Inefficient Unemployment and Bargaining Friction*

C. Y. Kelvin Yuen†

Hong Kong University of Science and Technology (Guangzhou);

Hong Kong University of Science and Technology

March 23, 2026

Abstract

This paper examines how bargaining frictions generate privately inefficient job separations and amplify unemployment fluctuations. I propose a simple specification of bargaining friction by including bargaining wedges in the standard Nash bargaining model. Such bargaining wedges arise when, for example, wages are determined by alternating offers bargaining, which is often used in the literature to generate real wage rigidity, or when there is asymmetric information about worker's productivity. I show that due to the misalignment between actual surpluses and bargaining surpluses, inefficient separations could be generated, which would in turn induce inefficient unemployment. I highlight a distinct amplification mechanism that operates through the separation margin. The existence of inefficient unemployment due to bargaining friction could accentuate the fluctuation of unemployment. Quantitatively, I find that inefficient unemployment accounts for up to 30% of the total unemployment volatility in the calibrated model.

JEL classification: C78, E24, E32, J63, J64

Keywords: inefficient unemployment, bargaining friction, inefficient separations, unemployment volatility, alternating offers bargaining, asymmetric information

---

*I would like to thank Rody Manuelli, Yongseok Shin, Ping Wang for their comments and advice. I am also grateful to an anonymous Associate Editor, two anonymous referees, participants at Midwest Macroeconomics Meetings in Madison, WI, Asian Meeting of the Econometric Society in Seoul, Korea, and internal seminars at WashU for helpful comments and suggestions.

†Urban Governance and Design Thrust, The Hong Kong University of Science and Technology (Guangzhou) and Department of Economics, The Hong Kong University of Science and Technology. Email: kelvinyuen@ust.hk

1 Introduction

Labor market dynamics are profoundly influenced by the nature of labor turnovers. An efficient job separation occurs when the employment relationship ends due to a negative joint surplus from matching.[1] If wages are fully flexible, then any inefficient separation can be avoided by wage renegotiation.[2] However, such Pareto improvement may be infeasible when there are labor market frictions.[3] Such inefficiency in labor turnovers is quantitatively substantial. Gielen and van Ours (2006) show that while inefficient quits are insignificant, almost half of all predicted layoffs are inefficient in an estimated structural model. They argue that the substantial amount of inefficient layoffs may be due to downward wage rigidity. Recently, Jäger et al. (2023) show more direct evidence of inefficient job separations by exploiting a temporary unemployment insurance (UI) extension in Austria. They find that those surviving jobs in the treatment group after a negative UI shock, presumed to have a higher joint surplus, subsequently behaved identically to a control group, which is inconsistent with the efficient-turnover hypothesis. In addition, inefficient separations may amplify macroeconomic volatility due to excessive labor turnovers.[4] Therefore, it is crucial to consider the possibility of inefficient job separations for a comprehensive understanding of unemployment dynamics.

Standard search and matching models á la Mortensen and Pissarides (1994), henceforth MP, with endogenous separation have assumed that wages are determined by Nash bargaining. One implication of the flexible wage rule is that job turnovers are always privately efficient and mutually agreed to. Hence, there are no involuntary layoffs or quits. As pointed out by the seminal critique by Shimer (2005), such models fail to explain the unemployment volatility observed in the data, since wages are too flexible in the model. Subsequently, researchers have attempted to generate a larger response of unemployment to productivity change by incorporating real wage stickiness (e.g. Hall, 2005c; Gertler and Trigari, 2009).

[1] In this paper, I focus on the private efficiency of job separations. There has been an extensive literature focusing on the social efficiency of unemployment in search and matching models, see, e.g., Hosios (1990) and Guerrieri (2008).

[2] McLaughlin (1991) shows that the quit-layoff distinction is inconsequential when all separations are efficient.

[3] For example, Hall and Lazear (1984) and Haltiwanger (1984) demonstrate that asymmetric information regarding worker productivity and outside options can lead to inefficient job turnovers, resulting in excessive layoffs and quits. Also, there could be inefficient separations when renegotiation is costly (Antel, 1985) and when there exists firm-specific capital (Becker, 1962) so that the worker would demand a higher wage to reap the return of the investment. A more recent study by Gerali et al. (2021) shows that excessive firing costs may also lead to inefficiency in job separations.

[4] For instance, Ramey and Watson (1997) and den Haan et al. (1999) show that contracting imperfections such as limited verifiability and liquidity constraints may lead to privately inefficient separations, which magnify the propagation of business cycle shocks through the fluctuations in job destruction.

However, they largely maintain the assumption, without justification, that all employment pairs are efficient.

Recently, replacing Nash bargaining with alternating offers bargaining (AOB) in the otherwise standard search and matching model, first proposed by Hall and Milgrom (2008), has achieved some success in explaining labor market dynamics. Due to the different bargaining protocol, wages become less sensitive to the labor productivity, generating real wage rigidity endogenously. In fact, Christiano et al. (2016) show that the AOB model outperforms both Nash bargaining and Calvo sticky wage models in matching the dynamic responses to macroeconomic shocks in the data.[5] While the real wage rigidity derived in the model can potentially produce inefficient job turnovers, the literature is mostly silent about the efficiency of separations in this class of models.

I construct an otherwise standard MP model with bargaining friction, where wages are determined by Nash bargaining with the existence of wedges between the actual surplus and the respective bargaining surplus. Such bargaining wedges can be considered as a reduced-form specification of some bargaining friction. In this model, workers and firms would still separate efficiently in response to a sufficiently low match-specific productivity, leading to a negative joint match surplus. However, due to the misalignment between actual surpluses and the bargaining surpluses, inefficient separations could be generated when the match-specific productivity lies in a certain range where inefficiency may arise. The inefficiency gap depends positively on the *net bargaining wedge* between the worker and the firm. If the net wedge is positive (resp. negative), there may be inefficient layoffs (resp. quits). As a result of the excessive job separations, inefficient unemployment may also arise.

Economically, a bargaining wedge captures situations in which the bargaining-relevant threat points differ from the pair's true continuation values. This can arise from protocols that distinguish between outside options and disagreement payoffs, from delay or renegotiation costs that make disagreement payoffs state dependent, or from informational frictions in which continuation values are not contractible or not commonly known. In all these cases, the wage bargain is formed using distorted "bargaining surpluses," which can shift wages and—crucially—can also shift the privately optimal separation decision away from the joint-surplus criterion. The quantitative calibration emphasizes that the required wedge is small relative to match surplus, making it plausible that modest bargaining frictions could have sizable aggregate consequences. The advantage of this generalized representation is that it isolates a single sufficient statistic—the net bargaining wedge—that governs the direction of

[5] Christiano et al. (2016) demonstrate that although Nash bargaining and AOB models produce similar likelihood functions in matching empirical impulse response functions, the estimated Nash model requires a high replacement ratio, as in Hagedorn and Manovskii (2008).

separation inefficiency and allows the quantitative analysis to proceed without committing to one specific bargaining micro-foundation.

To justify the bargaining wedge specification, I show that some bargaining protocols used in the literature to generate higher unemployment volatility belong to the same class. For example, the solution to the intra-period version of the AOB model by Hall and Milgrom (2008), also adopted by Christiano et al. (2016), can be considered as a micro-foundation of the model considered in this paper. Here, the inefficiency comes from the possibility of costly delay to reach an agreement, and the existence of flow benefit of the worker and flow cost of the firm during bargaining. Hall and Milgrom (2008) distinguish between outside-option payoff and disagreement payoff. This view is consistent with the existence of bargaining wedges when the actual surplus is different from the bargaining surplus. In this case, the net bargaining wedge is positive, resulting in some inefficient layoffs.

Such bargaining wedges may also arise when there is asymmetric information about the match-specific productivity. Kennan (2010) applies the neutral bargaining solution (NBS) of Myerson (1984) when the firm has private information about the productivity of the worker. However, he assumes the worker always proposes a pooling offer, resulting in efficient employment. By allowing the worker to choose optimally also the screening offer, there may be inefficient job separations. I show that in this case, the NBS is equivalent to the Nash bargaining solution when there is a positive net bargaining wedge, and the firm has all the bargaining power due to private information.

By allowing endogenous and potentially inefficient separations, bargaining frictions generate two distinct amplification mechanisms. The first is a *small-surplus channel* (Ljungqvist and Sargent (2017)): when the net bargaining wedge is positive, it effectively reduces the fundamental surplus that governs vacancy creation, making market tightness more sensitive to aggregate shocks. The second—and central contribution of this paper—is a *separation channel*. Because the bargaining wedge is not mechanically proportional to aggregate productivity y_t, the privately optimal separation region can move substantially in response to small changes in y_t. This creates large countercyclical fluctuations in separations, including *privately inefficient* layoffs (or quits), and thereby amplifies unemployment volatility beyond what the small-surplus channel alone can generate.

This separation channel is distinct from the role of endogenous job destruction in standard DMP models with flexible Nash bargaining. In those models, even when separations are endogenous and productivity is heterogeneous, the wage bargain aligns private continuation incentives with the joint surplus, implying privately efficient separations. In contrast, once bargaining surpluses differ from actual surpluses, the privately optimal separation cutoff generically differs from the efficient cutoff. This wedge-driven misalignment produces an additional propagation margin: inefficient separations rise in recessions at the same time as vacancy creation weakens, jointly generating much larger unemployment fluctuations.

Quantitatively, bargaining friction has a large impact on the volatility of the unemployment rate. Specifically, I show that the unemployment volatility in the economy increases with the net bargaining wedge. As a result, by including a relatively small amount of bargaining friction, the model can generate substantial labor market fluctuation without requiring a high replacement ratio calibration (as in Hagedorn and Manovskii (2008)). To put the magnitude in perspective, the calibrated bargaining friction is about 3.55% of the average joint surplus, yet it has large aggregate effects through the separation channel. Also, the small-surplus channel contributes only 6% of the total effects of bargaining friction, which shows that separation channel is the dominant force in affecting the unemployment volatility. I show that in the calibrated model, inefficient unemployment constitutes about 29% of the total unemployment volatility. Moreover, the peak of the unemployment rate in the US during the Great Recession would have been 3 percentage points lower if all separations were efficient. Lastly, I find that while the existence of bargaining friction has a relatively minor effect on job finding rate, it substantially increases the volatility of the separation rate.

This paper focuses on the role of job separations in the labor market dynamics.[6] The literature has been divided on whether job separation rate contributes to a significant fraction of the unemployment volatility.[7] Moreover, the literature on the unemployment volatility mostly assumes that the separation rate is constant and exogenous.[8] Mortensen and Nagypàl (2007) and Fujita and Ramey (2012) show that by allowing endogenous separations, the MP model is able to generate a sizable, albeit not sufficient, amount of unemployment volatility. Hence, this paper adds to the literature on endogenous separations by considering also inefficient job separations.

Few studies explore privately inefficient separations within the standard MP framework.[9]

[6]Specifically, in this paper I consider job separations only as transitions from employment to unemployment. Other authors in the literature also consider separations into employment (i.e. job-to-job transitions) and inactive states.

[7]Hall (2005a) and Shimer (2012) argue that the job separation rate into unemployment is roughly acyclical over the business cycle. On the other hand, Elsby et al. (2009), Fujita and Ramey (2009), and Pissarides (2009) show that empirically job separations are countercyclical and contribute substantially to the unemployment volatility. More recently, Coles and Moghaddasi Kelishomi (2018) find, by relaxing the free entry condition, that job destructions are the main driver of unemployment volatility.

[8]Recently, Ljungqvist and Sargent (2017) survey the literature and find that it is the fundamental surplus that matters for the elasticity of market tightness. But they do not consider the case of endogenous separations.

[9]Hall (2005b) compares the performance of several types of models and concludes that the sticky-wage efficient-separations model of Hall (2005c) performs much better than its inefficient counterpart from the

In the model considered in this paper, wages can still be flexible, and inefficient separations are attributed to some bargaining friction. Also, this paper is related to the theoretical literature on inefficient bargaining.[10] To the best of my knowledge, this paper is the first to bring the notion of inefficient bargaining outcome to the equilibrium model of unemployment.

The rest of the paper is organized as follows. The baseline model is constructed in Section 2, along with the discussion of the efficiency of job separations and unemployment. In Section 3, I analyze the AOB model by Hall and Milgrom (2008), showing it falls within the same model class. I introduce asymmetric information about the match-specific productivity, as in Kennan (2010), in Section 4. Then, in Section 5, I discuss how bargaining friction impacts the volatility of cyclical unemployment. In Section 6, I calibrate the model to the US economy, and evaluate quantitatively the impacts of bargaining friction on the labor market dynamics. Section 7 concludes.

2 The model economy

In this section, I build a standard search and matching model á la Mortensen and Pissarides (1994) with endogenous separations and ex-post heterogeneity. Diverging from their framework, my model introduces bargaining wedges within the Nash bargaining process. I show that the wedges would generate inefficient separations. Models with alternating offers bargaining and asymmetric information will be shown to belong to the same class. Lastly, I discuss how inefficient separations would induce a higher-than-efficient level of unemployment, or *inefficient unemployment*.

