Lecture Notes
Search in Macroeconomics

Lukas Nord*

This version: May 17, 2022

*These lecture notes are based on Edouard Challe's lectures for *Macroeconomics III – Search Theory* at the EUI. They build on a previous version by Alessandro Ferrari, who I thank for making his work available. All remaining errors are mine. Please report typos, mistakes, and comments to lukas.nord@eui.eu.

Contents

1 Introduction

The purpose of these notes is to provide an introduction to basic search theory and its applications to Macroeconomic phenomena.

In previous Macro classes, you have mostly relied on the assumption of centralized markets, cleared by a Walrasian auctioneer. While this is a useful and easily tractable representation of markets and in many contexts sufficient, it assumes frictionless information about all trading opportunities and an instant matching of supply and demand. Search frictions, on the other hand, are a formal representation of markets where supply and demand cannot meet instantly to trade. In reality this is often the case: It can be difficult to observe all posted prices or job openings simultaneously and to find trading partners for a transaction, even if both parties – seller and buyer, employer and employee – would find it beneficial to trade.

Some features are characteristic for models with search frictions: A search friction is usually an information friction on the transaction process (Where or with whom can I make a given trade?) but mostly – at least for the purpose of these notes – assumes full information on the potential trading opportunities (the distribution of posted prices or job offers is common knowledge). In addition, markets with search frictions often feature longterm relationships between buyers and sellers or employers and employees. When matching demand and supply is frictionless (Walrasian auctioneer) existing relationships have no continuation value as it is costless to destroy and rebuild them. When markets feature trading frictions and forming relationships requires some form of costly search, existing relationships become valuable.

The potential applications of search markets are far reaching. They include not only labor markets and job search, certainly the most prominent search problem in economics. In these notes we will also study how search introduces a role for money in trade and the problem of search for trading opportunities in over-the-counter (OTC) asset markets. Additionally, frameworks with search frictions are used to analyze a variety of other topics, from frictional goods markets (search for prices, varieties or quantities) and housing markets to the formation of romantic relationships, marriage and disease transmission.

The first part of the notes focusses on labor market search. We introduce the basic job search problem of a worker and solve for worker's optimal search behavior in a frictional labor market. Extending the baseline model to search on the job, we study wage dispersion and derive an endogenous wage offer distribution. We finish the first part of the notes studying vacancy creation and efficiency of labor market equilibria in the Diamond-Mortensen-Pissarides framework.

The second part of the notes introduces money in search economies. We study the problem of double coincidence of wants and show how money can improve upon economic arrangements relying on barter alone.

The third part of the notes discusses search in asset markets. We show how buyers and sellers interact via intermediaries in decentralized OTC markets and study how search frictions can affect market prices.

Disclaimer: These notes follow closely the *Macro III – Search Theory* lecture by Edouard Challe. They should be seen as a complement to the lecture, not as a substitute, and I hope they will help you in studying the lecture's topics. The main part of the notes covers the material discussed in class, the appendix introduces some additional topics not directly relevant to the course. If there are any mistakes or typos in the notes or conflicts with information provided in the lecture, please drop me an email to lukas.nord@eui.eu. When in doubt (or for what is relevant in the exam), always rely on the lecture material provided by Edouard.

As the lecture, these notes are loosely build around a number of references. Useful books with (partial) treatment of the material covered here are e.g. Cahuc et al. (2014), Mortensen (2003), Nosal and Rocheteau (2011), Petrosky-Nadeau and Wasmer (2017), Pissarides (2000), or (parts of) Chapters 6, 29, and 30 in Ljungqvist and Sargent (2018). In addition, we will cite relevant research papers where applicable.

2 A Basic Job Search Model

In this section we introduce the basic job search problem of McCall (1970). For the baseline setup, assume that unemployed workers receive per period replacement income b and job offers at a constant, exogenous arrival rate λ.[1] If they receive a job offer, the corresponding wage w is drawn from exogenous CDF $F(w)$. If the worker accepts a job offer at wage w, she will maintain this wage throughout the tenure of the job. All employed workers supply a fixed amount of labor, i.e. there is no labor supply choice at the intensive margin, and are exogenously separated from their jobs at rate q. During the course of these notes, unless specified otherwise, we will assume that workers are infinitely lived, risk neutral (their utility is linear, $u(c) = c$), and live hand to mouth (there are no savings, $c = income$).

2.1 From Discrete to Continuous Time

In previous classes you have mostly worked with problems in discrete time. In this class, however, we will study problems which are set up in continuous time. Continuous time has some advantage when studying questions concerned with the flow of agents across states, often yielding problems more tractable than their discrete time counterparts, without any loss of economic intuition. The continuous time setup of a problem can in most cases be derived as the limiting case of its discrete time counterpart, letting (discrete) time intervals between periods go to zero. Intuitively, the advantage of continuous time problems often arises from the fact that very little will happen in an arbitrarily small amount of time.[2]

To show how you can move between discrete and continuous time problems, we will set up the value function of a worker employed at wage w in discrete time and derive the continuous time representation as an example. In discrete time the value of being employed at wage w at time t is given as

$$V_e(w,t) = w + \beta\left[(1-q)V_e(w,t+1) + qV_u(t+1)\right].$$

The equation above is specified for time intervals of length 1. It can be generalized to arbitrary time intervals of length Δ. We assume wage payments and separations are uniformly distributed within time periods such that for a time period of length Δ, a worker will receive wage $w\Delta$ and the rate at which workers are separated is $q\Delta$. Time is discounted with β^{Δ}. This yields

$$V_e(w,t) = w\Delta + \beta^{\Delta}\left[(1-q\Delta)V_e(w,t+\Delta) + q\Delta V_u(t+\Delta)\right]$$

[1] Replacement income can be broadly defined and can include e.g. unemployment benefits, additional leisure time, or home production. It has been shown that what is in included in a calibration of b can influence e.g. the prediction of the model with respect to unemployment fluctuations over the business cycle and provides a potential solution to the Shimer Puzzle (see Appendix B.3).

[2] For another, non-search related recent application of continuous time to simplify incomplete market models see e.g. Ahn et al. (2018).

Rearranging terms and subtracting $\beta^\Delta V_e(w,t)$ on both sides yields

$$(1-\beta^\Delta)V_e(w,t) = w\Delta+\beta^\Delta q\Delta\left[V_u(t+\Delta)-V_e(w,t+\Delta)\right]+\beta^\Delta(V_e(w,t+\Delta)-V_e(w,t))$$

Re-define the discount factor as $\beta = e^{-r}$ (for details see Appendix A.1). Dividing both sides of above equation by Δ yields

$$\frac{1-e^{-r\Delta}}{\Delta}V_e(w,t) = w+\beta^\Delta q\left[V_u(t+\Delta)-V_e(w,t+\Delta)\right]+\beta^\Delta\frac{V_e(w,t+\Delta)-V_e(w,t)}{\Delta}$$

To obtain the continuous time problem we need to let Δ become infinitesimally small, approaching 0 in the limit. $\lim_{\Delta\to 0}\frac{1-e^{-r\Delta}}{\Delta} = r$ can be obtained by applying L'Hospital's rule. $\lim_{\Delta\to 0}\beta^\Delta = 1$ and $\lim_{\Delta\to 0}V(t+\Delta) = V(t)$. The last term on the right hand side is the time derivative of the value function. Assuming a stationary solution, this term is equal to zero, i.e. value functions do not change with time.

This leaves us with

$$rV_e(w) = w + q\left(V_u - V_e(w)\right), \tag{1}$$

which implicitly defines the value function of an employed worker in continuous time. The corresponding continuous time representation for the value function of an unemployed worker can be derived accordingly.

2.2 The Reservation Wage

Equation (1) implicitly defines the value function of a worker currently employed at wage w, $V_e(w)$. V_u is the value to the worker of being unemployed (defined in detail below), r is the continuous time discount rate of the household and $rV_e(w)$ is called the *flow value* of a job at wage w. It corresponds to the current wage payment, adjusted for the potential loss in utility if the worker gets separated into unemployment.

It is natural to think of a job as an assets, paying a constant "dividend" w with probabilistic duration (until a match get separated). Dividing (1) by $V_e(w)$ yields

$$r = \underbrace{\frac{w}{V_e(w)}}_{\text{dividend yield}} + \underbrace{q\frac{V_u - V_e(w)}{V_e(w)}}_{\text{expected capital gain } (<0)}.$$

You can think of finding the value of employment similarly to finding the value (price) of an asset. This implies that the value of a job has to equalize its expected return (dividend yield plus expected capital gain) with households' time preferences, very similar to holding a bond or shares.

Before we turn to defining the value of unemployment, let us focus on which kind of wages a worker would accept. Clearly, a worker should accept any job offer $\hat{w}$ for which the flow value of employment at that wage is at least as large as the flow value of being unemployed, i.e. accept iff $rV_e(\hat{w}) - rV_u \geq 0$. As $V_e(w)$ is strictly increasing in w there exists a threshold wage level that makes the worker indifferent between accepting an offer or not, with all higher offers being strictly accepted and all lower offers being strictly declined. Define the reservation wage R as the wage at which the worker is just indifferent, s.t. $rV_e(R) - rV_u = 0$. Using (1), we can also define it solving

$$V_e(R) - V_u = \frac{R - rV_u}{r + q} = 0 \tag{2}$$

which defines the reservation wages as $R = rV_u = rV_e(R)$. It is the wage level that exactly compensates the worker for the flow value of unemployment which she has to give up in order to accept a job offer.

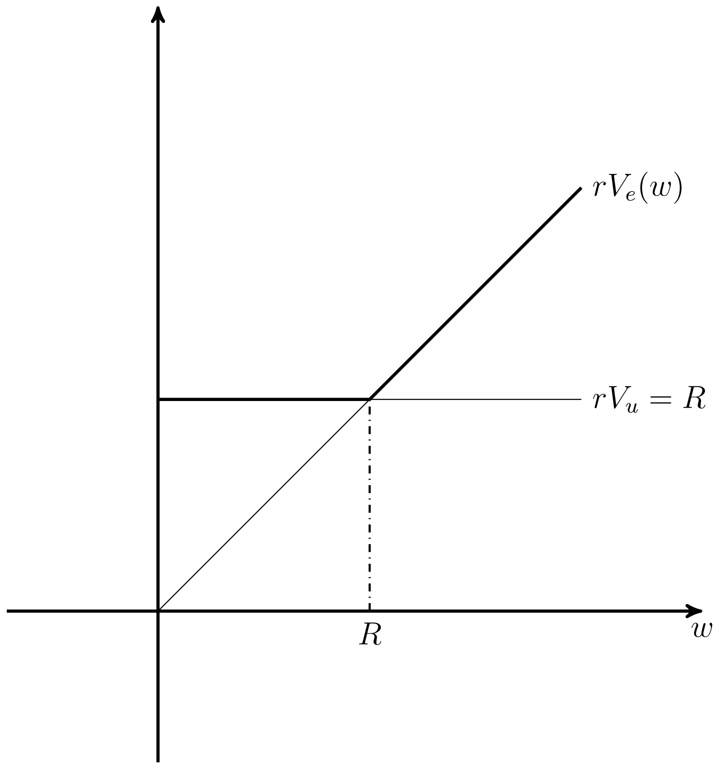


Figure 1 – Value Functions and Reservation Wage

Graphically, the reservation wage strategy is plotted in Figure 1. Again, for any wage $w < R$ we see $rV_u > V_e(w)$ and hence the worker should prefer to remain unemployed, while for wages $w > R$ we get $rV_u < V_e(w)$ and hence the worker should choose to accept a job offer.

To determine the reservation wage formally, we need to define the value to a worker of being unemployed. With our definition of R as the threshold value of accepting a job, we can define the value to an unemployed worker of receiving a job offer, but not yet knowing the attached wage, as

$$\hat{V} = \underbrace{\int_0^R V_u dF(w)}_{\text{offer will be rejected}} + \underbrace{\int_R^\infty V_e(w) dF(w)}_{\text{offer will be accepted}}. \tag{3}$$

Graphically, computing $\hat{V}$ is taking expectations over the thick line in Figure 1 with respect to the wage offer w.

With this in mind, we can define the value of being unemployed implicitly from

$$rV_u = b + \lambda(\hat{V} - V_u). \tag{4}$$

Similar to the the employed workers value function, the flow value of unemployment consists of an instantaneous payment b adjusted for the expected change in value due to potential job offers.

Using (2), (3) and (4), we can implicitly solve for the value of unemployment as a function of parameters:

$$\begin{aligned} rV_u &= b + \lambda\Big(\underbrace{\int_0^R V_u dF(w) + \int_R^\infty V_e(w) dF(w)}_{\hat{V}} - \underbrace{\int_0^R V_u dF(w) - \int_R^\infty V_u dF(w)}_{V_u}\Big) \\ &= b + \lambda \int_R^\infty (V_e(w) - V_u) dF(w) \\ &= b + \lambda \int_R^\infty \left(\frac{w - rV_u}{r+q}\right) dF(w) \end{aligned}$$

Now using the fact that $R = rV_u$, we get

$$R = b + \frac{\lambda}{r+q} \int_R^\infty (w - R) dF(w) \tag{5}$$

which defines the reservation wage as an implicit function of replacement income, time preferences, arrival and separation rates, and the wage offer distribution. It can be shown that the solution to this equation is unique. This proof evolves in two steps: 1.) Note that the LHS of the equation is increasing in R while the RHS is decreasing, hence there can be at most one intersection. 2.) For $R = 0$ the RHS is clearly positive while the LHS is zero, hence an intersection exists.

Intuitively, the reservation wage is the solution to an optimal stopping time problem: An unemployed worker receiving a job offer trades off the value of getting a job right away against the value of waiting and potentially receiving an even better offer in the future. The latter is often referred to as the *option value of waiting*. Waiting for a better offer has value as the worker will be stuck with any job she accepts until the match is exogenously separated (at rate q). The option value is why as soon as there is *any* wage offered higher than b, we get that $R > b$, i.e. workers need to be *always* offered a wage higher then their replacement income to accept a job.

Applying the implicit function theorem to (5), it can be shown that the reservation wage is increasing in replacement income b and job arrival rate λ, but decreasing in time discount r and the separation probability q. Along the tradeoff highlighted above it is clear that a higher arrival rate increases the option value of waiting (because it is more likely to get another, better offer tomorrow) while a higher b makes waiting less costly. A larger time discount factor r implies that the worker cares less about the future and is hence less likely to forgo a certain current wage for an uncertain higher wage at some point in the future. A higher separation rate reduces the expected job tenure, disproportionately affecting the value of jobs at higher wages w. This again reduces the option value of waiting, which depends on absolute, not relative differences in the value of employment at different wages.

Knowing the reservation wage is sufficient to fully characterize the solution to the workers' problem. We can rewrite the value functions as a function only of the reservation wage and parameters:

$$V_u = \frac{R}{r}, \quad V_e(w) = \frac{w + qR/r}{r+q}, \quad \hat{V} = \frac{(r+\lambda)R/r - b}{\lambda}$$

where the expression for $\hat{V}$ is obtained by rearranging 4. The probability of receiving and accepting a job offer, i.e. the probability of leaving unemployment, is independent of the time spent without a job and is the *hazard rate* $\lambda^* = \lambda(1-F(R))$, where $F(\cdot)$ denotes the cumulative density function of wage offers. The hazard rate is the probability of receiving an offer λ times the probability of accepting a wage offer randomly drawn from $F(w)$, $(1-F(R))$. From λ^* the duration of unemployment is $\frac{1}{\lambda^*}$, see Appendix A.2 for the derivation.

