$