2.1 The MP environment

Workers and firms are risk-neutral and infinitely lived with a common discount factor β. They randomly match in the job market subject to search friction. Once matched, the firm-worker pair produces a flow value of output $p_t(\varepsilon_t)$, where ε_t is the idiosyncratic, or match-specific, productivity specific only to the pair and $p_t'(\cdot) > 0$. To model the persistence

behavior of the flows in the labor market. However, as pointed out by Robert Shimer in the discussion of the lecture, Hall (2005b) considers only the inefficiency arising from a constant wage model. Also, Blanchard and Galì (2010) generate inefficient unemployment fluctuations in a model with nominal and real wage rigidities. However, their model does not allow for endogenous separations.

[10] For example, Chatterjee and Samuelson (1987), Kennan and Wilson (1993), and the references therein find that incomplete information may lead to inefficient bargaining outcomes. Also, Busch and Wen (1995), and Anderlini and Felli (2001) find that inefficiencies may arise when there are transaction costs in the process of bargaining.

of the idiosyncratic shock, I follow Mortensen and Pissarides (1994) and Fujita and Ramey (2012) to assume a probability λ of switching productivity. Hence, with probability $1-\lambda$, we have $\varepsilon_{t+1}=\varepsilon_t$. Otherwise, ε_{t+1} will be drawn from the stationary distribution $G(\cdot)$ on a compact support $[\varepsilon_{\min},\varepsilon_{\max}]$. There is a probability s of exogenous separation. Otherwise, in each period, firms and workers can decide endogenously whether to dissolve the employment relationship. There is also a standard matching function $m(u_t,v_t)$ where u_t and v_t are the measures of unemployed workers and vacancies respectively. The market tightness at time t is $\theta_t=\frac{v_t}{u_t}$. Hence, the probability of an unemployed worker meeting with a vacancy is $\phi(\theta_t)=\frac{m(u_t,v_t)}{u_t}=m(1,\theta_t)$, and that of a vacancy meeting with an unemployed worker is $q(\theta_t)=\frac{\phi(\theta_t)}{\theta_t}$.

2.1.1 Value functions

Firms post vacancies with a flow cost c. Let P_t be the (present discounted) value of production and W_t be the expected present value of wages during the time of the production.[11] Hence, a productive and working firm would receive a value of $P_t(\varepsilon_t)-W_t(\varepsilon_t)$. On the other hand, an unemployed worker earns a value of U_t. Once employed in a firm, the worker would receive the value of wage W_t and the value of working V_t.[12] Therefore, the firm would fire the worker when $P_t(\varepsilon_t)\le W_t(\varepsilon_t)$, and the worker would quit the job if $W_t(\varepsilon_t)+V_t(\varepsilon_t)\le U_t$.

Let $\chi_t(\varepsilon_t)$ is the indicator function of endogenous separation at time t, i.e.

$$\chi_t(\varepsilon_t)=\begin{cases}1 & \text{when } \underbrace{P_t(\varepsilon_t)\le W_t(\varepsilon_t)}_{\text{layoff}} \text{ or } \underbrace{W_t(\varepsilon_t)+V_t(\varepsilon_t)\le U_t}_{\text{quit}}\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

Hence, the value of production is given by

$$P_t(\varepsilon_t)=p_t(\varepsilon_t)+\beta(1-s)\mathbb{E}_t\begin{bmatrix}\lambda\left(1-\chi_{t+1}(\varepsilon_{t+1})\right)P_{t+1}(\varepsilon_{t+1})\\ +(1-\lambda)\left(1-\chi_{t+1}(\varepsilon_t)\right)P_{t+1}(\varepsilon_t)\end{bmatrix} \tag{2}$$

[11] Here I follow Hall and Milgrom (2008) and Christiano et al. (2016) to use the discounted value of wages instead of the flow value of wages which would simplify the notations later on when I consider alternating offers bargaining.

[12] It is worth noting that since $W_t(\varepsilon_t)$ denotes the present discounted value of wages, it is useful to separate the worker's employment value into a wage component and a non-wage component. Formally, if $E_t(\varepsilon_t)$ denotes the standard MP value of being employed (in flow-wage notation), then $E_t(\varepsilon_t)=W_t(\varepsilon_t)+V_t(\varepsilon_t)$, so $V_t(\varepsilon_t)$ can be interpreted as the value of being employed net of the present value of wages (i.e. the value of career only). This decomposition is convenient for bargaining protocols that negotiate over W_t while taking continuation (career) values as given. Appendix A.1 provides the mapping between the present-value wage formulation and the conventional flow-wage formulation.

Free entry of the firms implies that

$$\beta q\left(\theta_{t}\right) \mathbb{E}_{t}\left[\left(1-\chi_{t+1}\left(\varepsilon_{t+1}\right)\right)\left(P_{t+1}\left(\varepsilon_{t+1}\right)-W_{t+1}\left(\varepsilon_{t+1}\right)\right)\right]=c \tag{3}$$

where ε_{t+1} denotes the idiosyncratic productivity at time $t+1$ which is random at time t. The free entry condition states that the flow cost equals expected profit of the firm in the future. Note that there is a possibility of immediate separation, should the value of production be not enough to cover that of the wage, or the worker is rejecting the job offer.

Unemployed workers receive a flow value b of non-market activity, which may include such things as leisure and unemployment benefit. Hence, the value of an unemployed worker is given by

$$U_{t}=b+\beta \mathbb{E}_{t}\left[\begin{array}{c}\phi\left(\theta_{t}\right)\left(1-\chi_{t+1}\left(\varepsilon_{t+1}\right)\right)\left(W_{t+1}\left(\varepsilon_{t+1}\right)+V_{t+1}\left(\varepsilon_{t+1}\right)\right) \\ +\left[\left(1-\phi\left(\theta_{t}\right)\right)+\phi\left(\theta_{t}\right) \chi_{t+1}\left(\varepsilon_{t+1}\right)\right] U_{t+1}\end{array}\right] \tag{4}$$

and the value of working is given by[13]

$$V_{t}\left(\varepsilon_{t}\right)=\beta \mathbb{E}_{t}\left\{\begin{array}{c}\left[s+(1-s)\left(\lambda \chi_{t+1}\left(\varepsilon_{t+1}\right)+(1-\lambda) \chi_{t+1}\left(\varepsilon_{t}\right)\right)\right] U_{t+1} \\ +(1-s)\left[\begin{array}{c}\lambda\left(1-\chi_{t+1}\left(\varepsilon_{t+1}\right)\right) V_{t+1}\left(\varepsilon_{t+1}\right) \\ +(1-\lambda)\left(1-\chi_{t+1}\left(\varepsilon_{t}\right)\right) V_{t+1}\left(\varepsilon_{t}\right)\end{array}\right]\end{array}\right\} \tag{5}$$

2.1.2 Match surplus and thresholds

The joint surplus, worker's surplus, and firm's surplus are given, respectively, by

$$S_{t}\left(\varepsilon_{t}\right)=P_{t}\left(\varepsilon_{t}\right)+V_{t}\left(\varepsilon_{t}\right)-U_{t} \tag{6}$$

$$J_{t}\left(\varepsilon_{t}\right)=V_{t}\left(\varepsilon_{t}\right)+W_{t}\left(\varepsilon_{t}\right)-U_{t} \tag{7}$$

$$F_{t}\left(\varepsilon_{t}\right)=P_{t}\left(\varepsilon_{t}\right)-W_{t}\left(\varepsilon_{t}\right) \tag{8}$$

whenever they are positive, and zero otherwise. I assume the wage rule is such that the above surplus functions are all strictly increasing and continuous in ε_t (which would apply to the Nash bargaining model I consider in the next subsection). Hence, there exist cutoff

[13] Alternatively, the value of working can be interpreted as the difference between the value of being employed and the value of wages. See Appendix A.1 for a discussion.

values $\varepsilon_t^e, \varepsilon_t^w$ and ε_t^f such that

$$S_t\left(\varepsilon_t^e\right) = 0 \tag{9}$$

$$J_t\left(\varepsilon_t^w\right) = 0 \tag{10}$$

$$F_t\left(\varepsilon_t^f\right) = 0 \tag{11}$$

Intuitively, the worker will choose to quit the job if she observes a productivity $\varepsilon < \varepsilon_t^w$. Similarly, the firm would fire the worker if $\varepsilon < \varepsilon_t^f$. Finally, ε_t^e denotes the efficient cutoff productivity. Therefore, any job separation happening at a productivity $\varepsilon_t > \varepsilon_t^e$ is privately inefficient, since the employment pair would have had a positive joint surplus if they were to choose not to separate.

2.2 Nash bargaining with bargaining wedges (NBBW) and inefficient separations

Let $\eta \in [0,1]$ be the bargaining power of the worker. Wages are determined by maximizing the generalized Nash product

$$\begin{aligned} & \max_{W_t}\left(J_t\left(\varepsilon_t\right)-\Omega_t^w\right)^\eta\left(F_t\left(\varepsilon_t\right)-\Omega_t^f\right)^{1-\eta} \qquad (12) \\ = & \max_{W_t}\left(V_t\left(\varepsilon_t\right)+W_t\left(\varepsilon_t\right)-U_t-\Omega_t^w\right)^\eta\left(P_t\left(\varepsilon_t\right)-W_t\left(\varepsilon_t\right)-\Omega_t^f\right)^{1-\eta} \end{aligned}$$

where Ω_t^w and Ω_t^f, which are exogenous to the worker and the firm, are the bargaining wedges of the worker and the firm respectively. A bargaining wedge is the difference between the actual surplus and the bargaining surplus.

We can interpret the wedges Ω_t^w and Ω_t^f as reduced-form objects that summarize the difference between actual continuation values and the values that effectively discipline bargaining in the protocol at hand. They can reflect (i) disagreement payoffs that include flows during bargaining (Section 3), (ii) informational rents and screening distortions under private information (Section 4), or (iii) other contracting imperfections (e.g., limited verifiability or costly renegotiation) that create a gap between what is feasible ex post and what can be credibly threatened during negotiation. The key sufficient statistic for inefficiency is the net wedge (Ω_t^{net}; defined below), which governs both the wage shift and the ordering of separation cutoffs in Proposition 1. If $\Omega_t^w > 0$, for example, then the actual surplus the worker receives ex-post will be higher than the surplus perceived when the worker is bargaining with the firm. Note that when $\Omega_t^w = \Omega_t^f = 0$, it reduces to the standard Nash bargaining problem.

The solution to the above problem is given by

$$\eta\left(P_t(\varepsilon_t)-W_t(\varepsilon_t)-\Omega_t^f\right)=(1-\eta)\left(V_t(\varepsilon_t)+W_t(\varepsilon_t)-U_t-\Omega_t^w\right) \tag{13}$$

and by rearranging we have the wage rule

$$W_t(\varepsilon_t)=\eta P_t(\varepsilon_t)+(1-\eta)\left(U_t-V_t(\varepsilon_t)\right)+(1-\eta)\Omega_t^w-\eta\Omega_t^f \tag{14}$$

Note that other things equal, the wage is increasing with Ω_t^w and decreasing with Ω_t^f. In what follows I define

$$\Omega_t^{net}\equiv(1-\eta)\Omega_t^w-\eta\Omega_t^f \tag{15}$$

as the *net bargaining wedge* between the worker and the firm. It measures the weighted difference between the bargaining wedge of the worker and that of the firm. Hence, it represents the net distortionary effect of the bargaining wedges on the wage rule. It turns out that the value of the net bargaining wedge has important implications for the efficiency of separation as well, as demonstrated in the following proposition.

Proposition 1 *Suppose wages are determined by the NBBW model.*

(i) If $\Omega_t^{net}=0$, *then* $\varepsilon_t^w=\varepsilon_t^e=\varepsilon_t^f$.

(ii) If $\Omega_t^{net}>0$, *then* $\varepsilon_t^w<\varepsilon_t^e<\varepsilon_t^f$.

(iii) If $\Omega_t^{net}<0$, *then* $\varepsilon_t^f<\varepsilon_t^e<\varepsilon_t^w$.

Proof. See Appendix A. ■

The intuition of Proposition 1 is summarized in Figure 1. When $\Omega_t^{net}=0$, the three cutoff productivities are the same, hence job separation is always efficient in this case. This implies the classical result that employment and separation under standard Nash bargaining are always privately efficient. On the other hand, inefficient separation may arise when the net bargaining wedge is non-zero. When $\Omega_t^{net}>0$, wages are excessively high relative to the standard bargaining solution. As a result, the cutoff productivity ε_t^f for layoff decision by the firm is larger than the efficient cutoff productivity ε_t^e. In this case, *inefficient layoffs* would be possible. For any productivity that lies between ε_t^e and ε_t^f, the firm would choose to fire the worker even when the joint surplus is positive. Efficient separation occurs only if productivity is sufficiently low (i.e. less than ε_t^e). Conversely, when $\Omega_t^{net}<0$, wages are too low relative to the standard case. So the cutoff productivity ε_t^w for quitting decision by the worker is larger than the efficient cutoff productivity ε_t^e. Similarly, *inefficient quits* would occur when the productivity falls between ε_t^e and ε_t^w.