Using this simple model, we are able to describe worker flows in and out of unemployment. Assuming a total mass one of workers, the instantaneous change in the unemployment rate u can be defined as

$$\dot{u} = -u\lambda(1-F(R)) + (1-u)q.$$

The first term on the RHS captures those workers accepting a job offer and leaving unemployment, while the second term captures those employed workers (of total mass $1-u$) which are separated. In a stationary equilibrium $\dot{u} = 0$ and hence the unemployment rate is given by

$$u = \frac{q}{\lambda(1-F(R)) + q}. \tag{6}$$

It is easy to see that the unemployment rate is increasing in the reservation wage R. The more workers decline offers and choose to exercise their option to wait (for better offers) the more will be unemployed. As both λ and q affect u directly but also affect reservation wage R (see above), a conclusion as to their effect on equilibrium u is less straight forward and depends also on the offered wage distribution $F(w)$.

The CDF of the distribution of wages, conditional on employment is given as

$$G(w) = \begin{cases} \frac{F(w)}{1-F(R)} & w \geq R \\ 0 & w < R \end{cases}$$

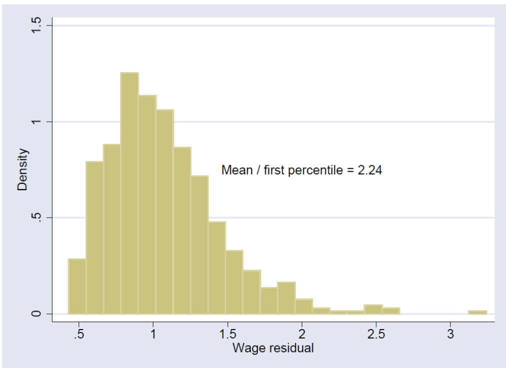
and the average observed wage can be computed as

$$w^* = \mathbb{E}[w|w \geq R] = \int_0^\infty w dG(w) = \int_R^\infty w dG(w) = \frac{\int_R^\infty w dF(w)}{1 - F(R)} \geq \int_0^\infty w dF(w).$$

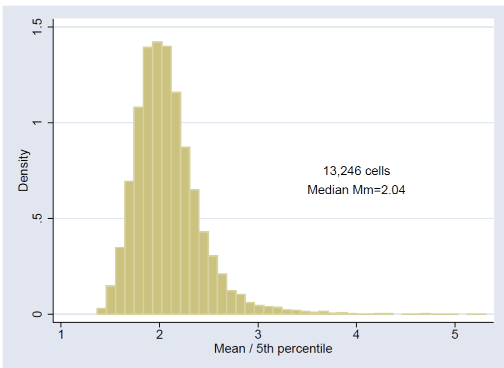
To due the reservation wage strategy and workers' choice to potentially decline job offers at some (low) wages, the average observed wage can be higher than the average of posted wages. Formally this happens when R exceeds the lower bound of the support of $F(w)$.

2.3 A Too Simple Model? - Part I: Wage Dispersion

A key empirical feature of the wage distribution is, that it features substantial dispersion in wages even for jobs that are very similar to each other. One can argue about productivity differences and scarcity of skills to reason why a manager earns a different wage than a janitor. However, dispersion among janitors is less easily justified since they complete a fairly homogeneous task for which little education and acquired skills are necessary.



(a) Residual Wage for Janitors in Philadelphia



(b) Mp5 across US labour markets

Figure 2 – Wage Dispersion

As an example, Figure 2a plots the observed distribution of wages for janitors in Philadelphia, which exhibits substantial dispersion.

To compare the empirical wage dispersion to the predictions of our basic search model, Hornstein et al. (2011) propose to look at the ratio between the mean and the minimum wage (*Mm-ratio*). Figure 2b plots the distribution of this ratio across many labor markets (job interacted with geographic area) in the US. The median Mm-ratio is around 2.

We can compute the equivalent model implied statistics from our results above. We have already computed the average wage w^* and we know that the minimum wage observed will correspond to the reservation wage. Hence we can compute $Mm = \frac{w^*}{R}$. To do so, we slightly rearrange the reservation wage equation to obtain:

$$\begin{aligned} R &= b + \frac{\lambda}{r+q} \int_R^\infty (w - R) dF(w) \\ &= b + \frac{\lambda(1 - F(R))}{r+q} \left(\frac{\int_R^\infty w dF(w)}{1 - F(R)} - R \frac{\int_R^\infty dF(w)}{1 - F(R)} \right) \\ &= b + \frac{\lambda^*}{r+q}(w^* - R) \end{aligned}$$

Assuming without loss of generality that $b = \rho w^*$, we get that

$$Mm \equiv \frac{w^*}{R} = \frac{\frac{\lambda^*}{r+q} + 1}{\frac{\lambda^*}{r+q} + \rho}.$$

Rearranging is necessary to obtain a formulation of the Mm-ratio as a function of variables we can observe in the data: It is hard to disentangle $F(w)$ from λ, but we can determine λ^* easily, which is simply the rate at which people transition from unemployment to employment.

Using the calibration of Hornstein et al. (2011) and setting $\lambda^* = 0.43$, $r + q = 0.034$, $\rho = 0.4$, we get a model-implied Mm-ratio of approx. 1.05. This is much lower than the Mm-ratio of around 2 we observe in the data. Hence, for realistic calibrations of parameters our simple model *fails* to capture the empirically observed dispersion in wages.

The failure of the simple model to capture the empirically observed wage dispersion is intuitively due to the option value logic around workers decision to accept a job offer: With higher dispersion in the offered distribution $F(w)$ (necessarily due to a longer upper tail as we assume wages to be strictly positive), the option value to wait for better future offers increases. This implies a rise in the reservation wage and a lower λ^*, which offset the rise in dispersion of posted prices.

2.4 A Too Simple Model? - Part II: Wage Offer Distribution

So far, we have taken the distribution of offered wages $F(w)$ as entirely exogenous. Before we turn to potential extensions of the basic job search model, it is useful to think about what our simple model of worker behavior implies for firms' optimal wage posting strategies. This yields two paradoxes:

The first is called the *Rothschild-paradox* and poses that in equilibrium, all firms should offer the same wage $w = R$. Take as given that households accept every job offer with wage $w \geq R$ and treat R as exogenous for now. Then a firm posting a wage $w' > R$ could decrease the wage and get strictly higher profits: it would still hire the worker but pay her strictly less over the course of the job tenure. Hence it must be that, given R, firms' best response is to post only wages $w = R$.

The second is called the *Diamond-paradox* and poses that the unique equilibrium wage should be $w = R = b$. To see this, start from the reservation wage equation (5) and impose a singleton offer distribution with positive mass only at $w = R$. This implies that

$$R = b + \frac{\lambda}{r + q}(R - R) = b.$$

Hence, the failure of the simple search model to capture the empirically observed dispersion in wages goes beyond the shortfall described in the previous section: Not only does workers optimal search behavior limit observed dispersion for any given distribution of offers $F(w)$, firms optimal response to a reservation wage strategy in the simple search problem will yield that there is no dispersion in offered wages at all!

The failure of the basic model to generate empirically observed wage dispersion can be explained by missing market power of workers. If firms can post take it or leave it offers and workers only outside option is unemployment, all value added of forming a match will be claimed by the firm. To break the paradoxes described above, we need to introduce some sort of market power on the worker side. The literature has found three ways to do so: On-the-job search, bilateral bargaining, and competitive search. We will take a look at each of these in turn in the following sections.

3 On-the-Job Search and Endogenous Wage Dispersion

In the previous section we have assumed that workers will only leave a job when they are (exogenously) separated and fall back into unemployment. In reality, however, most labor market transitions happen directly from one job to another. In this section, we will introduce on-the-job search, i.e. allow workers to receive job offers while employed and move directly from one job to another. As we will see, this simple extension goes a long way in fixing the lack of wage dispersion in the basic framework.

For this section, we will assume that employed workers receive job offers at rate λ_e, which is potentially different from the arrival rate for unemployed workers λ_u. We still assume, however, that conditional on receiving a job offer both employed and unemployed workers draw the associated wage from the same distribution $F(w)$, i.e. firms' wage offers do not differentiate between previously employed and unemployed workers.

3.1 Reservation Wages with On-the-Job Search

We begin by studying how the possibility of on-the-job search affects households reservation wages. Analogue to before, we can set up the value of being employed at wage $\tilde{w}$ as

$$rV_e(\tilde{w}) = \tilde{w} + q\left(V_u - V_e(\tilde{w})\right) + \lambda_e \underbrace{\int_{\tilde{w}}^{\bar{w}} \left(V_e\left(w\right) - V_e(\tilde{w})\right) dF(w)}_{=\mathbb{E}\max[V_e(w)-V_e(\tilde{w}),0]} . \tag{7}$$

It is straight forward that a worker employed at wage $\tilde{w}$ will accept any offer satisfying $w \geq \tilde{w}$. The last term in (7) now captures this possibility that the worker finds a better paying job. This effect of step-by-step moving to higher paying wages while employed is referred to as *job ladder* in the literature. In its simplest form here, it already captures the empirical facts that wages are higher for workers who have been continuously employed for a longer time period and that wages recover only slowly to their previous levels after an unemployment spell.[3]

The value of being unemployed can again be written as

$$rV_u = b + \lambda_u \underbrace{\int_{R}^{\bar{w}} \left(V_e(w) - V_u\right) dF(w)}_{=\mathbb{E}\max[V_e(w)-V_u,0]}, \tag{8}$$

where R is the reservation wage. This equation is exactly the same as (4) above for the case without on the job search. Nevertheless, the adjustment to (7) is sufficient

[3]For more work on the job ladder, especially its effect on the (re-)bargaining of wages, see e.g. Postel-Vinay and Robin (2002), Cahuc et al. (2006), Moscarini and Postel-Vinay (2018), Jarosch (2021).

to change the reservation wage substantially. We can show this by noting that again the reservation wage is defined by $V_e(R) = V_u$, evaluating (7) at R and subtracting rV_u from both sides to obtain

$$R = b + (\lambda_u - \lambda_e) \int_R^{\bar{w}} (V_e(w) - V_e(R))\, dF(w). \tag{9}$$

Remember that in the previous section we had $R > b$. With on-the-job search this only holds iff $\lambda_u > \lambda_e$. Intuitively, the option value of waiting and remaining unemployed in the previous section was due to the fact that workers could receive better job offers only during unemployment as wages were fixed during employment spells. This case is nested here, we can set $\lambda_e = 0$ to recover the previous solution. Now, however, workers can increase their wage also while employed when a better job offer arises. The option value to wait in unemployment therefore only remains if (better) job opportunities arise *faster* out of unemployment, which is the case iff $\lambda_u > \lambda_e$. On the other hand, if $\lambda_u < \lambda_e$ we obtain $R < b$, i.e. workers are willing to give up on instantaneous payments to get better (faster) access to high paying job offers on-the-job. That this can be an empirically relevant case is no surprise to anyone who has ever worked for low wages (as low as zero) as a trainee or intern.

3.2 Hornstein, Krusell, Violante Revisited

With the on-the-job search extension, we can revisit the Mm-ratio of the model. Assume for now that $\lambda_u = \lambda_e$. In this case from (9) clearly $R = b$ and the Mm-ratio is given as $Mm = \frac{w^*}{R} = \frac{1}{\rho}$. With the calibration of Hornstein et al. (2011) $\rho = 0.4$ and we get a Mn-ratio of 2.5, not too far from its empirical counterpart of around 2. Intuitively, when $\lambda_u = \lambda_e$ there are no opportunity cost of leaving unemployment. If a worker is as likely to receive (the same) job offers on-the-job as when unemployed, she is willing to accept lower wages, hence increasing the observed wage dispersion.

For the more general case of $\lambda_u \neq \lambda_e$, we can derive the Mm-ratio as before by solving the reservation wage equation (9) as a function of only parameters. Detailed derivations are presented in Appendix B.1 for reference.

3.3 The Diamond-Paradox Revisited

The extension of the model to on the job search also allows us to revisit the implications of the Diamond-Paradox. We will do so by studying equilibrium wage dispersion with on-the-job search in a model similar to Burdett and Mortensen (1998).

To take the determination of wage postings into account and study equilibrium wage disperison, we need to specify more carefully a firm side of the basic labor search model. Assume that there is an (exogenously) fixed, large number of atomistic firms in the economy, all producing output at common productivity p. To produce one unit of output a firm needs one worker, which it hires by posting a wage w in a frictional labor market. It is straightforward to see that all posted wages will satisfy $w \in S = [R, p]$ in equilibrium, as no firm would post an offer she knows will be rejected ($w \geq R$) and posting wages above productivity would yield losses if a match is formed ($w \leq p$). However, as we do not allow firms to distinguish between unemployed and employed workers, some wage offers made will be rejected by workers already employed at $w_0 > w$. This implies that even though the support of the offer distribution $F(w)$ lies strictly above the reservation wage R, the distribution of wages offered will still differ from the distribution of offers accepted (= the distribution of wages paid) $G(w)$. To study equilibrium price dispersion with on-the-job search, we need to determine jointly $F(w)$ and $G(w)$.

To simplify derivations and without loss of generality, for this section assume that $\lambda_u = \lambda_e \equiv \lambda$ and that $r = 0$, i.e. offers arrive at equal rate on- and off-the-job and there is no discounting.

We begin by characterizing labor market flows. We study a stationary equilibrium, implying that worker flows in and out of unemployment as well as in and out of any given wage level have to exactly offset each other. This yields two equilibrium conditions linking the unemployment rate u to the CDF of wages posted $F(w)$ and the CDF of wages paid $G(w)$ as

$$
\begin{aligned}
q(1-u) &= \lambda u \quad \Rightarrow u = \frac{q}{\lambda + q} \\
\lambda F(w) u &= \left[q + \lambda(1 - F(w))\right] G(w)(1-u) \;\; \forall w
\end{aligned}
$$

where for the first equation we use the fact that all wage offers are weakly larger than R, i.e. $F(R) = 0$, and the second applies the wage ladder logic (moving up in wage if receiving a better offer on the job) to every percentile of the wage distribution. The LHS of each equation characterizes the inflow into unemployment or wages below w respectively, while the RHS characterizes outflows. For unemployment, it is easy to see that inflows occur if workers are exogenously separated (at rate q) and outflows happen when workers receive job offers (at rate λ, all accepted as all posted wages above R). For the second equation, inflows into wages paid below w arise if unemployed workers receive a job offer below w. Outflows arise if employed workers are exogenously separated, or if workers employed at wages below w receive a job offer above w and change jobs ($\lambda(1-F(w))G(w)(1-u)$). Note that there is no inflow from on-the-job search here as we are considering the CDF, and on-the-job poaching of workers towards jobs paying less than w can only happen with workers previously already earning less than w.

Since we exogenously fix the mass of firms, assuming that each firm employs exactly one worker (*single job* definition of a firm) and normalizing the mass of workers to 1, we need to fix the mass of firms in the economy to $(1-u) = \frac{\lambda}{\lambda+q}$ for given λ and q. This ensures that in equilibrium there will always be exactly one employed worker per firm (job).[4]

[4] In later sections we will endogenize the mass of firms as well as λ by assuming costly vacancy creation.

Plugging the equilibrium unemployment rate into the second equation above, we get the relationship between the distribution of wages paid and wages posted as

$$G(w) = \frac{qF(w)}{q + \lambda(1 - F(w))} < F(w).$$

formally denoted as *first order stochastic dominance* of $G(w)$ over $F(w)$. As discussed, $G(w) < F(w)$ is due to the job ladder effect, i.e. because workers already employed will reject offers received below their current wage.