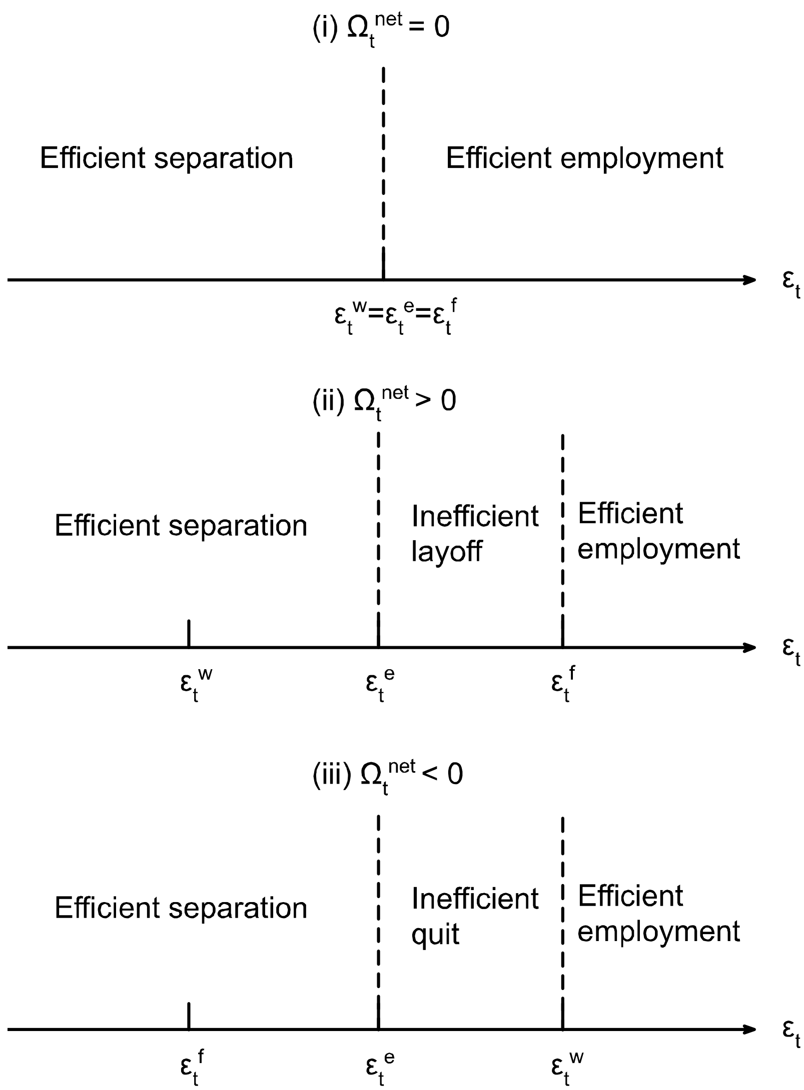

Figure 1: Efficiency of job separations

Sections 3 and 4 show that specific bargaining environments—such as alternating-offers bargaining with delay and bargaining under private information—can generate nonzero bargaining wedges. But the purpose of the generalized model in Section 2 is not merely to nest those cases. Its main value is to isolate a common sufficient statistic, the net bargaining wedge Ω_t^{net}, that governs whether separations are inefficient and in which direction. This reduced-form representation delivers several insights that are difficult to obtain cleanly from any single bargaining protocol. First, it separates the source of inefficiency from the particular institutional details of bargaining, making clear that inefficient layoffs and quits arise whenever bargaining surpluses differ from actual surpluses in a way summarized by Ω_t^{net}. Second, it yields transparent comparative statics: the sign and magnitude of Ω_t^{net} determine the ordering of the separation thresholds and therefore whether the economy exhibits inefficient layoffs or inefficient quits. Third, it makes it possible to carry out a quantitative decomposition of unemployment into efficient and inefficient components without committing the entire analysis to one specific bargaining micro-foundation.

It may be interesting to see what factors affect the *inefficiency gaps* $\varepsilon_t^f - \varepsilon_t^e$ and $\varepsilon_t^w - \varepsilon_t^e$ in (ii) and (iii). For illustration, consider a special case when there is no persistence between productivity shocks (i.e. $\lambda = 1$), and the total productivity p_t is a product of aggregate productivity y_t and individual productivity ε_t (i.e. $p_t(\varepsilon_t) = y_t\varepsilon_t$). In this case, we have[14]

$$\varepsilon_t^f - \varepsilon_t^e = \frac{\Omega_t^{net}}{y_t(1-\eta)} \tag{16}$$

$$\varepsilon_t^w - \varepsilon_t^e = \frac{-\Omega_t^{net}}{y_t\eta} \tag{17}$$

This special case demonstrates that the inefficiency gap is proportional to the magnitude of the net bargaining wedge, and inversely proportional to the aggregate productivity. It also depends negatively on the bargaining power of the party that initiates the inefficient separation. When $\Omega_t^{net} > 0$, for example, the inefficiency gap for inefficient layoff depends negatively on the bargaining power of the firm.

Equations (16)–(17) are derived as a special case to provide a transparent illustration of how the inefficiency gap scales with the net wedge and aggregate productivity. In the calibrated quantitative model, idiosyncratic productivity is persistent (i.e. $\lambda < 1$),[15] so the cutoff values are determined by fixed-point conditions that incorporate continuation values. Nevertheless, the key qualitative implication is robust: when Ω_t^{net} is less cyclical than the

[14] See Appendix A for a proof.

[15] The homogeneous benchmark sets $\lambda = 0$ by definition; the baseline heterogeneous calibration features $\lambda \in (0, 1)$.

joint surplus, the wedge-induced component of the separation cutoff becomes relatively more important in recessions, making the inefficiency gap (e.g., $\varepsilon_t^f - \varepsilon_t^e$) countercyclical when $\Omega_t^{net} > 0$.

There are some implications following this expression. First, the effect of the net bargaining wedge is magnified in bad times, implying that the inefficiency gap tends to be larger during recessions. Intuitively, the job separation decision is made only when the total productivity is low enough. Hence, the labor market will be more sensitive to the distortionary effect during bad times, when the aggregate productivity is low. This generates more inefficient job separations and could potentially explain the mass layoffs observed in the recessions. Second, a higher bargaining power of the firm could mitigate the distortion when the net bargaining wedge is positive, and similarly for that of the worker when net bargaining wedge is negative. Intuitively, when the weighted bargaining wedge of the worker is larger than that of the firm (i.e. $\Omega_t^{net} > 0$), the wage rule would be distorted in favor of the worker, which would entail a higher-than-efficient level of wage rate. A higher bargaining power of the firm would mean the firm would be able to bargain for a lower wage, hence offsetting part of the distortion by the bargaining wedge.

It is worth noting the above analysis assumes that the bargaining wedge is exogenous and fixed. In general, however, Ω_t^{net} may depend on some parameters in the model and may even fluctuate over the business cycle. Hence, the above comparative statics results may be amplified or dampened depending on the endogeneity of Ω_t^{net}. In the following sections when I provide micro-foundations to Ω_t^{net}, I will explicitly internalize Ω_t^{net} when performing the comparative statics. I will also consider the cyclicality of Ω_t^{net} in the quantitative analysis.

2.3 Worker flows and inefficient unemployment

I measure the worker flows as follows. Let $e_t(\varepsilon)$ be the distribution function of the employed workers, and u_t be the unemployment rate. Hence, $e_t = e_t(\varepsilon_{\max}) = 1 - u_t$ will be the employment rate. Also, denote ε_t^s as the separation cutoff value of the individual productivity. i.e.

$$\varepsilon_t^s = \max\left\{\varepsilon_t^w, \varepsilon_t^f\right\} \tag{18}$$

Note that Proposition 1 implies that $\varepsilon_t^s \geq \varepsilon_t^e$. Then the flow equation of the employment distribution is given by

$$\begin{aligned} e_{t+1}(\varepsilon) &= (1-s)\left\{\lambda e_t\left[G(\varepsilon) - G\left(\varepsilon_{t+1}^s\right)\right] + (1-\lambda)\left[e_t(\varepsilon) - e_t\left(\varepsilon_{t+1}^s\right)\right]\right\} \\ &\quad + u_t\phi(\theta_t)\left[G(\varepsilon) - G\left(\varepsilon_{t+1}^s\right)\right] \end{aligned} \tag{19}$$

Hence, the measure of employed workers having productivity less than or equal to ε at time $t+1$ comes from three groups of workers at time t, namely (i) those employed at time t who are switching productivity, and drawing a productivity from the interval $\left[\varepsilon_{t+1}^{s}, \varepsilon\right]$, (ii) those employed at time t who have a productivity in $\left[\varepsilon_{t+1}^{s}, \varepsilon\right]$ and not switching their productivity, and (iii) those unemployed at time t, having matched to a firm, and drawing a productivity from the interval $\left[\varepsilon_{t+1}^{s}, \varepsilon\right]$.

Similarly, the flow to unemployment is given by

$$u_{t+1} = e_t \left\{ s + (1-s)\lambda G\left(\varepsilon_{t+1}^{s}\right) \right\} + e_t\left(\varepsilon_{t+1}^{s}\right)(1-s)(1-\lambda) + u_t\left[1 - \phi\left(\theta_t\right)\left(1 - G\left(\varepsilon_{t+1}^{s}\right)\right)\right] \tag{20}$$

Therefore, those employed workers at time t whose jobs are destroyed for exogenous reason and whose productivities are too low, and those unemployed at time t who are not able to match a firm and to draw a high enough productivity, will be unemployed at time $t+1$.

To see the effect of inefficient job separations on the efficiency of unemployment, consider the steady state unemployment rate,

$$u = \frac{s + (1-s)\lambda G\left(\varepsilon^{s}\right)}{\phi(\theta)\left(1 - G\left(\varepsilon^{s}\right)\right) + s + (1-s)\lambda G\left(\varepsilon^{s}\right)} \tag{21}$$

and its efficient counterpart u^e, when ε^s is replaced by ε^e. Since $\varepsilon^s \geq \varepsilon^e$ and u is increasing with ε^s, we have $u \geq u^e$.[16] This shows that the existence of bargaining friction induces a higher level of unemployment than the standard model where all separations are efficient. Hereafter I shall call the extra level of unemployment the *inefficient unemployment*.

I derive the flow equations of an efficient level of unemployment u_t^e by replacing ε_t^s in the flow equations (19) and (20) with ε_t^e. Hence, efficient unemployment corresponds to the level of unemployment when all separations are efficient. Inefficient unemployment will then be given by the difference $u_t^i = u_t - u_t^e$. This simple decomposition allows me to measure the contribution of unemployment volatility by the inefficient separations later in the quantitative analysis. Section 6 returns to this decomposition and shows that the inefficient component becomes especially important in recessions, when the gap between total and efficient unemployment widens substantially. Also, in the calibrated model, inefficient unemployment explains a substantial part of total unemployment volatility.

[16] Note that I have assumed the steady-state value of the market tightness is fixed and determined how unemployment rate varies with the cutoff productivity.

The EU and UE transitions are given respectively as follows:

$$EU_{t+1} = e_t \left\{ s + (1-s)\lambda G\left(\varepsilon^s_{t+1}\right) \right\} + e_t\left(\varepsilon^s_{t+1}\right)(1-s)(1-\lambda) \tag{22}$$

$$UE_{t+1} = \phi(\theta_t)\left(1 - G\left(\varepsilon^s_{t+1}\right)\right) u_t \tag{23}$$

Finally, the job separation and job finding rates are given by

$$SR_{t+1} = \frac{EU_{t+1}}{e_t} = s + (1-s)\lambda G\left(\varepsilon^s_{t+1}\right) + \frac{e_t\left(\varepsilon^s_{t+1}\right)}{e_t}(1-s)(1-\lambda) \tag{24}$$

$$FR_{t+1} = \frac{UE_{t+1}}{u_t} = \phi(\theta_t)\left(1 - G\left(\varepsilon^s_{t+1}\right)\right) \tag{25}$$

As we will see later, a relatively acyclical bargaining wedge would induce a higher volatility in the cutoff productivity ε^s_t and hence in the workers flow in the labor market. Quantitatively, however, the effect on the job finding rate is limited, since the cutoff productivity only affects the job finding rate through the distribution function $G(\cdot)$.

2.4 Comparison with the canonical MP model

Since the canonical MP model is typically written in flow terms, it is useful to translate the NBBW wage rule into a flow-wage equation. To facilitate comparison, consider a steady-state benchmark with a constant bargaining wedge. Appendix A.2 provides the derivations.

Let w^{ss} denote the steady-state flow wage and ε^s be the steady-state separation cutoff, then it can be shown that

$$w^{ss}(\varepsilon) = (1-\eta)b + \eta\left(p(\varepsilon) + c\theta^{ss}\right) + \begin{bmatrix} 1 - \beta(1-s)\left(1 - \lambda G(\varepsilon^s)\right) \\ +\beta\theta^{ss} q(\theta^{ss})\left(1 - G(\varepsilon^s)\right) \end{bmatrix} \Omega^{net} \tag{26}$$

where θ^{ss} is the steady-state market tightness. Thus, the bargaining wedge shifts the standard Nash wage by an additive term. The corresponding steady-state free-entry condition is

$$\frac{c}{\beta q(\theta^{ss})} = \frac{\int_{\varepsilon^s}^{\varepsilon_{\max}} \left[p(\varepsilon) - w^{ss}(\varepsilon)\right] dG(\varepsilon)}{1 - \beta(1-s)\left(1 - \lambda G(\varepsilon^s)\right)} \tag{27}$$

These expressions show that there are two roles of Ω^{net}. First, a positive net bargaining wedge raises wages directly. Through the free-entry condition, this wage effect reduces the expected profit from a match and therefore weakens job creation. Second, Ω^{net} affects free entry indirectly through the separation margin. A higher Ω^{net} raises the separation cutoff ε^s, so fewer newly matched workers draw acceptable productivity ($1 - G_s$ falls) and existing matches are less likely to survive a redraw ($1-\lambda G_s$ falls). Hence, bargaining friction depresses vacancy creation because it shortens expected match duration. This is why the bargaining wedge that distorts wages also amplifies unemployment through the separation channel.

3 Alternating offers bargaining (AOB)

Hall and Milgrom (2008) apply the AOB model of Rubinstein (1982) and Binmore et al. (1986) to the MP environment, where they distinguish between the *outside-option payoff* and the *disagreement payoff*. In this bargaining situation, when the parties fail to reach an agreement, they continue to bargain through a sequence of offers and counter-offers, instead of quitting the negotiation immediately. In fact, at the unique subgame perfect equilibrium in the AOB model, both parties would move to an agreement immediately. Because workers and firms are mainly considering the disagreement payoff when bargaining, the wage solution can be less sensitive to productivity change than the standard Nash bargaining model. Hence, this generates real wage rigidity endogenously.

The model is described as follows. Following Christiano et al. (2016), I assume the whole bargaining process lasts within one period.[17] In each period, the firm would make a first offer W_t in the first subperiod.[18] If the worker accepts the offer, she will receive a value of $W_t + V_t$. Otherwise, the worker can propose a counter-offer W'_t in the next subperiod, when the firm can similarly choose to accept the counter-offer with the value $P_t - W'_t$ or reject the offer and propose another offer in the next round, and so on *ad infinitum*. In addition, there is a probability δ in each subperiod that the bargaining would break down and both parties would get their respectively outside-option payoff. While bargaining, the worker receives a flow benefit z_t,[19] and the firm incurs a cost γ_t for maintenance. I assume, for simplicity, no party can leave voluntarily once entered the negotiation.[20]

[17] Christiano et al. (2016) assume the number of subperiods in a period is finite. Hence, the solution obtained here can be considered as the limiting case when the number of subperiods tends to infinity. The results below would still follow when number is finite instead. Hall and Milgrom (2008) assume bargaining lasts for multiple periods. The bargaining setting considered here is the intra-period version of their model.

[18] As in the NBBW model, I assume the worker and the firm are bargaining only over the present value of wages W_t, instead of over other values such that V_t and P_t. Theoretically, the firm can change V_t (and similarly the worker can also affect P_t) by choosing the timing of separation. I ignore this possibility since in all practical bargaining situation, workers and firms mainly bargain over wages only.

[19] Intuitively, z_t is a flow payoff to the worker during disagreement (e.g., the value of temporary non-market activity, leisure, or public benefits received while bargaining continues). z_t captures the worker-side component of the disagreement payoff.

[20] Since the process of bargaining is costly, it can be shown that if one party chooses to initiate separation during the bargaining process, the party would not have entered the bargaining in the first place.

At the unique subgame perfect equilibrium, the firm would always propose the lowest wage that the worker would accept, and similarly the worker would always propose the highest wage that the firm would accept. Hence, we have the following indifference conditions:

$$W_t(\varepsilon_t) + V_t(\varepsilon_t) = \delta U_t + (1-\delta)\left[z_t + \left(W_t'(\varepsilon_t) + V_t(\varepsilon_t)\right)\right] \tag{28}$$

$$P_t(\varepsilon_t) - W_t'(\varepsilon_t) = (1-\delta)\left[-\gamma_t + \left(P_t(\varepsilon_t) - W_t(\varepsilon_t)\right)\right] \tag{29}$$

The worker receives the value of $W_t(\varepsilon_t) + V_t(\varepsilon_t)$ by accepting the offer. On the other hand, if the worker rejects the offer, there is a probability δ that the worker would be unemployed; otherwise, the worker earns a flow benefit z_t and proposes a counter-offer W_t'. Hence, (28) shows that the firm would choose W_t such that the worker is indifferent between accepting and rejecting the offer. The same is true for the firm, and it is reflected in (29). Since the firm is assumed to make the first offer, the worker would accept it by indifference, and so W_t would be the wage solution in this model. In fact, it can be shown that it is a special case of the NBBW model considered in the previous section.

Proposition 2 *The unique subgame perfect Nash equilibrium in the AOB model is equivalent to the solution to the NBBW model with*

$$\eta = \frac{1-\delta}{2-\delta}$$
$$\Omega_t^w = \frac{1-\delta}{\delta} z_t$$
$$\Omega_t^f = -\frac{1-\delta}{\delta}\gamma_t$$

Proof. See Appendix A. ■

Intuitively, under the AOB protocol, while the outside-option payoff of the worker is still U_t, the disagreement payoff is

$$\sum_{i=0}^{\infty}(1-\delta)^i\left[\delta U_t + (1-\delta) z_t\right] = U_t + \frac{1-\delta}{\delta} z_t$$

which is larger than the outside-option payoff. Hence, the difference $\frac{1-\delta}{\delta} z_t$ would be the bargaining wedge. Similarly, the outside-option payoff of the firm is zero due to free entry, but the disagreement payoff is

$$\sum_{i=1}^{\infty}(1-\delta)^i(-\gamma_t) = -\frac{1-\delta}{\delta}\gamma_t$$

which would also be the bargaining wedge. Note that the bargaining power of the worker is no larger than $\frac{1}{2}$, which shows the first-mover advantage of the firm by proposing an offer first.

In this case, the net bargaining wedge is given by

$$\Omega_t^{net} = (1-\eta)\,\Omega_t^w - \eta\Omega_t^f = \frac{1}{2-\delta}\frac{1-\delta}{\delta}\left[z_t + (1-\delta)\,\gamma_t\right] > 0 \tag{30}$$

and so by Proposition 1, there could be inefficient layoffs. Here the inefficiency comes from the cost of delay. Although the negotiation happens within a period, hence there is no discounting, the probability of a breakdown, along with the flow benefit and cost during bargaining creates incentive problem for both parties.

Finally, the comparative static analysis of the inefficiency gap $\varepsilon_t^f - \varepsilon_t^e$ for the special case considered previously is as follows.

$$\frac{\partial\left(\varepsilon_t^f - \varepsilon_t^e\right)}{\partial z_t} > 0 \tag{31}$$

$$\frac{\partial\left(\varepsilon_t^f - \varepsilon_t^e\right)}{\partial \gamma_t} > 0 \tag{32}$$

$$\frac{\partial\left(\varepsilon_t^f - \varepsilon_t^e\right)}{\partial \delta} < 0 \tag{33}$$

The intuition of the above is straightforward. If the flow benefit during bargaining increases, the worker would be more willing to disagree under AOB, and hence would ask for a higher wage than the standard Nash bargaining model. This could possibly generate more inefficient layoffs. Also, if the maintenance cost during bargaining increases, the firm would be more likely to agree with the worker, and hence would be willing to accept a higher wage than the standard model. Again it could generate more inefficient layoffs. Lastly, when the breakdown probability increases, the expected duration of disagreement decreases, and hence would reduce the effects of the flow benefit and the flow cost on the wage. This would mitigate the inefficiency. In particular, if $\delta \to 1$, then we have $\varepsilon_t^f \to \varepsilon_t^e$ and there is no inefficiency.

The role of this section is to show that an alternating-offers bargaining environment can be mapped into the wedge representation developed in Section 2. This mapping is useful because it provides a concrete micro-foundation for a positive net bargaining wedge, but the paper's broader mechanism does not depend on the AOB protocol itself. What matters for separation inefficiency is the implied net wedge Ω_t^{net}, not the particular institutional details that generate it. The generalized model is therefore informative beyond the AOB case: it identifies which features of the bargaining environment matter for inefficient separations and which are inessential for the aggregate mechanism.

4 Asymmetric information and neutral bargaining solution (NBS)

In this section, I introduce asymmetric information into the MP model. I consider the simple case when only the firm has private information about the match-specific productivity. This case has been studied by Kennan (2010), who employs the neutral bargaining solution (NBS) of Myerson (1984) to solve the model. However, he assumes the worker would always propose a pooling offer, resulting in efficient employment. Here I relax this assumption, which allows the possibility of inefficient separations.

Suppose now the firm has private information about ε_t and the worker has no way to verify it. Myerson (1984) proposes the NBS as a generalization of the Nash bargaining solution in the presence of private information. He shows that the NBS to the bargaining problem in this case always exists, and can be implemented by a Random Dictator mechanism. Specifically, suppose upon matching with the firm, there is a probability ν that the worker is chosen by the Dictator to propose a take-it-or-leave-it offer. Otherwise, the firm would be chosen to make the offer. Since the firm observes ε_t, its offer would extract all the match surplus $S_t(\varepsilon_t)$, leaving the worker with the value of unemployment and zero surplus. On the other hand, since the worker have no knowledge about the size of the surplus, the offer cannot depend on ε_t. If the worker demands a surplus of $S_t(\bar{\varepsilon})$, then there is a probability $G(\bar{\varepsilon})$ that the firm would dissolve the employment relationship, when the total surplus $S_t(\varepsilon_t)$ is no greater than the worker's demand. Hence, the worker chooses $\bar{\varepsilon}_t$ to maximizes the expected surplus:

$$\bar{\varepsilon}_t = \arg\max_{\varepsilon} (1 - G(\varepsilon)) S_t(\varepsilon) \tag{34}$$

Under the Random Dictator mechanism, the worker would receive the ν fraction of the above surplus and similarly for the firm as follows.

$$J_t(\varepsilon_t) = \nu (1 - G(\bar{\varepsilon}_t)) S_t(\bar{\varepsilon}_t) \tag{35}$$

$$F_t(\varepsilon_t) = \nu (1 - G(\bar{\varepsilon}_t)) (S_t(\varepsilon_t) - S_t(\bar{\varepsilon}_t)) + (1 - \nu) S_t(\varepsilon_t) \tag{36}$$

Intuitively, with probability $\nu(1-G(\bar{\varepsilon}_t))$, the worker is picked to make the offer $S_t(\bar{\varepsilon}_t)$, and the total surplus is enough (i.e. when $\varepsilon_t \geq \bar{\varepsilon}_t$) to cover the offer. Hence, the firm would get $S_t(\varepsilon_t) - S_t(\bar{\varepsilon}_t)$ in this case. With probability $1-\nu$, the firm would propose to get the whole surplus, in which case the worker would get zero surplus. The Random Dictator mechanism stipulates that each party gets its corresponding expected payoff. It is clear that in the above specification, the strategies of the worker and the firm are incentive-efficient. Hence, the NBS can be implemented by the Random Dictator mechanism.

The following proposition follows immediately by the comparison with the bargaining solution (13) in Section 2, and its proof is thus omitted.

Proposition 3 *The NBS to the asymmetric information model is equivalent to the solution to the NBBW model with*

$$\begin{aligned}
\eta &= 0 \\
\Omega_t^w &= \nu(1-G(\bar{\varepsilon}_t))\,S_t(\bar{\varepsilon}_t) \\
\Omega_t^f &= -\nu(1-G(\bar{\varepsilon}_t))\,S_t(\bar{\varepsilon}_t)
\end{aligned}$$

First, under asymmetric information, the effective bargaining power of worker is zero. The private information about the match-specific productivity effectively gives the firm all the bargaining power. This is in contrast with the complete information case, when the bargaining power is given by the probability of making offer (ν). It is because the worker, not knowing the actual size of the match surplus, is not able to bargain with the firm over the actual surplus. Instead, the worker demands a fixed amount of surplus which would then cause inefficiency.

Note that in this case the net bargaining wedge is given by the worker's surplus.

$$\Omega_t^{net} = \nu(1-G(\bar{\varepsilon}_t))\,S_t(\bar{\varepsilon}_t) \geq 0 \tag{37}$$

Since the worker demands the same surplus regardless of the realization of ε_t, inefficient layoff occurs when the realized total surplus is positive but lower than certain level so that the firm's surplus is negative. Since only the firm observes the productivity, the layoff decision is efficient if the total match surplus is in line with the firm's surplus. This happens if the firm is always making the offer, i.e. when $\nu = 0$.