The value of an existing match with a worker to a firm can be characterized as

$$rJ(w) = p - w + [q + \lambda(1 - F(w))](0 - J(w))$$

where $p - w$ is the flow payoff to the firm, $q + \lambda(1 - F(w))$ is the rate at which matches are separated exogenously (q) or because workers quit to better paying jobs ($\lambda(1 - F(w))$). By our assumption that $r = 0$ we get

$$J(w) = \frac{p - w}{q + \lambda(1 - F(w))}. \tag{10}$$

From this equation we can see that increasing the wage w has an ambiguous effect on firm value: It lowers the flow profit $p - w$, reducing firm value. It also decreases the likelihood that a worker is poached by a better paying competitor($\lambda(1 - F(w))$), increasing the expected duration of a match and, hence, firm value.

The firm will not separate voluntarily from a worker. It only decides about a wage to post. We assume that each job offer posted will meet exactly one worker, who can then decide whether to accept or not. To decide what is the payoff to a firm of posting a wage w we need to determine the likelihood with which it is accepted. This probability is equal to 1 if the worker met is unemployed. It is equal to $G(w)$ if the worker is employed, as only workers currently earning less than w would be willing to switch jobs. The probability that a worker met is unemployed is equal to u and that she is employed equal to $(1-u)$ respectively. Hence, the overall probability that a job offer at wage w is accepted is equal to $u + G(w)(1 - u)$. The expected profit from posting a wage w is then given as

$$\begin{aligned}
\pi(w) &= [u + (1 - u)G(w)] \times [J(w) - 0] \\
&= \left[\frac{q}{\lambda + q} + \frac{\lambda q F(w)}{(\lambda + q)(q + \lambda(1 - F(w)))}\right] \times \frac{p - w}{q + \lambda(1 - F(w))} \\
&= \frac{q(p - w)}{[q + \lambda(1 - F(w))]^2}
\end{aligned}$$

As for firm value above, a higher wage w decreases net earnings (flow profits) but increases both the likelihood that a worker accepts the offer and that he will stay with the firm even when receiving on-the-job offers.

An endogenous equilibrium wage offer distribution $F(w)$ must satisfy four conditions:

1. **There can be no mass points at any $w < p$.** Else, a firm could marginally increase the posted wage and cause a discrete jump in acceptance / retention rates, yielding strictly higher profits. This logic is offset at $w = p$ where an increase in wages would yield negative profits.

2. **$b \in S$.** If the replacement income is not part of the support of the wage distribution, any firm posting the lowest wage (for which $F(w) = 0$) could increase profits by lowering its offer to b, increasing flow profits without any change in acceptance / retention rates.

3. **Firms have to be indifferent between all posted wages in equilibrium.** This implies that all posted wages have to yield exactly the same expected profits $\pi(w) = \bar{\pi}$. Assume they do not. In this case, firms would prefer to switch to posting the wage with the highest profits and the wage distribution would not be an equilibrium.

4. **$p \notin S$.** Posting a wage $w \geq p$ would yield expected profits $\pi(w) \leq 0$, while posting a wage $w = b$ yields strictly positive expected profits. Hence, there can be no wage offers with $w \geq p$ if firms are to be indifferent.

Note that 1. (and 2.) above already rule out the Diamond-Paradox. The potential of on-the-job poaching by competitors gives workers some market power due to an outside option beyond unemployment and hence yields them a share in the value added from forming a match.

From 3. above, the effects of changing the wage on flow profits and acceptance / retention rates have to exactly offset each other. Using 2. and the fact that workers will not accept any offer with $w < b$ (since we assume equal arrival rates on- and off-the-job $R = b$ from before) makes b the lowest wage posted in equilibrium. For the lowest wage posted it has to hold that $F(b) = 0$ and hence

$$\pi(b) = \frac{q(p-b)}{(q+\lambda)^2} = \bar{\pi}.$$

From 3. above, we can than get an equation pinning down the equilibrium wage offer distribution as

$$\frac{q(p-w)}{[q+\lambda(1-F(w))]^2} = \frac{q(p-b)}{(q+\lambda)^2}$$

$$\Rightarrow F(w) = \frac{\lambda+q}{\lambda}\left(1-\sqrt{\frac{p-w}{p-b}}\right)$$

$F(\cdot)$ is continuous with connected support $[b, \bar{w}]$. The upper bound of the distribution is given by $\bar{w} = F^{-1}(1) = \alpha b + (1-\alpha)p < p, \quad$ where $\alpha \equiv \left(\frac{q}{q+\lambda}\right)^2 \in (0,1)$.[5].

[5] Alternatively, show this by noting that $F(\bar{w}) = 1$ and solving $\pi(\bar{w}) = p - w = \frac{q(p-b)}{(q+\lambda)^2} = \bar{\pi}$.

What we observe in the data is not $F(w)$ but the distribution of wages paid $G(w)$ (or its density $g(w) = \frac{\partial G(w)}{\partial w}$). Using our result for $F(w)$, you can show that in the model the associated PDF $g(w)$ is increasing and convex. This is still somewhat add odds with the data, where we observe a hump shaped distribution of wages. However, further extensions such as heterogeneity in productivity p are able to align the shape of the model implied and empirically observed wage distribution.[6] This lets us conclude, that on-the-job search is a crucial feature of any quantitative theory of wage dispersion. Nevertheless, a simple model of labor market choices with search frictions can go a long way in explaining the observed distribution of wages.

[6] See e.g. Mortensen (2003), Chapter 3.

4 Job Creation and Endogenous Arrival Rates

For the previous sections, we have taken the rate at which workers find jobs as exogenously given parameters. In this section, we will work towards endogenizing these arrival rates. As in reality, we will allow firms to choose how many jobs to create and study how this affects labor market outcomes and what determines how many jobs will be created. We will see how firms' vacancy creation affects the rates at which households find jobs and at which firms hire workers for vacant positions as well as how the vacancy posting of one firm affects the likelihood of all others to hire a worker. To study labor market equilibria with endogenous arrival rates, we will rely on the *Diamond-Mortensen-Pissarides Model*, which has become the workhorse model of macro labor.[7] Based on the market technology assumed to bring workers and firms together, such models are generally referred to as *matching models*.

4.1 Matching Function and the Beveridge Curve

To formulate a theory of endogenous arrival rates, we have to start by discussing how workers and firms meet in the labor market. For this we will introduce the concept of a *matching function*. You can think of this as something similar to a production function, taking unemployed workers and unfilled jobs (also called *vacancies*) as inputs and producing a number of meetings between workers and firms willing to hire as output.

More formally, denote as u the number of unemployed workers and as v the number of open vacancies. Normalizing the mass of workers to 1, u is also equal to the unemployment rate in the economy and v is the vacancy rate (open positions per worker). We define the matching function as $m(u, v)$ and assume it has the following properties:

1. $m(u, 0) = m(0, v) = 0$. If there are no unemployed workers or no vacant jobs, no meetings can take place.
2. $m_u > 0$ and $m_v > 0$. The more unemployed workers or the more vacancies out there, the more meetings can take place.
3. $m_{uu} < 0$ and $m_{vv} < 0$, i.e. there are decreasing returns to increasing the number of unemployed workers or the number of vacancies separately.
4. $m(ku, kv) = km(u, v) \quad \forall k > 0$. The matching functions is constant returns to scale (CRS), if we increase the number of unemployed and the number of vacancies proportionately so will the number of matches.

[7]Peter A. Diamond, Dale T. Mortensen and Christopher A. Pissarides won the 2010 Nobel Memorial Prize in Economics for their work on this framework and search frictions more generally.

The last property is an important one and implies that the elasticities of the matching function with respect to u and v will sum up to one, i.e.

$$1 = \underbrace{u\frac{m_u(u,v)}{m(u,v)}}_{\eta_u \geq 0} + \underbrace{v\frac{m_v(u,v)}{m(u,v)}}_{\eta_v \geq 0}.$$

You can show this by noting that for any CRS function, by Euler's Theorem, it has to hold that $m(u,v) = um_u(u,v) + vm_v(u,v)$. One specific functional form satisfying all assumptions above is the Cobb-Douglas function.

The assumption of a CRS matching function is quite an important one as Diamond (1982) shows that deviations can lead to multiplicity of equilibria. Petrongolo and Pissarides (2001) survey the literature on matching functions and summarize empirical evidence on whether the matching function is CRS, finding overall support but substantial heterogeneity in estimates. For the remainder of these notes we will take the assumptions above as given.[8]

We can rewrite the matching function in terms of *market tightness*, which we define as $\theta \equiv v/u$. It captures the number of open vacancies per unemployed worker. The CRS property of the matching function allows us to define the *contact rates* at which unemployed workers find jobs and open vacancies meet unemployed workers as

$$\begin{aligned}
\lambda(\theta) &\equiv \frac{m(u,v)}{u} = m\left(\frac{u}{u}, \frac{v}{u}\right) = m(1,\theta) \\
\delta(\theta) &\equiv \frac{m(u,v)}{v} = m\left(\frac{u}{v}, \frac{v}{v}\right) = m\left(\frac{1}{\theta}, 1\right) = \frac{\lambda(\theta)}{\theta}
\end{aligned}$$

where the equalities exploit the CRS property of $m(\cdot)$. If θ is high, we speak of a *tight* labor market. In a tight labor market it is easier for workers to find jobs (λ is increasing in θ) but harder for firms to fill open positions (δ is decreasing in θ). In equilibrium, $\lambda(\theta)$ and $\delta(\theta)$ will not only be contact rates but also the job-finding and vacancy-filling rates as all unemployed workers will accept any job they are offered.

As before, in steady state worker flows in and out of unemployment will have to exactly offset each other. The flow out of unemployment is given by the number of matches $m(u,v) = \lambda(\theta)u$ and the inflow into unemployment as before by the number of exogenous separations $q(1-u)$. The resulting steady state condition for unemployment

$$u = \frac{q}{q + \lambda(\theta)}$$

implicitly defines any stationary relationship between equilibrium unemployment and vacancies as

$$v(u) = u\lambda^{-1}\left(q/u - q\right) \tag{11}$$

where we use the fact that $\theta = v/u$.

The graph of $v(u)$ is one of the most important objects in labor economics and is called the *Beveridge Curve*. Taking the total differential of the equality connecting

[8] A recent extension beyond the CRS matching function is provided e.g. in Davis et al. (2013).

the flows into and out of unemployment $m(u,v) = q(1-u)$, we can show that under our assumptions for $m(\cdot)$ the theoretical Beveridge Curve has negative slope

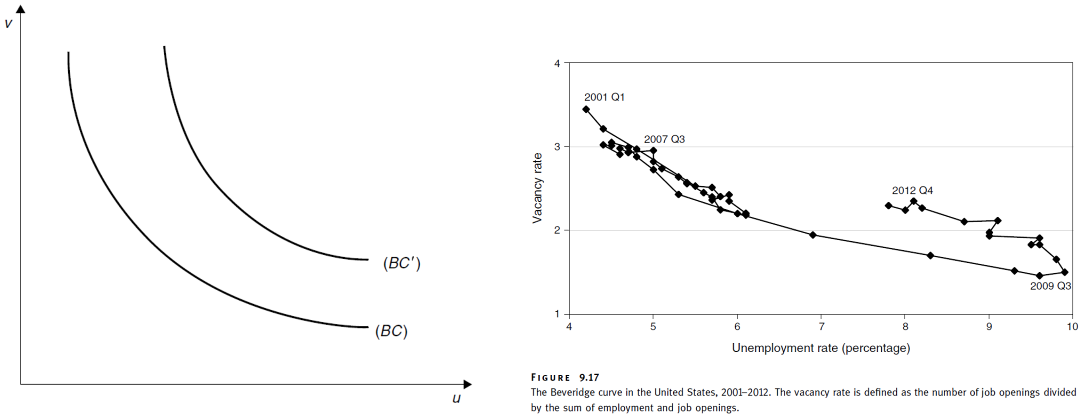
$$m_u du + m_v dv = -q du \quad \Rightarrow v'(u) = -\frac{q + m_u(u,v)}{m_v(u,v)} < 0$$

and is convex

$$v''(u) = -\frac{m_{uu}m_v - (q + m_u)m_{uv}}{m_v^2} > 0.$$

The former is due to assuming positive marginal returns in match creation to vacancies and unemployed and the latter due to the assumption of decreasing marginal returns.

The negative slope of the Beveridge Curve is intuitive: The more vacancies there are, the more meetings happen per unemployed worker and the easier it is to find jobs. Hence workers will transition fast out of unemployment, decreasing the equilibrium unemployment rate for given inflows. Furthermore, the Beveridge curve is shifted outwards by q. With a higher rate of separations there need to be more vacancies per unemployed worker in order to sustain the same equilibrium unemployment rate.



(a) Theory **(b)** Empirics

Figure 3 – Beveridge Curve

Figure 3b is taken from Cahuc et al. (2014).

Figure 3 plots the theoretical Beverdige Curve as well as its empirical counterpart for the US. The decreasing slope is a prominent feature of the empirical Beveridge Curve.

4.2 A Basic Diamond-Mortensen-Pissarides Model

While the Beveridge Curve provides a menu of stable (stationary) combinations of unemployment and vacancy rates, it does not determine which combination will be selected as an equilibrium. To understand how outcomes in a frictional labor market with endogenous arrival rates are determined, we need to take a closer look at how firms decide how many vacancies to create.

Suppose there is an infinite supply of potential single-job firms. Each firm that is currently not matched with a worker can decide to post a vacancy. Posting a vacancy for one period comes at cost $c > 0$, and a filled vacancy (call this a *job*) produces output worth $p > 0$ per period.

The value to the firm of a filled job is then given as

$$rJ(w) = p - w - qJ(w) \quad \Rightarrow J(w) = \frac{p - w}{r + q}$$

where w is the single equilibrium wage paid to the worker per period and q as before the exogenous rate of separation. Exogenous separation will be the only source of job destruction as long as $J(w) \geq 0 \Leftrightarrow w \leq p$, since this yields positive profits to the firm and we abstract here from on-the-job search.

How does the firm decide about creating a vacancy? If a vacancy is filled the firm will get profit $J(w)$. The firms knows that the hiring rate is $\delta(\theta)$ (on average a firm will hire $\delta(\theta)$ workers per vacancy, where it is generally possible that $\delta(\theta) > 1$), so the expected profit from posting a vacancy is $\delta(\theta)J(w)$. Since a vacancy comes at cost c, it is straight forward to see that firms optimal choice warrants posting a vacancy iff

$$\delta(\theta)J(w) \geq c. \tag{12}$$

This condition is known as the *free entry condition.* For a solution with finite but positive number of vacancies it has to hold with strict equality in equilibrium, making firms indifferent between posting an additional vacancy or staying out of the market.

On the other side of the labor market we assume the simple job-search worker problem without on the job search from before, which gives us

$$rV_e(w) = w + q(V_u - V_e(w))$$

$$rV_u = b + \lambda(\theta)(V_e(w) - V_u)$$

combined to

$$V_e(w) - V_u = \frac{w - b}{r + q + \lambda(\theta)}.$$

Unemployed workers will accept jobs iff $V_e(w) \geq V_u$, i.e. $w \geq b$. Note that we assume workers to be entirely passive here except for the decision to accept or reject a job offer. In a more evolved setup, we could add a worker decision on the optimal search effort, giving workers an active influence on labor market tightness.

For a first pass on an equilibrium, assume that the wage is an exogenous parameter satisfying *bilateral efficiency*, i.e. $w \in [b, p)$. In words, the wage is such that it compensates workers for their outside option (unemployment) but still yields positive profits to the firm (less than p) so both worker and firm would agree to a mutually beneficial match.[9]

[9] We will endogenize the wage below.