This section shows that asymmetric information can also be represented through the same wedge logic as in Section 2. As is the case with the AOB model, the purpose here is not to claim that the quantitative model must literally be interpreted as a neutral bargaining solution under private information. Rather, the purpose of this asymmetric-information example is to illustrate how informational rents can be represented as bargaining wedges and can generate inefficient separations. In the quantitative analysis, we do not require this exact NBS parameterization, and the aggregate wedge Ω_t^{net} identified can be due to a combination of a broader set of bargaining frictions, including (but not limited to) the NBS case.

4.1 Discussion

Viewed through the wedge representation, the AOB and NBS environments are best seen as two special cases rather than as exhaustive alternatives to the baseline model. In the AOB case, the implied net bargaining wedge is positive, so the model predicts inefficient layoffs generated by costly delay. In the NBS case, the implied net wedge is also non-negative, but the comparison with the wedge model requires $\eta = 0$, so the firm effectively has all bargaining power. By contrast, the generalized wedge model relaxes both restrictions. First, it allows Ω_t^{net} to be either positive or negative, so it can generate inefficient quits as well as inefficient layoffs. Second, it allows Ω_t^{net} to vary cyclically in a way that is not tied to a single bargaining protocol. This matters quantitatively: an acyclical bargaining wedge explains about 22% of unemployment volatility in the data, while allowing wedge cyclicality raises the explanatory power to about 29%. Third, the generalized model preserves η as a free parameter, so bargaining-power shifts have independent implications for separation inefficiency. These predictions are absent from the NBS special case, where $\eta = 0$ by construction. Hence, the value of the wedge approach is not merely that it nests AOB and NBS, but that it allows multiple bargaining frictions to operate jointly and yields comparative statics and quantitative predictions that are not tied to either case.

5 Implications for the volatility of cyclical unemployment

Before turning to the quantitative analysis, I discuss the implications of bargaining friction for the unemployment volatility. I first compare the specification of the bargaining wedges with the concept of "fundamental surplus" in Ljungqvist and Sargent (2017). Then I discuss the cyclicality of worker's share of the match surplus in the NBBW model, and its implications for the unemployment volatility.

5.1 Relationship with the fundamental surplus in Ljungqvist and Sargent (2017)

Ljungqvist and Sargent (2017) show that in a steady state MP model with exogenous separations and homogeneous productivity, the elasticity of market tightness θ with respect to productivity y depends crucially on a factor which they refer to as the *fundamental surplus*. Intuitively, fundamental surplus refers to the flow surplus from creating a match net of the worker's flow value in unemployment and related terms that shift wages; it is the object that governs the elasticity of market tightness with respect to productivity in the standard MP model. Ljungqvist and Sargent (2017) show that in order for the elasticity of market tightness to be large, the fundamental surplus has to be small. Here I incorporate the bargaining wedges described previously into their model, and examine how they are related to the concept of fundamental surplus.

Following their derivation (details are in the Appendix A), it can be shown that the elasticity of market tightness in this case is given by

$$\epsilon_{\theta,y} = \Upsilon\left(\Omega^{net}\right) \frac{y}{y - z - \left[\frac{1-\beta(1-s)}{1-\eta}\right]\Omega^{net}} \tag{38}$$

where

$$\Upsilon\left(\Omega^{net}\right) = \frac{r + s + \eta\theta q\left(\theta\right)\left(1 + \beta q\left(\theta\right)\frac{\Omega^{net}}{c\eta}\right)}{\alpha\left(r+s\right) + \eta\theta q\left(\theta\right)\left(1 + \left(1-\alpha\right)\beta q\left(\theta\right)\frac{\Omega^{net}}{c\eta}\right)} \tag{39}$$

is the factor with an upper bound $\max\left\{\frac{1}{\alpha}, \frac{1}{1-\alpha}\right\}$ and $\alpha \equiv \frac{-q'(\theta)\theta}{q(\theta)}$ is the elasticity of the matching function. Here the fundamental surplus is $y - z - \left[\frac{1-\beta(1-s)}{1-\eta}\right]\Omega^{net}$. Therefore, the existence of bargaining wedges will increase the elasticity of market tightness if $\Omega^{net} > 0$. As in many cases studied by Ljungqvist and Sargent (2017), bargaining friction can also potentially increase the unemployment volatility through raising the elasticity of market tightness.

There are problems with the steady state analysis of unemployment volatility by Ljungqvist and Sargent (2017). First, as mentioned in detail in the Appendix of Christiano et al. (2016), steady state models are generally misleading about the dynamic effects of a persistent shock. This point is made similarly in Petrosky-Nadeau and Zhang (2017). Second, and more relevant to this paper, the sole focus on the elasticity of market tightness to explain unemployment dynamics relies on the assumption that the elasticity of the separation rate is zero. However, there is evidence that the separation rate is counter-cyclical.[21] As shown in these studies, the elasticity of the separation rate relevant to the unemployment dynamics is indeed non-trivial. In fact, as I will show in the quantitative analysis, the main effect of the bargaining friction is on the volatility of separation rate, which is ignored by the analysis in much of the literature.

In what follows, I shall refer to the above mechanism as the *small surplus channel*. In models with exogenous separations and homogeneous productivity, the small surplus channel is the only channel through which bargaining friction can affect unemployment volatility.

5.2 Cyclicality of worker's share of the match surplus

In the standard Nash bargaining model, the worker's share of the match surplus, defined as

$$\pi_t(\varepsilon_t) = \frac{J_t(\varepsilon_t)}{S_t(\varepsilon_t)} \tag{40}$$

is equal to their (constant) bargaining power η, and thus is acyclical. Hence, the firm's share $1-\pi_t(\varepsilon_t) = 1-\eta$ is also unresponsive to aggregate fluctuation. This explains the flexibility of wages under Nash bargain, as the worker's surplus moves proportionally with the match surplus. Under the existence of bargaining friction, however, the worker's share can be responsive to the business cycle. By rearranging the wage rule (14), we have the following[22].

Proposition 4 *The worker's share of the match surplus in the NBBW model is given by*

$$\pi_t(\varepsilon_t) = \eta + \frac{\Omega_t^{net}}{S_t(\varepsilon_t)}$$

Hence, if Ω_t^{net} is less cyclical than $S_t(\varepsilon_t)$ (i.e. $\frac{\Omega_t^{net}}{S_t(\varepsilon_t)}$ is countercyclical), then the worker's share is countercyclical (procyclical) if $\Omega_t^{net} > 0$ (< 0).

[21] See, e.g., Elsby et al. (2009), Fujita and Ramey (2009), and, more recently, Coles and Moghaddasi Kelishomi (2018). This fact is also shown in the discussion of business cycle statistics in Appendix B.

[22] Clearly, if Ω_t^{net} is large or $S_t(\varepsilon_t)$ is small enough, it is possible that $\eta + \frac{\Omega_t^{net}}{S_t(\varepsilon_t)} > 1$. But in this case, the firm's surplus $F_t(\varepsilon_t)$ is negative and hence the firm would choose to fire the worker. The case when the expression is negative is similar. Therefore, the worker's share is properly defined when there is continuing employment.

The intuition of the proposition can perhaps be better understood in the context of the AOB model discussed earlier. In this case, we have a constant $\Omega_t^{net} > 0$ and so the worker's share is countercyclical. Since the disagreement payoff for both parties is to continue for another round of bargain, the wage offers proposed would be less sensitive to the aggregate conditions[23]. This entails that, consistent with the findings in Hall and Milgrom (2008), wages move less than proportionally with the aggregate productivity. Equivalently, the worker's share of the match surplus increases in recessions. While the above is true for the AOB model, the NBBW model considered in this paper provides a more general framework to explain the increasing bargaining power of the workers in bad times. It is worth noting that the last result still holds even if Ω_t^{net} is either countercyclical, or procyclical but less so than the match surplus (in the sense that $\frac{\Omega_t^{net}}{S_t(\varepsilon_t)}$ is still countercyclical).

There are two consequences for the countercyclicality of the worker's share of the match surplus. First, this reduces the job creation incentive in the recession, compared with the standard Nash bargaining model. Also, a small change in y_t causes significant fluctuations in the cutoff productivity, and thus in the quantity of job separations. As a result, in recessions there would be more (inefficient) job separations, which only exists in the NBBW model. I call this effect the *separation channel*, which does not exist in homogeneous models like that in Ljungqvist and Sargent (2017). This separation channel is also distinct from standard heterogeneous MP models with endogenous job destruction under flexible Nash bargaining. In those models, the separation cutoff moves over the business cycle, but separations remain privately efficient because the wage bargain internalizes the joint surplus. In the present model, bargaining frictions distort the cutoff itself, so changes in separations reflect not only efficient reallocation but also inefficient layoffs or quits. Hence, bargaining frictions amplify unemployment not merely by reducing job creation incentives, but also by generating excess cyclical separations. Both the small surplus and the separation channels mentioned above could potentially generate excessive fluctuation of unemployment. In the next section, I will quantify these channels and show that Ω_t^{net} has a much stronger effect on the separation channel.

[23] It should be noted that, due to the bargaining friction, a large part of the wage rigidity comes from the fact that many matches end before a large wage cut can be carried out. The wages of ongoing matches may in fact still be relatively procyclical.

6 Quantitative analysis

In this section, I calibrate the model to the US labor market. Details of the computation strategy are in Appendix C. I will use the calibrated model to evaluate the impact of bargaining friction on the unemployment dynamics. Motivated by the examples in the previous sections, I consider only the case when the net bargaining wedge is non-negative, in which case there may be inefficient layoffs.

6.1 Specification and calibration

The matching function is of the standard Cobb-Douglas form.

$$m\left(u,v\right)=Au^{\alpha}v^{1-\alpha} \tag{41}$$

where α is the elasticity of the matching function and A measures the matching efficiency. This implies the probability of a vacancy meeting with an unemployed worker is $q\left(\theta\right)=\frac{m(u,v)}{v}=A\theta^{-\alpha}$ and that of an unemployed worker meeting a vacancy is $\phi\left(\theta\right)=\frac{m(u,v)}{u}=A\theta^{1-\alpha}$.

Following Fujita and Ramey (2012), I assume $p_t\left(\varepsilon_t\right)=y_t\varepsilon_t$, where y_t is the aggregate productivity following the stochastic process

$$\ln y_t=\rho\ln y_{t-1}+\zeta_t \tag{42}$$

where ζ_t is identically and independently distributed normal white noise with mean μ_ζ and standard deviation σ_ζ. Also, the distribution function $G\left(x\right)$ is taken to be truncated log-normal with support $\left[0,\varepsilon_{\max}\right]$.[24]

The model is calibrated using data for the US labor market data from 1976 to 2016. Each period in the model corresponds to one week.[25] The discount factor is taken to be $\beta=0.9992$ which is equivalent to an annual discount rate of 4%. The process of the aggregate productivity follows Hagedorn and Manovskii (2008). Specifically, we have $\rho=0.9895$, $\mu_\zeta=0$ and $\sigma_\zeta=0.0034$.[26] I set $s=0.000625$, which is consistent with the mean value of the monthly

[24] In Appendix D, I consider a version of the model where the idiosyncratic shocks follow a gamma distribution. The qualitative and quantitative results are similar.

[25] We use a weekly model period to avoid the time aggregation bias emphasized in Shimer (2012). When helpful for interpretation, we report annualized equivalents of key weekly parameters in Table 1.

[26] In Appendix D, I estimate the process using the cyclical productivity series and get $\rho=0.9889$ and $\sigma_\zeta=0.0040$. It is shown that the more volatile process implies that a lower level of bargaining friction is required to match the unemployment volatility in the data. To be conservative, I use Hagedorn and

rate of "other separations" in the Job Openings and Labor Turnover Survey (JOLTS).[27]

On the worker side, I set $b = 0.71$, which follows the derivation of Hall and Milgrom (2008) and is based on a replacement ratio of 0.25.[28] Also, the elasticity of the matching function α is set to be 0.5, which is consistent with the survey of Petrongolo and Pissarides (2001). I assume Hosios condition, which implies the bargaining power $\eta = \alpha$.

For each fixed value of the net bargaining wedge, the rest of the parameters are jointly calibrated to match the labor market flows in the US economy. The maximum value of match-specific productivity $\varepsilon_{\max}$ is calibrated such that the mean value of the productivity per worker is 1. The coefficient of matching efficiency A is calibrated to fit the model to an average monthly job finding rate of 32% in the Current Population Survey (CPS). I calibrate σ_ε to match an average monthly job separation rate 2% in the CPS. These calibrations imply a steady state unemployment rate of 5.9%, which is consistent with the long-run average of the data. The flow cost of vacancy c is identified by matching the average labor market tightness θ, which is 0.548 in the sample.[29] Finally, following Fujita and Ramey (2012), the switching rate of match-specific productivity λ is calibrated to match the persistence (i.e. autocorrelation) of the separation rate in the CPS.

For later comparison, I consider also a homogeneous version of the model where I assume workers are homogeneous in match-specific productivity with no endogenous separation decision. Specifically, I set $\lambda = 0$ and $\chi_t(\varepsilon_t) = 0$ for all ε_t. The model is then calibrated similarly as above, except that now we have $\sigma_\varepsilon = 0$ and s is calibrated to match the average monthly job separation rate in the data.

A summary of the calibration values is given by Table 1. The baseline calibration is shown in the third column, where I calibrate the net bargaining wedge to match the unemployment volatility in the data, see the next subsection for details. The fourth column shows the estimated parameter values for the efficient model, where the net bargaining wedge is set to be zero. In this case, all job separations are privately efficient. Finally, the fifth column shows the calibration of the homogeneous model described above with $\Omega^{net} = 0$. In addition, I have performed a sensitivity analysis for $\{b, \eta, \lambda\}$ in Table D.1 which shows that main quantitative conclusions are robust to changes in these parameters.

Manovskii (2008)'s estimation in the baseline.

[27] In the JOLTS, the total separation rate is decomposed into quits, layoffs and discharges, and other separations. I consider other separations corresponding to exogenous separations in the model. It can be shown that this measure of separations is relatively acyclical.

[28] To calculate b, they take the unemployment benefit $z = 0.25$, and apply it to a linear approximation of the utility function with leisure with reasonable parameters.

[29] The vacancy data is from the updated series by Barnichon (2010).

Table 1: Calibration Parameters

Parameters	Meaning	Baseline	Efficient	Homogenous	Target/remark
b	flow value of unemployment	0.71			Hall and Milgrom (2008)
β	discount factor	0.99925			Annual discount rate = 4%
ρ	persistence of the aggregate productivity	0.9895			Hagedorn and Manovskii (2008); Annualized: 0.5776
σ_ζ	standard deviation of the aggregate productivity	0.0034			Hagedorn and Manovskii (2008); Annualized: 0.0245
α	elasticity of the matching function	0.5			Petrongolo and Pissarides (2001)
η	worker's bargaining power	0.5			Hosios (1990)
s	exogenous separation rate	0.0006255		0.0055	JOLTS/average job separation rate; Annualized: 0.032/0.2493
λ	switching rate of match-specific productivity	0.1040	0.1068	0	Autocorrelation of separation rate
c	flow cost of posting vacancy	0.4069	0.4647	0.4648	Average labor market tightness
$\varepsilon_{\max}$	maximum value of match-specific productivity	1.1080	1.3287	0.9991	Average Productivity = 1
σ_ε	standard deviation of match-specific productivity	0.1396	0.5524	0	Average job separation rate
A	coefficient of matching efficiency	0.1284	0.1283	0.1225	Average job finding rate
Ω^{net}	Net bargaining wedge	0.1881	0	0	Unemployment volatility

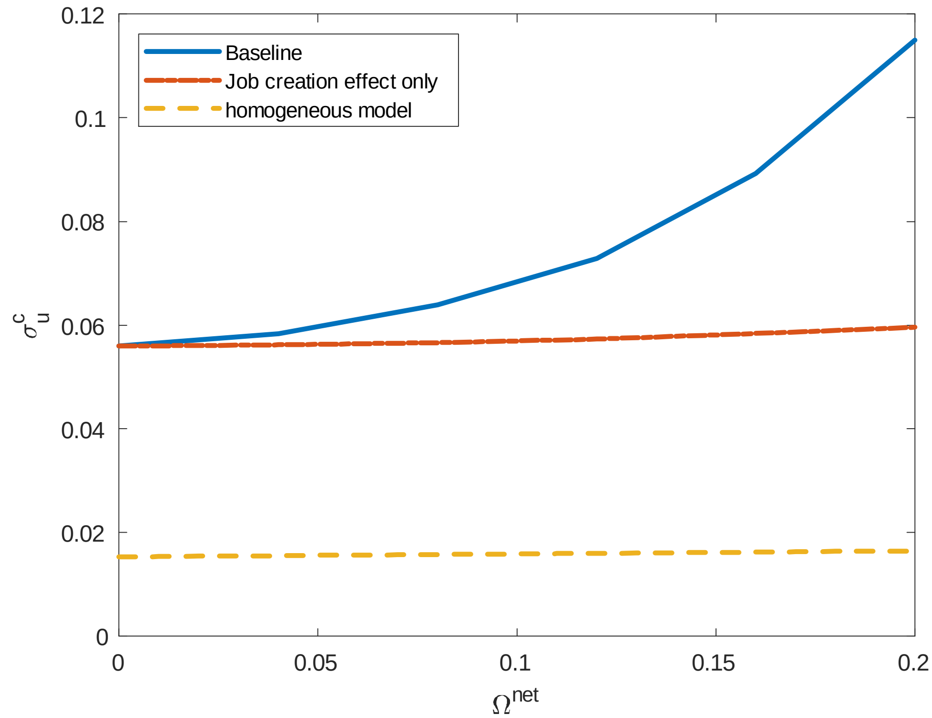

Figure 2: Unemployment volatility and bargaining friction

6.2 Unemployment volatility and bargaining friction

We have shown in the discussion of the model that the wedge would induce inefficient job separations and hence inefficient unemployment. Quantitatively, how would the bargaining wedge affect unemployment volatility?

Figure 2 shows the relationship between unemployment volatility and the net bargaining wedge.[30] The unemployment volatility is measured by the standard deviation of quarterly cyclical unemployment rate over a sample simulation of 1000 quarters. As we can see, unemployment volatility is monotonically increasing in the bargaining wedge. This is due to a combination of the small surplus and the separation channels described earlier: as the bargaining wedge increases, the wages are less sensitive to the aggregate conditions and the firm becomes more likely to fire the worker (inefficiently), hence there are more job separations in the economy, leading to higher level of unemployment fluctuation. In other words, while the workers' productivity varies with the aggregate productivity y_t, the bargaining wedge is relatively acyclical. As a result, a small change in y_t causes significant fluctuations in the cutoff productivity, and hence the quantity of job separations. The additional movement in job separations generates excessive unemployment volatility. Note

[30] Recall that for each value of Ω^{net}, the model is calibrated to the target moments.

that in principle, an increase in the bargaining wedge keeping other parameters unchanged would increase both the level of unemployment as well as its volatility. The procedure of the calibration described above, however, restricts the focus on the effect on unemployment volatility, since the steady state unemployment rate is kept fixed.

The average volatility of the cyclical unemployment rate in the data is about 10.9%. If all the unemployment volatility not explained by the standard model is assumed to be caused by a constant bargaining friction, this would imply a bargaining wedge of 0.188. I will use this value in the baseline calibration. See the third column of Table 1 for the calibrated parameter values.[31] It can be argued, however, that the bargaining wedge is moving with the business cycle. Hence, later in the decomposition of unemployment, I will allow for the cyclicality of the bargaining wedge.

To put the size of the bargaining wedge in perspective, the mean value of the total joint surplus in the simulation is about 5.29. Hence, the level of the bargaining wedge required to produce a realistic unemployment volatility is only 3.55% of the joint surplus, which is hardly a large number. This shows that the unemployment dynamics is in fact very sensitive to the bargaining friction in the model, hence even a small value of the bargaining wedge can generate a large unemployment volatility.

To quantify the previously discussed two channels, I perform a quantitative exercise in which the separation cutoff productivity ε_t^s is fixed while Ω^{net} varies. In this exercise, the bargaining wedge continues to affect wages, the fundamental surplus, and hence vacancy creation, but it cannot distort the separation margin. The resulting line in Figure 2, labeled "Job creation effect only," therefore isolates the small-surplus (job-creation) channel. It is clear that this channel is much less sensitive to bargaining friction than the baseline economy. As Ω^{net} rises from 0 to 0.188, cyclical unemployment volatility increases only from 5.60% to 5.91%. Relative to the total increase in unemployment volatility over the same range of Ω^{net} in the baseline economy, this implies that the small-surplus channel accounts for only about 6% of the overall amplification. Thus, the dominant source of amplification comes from the separation channel.

A related benchmark is the homogeneous productivity model, labeled "homogeneous model" in the same figure. Two observations are in order. First, the homogeneous model

[31] The business cycle statistics of the baseline calibration is shown and analyzed in Appendix B. As in Shimer (2005) and much of the subsequent literature, the baseline MP structure continues to underpredict vacancy and market tightness volatility relative to the data. Since the focus of this paper is the separation-based amplification mechanism induced by bargaining wedges, we do not introduce additional vacancy-creation frictions here. However, the wedge/separation mechanism is complementary to well-known approaches that increase tightness volatility, such as relaxing free entry or introducing hiring and vacancy-creation costs (e.g., Fujita and Ramey (2007); Coles and Moghaddasi Kelishomi (2018)).

has a much lower level of unemployment volatility even in the absence of bargaining friction, consistent with Fujita and Ramey (2012). Second, it is also much less sensitive to bargaining friction than the baseline heterogeneous economy. Taken together, these two counterfactual exercises show that, for empirically relevant values of Ω^{net}, the small-surplus channel is much weaker than the separation channel.[32]

Therefore, I conclude that much of the increase in unemployment volatility in the baseline model is due to the separation channel. Once bargaining frictions distort the separation cutoff, unemployment fluctuations are amplified through excess cyclical separations, rather than through the small-surplus/job-creation channel alone.

6.3 Inefficient unemployment

How much of the unemployment volatility is explained by the inefficient unemployment? To answer this question, I first use the baseline calibration as described above. The bargaining wedge is subject to idiosyncratic shocks. I identify the shocks to the wedge by matching the cyclical real wage over the same time period.[33] I also identify the shocks to the aggregate productivity process (42) by matching the unemployment rate in the data.[34] The matching of the unemployment rate from 1976 to 2016 is given in panel (a) of Figure 3 and the cyclical real wage in panel (b). Note that recessions defined by the NBER are in gray areas. To see the fitness of the model, I have also shown the cyclical component of labor productivity in panel (a) of Figure 4. In general, the model does well in matching the relationship between unemployment and cyclical productivity.[35] The estimated bargaining wedge is shown in panel (b) of Figure 4. We can see that the bargaining wedge closely follows the cyclical productivity ($corr(\Omega_t^{net}, p_t^c) = 0.66$). Consider the costly AOB model for intuition. When labor productivity is high, the opportunity costs of both the workers and the firms are higher. Hence, it would be more costly to bargain, and thus there would be a higher level of

[32]The small-surplus channel can become quantitatively significant when Ω^{net} is sufficiently large. Yet the range of values of Ω^{net} that is consistent with a realistic endogenous separation rate is too low to induce any meaningful amplification of unemployment volatility through this channel alone.

[33]I use the real hourly compensation for the non-farm business sector from the Bureau of Labor Statistics. In Appendix D I have also considered the case when the bargaining wedge is constant over the business cycle.

[34]In practice, I choose the values of y_t and Ω_t^{net} in the grids to jointly minimize the distance between the observed and simulated unemployment rates and cyclical real wage. To make sure the matched y_t is reasonable, I estimate an AR(1) process for the resulting y_t process. I obtain a sample estimate of $\hat{\rho} = 0.9913$ and $\hat{\sigma}_\zeta = 0.0040$, which are reasonably close to the population counterparts in the calibration.

[35]Table B.2 shows the moments of the model-implied wage time series compared with that in the data. It shows that the model matches the overall cyclical volatility and persistence of wages closely, but the model-implied wage is more responsive to productivity and unemployment than in the data.

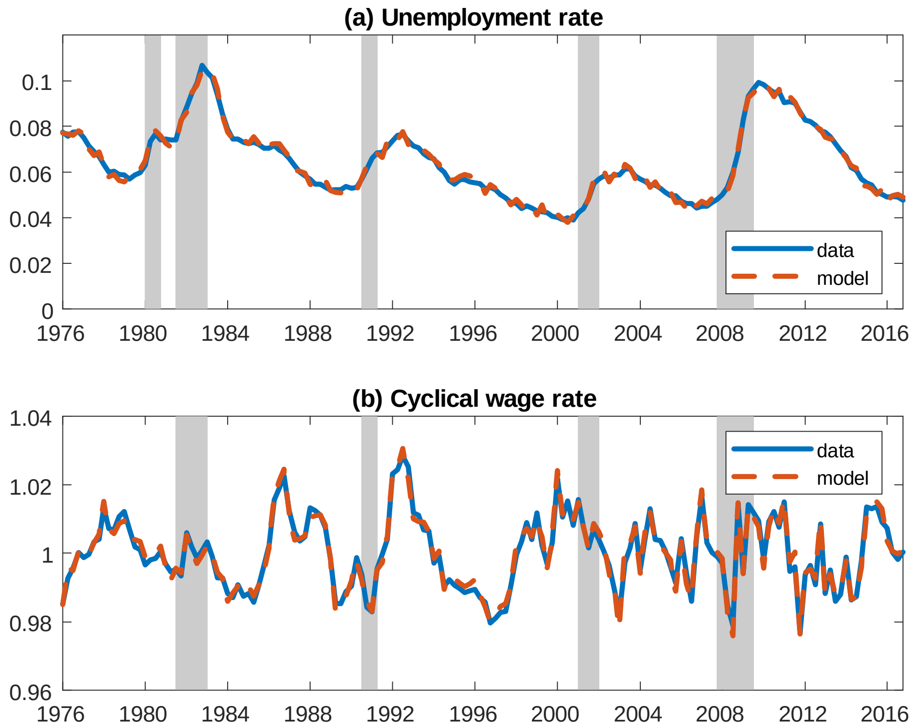

Figure 3: Unemployment and cyclical wage rate: data vs. model

bargaining friction.