Finding an equilibrium boils down to finding a fixed point in market tightness θ. Firms are *atomistic* (have no influence on equilibrium objects individually), and take (u, v) and hence $\delta(\theta)$ as given when deciding on whether or not to post a vacancy. In the aggregate, firms posting decisions will determine v and hence $\delta(\theta)$. So we are looking for a θ such that given this tightness, firms create exactly enough vacancies to yield this tightness as an equilibrium outcome.

As mentioned before, for a finite but positive number of vacancies in equilibrium we must have that $c = \delta(\theta)J(w)$. If expected profits where larger than the cost of posting, all firms would create vacancies, this increases v and decreases $\delta(\theta)$ until the condition holds with equality. If expected profits where less than the cost no firm would choose to post any vacancy, this decreases v and increases $\delta(\theta)$ until the condition holds with equality.

An equilibrium combination of (u, v) and implied market tightness θ has to satisfy both free entry 12 and be consistent with the dynamics of the Beveridge Curve 11 from above. The former guarantees firm optimality, while the latter restricts the set of feasible combinations of (u, v) to those consistent with implied labor market flows. Solving the free entry condition of $v(u)$, we are looking for an intersection of

$$v(u) = u\delta^{-1}(c/J(w)) \quad \text{(free entry)}$$

and

$$v(u) = u\lambda^{-1}\left(q/u - q\right) \quad \text{(Beveridge Curve)}$$

The former is clearly strictly increasing in u and gets shifted by q, r, c, p. The latter strictly decreases in u as discussed before (conditional on assumptions on the matching function), and hence an equilibrium exists and is unique. Figure 4 gives a graphical representation of the equilibrium. Note that $\theta = \delta^{-1}(c/J(w))$ gives the slope of the free-entry / job-creation curve.

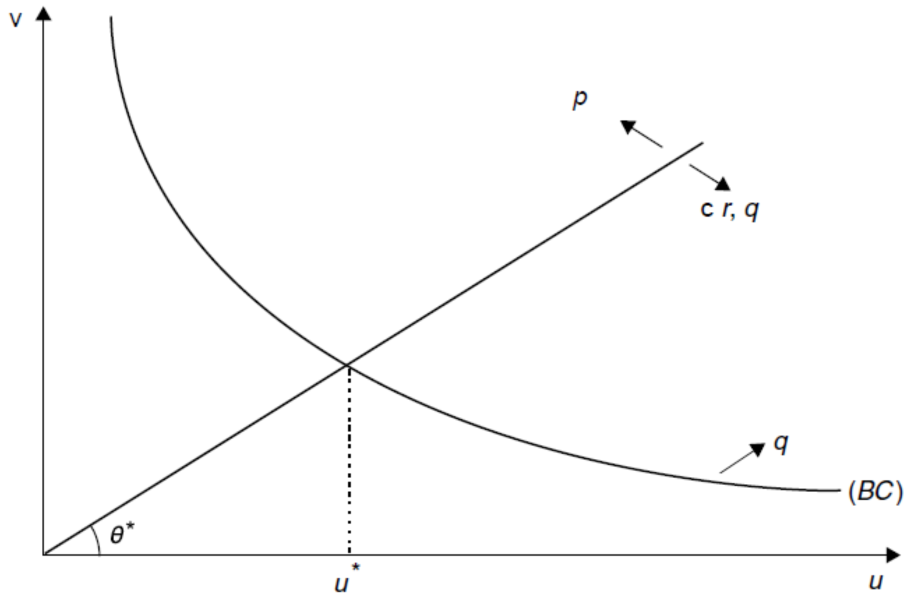


Figure 4 – Equilibrium Job Creation

Intuitively, the free entry condition says that if unemployment is high it is worthwhile for firms to post more vacancies since these get filled easier. More vacancies will increase labor market tightness and lead to more hiring out of unemployment so from the Beveridge Curve unemployment decreases, these forces bring the frictional labor market towards equilibrium. They also highlight the impact of what is known as *congestion externality:* If there are many offers on my own side of the market (i.e. many other vacancies posted if I am a firm), it is harder for me to find a trading partner in a market with search frictions. Therefore actions are *strategic substitutes* in the spirit of Cooper and John (1988), the more vacancies other firms post the lower the job-filling rate and the less likely I am to post additional vacancies. Another (non-labor) real-life example of a congestion externality is the ratio of men to women on most dating platforms.

4.3 Wage Determination and Nash Bargaining

So far we have taken the equilibrium wage w as exogenously given. Finding wage w endogenously imposes a challenge, since it is indeterminate in the *bargaining set* $[b, p)$. To see this, note that workers would accept any wage above their replacement income b as it makes them strictly better off, while firms would be willing to pay any wage less than p, leaving them with positive profits.

Unlike in Walrasian GE theory or in our model with on-the-job search, wage determination therefore requires imposing a (somewhat arbitrary) wage-setting mechanism. We will focus here on the most prominent one, *Nash-Bargaining*, but other forms of bargaining (Kalai, alternative offers,...) or self-fulfilling conventions (Hall, 2005) would work as well.

Determining the wage is equivalent to determining how the *match surplus* is shared between workers and firms, which is the sum of firm surplus $J(w)$ and worker surplus $V_e(w) - V_u$. It can be interpreted as the total additional value added that is realized when a match is formed.

The solution to a Nash-Bargaining game with workers bargaining power $\phi \in [0, 1]$ will yield a wage that satisfies

$$V_e(w) - V_u = \phi(J(w) + V_e(w) - V_u)$$

and

$$J(w) = (1 - \phi)(J(w) + V_e(w) - V_u).$$

Special cases are given by $\phi = \frac{1}{2}$ (standard Nash-Bargaining), $\phi = 0$ (take-it-or-leave it offer by firm to worker) and $\phi = 1$ (take-it-or-leave it offer by worker to firm).[10] The latter has no equilibrium since it yields zero profits of a match to firms, which would imply no vacancies in equilibrium with $c > 0$.

Either of the above equations can be solved for the *wage curve* $w(\theta)$, given by

$$\frac{V_e(w) - V_u}{J(w)} = \frac{\phi}{1 - \phi}$$

[10] For a derivation of the Nash-Bargaining solution see Appendix A.3.

where $J(w) = \frac{p-w}{r+q}$ and $V_e(w) - V_u = \frac{w-b}{r+q+\lambda(\theta)}$ from before.

Substituting the value functions and rearranging, you can show that the derivative $w'(\theta) \geq 0$, i.e. wages are higher in a tighter labor market.

Together with the Beveridge Curve and the job-creation curve (determining labor demand from the free entry condition), the wage curve fully determines the equilibrium (u, v, w) in the standard Diamond-Mortensen-Pissarides model with endogenous vacancy creation and endogenous wages. Figure 5 displays the graphical representation of the equilibrium intersection of job-creation and wage curve in the (w, θ) plane, complementing Figure 4 for a full graphical characterization of an equilibrium.

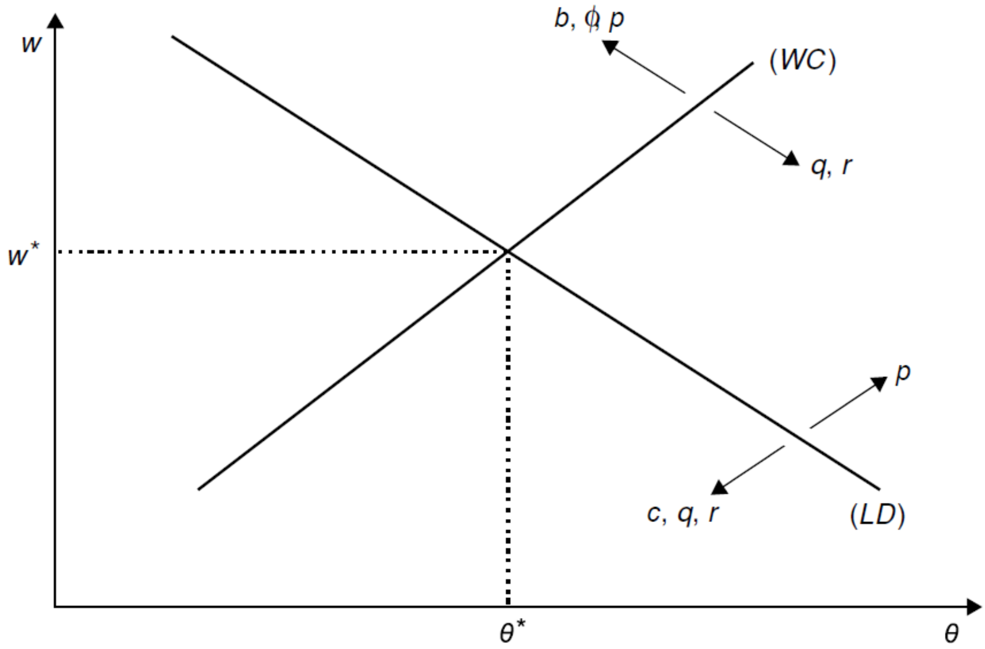


Figure 5 – Equilibrium Wage Determination

In this section we have focussed on a steady state labor market equilibirum. In Appendix B.2 we show how an economy adjusts towards this steady state when starting away from it. Appendix B.3 extends the model to incorporate aggregate risk and shows that the baseline model generates excessive volatility in wages over the business cycle, a feature known as the *Shimer Puzzle*.[11]

4.4 Efficiency: The Hosios Condition and Competitive Search

It is natural to ask whether the equilibrium levels of market tightness and vacancy creation are efficient from a social planner point of view. As mentioned before, the congestion externality could prevent the market outcome from reaching the efficient

[11]See Shimer (2005).

benchmark. Here it is important to note that the externality can play out both ways: *Within* groups, the externality imposes negative external effects. The higher the rate of unemployment the harder it is for any unemployed worker to find a job. On the other hand, *across* groups, the congestion externality has positive effects. The higher the rate of unemployment, the easier it is for firms to hire workers. Second best (constrained by search frictions) efficiency requires that these positive and negative effects exactly balance at the margin.

However, in our setup with Nash-Bargaining there is not really a market price incorporating the externalities to firms and workers of a higher / lower tightness θ, as the wage w is not a freely adjusting equilibrium object but pinnend down by (exogenous) bargaining weights. So what would be the efficient level of job creation? And is there a market arrangement sustaining the efficient equilibrium?

To answer these questions, we will look at a one period version of the model outlined above. Beginning of period unemployment u_0 is taken as given, vacancies v (and hence tightness $\theta = \frac{v}{u_0}$) are determined endogenously. The level of vacancies induces a flow from unemployment into employment of $m = m(u_0, v)$ and therefore $\lambda(\theta) = \frac{m(u_0,v)}{u_0}$ and $\delta(\theta) = \frac{m(u_0,v)}{v}$. End of period unemployment (during production) is $u = u_0(1 - \lambda(\theta)) + (1 - u_0)q$ and hence the total welfare of the economy measured as production and replacement income net of cost is given by

$$y = (1 - u)p + ub - cv.$$

We start with solving the planner problem as an efficient benchmark. A planner can freely choose equilibrium market tightness θ to maximize

$$\begin{aligned} W &= \max_{\theta} y = \max_{\theta} \left[\underbrace{(1-u)p + ub}_{\text{output}} - \underbrace{cv}_{\text{vacancy costs}} \right] \\ &= \max_{\theta} \left[p + (b - p) + u_0(b - p)[(1 - \lambda(\theta)) - q - c\theta] \right] \end{aligned}$$

which yields a first order condition for the planner solution θ^* as

$$\lambda'(\theta^*)(p - b) = c$$

or, equivalently, since $\lambda(\theta) = m(1, \theta)$

$$\underbrace{m_v(1, \theta^*)(p - b)}_{\text{social marginal return to a vacancy}} = c.$$

The left hand side of above equation is denoted as social marginal value of a vacancy as it multiplies the marginal increase in the number of matches from an additional vacancy with the additional output of an additional job. For an efficient solution the social marginal value of a vacancy has to be equalized to the resource cost of creating a vacancy, i.e. the marginal vacancy has to add exactly as much to aggregate output as it costs to create it.

We can now see how the decentralized economy compares to the planner solution to evaluate efficiency of the market outcome. In the one period economy, firms' surplus is given as $J(w) = p - w$ and workers' surplus as $w - b$, for a total match surplus of $p - b$. Surplus sharing according to Nash-Bargaining implies that $J(w) = (1 - \phi)(p - b)$ and $w = \phi p + (1 - \phi)b$, while free entry gives $c = \delta(\theta)J(w)$. Firm surplus and free entry together yield

$$\delta(\theta)(1 - \phi)(p - b) = c.$$

Obviously, the θ implied here is equivalent to the planner allocation θ^* iff

$$m_v(1, \theta) = \delta(\theta)(1 - \phi)$$

rearranged to

$$m_v(1, \theta) = \frac{m(1, \theta)}{\theta}(1 - \phi)$$

or

$$\underbrace{\frac{\theta m_v(1, \theta)}{m(1, \theta)}}_{=\eta_v} = 1 - \phi.$$

This well known equality is called the *Hosios condition* and was first described in Hosios (1990). It tells us that the decentralized solution with Nash-Bargaining is efficient if and only if the bargaining weight of the firm $(1 - \phi)$ is equal to the elasticity of the matching function with respect to vacancies.[12] You can interpret this as firms' relative contribution to job creation balancing their relative surplus share. If ϕ is too low, firms' are overcompensated and will create too many vacancies. On the other hand, if ϕ is too high firms are not compensated enough and will create too little vacancies, leading to inefficiently high unemployment.

Competitive Search

So far, we have focussed on a market structure called *random search*. Random, because both workers' and firms' have no control over who they meet, there is one single market in which both sides interact. The inefficiency occurring in this market when the Hosios condition is violated is partially due to this randomness. There is no price signal steering workers and firms to form matches one way or another.

In this section, we study how efficiency properties change when we consider a different market arrangement, commonly know as *competitive search* or *directed search*. This framework goes back to the work of Moen (1997), for a great overview / introductory treatment see Wright et al. (2021). We will see, that this market arrangement can bring back the signaling power of prices and yield efficient outcomes independent of further assumptions like the Hosios condition.

[12] Imposing Euler's Theorem on the matching function you can show that this is equivalent to imposing $\eta_u = \phi$.

Under competitive search, we assume that firm i can post a combination of wage and market tightness (w_i, θ_i) (implying arrival rates $\lambda(\theta_i), \delta(\theta_i)$). A higher wage will attract more workers, which lowers θ_i and $\lambda(\theta_i)$ but increases $\delta(\theta_i)$. In equilibrium, for multiple combinations of (w_i, θ_i) to be sustained in parallel, workers need to be indifferent.[13] This yields the condition that for the offers of any two firms i, j it has to hold that

$$\lambda(\theta_i)w_i + [1 - \lambda(\theta_i)]b = \lambda(\theta_j)w_j + [1 - \lambda(\theta_j)]b \ \forall i, j$$

You can think of this as firms posting only the wage and workers queueing optimally in response, the marginal worker being indifferent which queue to enter. Suppose that workers would not be indifferent, then all workers would direct their search towards the preferred pair of (w_i, θ_i) and all firms posting deviating offers would face infinite tightness. This clearly cannot be an equilibrium as infinite tightness would guarantee immediate matching for workers and would induce them to prefer these queues.

Firms take a required minimum utility level to be delivered to the worker V_u^* as given, which implies that any offer (w_i, θ_i) has to satisfy

$$\lambda(\theta_i)w_i + [1 - \lambda(\theta_i)]b = V_u^* \ \forall i.$$

As firms are maximizing profits, they will offer workers exactly V_u^* and not more. Offering them any higher value would yield strictly lower profits and can hence not be optimal. The full directed search equilibrium has V_u^* as an endogenous object, allowing workers to decide on which market to enter, and would require to solve also a worker problem taking firms minimum profit as given (dual approach). We omit this here for simplicity and focus on the implications of directed search for the efficiency of allocations.[14]

A firm chooses an optimal pair (w_i, θ_i) in order to maximize expected profits from the vacancy

$$\delta(\theta_i)(p - w_i)$$

conditional on providing V_u^* to visiting workers. The solution will be symmetric across firms so we can drop the index i.

We can rewrite the objective function as

$$\max_{w,\theta} \frac{m(1,\theta)}{\theta}(p - w) \quad \text{s.t.} \quad m(1,\theta)(w - b) = V_u^* - b$$

and define the Lagrangian

$$\mathcal{L} = \frac{m(1,\theta)}{\theta}(p - w) + \mu[V_u^* - b - m(1,\theta)(w - b)]$$

[13]You should be able to derive the same equilibrium imposing indifference for firms (free entry) and letting the worker optimize.

[14]For further details on directed search equilibria see e.g. Wright et al. (2021).

to get first order conditions

$$\begin{aligned}\left[\frac{m_v(1,\theta)}{\theta} - \frac{m(1,\theta)}{\theta^2}\right](p-w) - \mu m_v(1,\theta)(w-b) &= 0 \\ m(1,\theta)\left(\frac{1}{\theta}+\mu\right) &= 0\end{aligned}$$

which can be solved for

$$p - w = \underbrace{\frac{\theta m_v(1,\theta)}{m(1,\theta)}}_{=\eta_v}(p-b).$$

Combining this with free entry of firms, we get that

$$c = \frac{m(1,\theta)}{\theta}(p-w) = m_v(1,\theta)(p-b)$$

which satisfies exactly the optimality condition of the planner. Hence, competitive search yields the efficient outcome without any further assumptions like the Hosios condition.

Intuitively, the ability for workers to choose from a menu of (w_i, θ_i) restores the signaling ability of prices and guides the market towards the (constrained) efficient allocation. Firms internalize the externality that long queues (little vacancies per unemployed, low θ) impose on workers, because they have to compensate them in wage to keep them indifferent between entering a market with (w_i, θ_i) or looking for a job elsewhere. This is how a competitive search protocol is able to decentralize efficiency. Note that under Nash-Bargaining we had that

$$p - w = (1-\phi)(p-b)$$

which again resembles the condition from competitive search iff $\eta_v = 1 - \phi$, i.e. iff the Hosios condition holds.

5 Money in Search Equilibria

Fiat money (money without intrinsic value) is a predominant feature of all modern economies. The question of why economic agents are willing to accept a piece of paper in exchange for real valued objects is an essential one in economic theory. Standard approaches to justify fiat money include cash-in-advance constraints or money-in-the-utility function frameworks, introducing implicit liquidity considerations in reduced form.[15] The *New Monetarist* school of economics has provided a micro-foundation of the role of money by introducing (search) frictions in the trading process. In contrast to the Walrasian approach, which matches demand and supply instantly and frictionlessly, a detailed model of the exchange process can yield new insights into why money is held. Their conclusion, that money has a role in allowing economic agents to achieve allocations impossible to reach without it, is what we study in this section. The New Monetarist school can be broadly categorized into three generations of models: A first, studying frameworks with one unit of money and one unit of good per agent / trade (Kiyotaki and Wright, 1993). A second, with one unit of money and endogenous units of goods (Trejos and Wright, 1995). And a third, studying frameworks with endogenous units of both money and goods (Lagos and Wright, 2005). We will focus on the first and second generation and, for most of this section, use a simplified version of the setup in Rupert et al. (2000). Naturally, we can only provide an introduction into some basic models here. For a recent, more extensive survey of the New Monetarist perspective to money with many extensions and applications see e.g. Lagos et al. (2017).

5.1 Trading Frictions and the Double Coincidence of Wants

We begin by describing the market structure in which we operate. Assume a continuum of agents on $[0, 1]$. Further, assume a continuum of goods on $[0, 1]$ and assume that agent i is *specialized* and has the unique ability to produce good i. Goods are non-storable (no commodity money) and need to be consumed instantly after trading. Producing a unit of any good comes at utility cost $c \geq 0$.

To generate a potential for gains from trade we need to assume some heterogeneity in tastes. Read iWj as "i wants to consume the good that j produces". If $iWj = 1$, then agent i receives utility $u > 0$ from consuming one unit of good j, and u is constant across all agents and goods. With regard to preference operator W we make the following assumptions:

1. $Pr(iWi = 1) = 0$. No agent derives utility from consuming the good she produces herself.

2. $Pr(iWj = 1) = x \; \forall \; i \neq j$. The probability that an agents derives utility from consuming any good that she does not produce is x.

[15]See e.g. Chapters 2. and 3. in Walsh (2010).

3. $Pr(jWi = 1|iWj = 1) = y \; \forall \; i \neq j$. The probability that another agent derives utility from the good I produce conditional on me deriving utility from her good is y.

This immediately implies that for any random meeting of two agents the probability that both derive utility from each others good is $Pr(iWj = 1) \cdot Pr(jWi = 1|iWj = 1) = xy$. The first assumption above introduces gains from trade. The second assumption governs the extend of the search friction in the market, i.e. how likely it is that an agent meets someone with whom she would want to trade. The third assumption characterizes what is know as *double coincidence of wants*, i.e. how likely it is to find someone who I can actually trade with because each of us has something the respective other wants. We will see below that this assumption is crucial in introducing a role for money.[16]

A real world example for a market structured like this are e.g. services. Take a hairdresser and a nurse: Both a haircut and medical services are non-storable. The hairdresser does not do his own hair and the nurse should not self-medicate, i.e. both do not consume their own product. The nurse does not always need a haircut and the hairdresser does not generally need medical assistance. Double coincidence of wants is rarely given, when the nurse needs a haircut the hairdresser won't require medical assistance at the same time.

Suppose each agent can hold at most one unit of fiat money, which is assumed to be indivisible, storable, and costless to hold, but does not have any intrinsic value, i.e. money cannot be consumed directly and does not provide utility in and of itself. There is a total quantity $M \in (0, 1)$ of money in the economy which is initially distributed randomly to a measure M of agents, every one of them receiving one unit each. We assume that an agent holding money is unable to produce (purchases and consumption must take place before producing again) and that agents cannot freely dispose of money.

Meetings between two agents happen at exogenous rate α. We study a setup in continuous time and, hence, an agent will meet at most one other agent at the same time. Suppose two agents meet. They will never accept a good they do not like in trade. If two producers (non-money-holders) meet and both like each others product they barter, else they depart without trade. If a money holder meets a producer of a good she does not like, they depart without trade. If the money holder likes the good of the producer, *monetary exchange* may take place, trading the good of the producer for money. If two money holders meet they can never trade as they are always on the same side of the market (neither of them is able to produce).

Define as π_0 the probability that a producer accepts money in exchange for a good. Further, define as π_1 the probability that an agent holding money is willing to trade money for a good (conditional on her liking the good). Then $\pi = \pi_0\pi_1$ is the probability that monetary exchange will take place conditional on it being possible (meeting between producer and money-holder and money-holder wants the good of

[16]The double coincidence of wants is a fundamental problem in economics and dates back to Jevons (1875).

the producer). π, π_0, and π_1 are equilibrium objects which we will solve for below. We are interested in finding conditions under which money *circulates*, i.e. $\pi > 0$ and monetary exchange takes place.

It is straightforward to describe the value function of an agent holding money as

$$rV_1 = \alpha x(1-M)\pi \underbrace{(u + V_0 - V_1)}_{\equiv \Delta_1} \tag{13}$$

The first part of the expression is the the probability the money holder meets another agent (α) whose good she likes (x) and who is is not a money holder ($1-M$). With probability π both are willing to trade the good for money. In case of trade the money holder receives utility u and becomes a producer again (V_1 is replaced with V_0). We define $\Delta_1 \equiv u + V_0 - V_1$ as gains from trade from monetary exchange for a money-holder.

The value function of a producer (non-money-holder) is defined as

$$rV_0 = \alpha xy(1-M)(u-c) + \alpha xM\pi \underbrace{(V_1 - V_0 - c)}_{\equiv \Delta_0} \tag{14}$$

The first term is the probability of meeting another producer ($\alpha(1-M)$) and both producers liking each others good (xy) times the utility (net of cost of production) the producers gets if barter with another producer occurs ($u-c$). The second term captures the possibility of trade with a money holder (occurring with probability $\alpha xM\pi$), in which case the agent has to produce at cost c but does not receive a unit of good to consume in return. She only receives money and becomes a money-holder going forward. Here again $\Delta_0 \equiv V_1 - V_0 - c$ is the gains from monetary exchange for a producer.

It is easy to see that the sum of each agent's gains from monetary trade $\Delta_1 + \Delta_0 = u - c > 0$ and hence the overall *surplus* of monetary exchange (conditional on the possibility for it to take place) is always positive. It is, however, not ex-ante clear whether both agents need to profit from monetary exchange. Since barter among producers who like each others good always takes place, finding an equilibrium essentially boils down to determine the probability that monetary exchange will take place.

5.2 Equilibrium with Indivisible Goods (1st Generation)

We focus on stationary, symmetric and robust (i.e. stable) equilibria. By noting that $r\Delta_1 = ru + rV_0 - rV_1$ and using equations (13) and (14), one can show that the respective gains from monetary exchange for money-holders and produces can be described as:

$$\Delta_1 = \frac{\alpha x[M\pi + (1-M)y](u-c) + ru}{r + \alpha x\pi} \tag{15}$$

$$\Delta_0 = \frac{\alpha x(1-M)(\pi - y)(u-c) - rc}{r + \alpha x\pi} \tag{16}$$

The gains from trade are a function of exogenous parameters and the endogenous probability π that monetary exchange will take place when it is possible.

In return, the equilibrium trading probabilities π, π_0, and π_1 will depend on the gains from trade such that

$$\pi_j \begin{cases} = 1 \\ \in [0,1] \\ = 0 \end{cases} \text{as } \Delta_j \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases}$$

If the gains from trade are positive, an agent will always want to trade goods against money. If gains from trade are negative she will never want to trade. If gains from trade are zero, the agent should be indifferent between any probability of trade.

An equilibrium in this economy is a fixed point in (Δ_0, Δ_1) and (π_0, π_1), i.e. the gain of trade induced by a set of trading probabilities have to be such that these tradings probabilities are agents best response to the gains from trade. Note that this fixed point is not guaranteed to be unique.

It is straightforward from (15) that money holders always gain from trade as $\Delta_1 > 0 \ \forall\ \pi$ and, hence, in any equilibrium it holds that $\pi_1 = 1$. From (16), the gains of trade for a producer can be positive or negative and depend on the equilibrium value of π_0. This is determined by a game between producers. Here, we focus on symmetric Nash-Equilibria (SNE) of this game.

For a given equilibrium acceptance rate of π_0, take the surplus of a producer i as $\Delta_{0,i}(\pi_0)$, computed as above. The agent's best response in acceptance probabilities is defined as $\pi_{0,i}(\Delta_{0,i})$. In any SNE it then has to hold that $\pi_{0,i}(\Delta_{0,i}(\pi_0)) = \pi_0$. Crucially, from (16) and the fact that acceptance probabilities are weakly increasing in surplus, note that $\frac{\partial \pi_{0,i}}{\partial \pi_0} = \frac{\partial \pi_{0,i}}{\partial \Delta_{0,i}} \frac{\partial \Delta_{0,i}}{\partial \pi} > 0$, i.e. agents' actions are *strategic complements* in the spirit of Cooper and John (1988). The higher the likelihood that other agents are willing to trade, the higher I will set my own likelihood to trade.

It is well known that strategic complementarities can lead to multiple equilibria, and in our case three equilibria can exist:

1. **A pure-strategy, non-monetary equilibrium.** This equilibrium always exists. In a non-monetary equilibrium $\pi_0 = 0 = \pi$, i.e. money does not circulate. To prove its existence note that $\pi = 0 \Rightarrow \Delta_0(0) < 0 \Rightarrow \pi_{0,i}(\Delta_{0,i}(0)) = 0$.

2. **A pure-strategy, monetary equilibrium.** This equilibrium exists iff $\pi_0 = 1 = \pi \Rightarrow \Delta_0(1) > 0$, which is the case iff $\frac{c}{u} < \frac{\alpha x(1-M)(1-y)}{r+\alpha x(1-M)(1-y)}$.

3. **A mixed-strategy, monetary equilibrium.** This equilibrium exists iff $\exists\ \pi_0 = \tilde{\pi} \in (0,1)$ s.th. $\Delta_{0,i}(\tilde{\pi}) = 0$.

Figure 6 plots the three possible equilibria for an example with $c = 0$, $y = r = 1/4$, $\alpha x = 1$, $M = 1/3$, $u = 3/2$. The red line represents the surplus equation $\Delta_{0,i} = \frac{4\pi_0 - 1}{4\pi_0 + 1} \Leftrightarrow \pi_0 = \frac{1+\Delta_{0,i}}{4(1-\Delta_{0,i})}$ and the black line is the trading-probability best response function. We can solve above for $\tilde{\pi} = 1/4$ setting $\Delta_{0,i} = 0$. As c increases,

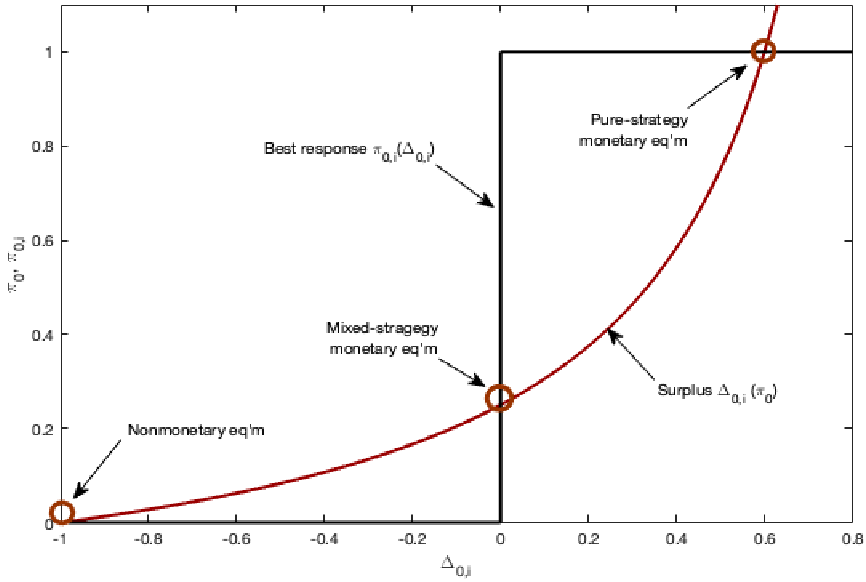


Figure 6 – Multiple Equilibria (Numerical Example)

the $\Delta_{0,i}(\pi_0)$-line shifts up, causing $\Delta_{0,i}(1)$ to shift down. Eventually, both the pure and mixed strategy monetary SNE vanish and only the non-monetary equilibrium remains.