Finally, I decompose the total unemployment into efficient and inefficient parts, as defined in Section 2. Figure 5 shows the decomposition of the total unemployment into efficient and inefficient unemployment rates. First, note that the inefficient unemployment appears to be much smaller in value, but is as volatile as the efficient unemployment. In fact, the average inefficient unemployment rate is about 1.6%, compared to the average efficient unemployment of 4.8%. The standard deviation of inefficient unemployment rate is 0.5%, compared to 1.1% for the efficient unemployment. Also, in good times when the total unemployment rate is low, the inefficient-efficient unemployment ratio is also low relative to that in recessions. In fact, the inefficient unemployment rate is less than 1% around 2000. On the other hand, the level of inefficient unemployment rate can be as high as 3% in the Great Recession. Hence, the peak of the unemployment rate in the US during the Great Recession would have been about 3 percentage points lower if all separations were efficient. The efficient unemployment reflects the effect of efficient separations only, without changing the value of the bargaining wedge. Hence, this shows that inefficient separation is a significant source of unemployment volatility in the estimated model.

To see quantitatively how much of the volatility is explained by the inefficient unemploy-

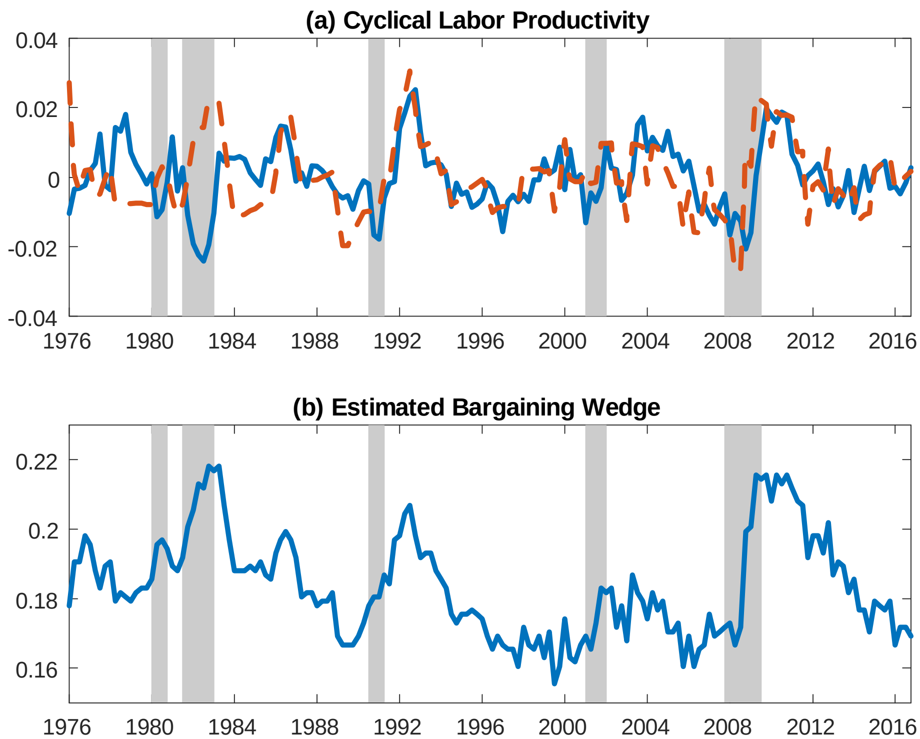

Figure 4: Cyclical labor productivity and estimated bargaining wedge

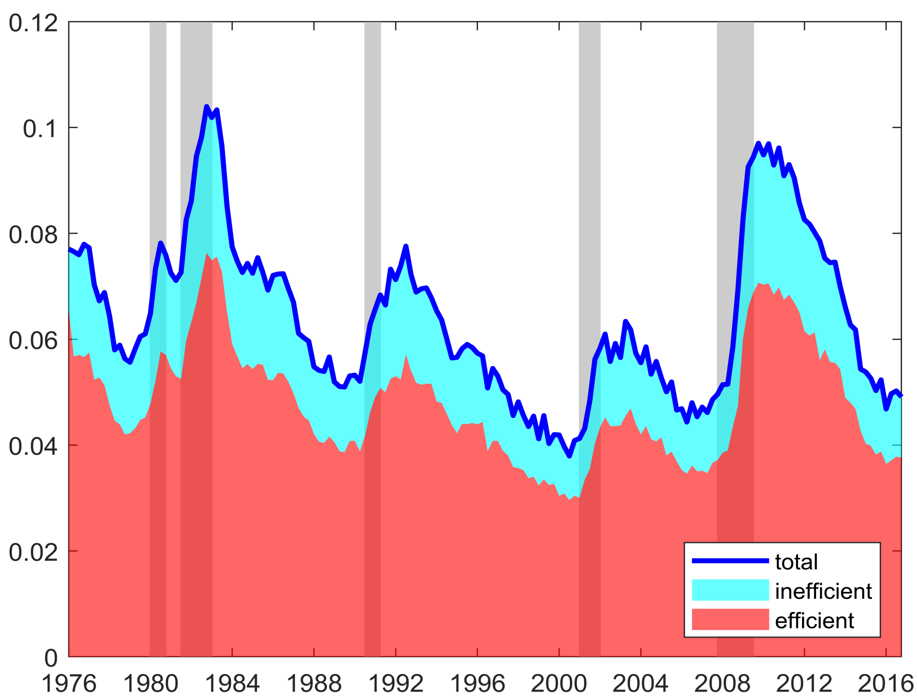

Figure 5: Decomposition of unemployment rate

ment, I can write, by the decomposition of the unemployment rate,

$$\begin{aligned} var\left(u_t\right) &= cov(u_t, u_t^e + u_t^i) \\ &= cov\left(u_t, u_t^e\right) + cov\left(u_t, u_t^i\right) \end{aligned} \tag{43}$$

Hence, define

$$\gamma = \frac{cov\left(u_t, u_t^i\right)}{var\left(u_t\right)} \tag{44}$$

as the fraction of unemployment volatility explained by the inefficient job separations, and $1-\gamma$ would be that by efficient job separations. In the baseline case, I have $\gamma = 0.291$. Hence, about 30% of the unemployment volatility can be explained by inefficient job separations, and the remaining 70% is due to efficient separations.[36]

6.4 Worker flows

Recently, there has been increasing attention in the literature on the worker flows in the business cycle frequency[37]. How would the bargaining friction affect the labor market flows? To see this, I use the definition in Section 2 to compute job finding and separation rates.

Figure 6 shows the labor market flows for the standard model and for the baseline calibration in a simulation. Note that in both cases, I have calibrated the model to the average values of the job finding and separation rates. Hence, the focus here is on the volatility of the flows. First, job finding rate is procyclical and the job separation rate is countercyclical. These are consistent with the evidence in Fujita and Ramey (2009). Next, the effect of the bargaining wedge on the job finding rate appears to be small, except when the aggregate shock is large. The job finding rate in the baseline case is slightly more volatile than the efficient model.

The main difference between the standard model and the baseline case lies in the volatility of separation rate. In fact, the cyclical volatility of separation rate in the baseline case is about 10% which is much higher than 5% in the efficient model. On the other hand, the effect on job finding rate is relatively insignificant (3.0% vs. 2.0%). Hence, the existence of bargaining friction affects mainly the separation rate in the labor market.

Note that the above result implies that, at the business cycle frequency, the job separation

[36] By construction, it is assumed that inefficient separation is the explanation for all the unemployment volatility not explained by the standard model. Hence, this can be considered as an upper bound of the explanatory power of bargaining friction.

[37] See, e.g., Davis et al. (2012) and Fujita and Nakajima (2016).

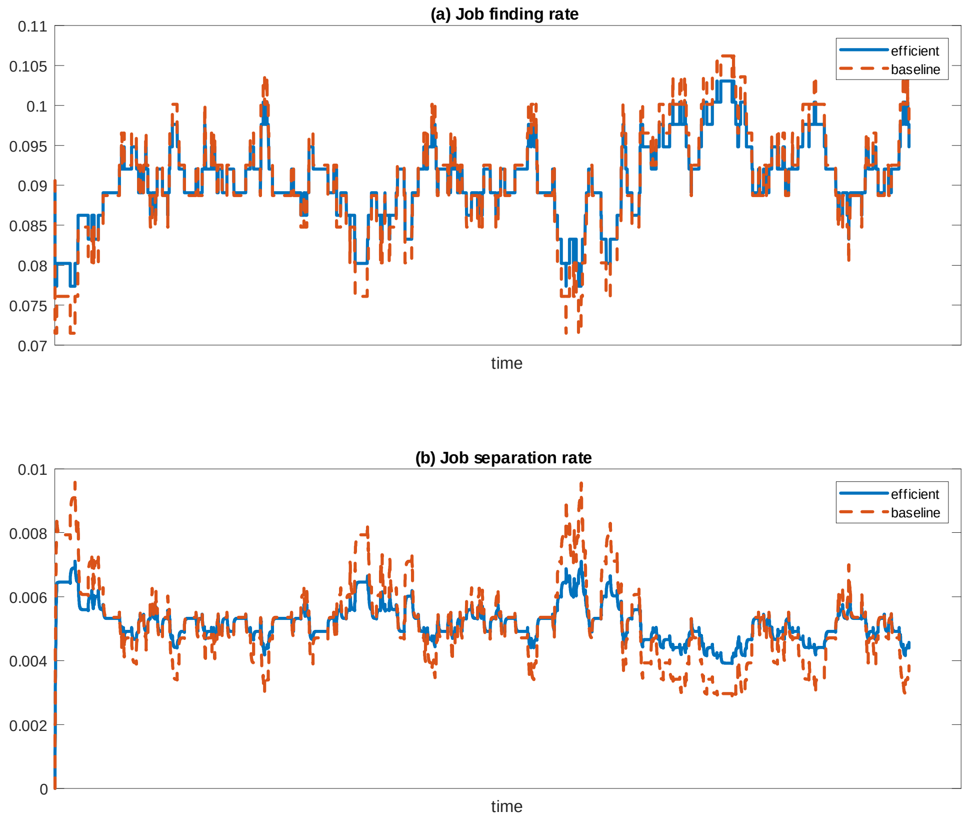

Figure 6: Simulated job finding and separation rates

rate is as important as the job finding rate in driving the fluctuation of unemployment. This stands in contrast to the evidence in Shimer (2012), who finds that the job separation rate is not significant in explaining unemployment volatility. However, Elsby et al. (2009) and Fujita and Ramey (2009) show that job separations are countercyclical and contribute substantially to the unemployment volatility. More recently, Coles and Moghaddasi Kelishomi (2018) estimate that job destructions have been the main driver of unemployment volatility if we relax the free entry assumption. This paper makes no attempt to resolve the debate on the relative importance of the ins and outs of unemployment flows. The message here is clear: by allowing inefficient separations by bargaining friction, the separation rate and the worker flows/transitions become much more volatile, while the effect on job finding rate is insignificant.

This quantitative exercise shows that even a small amount of bargaining friction could have a significant impact on the volatility of job separations, and thus on the unemployment volatility through the separation channel.

7 Conclusion

In this paper, I have shown that bargaining frictions can have a profound effect on unemployment dynamics by distorting the efficiency of job separations. The paper's central contribution is to identify a distinct separation-based amplification mechanism: when bargaining surpluses differ from actual surpluses, the privately optimal separation cutoff differs from the efficient cutoff, so recessions generate not only weaker job creation but also excessive separations. This distinguishes the model from standard search-and-matching frameworks with endogenous job destruction and flexible Nash bargaining, where separations may vary over the cycle but remain privately efficient.

Using a reduced-form wedge specification, I show that several bargaining environments—including alternating-offers bargaining and asymmetric information about worker productivity—can be mapped into the same class of models. This mapping is useful not only because it nests these frameworks, but because it clarifies a common sufficient statistic, the net bargaining wedge, that governs wage distortion, inefficient separations, and the contribution of inefficient unemployment to aggregate volatility. Quantitatively, I find that even a small bargaining wedge relative to average surplus can substantially increase unemployment volatility, primarily through the separation channel.

I have shown that the alternating-offers bargaining model and the asymmetric-information environment can both be represented within the same class of models as Nash bargaining with bargaining wedges. These micro-foundations are, by no means, exhaustive. Introducing a firing cost, for example, would also create some bargaining wedges. Instead of setting up a complicated bargaining environment to generate high unemployment volatility, this paper shows that one can simply introduce bargaining wedges to the standard Nash bargaining problem. In addition, the bargaining wedge specification goes beyond pure generalization of Hall and Milgrom (2008) and Kennan (2010) and gives important insights about inefficient separations and unemployment volatility.

Further research can be done to improve the business cycle performance of the model. For example, the common failure of the MP model to produce enough volatility of the vacancy, and hence market tightness, also exists in the presence of bargaining friction. To address the problem, one can model the vacancy creation process by relaxing the free entry condition in the spirit of Fujita and Ramey (2007) and Coles and Moghaddasi Kelishomi (2018). Another extension of the model would be to consider also the on-the-job search. This would likely produce the Beveridge curve observed in the data, as shown in Fujita and Ramey (2012).