What is the intuition behind this result? It is clear why the money-holder would always want to trade: She receives immediate utility from a unit of consumption and regains the opportunity to produce (and barter) in the future. But why would the producer accept money under some conditions but not all? Essentially, the producers has to balance two forces: On the one hand, she has to incur a cost of production without getting an immediate utility from consuming a good in return. Additionally, she will not be able to produce again before trading and therefore needs to find a producer willing to accept money for goods. On the other hand, she no-longer relies on finding another producer who wants exactly what she produces (as she now offers money in return). From (16) we can see this tradeoff: rc is the cost incurred right away. The first term of the numerator describes the tradeoff between offering ones own good vs. offering money in the future. If π is high relative to y, it is more likely to find a trading partner accepting money than accepting ones own good and the surplus from trading the good for money is high for the producer. On the other hand if y is high relative to π, it is easy to trade for ones own good and the value of giving a good up today to hold money in the future is low.

From this argument it is clear how the double coincidence of wants impacts the role of money in the economy. If the double-coincidence problem is strong (low y, tough to find a trading parter accepting ones own good), money can get around this issue by offering a unit of account to make more transactions possible. For this it is also important that goods are non-storable and that there is no alternative way of keeping track of agents history of past trades.

Stability

To see how stable the equilibria defined above are, we study how robust they are to small variations in the behavior of other agents. More formally, we analyze how an agents best response $\pi_{0,i}$ changes if all other agents adjust their behavior (π_0) by a marginal amount ϵ.

For the pure-strategy non-monetary equilibrium assume all other agents change their policy to $\pi_0 = \epsilon$. Then from (16) we get that for marginal ϵ still $\Delta_{0,i} < 0$ and hence $\pi_{0,i} = 0$. Therefore the pure-strategy non-monetary SNE is stable.

For the pure-strategy monetary equilibrium assume all other agents change their policy to $\pi_0 = 1 - \epsilon$. Then from (16) we get that for marginal ϵ still $\Delta_{0,i} > 0$ and hence $\pi_{0,i} = 1$. Therefore the pure-strategy monetary SNE is stable.

For the mixed-strategy monetary equilibrium assume all other agents change their policy to $\pi_0 = \tilde{\pi} \pm \epsilon$. Then from (16) we get that even for marginal ϵ $\Delta_{0,i} > 0$ or $\Delta_{0,i} < 0$ and hence $\pi_{0,i} = 1$ or $\pi_{0,i} = 0$. Therefore the mixed-strategy monetary SNE is unstable.

Welfare

To study implications for welfare, we first need to define a first best solution as benchmark. We do so by setting up the problem for a planner who can force production and exchange whenever a meeting takes place and *single coincidence* prevails, i.e. at least one agent likes the product the other has to offer.

In this planner-economy, an agent producing good i and liking good j meets another producing good j and/or liking good i both with probability αx respectively. The welfare of each agent in this economy is hence given as

$$rW = \alpha x u - \alpha x c = \alpha x(u - c)$$

Similar to what you might have seen in previous classes, this first best solution can be decentralized in case of a frictionless, multilateral credit arrangement equivalent to assuming complete asset markets.

We abstract from this possibility here and turn to the equilibrium with money described above. To find a second best solution (constraining the feasible set of the planner) we analyze the problem of a central bank setting monetary policy by choosing M. The central bank maximizes social welfare

$$\begin{aligned} rW &= MrV_1 + (1 - M)rV_0 \\ &= \kappa \times (1 - M)[\pi M + y(1 - M)] \end{aligned}$$

where κ is a positive constant depending on model parameters but independent of M and π.

It can be shown that a monetary equilibrium in which the central bank sets M weakly dominates the basic economy without money, where transactions can take place only in case of double coincidence of wants. On the other hand, the monetary equilibrium is dominated by the first best as defined above, where every transaction satisfying single coincidence of wants takes place. Even though money fails to reach the first best, it allows agents to achieve allocations beyond those feasible with barter alone.

Taking the derivative of the social welfare function above with respect to M, the optimal monetary policy satisfies

$$\begin{aligned} M^* &= \frac{1-2y}{2-2y} \; (< 1/2) \text{ if } y < 1/2 \\ M^* &= 0 \text{ if } y \geq 1/2 \end{aligned}$$

As relying exclusively on barter is a nested case of the central bank's problem the fact that for certain y it is optimal to choose positive M proves the weak dominance of the monetary economy with optimal M.

Intuitively, monetary exchange is beneficial as it allows to get around the need for double coincidence of wants. However, as we restrict agents holding money to not produce money does crowd out some transactions for which double coincidence is satisfied. These are transactions where a money holder meets a producer and the producer would like to purchase the good of the money-holder but cannot as the latter is restrained from producing. This crowding out effect increases in y which is why M^* decreases in y and why there is a threshold for y above which the central bank chooses to optimally not supply any money.[17]

5.3 Equilibrium with Divisible Goods (2nd Generation)

So far, we have assumed that one unit of money buys exactly one unit of good and that units of both money and goods are indivisible. In this section, we relax the assumption on the good side. Assume that goods are divisible and that one unit of money buys q units of good, hence the nominal price (price in units of money) of one unit of good is $p = \frac{1}{q}$. Consumption yields utility $u(q)$, where we make the common assumptions that $u(0) = 0, u'(q) > 0, u''(q) \leq 0$. Production is costly and causes a utility loss $c(q)$, where $c(0) = 0, c'(q) > 0, c''(q) \geq 0$.

We will study an economy where $p = \frac{1}{q}$ is determined by bargaining upon the formation of a suitable match. To simplify exposition we will assume $y = 0$ (no barter takes place) and $\alpha x = 1$ (every money holder meets a producer with probability $(1 - M)$ every period) without loss of generality. Define q^* as the efficient level of production, satisfying $u'(q^*) = c'(q^*)$.

Bargaining introduces a game inside the game. Similar to the discussion about π_0 and best response $\pi_{0,i}$ above, assume that all agents expect all bargaining outcomes to lead to Q and ask what the outcome q of a particular bargaining game is taking Q as given. A SNE is then again a fixed point such that $q = Q$.

[17] Rupert et al. (2000) provide an extension relaxing the assumption that money holders cannot trade and show that this economy provides strictly higher welfare.

The value functions of money holders and producers are now defined as

$$rV_1(Q) = (1 - M)\left[u(Q) + V_0(Q) - V_1(Q)\right] \tag{17}$$
$$rV_0(Q) = M\left[V_1(Q) - V_0(Q) - c(Q)\right] \tag{18}$$

and depend on the (expected) equilibrium outcome of bargaining games Q.

We assume that money-holders can make take-it-or-leave-it offers to producers, giving money-holders all bargaining power. Under this assumption, money holders will capture the entire surplus from trade by setting q such that $\Delta_0 = V_1(Q) - V_0(Q) - c(q) = 0$. This immediately implies that $V_0(Q) = 0$ from (18) and, hence, $\Delta_0 = V_1(Q) - V_0(Q) - c(q) = 0 \Rightarrow V_1(Q) = c(q)$. Note that the latter includes $c(q)$ not $c(Q)$, because it is derive from the solution to the bargaining problem. From (17) $V_1(Q) = \left(\frac{1-M}{1-M+r}\right) u(Q)$.

As before, we are looking for a SNE, such that $q = Q$. From the solution to the bargaining problem and the value function we get agents best response to Q as

$$q(Q) = c^{-1}\left[\left(\frac{1-M}{1-M+r}\right) u(Q)\right] \quad \Rightarrow \frac{\partial q(Q)}{\partial Q} > 0$$

The worse the terms others are offering, the lower is my own offer ($p = \frac{1}{q}$). This essentially captures that money will be more valuable in the future because it gets more units of goods if Q is higher and hence the price of a good in money terms is lower.

There are two candidate SNE:

1. **Non-monetary equilibrium.** $q(0) = 0$. In this case there is no exchange at all (we excluded barter by assumption) and hence $V_1(0) = V_0(0) = 0$.
2. **Monetary equilibrium.** q^m, satisfying $c(q^m) = \left(\frac{1-M}{1-M+r}\right) u(q^m)$.

There are as many monetary SNE as there are solutions to the second equation, which in turn depends on the exact functional form for $c(\cdot)$ and $q(\cdot)$.

Welfare

For the first best we again assume that a planner can set q and force every agent to produce upon meeting another who derives utility from the good. Under our assumption that $\alpha x = 1$, the planner maximizes welfare defined by

$$rW^{FB} = \max_q u(q) - c(q) \quad \Rightarrow q = q^*$$

For the second best we again assume a central bank can set an optimal quantity of money M. Under the assumption that money-holders have all bargaining power, we know that $V_0 = 0$ for any value of M and hence the central bank optimizes

$$W^{CB} = \max_M MV_1(q^m) = \max_M Mc(q^m(M))$$

The monetary SNE obviously dominates the non-monetary SNE as in the the latter no trade takes place at all. However, the monetary SNE is necessarily inefficient relative to the 1st best as you need money for trade, but money (under the assumption that money-holders cannot produce) generates inefficient meetings between two money-holders in which no production and exchange will take place. For this reason, the second best will feature $q^m \neq q^*$.

6 Search in OTC Asset Markets

Financial markets are often considered as the most direct representation of the Walrasian auctioneer in the real world: They operate via centralized exchanges and are assumed to incorporate any new information within the blink of an eye, balancing supply and demand via price adjustments in real time. This common view of financial markets, however, does not apply to a wide range of financial securities, which are only traded in one-to-one transactions rather than via centralized exchanges. This class of securities is often referred to as *over-the-counter* or using its abbreviation – OTC. Figure 7 shows that the market volume of OTC securities turns out to easily exceed that of exchanged traded equities and treasury securities.

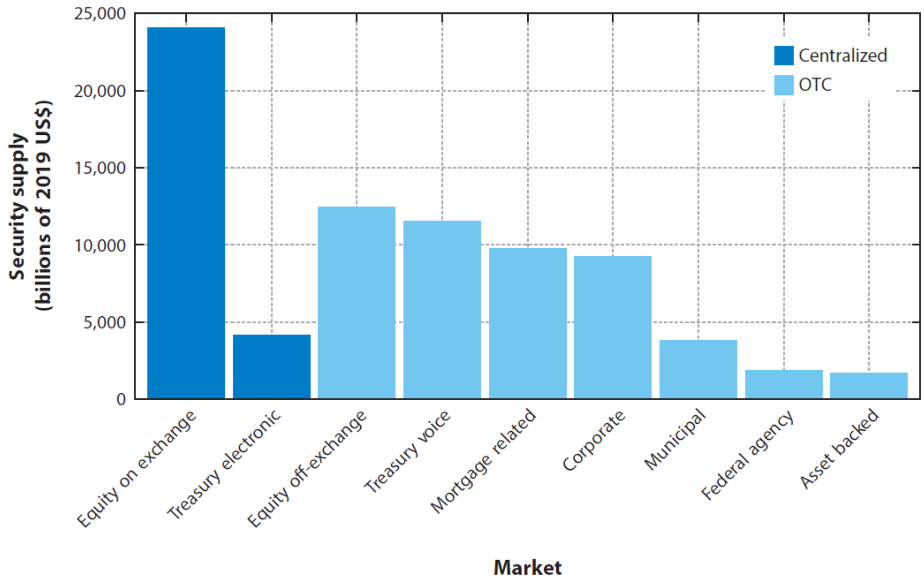


Figure 1

Security supply outstanding in 2018, broken down between centralized and over-the-counter (OTC) markets.

Figure 7 – Market volume of OTC securities

Figure taken from Weill (2020).

As known from other contexts, decentralized (OTC) financial markets are subject to microstructure frictions. In this section, we are going to see how search frictions provide a natural framework to think about frictional, decentralized asset markets and the role intermediation (market making) can play in alleviating these frictions. For most of the section we will follow the setup discussed in Weill (2020).

6.1 A Framework for Decentralized Asset Markets

Consider an environment with a unit mass of customers who can either hold 0 or 1 unit of an indivisible asset. The total supply of this asset in the economy is given as $s \in (0, 1)$. An agent holding the asset enjoys a flow utility δ. δ is stochastic and follows a two-stage process: It stays constant unless, with arrival rate γ, an adjustment shock hits. When the adjustment shock hits, the agents flow utility is drawn anew from a CDF $F(\delta)$ with support $\delta \in [0, 1]$ and mean $\bar{\delta}$. The new draw is independent of the previous δ, once the adjustment shock hits there is no persistence. The flow utility is a reduced form way to capture a variety of economic phenomena from housing services (if the asset is real estate), hedging services (swaps or other derivatives) or heterogeneous information (here δ can be seen as the perceived fundamental value / dividend stream of the asset). In addition to the customers, there is a mass of dealers who cannot store the assets and derive no utility from it.

The market environment of the economy is semi-centralized: Competitive dealers can trade the asset instantly in a centralized *dealer market*. In this market they can buy or sell at price P. The price at which they trade is endogenous and will reflect overall buying/selling pressure in the market. Customers do not have access to a centralized market and have to search for dealers to trade the asset. Upon meeting a dealer, customers bargain with them bilaterally about the price at which they buy/sell the asset. At any point in time, customers meet a dealer with arrival rate λ, but at most one at a time. On the other side, at any point in time each dealer is matched with a mass of customers. We do not allow for free entry in the dealer market – the mass of dealers is fixed exogenously – and hence dealers will be able to extract rents from customers in equilibrium.

Once a dealer and a customer meet, we assume that the price at which they trade is determined by Generalized Nash Bargaining. We will refer to the price at which a customer sells to the dealer as the *bid price* (B), whereas the price at which the customer buys from the dealer is referred to as *ask price* (A). Define as $V_q(\delta)$ the value to the customer of holding $q \in \{0, 1\}$ units of the asset when flow utility is δ.

Suppose that a customer with $q = 1$ meets a dealer. The bid price B at which he sells to the dealer has to satisfy

$$V_0(\delta) - V_1(\delta) + B \geq 0$$
$$P - B \geq 0$$

where the first equation is the surplus from trade for the customer and the second equation the surplus for the dealer. Both have to be weakly positive as otherwise either party could walk away from the trade. This can only be guaranteed if

$$P + V_0(\delta) - V_1(\delta) \geq 0$$

which ensures that the overall surplus from trading the asset is positive and, hence, is a necessary condition for trade to take place.

Denote as $\Delta V(\delta) \equiv V_1(\delta) - V_0(\delta)$ the customer's *reservation value*, which (from the customer surplus constraint above) is the minimum bid price B to ensure the customer is willing to make the trade. Further denote the bargaining weight of the dealer as θ. Then, from the definition of the dealer's surplus and the solution to Generalized Nash Bargaining, the bid price has to satisfy

$$P - B = \theta(P - \Delta V(\delta))$$

and hence

$$B(\delta) = \theta\Delta V(\delta) + (1-\theta)P.$$

Note here that the bid price will depend on the customers flow value from the the asset δ and – as the Nash bargaining solution – is the weighted average between the minimum ($\Delta V(\delta)$) and maximum (P) price to have both the customer and the trader agree to the transaction.

In a similar manner, for a customer with $q = 0$ and a dealer to trade upon meeting, the ask price A has to satisfy

$$\Delta V(\delta) - A \geq 0$$
$$A - P \geq 0$$

to have both the dealer and the customer agree to the transaction, with necessary conditions $\Delta V(\delta) - P \geq 0$ for positive total surplus. Combining the dealer surplus with the Nash bargaining solution as above, we get the ask price as

$$A(\delta) = \theta\Delta V(\delta) + (1-\theta)P$$

From the constraint on customer surplus you can see that the reservation value $\Delta V(\delta)$ is not only the minimum value for the customer to sell the asset, it is also the maximum value he would be willing to pay in order to purchase the asset. Hence, $\Delta V(\delta)$ captures the customers valuation of the asset. Additionally, note that the signs on $\Delta V(\delta)$ and P are flipped here compared to above as now the customer is buying from the dealer (which means the dealer needs to purchase the asset in the dealer market at price P) whereas before the dealer was buying from the customer (and then selling the asset at price P). We need $\Delta V(\delta) < P$ for a customer to sell the asset to the dealer but $\Delta V(\delta) > P$ for a customer to purchase the asset from a dealer.

We can summarize the setup in the equations for the flow values of customers with $q \in \{0, 1\}$ which are given as

$$rV_1(\delta) = \delta + \gamma[\bar{V_1} - V_1(\delta)] + \lambda \max\{B(\delta) - \Delta V(\delta), 0\}$$

and

$$rV_0(\delta) = \gamma[\bar{V_0} - V_0(\delta)] + \lambda \max\{\Delta V(\delta) - A(\delta), 0\}$$

where

$$\bar{V_1} = \int_0^1 V_1(\delta)dF(\delta) \quad \text{and} \quad \bar{V_0} = \int_0^1 V_0(\delta)dF(\delta)$$

capture the expectations over the change in the flow value conditional on an adjustment shock and are independent of the current δ as the new draw is i.i.d..

6.2 Equilibrium in Decentralized Asset Markets

The key equilibrium objects in the economy with decentralized asset markets are customers' reservation value $\Delta V(\delta)$ and the price in the dealer market P. To solve for an equilibrium we begin with the former. Note first that

$$\max\{B(\delta) - \Delta V(\delta), 0\} - \max\{\Delta V(\delta) - A(\delta), 0\} =$$
$$(1-\theta)\left[\max\{P - \Delta V(\delta), 0\} - \max\{\Delta V(\delta) - P, 0\}\right] = (1-\theta)(P - \Delta V(\delta))$$

We can use this expression to derive that

$$r\Delta V(\delta) = rV_1(\delta) - rV_0(\delta) = \delta + \gamma(\overline{\Delta V} - \Delta V(\delta)) + \lambda(1-\theta)(P - \Delta V(\delta))$$

where

$$\overline{\Delta V} = \bar{V}_1 - \bar{V}_0 = \int_0^1 V_1(\delta) dF(\delta) - \int_0^1 V_0(\delta) dF(\delta) = \int_0^1 \Delta V(\delta) dF(\delta)$$

is the average reservation value across all levels of δ. Taking expectations on both sides of $r\Delta V(\delta)$ we get that

$$r\overline{\Delta V} = \bar{\delta} + \gamma(\overline{\Delta V} - \overline{\Delta V}) + \lambda(1-\theta)(P - \overline{\Delta V})$$

and thus

$$\overline{\Delta V} = \frac{\bar{\delta} + \lambda(1-\theta)P}{r + \lambda(1-\theta)}$$

The average value the customers attach to the asset depends on the average flow utility they derive from holding it and the dealer market price P which impacts at which price the customers can sell the asset to the dealer. We can plug $\overline{\Delta V}$ back into the the expression for $r\Delta V(\delta)$ and rearrange to get that

$$\Delta V(\delta) = \frac{[r + \lambda(1-\theta)]\delta + \gamma\bar{\delta}}{[r + \lambda(1-\theta)][r + \gamma + \lambda(1-\theta)]} + \frac{\lambda(1-\theta)P}{r + \lambda(1-\theta)}$$

$\Delta V(\delta)$ is increasing in both in the current (actual, δ) as well as the expected future (potential, $\bar{\delta}$) utility from holding the asset and also in the dealer market price P.

To solve for the dealer market price P first note the following

1. A customer with $q = 0$ wants to buy the asset iff $\Delta V(\delta) \geq A(\delta) \Leftrightarrow \Delta V(\delta) \geq P$
2. A customer with $q = 1$ wants to sell the asset iff $\Delta V(\delta) \leq B(\delta) \Leftrightarrow \Delta V(\delta) \leq P$
3. $\Delta V(\delta)$ is strictly increasing in δ (see above).

This implies that there must be a *marginal customer* with flow utility δ^* who is indifferent btw holding the asset or not, i.e. $\Delta V(\delta^*) = P$. All customers with $\delta > \delta^*$ would buy the asset from the dealer if they do not own a unit of it already whereas all customers with $\delta < \delta^*$ would sell the asset to the dealer if they hold it and get the opportunity to trade.

At any point in time, a mass λs of customers holding the asset is matched with a dealer. Since it is purely random which agents meet a dealer at any point in time, a fraction $F(\delta^*)$ of those holding the asset will have a valuation low enough so that they would be willing to sell to the dealer. This yields total asset supply as $\lambda s F(\delta^*)$. On the other hand a mass $\lambda(1-s)$ of customers not holding an asset will meet the dealer, a fraction $(1-F(\delta^*))$ of which with a valuation high enough such that they are willing to buy. This gives total asset demand as $\lambda(1-s)(1-F(\delta^*))$. Note that in this argument we have made used of the fact that only customers can hold the asset and that every unit of the asset (with total supply of s) has to be held by someone.

To clear the asset market, because dealers cannot hold the asset themselves it has to hold at any point in time that demand equals supply and hence

$$\lambda s F(\delta^*) = \lambda(1-s)(1-F(\delta^*))$$

which we can re-arrange to solve for the utility of the marginal customer as

$$\delta^* = F^{-1}(1-s)$$

Note that this threshold is independent of the search friction λ and only depends on the distribution of flow utilities and the total supply of the asset. This implies that who will buy or sell the asset is not a function of market frictions. Who holds the asset, however, will be influenced by market frictions: The are a number of *frustrated customers* with either $q = 1$ and $\delta < \delta^*$ or $q = 0$ and $\delta > \delta^*$ who would be happy to trade the asset but cannot as they must first meet a dealer to be able to do so.

What we described above is an approach based on *net supply* and *net demand* of the asset, i.e. capturing only those agents who ultimately trade with the dealer. We could redo everything in terms of *gross supply* and *gross demand*. For this we would consider on both side of the market the mass $\lambda s(1-F(\delta^*))$ of agents who hold the asset (supply) but also want to keep the asset (demand). Adding this to our expressions for net supply and demand above, gross supply is given as λs and gross demand as $\lambda(1-F(\delta^*))$. Setting these two equal and solving for δ^* yields exactly the same condition as above. The gross approach is comparable to e.g. the Walrasian equilibrium in an endowment economy. Usually endowments are already distributed in the population but you sum them up to determine aggregate supply, which you then cross with the gross aggregate demand (the sum of the gross demands of all the agents) to get the price. Here you also implicitly assume that some agents "sell" their endowment to themselves.

Finally, we can solve for P by using the fact that $\Delta V(\delta^*) = P$ and for δ^* satisfying market clearing we get that

$$P = \frac{[r + \lambda(1-\theta)]\delta^* + \gamma\bar{\delta}}{r[r + \gamma + \lambda(1-\theta)]}$$

The impact of the arrival rate of trading opportunities λ on the dealer market price P is ambiguous: If trading opportunities are scarce, high valuation (high δ) are more eager to take any given opportunity, increasing the buying pressure and pushing P

upwards. On the other hand, scarcity of trading opportunities make low valuation (low δ) customers more eager to sell, putting downward pressure on P. The ultimate effect of λ on P depends on the balance of these two forces and is influenced also by the distributions of flow utilities and the total asset supply in the economy.

The model outlined here can be used to study e.g. price dispersion in financial markets (there is heterogeneity in the bid and ask prices of individual transactions), delays in trading or deviations of market clearing prices from their fundamental value. For further discussion of this framework, as well as possible extensions and applications – among others free entry of dealers, unrestricted asset holdings, multiple assets, or the dynamic market response to shocks – see Weill (2020).

References

Ahn, SeHyoun, Greg Kaplan, Benjamin Moll, Thomas Winberry, and Christian Wolf, "When inequality matters for macro and macro matters for inequality," *NBER macroeconomics annual*, 2018, *32* (1), 1–75.

Burdett, Kenneth and Dale T Mortensen, "Wage differentials, employer size, and unemployment," *International Economic Review*, 1998, pp. 257–273.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage bargaining with on-the-job search: Theory and evidence," *Econometrica*, 2006, *74* (2), 323–364.

_ , Stéphane Carcillo, and André Zylberberg, *Labor Economics*, MIT press, 2014.

Cooper, Russell and Andrew John, "Coordinating coordination failures in Keynesian models," *The Quarterly Journal of Economics*, 1988, *103* (3), 441–463.

Davis, Steven J, R Jason Faberman, and John C Haltiwanger, "The establishment-level behavior of vacancies and hiring," *The Quarterly Journal of Economics*, 2013, *128* (2), 581–622.

Diamond, Peter A, "Aggregate demand management in search equilibrium," *Journal of political Economy*, 1982, *90* (5), 881–894.

Hagedorn, Marcus and Iourii Manovskii, "The cyclical behavior of equilibrium unemployment and vacancies revisited," *American Economic Review*, 2008, *98* (4), 1692–1706.

Hall, Robert E, "Employment fluctuations with equilibrium wage stickiness," *American economic review*, 2005, *95* (1), 50–65.

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

Hosios, Arthur J, "On the efficiency of matching and related models of search and unemployment," *The Review of Economic Studies*, 1990, *57* (2), 279–298.

Jarosch, Gregor, "Searching for job security and the consequences of job loss," Technical Report, National Bureau of Economic Research 2021.

Jevons, William Stanley, "Money and the Mechanism ofExchange," *New York: D. Appleton & Co*, 1875.

Kiyotaki, Nobuhiro and Randall Wright, "A search-theoretic approach to monetary economics," *The American Economic Review*, 1993, pp. 63–77.

Lagos, Ricardo and Randall Wright, "A unified framework for monetary theory and policy analysis," *Journal of political Economy*, 2005, *113* (3), 463–484.

— , **Guillaume Rocheteau, and Randall Wright**, "Liquidity: A new monetarist perspective," *Journal of Economic Literature*, 2017, *55* (2), 371–440.

Ljungqvist, Lars and Thomas J Sargent, *Recursive Macroeconomic Theory*, MIT press, 2018.

McCall, John Joseph, "Economics of information and job search," *The Quarterly Journal of Economics*, 1970, pp. 113–126.

Moen, Espen R, "Competitive search equilibrium," *Journal of political Economy*, 1997, *105* (2), 385–411.

Mortensen, Dale, *Wage Dispersion: Why Are Similar Workers Paid Differently?*, MIT press, 2003.

Moscarini, Giuseppe and Fabien Postel-Vinay, "The cyclical job ladder," *Annual Review of Economics*, 2018, *10*, 165–188.

Nosal, Ed and Guillaume Rocheteau, *Money, Payments, and Liquidity*, MIT press, 2011.

Petrongolo, Barbara and Christopher A Pissarides, "Looking into the black box: A survey of the matching function," *Journal of Economic literature*, 2001, *39* (2), 390–431.

Petrosky-Nadeau, Nicolas and Etienne Wasmer, *Labor, Credit, and Goods Markets: The Macroeconomics of Search and Unemployment*, MIT Press, 2017.

Pissarides, Christopher A, *Equilibrium Unemployment Theory*, MIT press, 2000.

Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium wage dispersion with worker and employer heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

Rupert, Peter, Martin Schindler, Andrei Shevchenko, Randall Wright et al., "The search-theoretic approach to monetary economics: a primer," *Economic Review-Federal Reserve Bank of Cleveland*, 2000, *36* (4), 10–28.

Shimer, Robert, "The cyclical behavior of equilibrium unemployment and vacancies," *American economic review*, 2005, *95* (1), 25–49.

Trejos, Alberto and Randall Wright, "Search, bargaining, money, and prices," *Journal of political Economy*, 1995, *103* (1), 118–141.

Walsh, Carl E, *Monetary Theory and Policy*, MIT press, 2010.

Weill, Pierre-Olivier, "The Search Theory of Over-the-Counter Markets," *Annual Review of Economics*, 2020, *12*, 747–773.

Wright, Randall, Philipp Kircher, Benoît Julien, and Veronica Guerrieri, "Directed search and competitive search equilibrium: A guided tour," *Journal of Economic Literature*, 2021, *59* (1), 90–148.

A Technical Appendix

A.1 Discounting in Continuous Time

This brief section describes why, in continuous time, discounting is carried out in the form of $\beta = e^{-r}$. Start by recalling that given an interest rate r, $\beta = \frac{1}{1+r}$. Think of an asset with value v_t in discrete time, then

$$\begin{aligned} v_t(1+r) &= v_{t+1} \\ r &= \frac{v_{t+1} - v_t}{v_t} \end{aligned}$$

In continuous time the instantaneous rate of change of the value is $\frac{d}{dt}v(t)$, so that

$$r = \frac{\frac{d}{dt}v(t)}{v(t)}$$

or, equivalently

$$r = \frac{d}{dt}ln(v(t))$$

Use this definition to compare values that are Δ time apart, integrate both sides

$$\begin{aligned} \int_t^{t+\Delta} rds &= \int_t^{t+\Delta} \frac{d}{dt}ln(v(s))ds \\ r\Delta &= ln(v(t+\Delta)) - ln(v(t)) \\ r\Delta &= ln\left(\frac{v(t+\Delta)}{v(t)}\right) \end{aligned}$$

This implies that

$$e^{r\Delta} = \frac{v(t+\Delta)}{v(t)}$$

rearranging

$$v(t)e^{r\Delta} = v(t+\Delta)$$

This looks like the equation with started from in discrete time, if we set $\Delta = 1$ e^r. This implies that

$$\beta = e^{-r}$$

A.2 Duration

This section shows that the duration in the baseline McCall model is $\frac{1}{\lambda^*}$.

Start by noting that the probability an agent will be unemployed for arbitrary d periods from now is

$$Prob(D = d) = (1 - \lambda^*)^{d-1}\lambda^*$$

This is saying that the probability of being unemployed d periods is equal to the probability of rejecting $d - 1$ offers and accepting 1. From this, the cumulative distribution is

$$F(D) = \sum_{d=1}^{D}(1 - \lambda^*)^{d-1}\lambda^*$$

We know that (by the property of probability distributions) $F(\infty) = 1$. Redefine the cumulative, simply by having $t = d - 1$ and changing the starting point of the sum accordingly, as

$$F(\infty) = \sum_{t=0}^{\infty}(1 - \lambda^*)^{t}\lambda^* = 1$$

$F(\infty) = 1$ has to hold irrespective of λ^*, so the derivative with respect to λ^* satisfies

$$\begin{aligned}\frac{\partial F(\infty)}{\partial \lambda^*} &= \sum_{t=0}^{\infty}[-t(1 - \lambda^*)^{t-1}\lambda^* + (1 - \lambda^*)^{t}] = 0 \\ &\sum_{t=0}^{\infty} -t(1 - \lambda^*)^{t-1}\lambda^* = -\sum_{t=0}^{\infty}(1 - \lambda^*)^{t} \\ &\sum_{t=1}^{\infty} t(1 - \lambda^*)^{t-1}\lambda^* = \frac{1}{\lambda^*}\end{aligned}$$

Where the change of index in the last line is allowed cause the element of the summation for $t = 0$ is equal to 0.

Note that this is the expected duration. You can see this by noting that it takes the form $\mathbb{E}(x) = \sum xp(x)$. Hence $\mathbb{E}(d) = \frac{1}{\lambda^*}$.

A.3 Nash-Bargaining

Assume that a worker and a firm bargain over how to split a total surplus of S. Further assume that workers and firms have bargaining weights of ϕ and $1 - \phi$ respectively. The Nash-Bargaining outcome is defined as the solution to

$$\max_{S_w, S_f}(S_w)^{\phi}(S_f)^{1-\phi} \qquad s.t. \quad S = S_f + S_w$$

where S_f and S_w is the total surplus allocated to the firm and worker respectively.

The FOC for this problem implies that

$$\begin{aligned}&\phi(S_f) = (1 - \phi)S_w \\ &S_w = \phi S, \quad S_f = (1 - \phi)S\end{aligned}$$

B Additional Material – Labor Search

B.1 Mm-Ratio with Distinct Arrival Rates

In this section we derive the Mm-ratio for the basic search framework with on-the-job search and $\lambda_e \neq \lambda_u$. Assume for simplicity that the highest wage offered is normalized to 1, s.th. $F(1) = 1$. Integrating (9) by parts we have

$$
\begin{aligned}
R &= b + (\lambda_u - \lambda_e) \int_R^1 (V_e(w') - V_u) dF(w') \\
&= b + (\lambda_u - \lambda_e) \left[F(w')(V_e(w') - V_e(R))|_R^1 - \int_R^1 V_e'(w')F(w')dw' \right] \\
&= b + (\lambda_u - \lambda_e) \left[(F(1)V_e(1) - F(R)V_e(R) - F(1)V_e(R) + F(R)V_e(R)) - \int_R^1 V_e'(w')F(w')dw' \right] \\
&= b + (\lambda_u - \lambda_e) \left[(F(1)V_e(1) - F(1)V_e(R)) - \int_R^1 V_e'(w')F(w')dw' \right] \\
&= b + (\lambda_u - \lambda_e) \left[V_e(1) - V_e(R) - \int_R^1 V_e'(w')F(w')dw' \right] \\
&= b + (\lambda_u - \lambda_e) \left[\int_R^1 V_e'(w')dw' - \int_R^1 V_e'(w')F(w')dw' \right] \\
&= b + (\lambda_u - \lambda_e) \int_R^1 V_e'(w')(1 - F(w'))dw' \\
&= b + (\lambda_u - \lambda_e) \int_R^1 \frac{1 - F(w')}{r + q + \lambda_e(1 - F(w'))} dw' \qquad \text{(B1)}
\end{aligned}
$$

where for the last step we use the derivative of the value function with respect to the wage which is $V_e'(w') = \frac{1}{r+q+\lambda_e(1-F(w'))}$.[18] This gives the reservation wage solution as a function of only exogenous parameters. Note that when we impose that on the job search is not allowed ($\lambda_e = 0$), the problem collapses to the previous formulation of the model.

From this result it is possible to re-evaluate Hornstein et al. (2011). In this context we want to show that the mean-min ratio is

$$
Mm \approx \frac{\frac{\lambda_u - \lambda_e}{r+q+\lambda_e} + 1}{\frac{\lambda_u - \lambda_e}{r+q+\lambda_e} + \rho},
$$

provided that r is small relative to q.

To show that this is the case we make two assumptions that simplify the problem: There are no mass points in $F(w)$ and all offers are above the reservation wage. We start by working out how many workers are earning wages below or equal to

[18]This can be shown by reshuffling (7), applying Leibniz rule to the integral over future wages and noting that this integral is 0 at its lower bound.

w. Define the stock of agents employed at wages below w as $E(w)$. Then the flow equation of this mass is

$$\dot{E}(w) = u\lambda_u F(w) - E(w)(q + \lambda_e(1 - F(w)))$$

This equation describes the net change of the mass of workers $E(w)$. The first terms states that out the unemployed agents some, with probability λ_u will receive a wage offer, out of those $F(w)$ will receive (and accept since they are all above the reservation wage) an offer lower or equal to w; the second term states that out the agents employed at w or less some will have their match destroyed (with probability q) and some other will receive, with probability λ_e a wage offer that will be above w with probability $1 - F(w)$ and will therefore accept it. Note that here $E(w)$ is almost like a CDF of workers employed at wages below w, the difference being that it does not integrate to 1, i.e. $E(1) = 1 - u$. To obtain a CDF of the wages accepted define $G(w) = \frac{E(w)}{1-u}$ where $G(w)$ is the distribution of wages among the employed.

By substituting the latter result into the flow equation of employment and noting that in steady state $\dot{E} = 0$

$$\begin{aligned}
&0 = u\lambda_u F(w) - E(w)(q + \lambda_e(1 - F(w))) \\
&\frac{q}{\lambda_u + q}\lambda_u F(w) = E(w)(q + \lambda_e(1 - F(w))) \\
&\frac{q}{\lambda_u + q}\lambda_u F(w) = \frac{\lambda_u}{\lambda_u + q} G(w)(q + \lambda_e(1 - F(w))) \\
&qF(w) = G(w)(q + \lambda_e(1 - F(w))) \\
&G(w) = \frac{qF(w)}{q + \lambda_e(1 - F(w))}
\end{aligned}$$

where we use $u = \frac{q}{q+\lambda_u}$ when we assume $F(R) = 0$. This is a mapping between the distribution of wage offers and the distribution of wages of employed agents. Then it is possible to compute the distribution of wages better than w:

$$1 - G(w) = \frac{(q + \lambda_e)(1 - F(w))}{q + \lambda_e(1 - F(w))}$$

Here, making use of the assumption of r being small relative to q we can add it to both the numerator and the denominator r

$$1 - G(w) \approx \frac{(r + q + \lambda_e)(1 - F(w))}{r + q + \lambda_e(1 - F(w))} \tag{B2}$$

Now we can compute the average wage

$$
\begin{aligned}
w^* &= \int_R^1 w dG(w) \\
&= [wG(w)]_R^1 - \int_R^1 G(w)dw \\
&= 1 - RG(R) - \int_R^1 G(w)dw \\
&= \pm R + 1 - \int_R^1 G(w)dw \\
&= R + \int_R^1 (1 - G(w))dw
\end{aligned} \tag{B3}
$$

Where the 4th equation comes from $G(R) = 0$ due to the assumption that job offers are all above R and the last equation from $(1 - R) = \int_R^1 1dw$. Now go back to (B1):

$$
R = b + (\lambda_u - \lambda_e) \int_R^1 \frac{1 - F(w)}{r + q + \lambda_e(1 - F(w))} dw
$$

The object inside the integral, can be substituted using (B2)

$$
R \approx b + (\lambda_u - \lambda_e) \int_R^1 \frac{1 - G(w)}{r + q + \lambda_e} dw
$$

Which, making use of (B3), becomes

$$
R \approx b + \frac{\lambda_u - \lambda_e}{r + q + \lambda_e}(w^* - R)
$$

By assuming again that $b = \rho w^*$ we have that

$$
\frac{w^*}{R} \approx \frac{\frac{\lambda_u - \lambda_e}{r+q+\lambda_e} + 1}{\frac{\lambda_u - \lambda_e}{r+q+\lambda_e} + \rho}
$$

Which is what we wanted to show from Hornstein et al. (2011). By having $\lambda_e = 0$ we can retrieve the previous formulation of the Mm-ratio. Note that separation does not play a major role whereas quantitatively on-the-job search does matter (much fewer people become unemployed than move from one job directly to another). From the formula, the larger λ_u relative to λ_e, the closer the Mm-ratio gets to 1 as before. The closer λ_u and λ_e are, the closer the Mm-ratio gets to 2.5 (as in the case with $\lambda_u = \lambda_e$ discussed in the main text). Taking the empirical fact of a median Mm-ratio across labor markets of about 2, we need λ_u and λ_e close to each other for the model to match the data.

B.2 DMP-Equilibrium: Out-of-Steady-State Dynamics

The model described above was all in steady state, but it also possible to describe the behavior of the system out of the steady state.
Start by rewriting the values of the problem

$$
\begin{aligned}
& rV_u = b + \lambda(\theta)(V_e(w) - V_u) + \dot{V}_u \\
& rV_e(w) = w - q(V_e(w) - V_u) + \dot{V}_e(w) \\
& c = \delta(\theta)J(w) \\
& rJ(w) = p - w - qJ(w9 + \dot{J}(w)
\end{aligned}
$$

where the terms with the dot factor in derivations from steady state.

The surplus is the same as before ($S(w) = J(w) + V_e(w) - V_u$) and we assume that wages are continuously renegotiated.

Then we can reduce the problem to

$$
\begin{aligned}
\dot{S} - (r + \delta + \phi\lambda(\theta))S(w) + p - b = 0 \\
q(\theta)(1 - \phi)S(w) = c
\end{aligned}
$$

The solution of the steady state problem is still a solution (all dotted values equal to zero), the system however now allows for out of steady state analysis. In particular if one starts at $u \neq u^*$, then v adjusts immediately and θ jump to the steady state levels by free entry, whereas u adjust slowly according to the following dynamics

$$\dot{u} = \delta(1 - u) - \alpha(\theta^*)u.$$

θ is the only state variable which matter for firms vacancy posting decisions and workers value functions. The fact that it is a jump variable ensures that from any starting value of u we will always immediately jump to the equilibrium market tightness. Then (as Figure 8 shows below), we will move along the job-creation curve while u converges back to its steady state value. Even though it does not influence agents' decisions, u is a state of the aggregate economy as it impacts the future dynamics of unemployment.

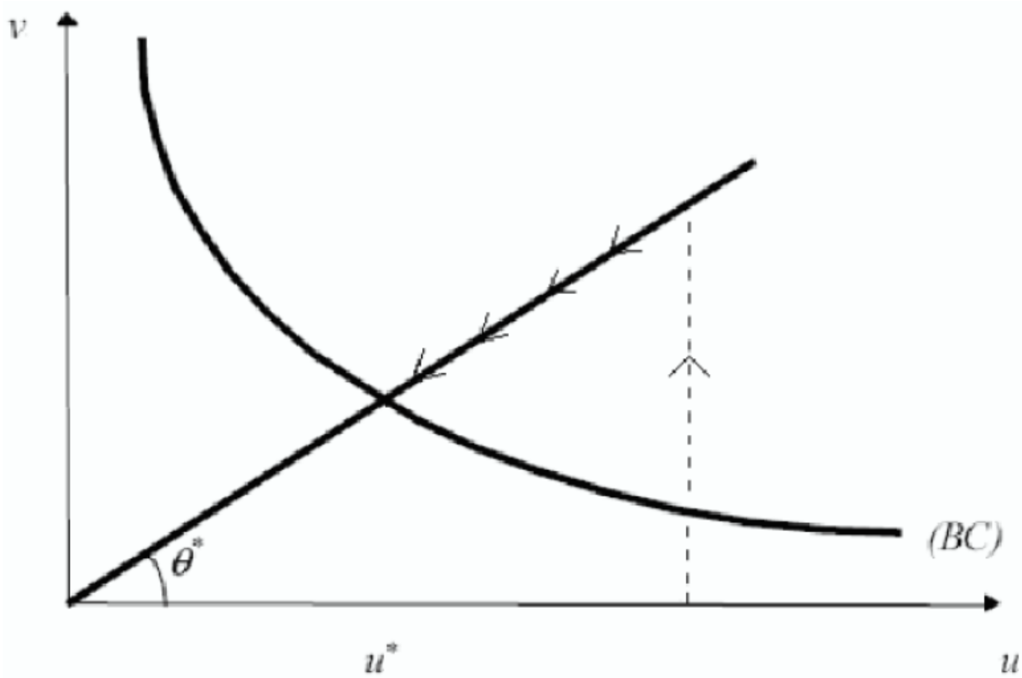


Figure 8 – Out of Steady State Dynamics

B.3 DMP-Equilibrium: Shimer Puzzle

In the DMP model described before firms productivity was exogenous and fixed, now by assuming that productivity follows a stochastic process, where $p_{t+1} = f(p_t)$, we can have the model speak to the behavior of labour markets along the business cycle.

Firms and workers now make optimal decisions in which the relevant state variables are p_t and u_t. However since θ_t is a jump variable and immediately adjusts to its equilibrium value (see out-of-SS dynamics above) then u_t ceases to be a state variable.

In discrete time the model has the following values, when p denotes today's productivity and p' next period's productivity:

$$V_u(p) = b + \beta[\lambda(\theta(p))E_pV_e(p') + (1 - \lambda(\theta_y))E_pV_u(p')]$$
$$V_e(p) = w(p) + \beta[(1 - q)E_pV_e(p') + qE_pV_u(p')]$$

And for the firm

$$c = \delta(\theta(p))E_pJ(p')$$
$$J(p) = p - w(p) + \beta[(1 - q)E_pJ(p')]$$

where the free entry condition is in expectation as the first production will occur in the next period under a potentially new p.

Wage determination occurs again through Nash bargaining over

$$S(p) = J(p) + V_e(p) - V_u(p)$$

and hence

$$J(p) = (1-\phi)S(p)$$
$$V_e(p) - V_u(p) = \phi S(p)$$

Then, from free entry

$$c = \delta(\theta(p))E_p(1-\phi)S(p')$$

and from the value of a job to the firm

$$(1-\phi)S(p) = p - w(p) + \frac{1-q}{\delta(\theta(p))}c$$

It is possible then to substitute into the definition of the surplus

$$\begin{aligned} S(p) =& p - b + \beta(1-q)E_pJ(p') \\ &+ \beta E_p[(1-q)V_e(p') + qV_u(p') - \lambda(\theta(p))E_pV_e(p') - (1-\lambda(\theta(p)))V_u(p')] \end{aligned}$$

Which can be rewritten as

$$S(p) = p - b + \beta(1-q)E_p[J(p') + V_e(p') - V_u(p')] - \beta\lambda(\theta(p))E_y[V_e(p') - V_u(p')]$$

Note that the first square bracket is $S(p')$ and the second one is $\phi S(p')$. By free entry we know that

$$E_yS(p') = \frac{c}{\beta\delta(\theta(p))(1-\phi)}$$

Hence

$$S(p) = p - b + [1 - q - \lambda(\theta(p))\phi]\frac{c}{\delta(\theta(p))(1-\phi)}$$

Finally from the firm's surplus

$$w(p) = \phi p + (1-\phi)b - c\phi\theta(p)$$

which is the wage equation. The first two elements show that the bargaining weights move the wage closer to the output or to the unemployment benefit. Also note that a low bargaining power for the worker keeps wages rigid since they move less with p and they move less with θ which is the variable that adjust the fastest.

In this context it is now possible to discuss the so called *Shimer Puzzle*. The Shimer Puzzle consists of the observation that in this model wages respond way too much to output fluctuations, thereby failing to mimic the observed volatility of vacancy posting through the business cycle. In particular, taking the Shimer (2005) parametrization (consistent with the Hosios efficiency condition), say that output p is normalized to 1 in normal times and assume that it goes to 0.98 in downturns and to 1.02 in upturns. Fix the unemployment benefit to 0.4 (ini the US the replacement rate of unemployment benefits is around 40% on average), then wages turn out to be 0.96

in downturns and 1 in upturns. Since the profit margin for firms is approximately the same along the cycle the model fails to replicate the volatility of vacancy posting and the implied tightness of the market.The reason for this behavior of the model is that the Hosios condition applied to the estimated matching functions require a $\phi \approx 0.7$ which implies that workers are able to keep the part of the surplus that goes to firm relatively low, having wages follow output closely.

Two solutions to this problem have been proposed by Hagedorn and Manovskii (2008) and Hall (2005). The former comes from the analysis of the final wage equation with aggregate risk. In particular note that in order to have wages respond less one would need to reduce the workers' bargaining power. However these models are always calibrated to match long run unemployment, with a lower ϕ the only way to keep matching it is by having a higher c which, by the wage equation would work against the goal of having a more stable wage and vacancies along the cycle. Therefore Hagedorn and Manovskii (2008) propose to increase the unemployment benefit to account for utility of leisure, in order to get the correct cycle of vacancies and tightness they have $b \approx 0.94$. Hall (2005)'s solution instead consists of sticky wages, which intuitively will make wages respond less to output fluctuations, generating the observed volatility of vacancies.