Finally, a more detailed identification of inefficient job separations can also be a worthwhile undertaking for future investigation. In the quantitative analysis, while I identify the bargaining wedge and thus inefficient unemployment by using aggregate labor market statistics, a more direct way would be to identify them from the micro-level data, which I will leave for further research. I believe the framework developed in this paper can be potentially useful to investigate both the quantity and the cause of inefficient job separations in the labor market.

References

Anderlini, L. and Felli, L. (2001). Costly bargaining and renegotiation. *Econometrica*, 69:377–411.

Antel, J. J. (1985). Costly employment contract renegotiation and the labor mobility of young men. *American Economic Review*, 75:976–991.

Barnichon, R. (2010). Building a composite help-wanted index. *Economics Letters*, 109:175–178.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70:9–49.

Binmore, K., Rubinstein, A., and Wolinsky, A. (1986). The nash bargaining solution in economic modelling. *The RAND Journal of Economics*, 17:176.

Blanchard, O. and Galì, J. (2010). Labor markets and monetary policy: A new keynesian model with unemployment. *American Economic Journal: Macroeconomics*, 2:1–30.

Busch, L.-A. and Wen, Q. (1995). Perfect equilibria in a negotiation model. *Econometrica*, 63:545.

Chatterjee, K. and Samuelson, L. (1987). Bargaining with two-sided incomplete information: An infinite horizon model with alternating offers. *The Review of Economic Studies*, 54:175.

Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2016). Unemployment and business cycles. *Econometrica*, 84:1523–1569.

Coles, Melvyn G, & Moghaddasi Kelishomi, Ali. 2018. Do Job Destruction Shocks Matter in the Theory of Unemployment. *American Economic Journal: Macroeconomics*, **10**(3), 118–136.

Davis, S. J., Faberman, R. J., and Haltiwanger, J. (2012). Labor market flows in the cross section and over time. *Journal of Monetary Economics*, 59:1–18.

den Haan, W. J., Ramey, G., and Watson, J. (1999). Contract-theoretic approaches to wages and displacement. *Review*, 81.

Elsby, M. W. L., Michaels, R., and Solon, G. (2009). The ins and outs of cyclical unemployment. *American Economic Journal: Macroeconomics*, 1:84–110.

Fujita, S. and Nakajima, M. (2016). Worker flows and job flows: A quantitative investigation. *Review of Economic Dynamics*, 22:1–20.

Fujita, S. and Ramey, G. (2007). Job matching and propagation. *Journal of Economic Dynamics and Control*, 31:3671–3698.

Fujita, S. and Ramey, G. (2009). The cyclicality of separation and job finding rates. *International Economic Review*, 50:415–430.

Fujita, S. and Ramey, G. (2012). Exogenous versus endogenous separation. *American Economic Journal: Macroeconomics*, 4:68–93.

Gerali, A., Guglielminetti, E., and Liberati, D. (2021). (in)efficient separations, firing costs and temporary contracts.

Gertler, M. and Trigari, A. (2009). Unemployment fluctuations with staggered nash wage bargaining. *Journal of Political Economy*, 117:38–86.

Gielen, A. C. and van Ours, J. C. (2006). Why do worker-firm matches dissolve? *IZA Discussion Papers*.

Guerrieri, V. (2008). Inefficient unemployment dynamics under asymmetric information. *Journal of Political Economy*, 116:667–708.

Hagedorn, M. and Manovskii, I. (2008). The cyclical behavior of equilibrium unemployment and vacancies revisited. *American Economic Review*, 98:1692–1706.

Hall, R. (2005a). *Job Loss, Job Finding, and Unemployment in the U.S. Economy Over the Past Fifty Years*, volume 20, pages 101–137.

Hall, R. E. (2005b). Employment efficiency and sticky wages: Evidence from flows in the labor market. *Review of Economics and Statistics*, 87:397–407.

Hall, R. E. (2005c). Employment fluctuations with equilibrium wage stickiness. *American Economic Review*, 95:50–65.

Hall, R. E. and Lazear, E. P. (1984). The excess sensitivity of layoffs and quits to demand. *Journal of Labor Economics*, 2:233–257.

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

Haltiwanger, J. C. (1984). Asymmetric information, multiperiod labor contracts, and inefficient job separations. *Southern Economic Journal*, 50:1005.

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *The Review of Economic Studies*, 57:279.

Jäger, S., Schoefer, B., and Zweimüller, J. (2023). Marginal jobs and job surplus: A test of the efficiency of separations. *The Review of Economic Studies*, 90:1265–1303.

Kennan, J. (2010). Private information, wage bargaining and employment fluctuations. *Review of Economic Studies*, 77:633–664.

Kennan, J. and Wilson, R. (1993). Bargaining with private information. *Journal of Economic Literature*, 31:45–104.

Ljungqvist, L. and Sargent, T. J. (2017). The fundamental surplus. *American Economic Review*, 107:2630–2665.

McLaughlin, K. J. (1991). A theory of quits and layoffs with efficient turnover. *Journal of Political Economy*, 99:1–29.

Mortensen, D. and Nagypàl, È. (2007). Labor-market volatility in matching models with endogenous separations. *The Scandinavian Journal of Economics*, 109:645–665.

Mortensen, D. T. and Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies*, 61:397–415.

Myerson, R. B. (1984). Two-person bargaining problems with incomplete information. *Econometrica*, 52:461.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39:390–431.

Petrosky-Nadeau, N. and Zhang, L. (2017). Solving the diamond-mortensen-pissarides model accurately. *Quantitative Economics*, 8:611–650.

Pissarides, C. A. (2009). The unemployment volatility puzzle: Is wage stickiness the answer? *Econometrica*, 77:1339–1369.

Ramey, G. and Watson, J. (1997). Contractual fragility, job destruction, and business cycles. *The Quarterly Journal of Economics*, 112:873–911.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 50:97.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, 95:25–49.

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

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20:177–181.

Online Appendix

A Derivations

A.1 Present-discounted value vs. flow value

The original Mortensen-Pissarides model uses the flow value of wages in defining the value of being employed, which in the model considered here is given by

$$E_t(\varepsilon_t) = w_t(\varepsilon_t) + \beta \mathbb{E}_t \left\{ \begin{array}{c} \left[s + (1-s) \left(\begin{array}{c} \lambda \chi_{t+1}(\varepsilon_{t+1}) \\ +(1-\lambda)\chi_{t+1}(\varepsilon_t) \end{array} \right) \right] U_{t+1} \\ +(1-s) \left[\begin{array}{c} \lambda \left(1-\chi_{t+1}(\varepsilon_{t+1})\right) E_{t+1}(\varepsilon_{t+1}) \\ +(1-\lambda)\left(1-\chi_{t+1}(\varepsilon_t)\right) E_{t+1}(\varepsilon_t) \end{array} \right] \end{array} \right\}$$

If we also define the present value of wages as

$$W_t(\varepsilon_t) = w_t(\varepsilon_t) + \beta(1-s)\mathbb{E}_t \left[\begin{array}{c} \lambda\left(1-\chi_{t+1}(\varepsilon_{t+1})\right) W_{t+1}(\varepsilon_{t+1}) \\ +(1-\lambda)\left(1-\chi_{t+1}(\varepsilon_t)\right) W_{t+1}(\varepsilon_t) \end{array} \right]$$

Then it is easy to see that

$$V_t(\varepsilon_t) = E_t(\varepsilon_t) - W_t(\varepsilon_t)$$

Moreover, the value of a matched firm is

$$\begin{aligned} F_t(\varepsilon_t) &= P_t(\varepsilon_t) - W_t(\varepsilon_t) \\ &= p_t(\varepsilon_t) - w_t(\varepsilon_t) \\ &\quad +\beta(1-s)\mathbb{E}_t\left[\lambda\left(1-\chi_{t+1}(\varepsilon_{t+1})\right)F_{t+1}(\varepsilon_{t+1}) + (1-\lambda)\left(1-\chi_{t+1}(\varepsilon_t)\right)F_{t+1}(\varepsilon_t)\right] \end{aligned}$$

which is consistent with the standard MP model with flow wages. Hence, all the analytical results still hold if wages are expressed in terms of present discounted value $W_t(\varepsilon_t)$ instead of flow value $w_t(\varepsilon_t)$.

A.2 Steady-state flow wage and free-entry condition

This subsection derives the steady-state flow wage equation and the corresponding free-entry condition with bargaining wedge.

For a continuing match with $\varepsilon \geq \varepsilon^s$, the firm's surplus satisfies

$$F(\varepsilon) = p(\varepsilon) - w^{ss}(\varepsilon) + \beta(1-s)\left[(1-\lambda)F(\varepsilon) + \lambda \int_{\varepsilon^s}^{\varepsilon_{\max}} F\left(\varepsilon'\right) dG\left(\varepsilon'\right)\right] \tag{A.1}$$

where $w^{ss}(\varepsilon)$ is the flow wage and ε^s is the separation cutoff. Free entry condition of the firm implies

$$c = \beta q\left(\theta^{ss}\right) \int_{\varepsilon^s}^{\varepsilon_{\max}} F\left(\varepsilon\right) dG\left(\varepsilon\right) \tag{A.2}$$

so that the cost of vacancy equals expected profit.

On the worker side, the employment value is

$$E(\varepsilon) = w^{ss}(\varepsilon) + \beta\{[s + (1-s)\lambda G(\varepsilon^s)]U + (1-s)[(1-\lambda)E(\varepsilon) + \lambda \int_{\varepsilon^s}^{\varepsilon_{\max}} E(\varepsilon')\, dG(\varepsilon')]\} \tag{A.3}$$

and the unemployment value is given by

$$U = b + \beta\{\theta^{ss} q\left(\theta^{ss}\right) \int_{\varepsilon^s}^{\varepsilon_{\max}} E(\varepsilon')\, dG(\varepsilon') + [1 - \theta^{ss} q\left(\theta^{ss}\right)\left(1 - G(\varepsilon^s)\right)]U\} \tag{A.4}$$

Subtracting (A.4) from (A.3) yields

$$E(\varepsilon) - U = w^{ss}(\varepsilon) - b + \beta(1-s)(1-\lambda)\left[E(\varepsilon) - U\right] + \beta\left[\lambda(1-s) - \theta^{ss} q\left(\theta^{ss}\right)\right] \int_{\varepsilon^s}^{\varepsilon_{\max}} \left(E(\varepsilon') - U\right)\, dG(\varepsilon')) \tag{A.5}$$

The generalized Nash bargaining problem in steady state implies

$$\left(1-\eta\right)\left(E(\varepsilon) - U - \Omega^w\right) = \eta\left(F(\varepsilon) - \Omega^f\right) \tag{A.6}$$

or

$$\left(1-\eta\right)\left(E(\varepsilon) - U\right) = \eta F(\varepsilon) + \Omega^{net} \tag{A.7}$$

Integrating over the continuation region gives

$$\left(1-\eta\right) \int_{\varepsilon^s}^{\varepsilon_{\max}} \left(E(\varepsilon) - U\right) dG(\varepsilon) = \eta \int_{\varepsilon^s}^{\varepsilon_{\max}} F(\varepsilon) dG(\varepsilon) + \Omega^{net}(1 - G(\varepsilon^s)) \tag{A.8}$$

Combining (A.1), (A.2), (A.5), (A.7), (A.8), and after rearranging, we arrive at

$$w^{ss}(\varepsilon) = \eta\left(p(\varepsilon) + \theta^{ss} c\right) + \left(1-\eta\right) b + \left[1 - \beta(1-s)\left[1 - \lambda G(\varepsilon^s)\right] + \beta\theta^{ss} q\left(\theta^{ss}\right)\left(1 - G(\varepsilon^s)\right)\right] \Omega^{net} \tag{A.9}$$

To derive the free-entry condition, integrating (A.1) yields

$$\int_{\varepsilon^s}^{\varepsilon_{\max}} F(\varepsilon)dG(\varepsilon) = \int_{\varepsilon^s}^{\varepsilon_{\max}} p(\varepsilon) - w^{ss}(\varepsilon)dG(\varepsilon) + \beta(1-s)\left[1-\lambda G(\varepsilon^s)\right]\int_{\varepsilon^s}^{\varepsilon_{\max}} F(\varepsilon)dG(\varepsilon) \tag{A.10}$$

Using (A.2), we have the free-entry condition for the flow model:

$$\frac{c}{\beta q\left(\theta^{ss}\right)} = \frac{\int_{\varepsilon^s}^{\varepsilon_{\max}} p(\varepsilon) - w^{ss}(\varepsilon)dG(\varepsilon)}{1-\beta(1-s)\left[1-\lambda G(\varepsilon^s)\right]} \tag{A.11}$$

A.3 Proof of Proposition 1

By putting $\varepsilon_t = \varepsilon_t^e$ in (13), we have

$$V_t\left(\varepsilon_t^e\right) + W_t\left(\varepsilon_t^e\right) - U_t = (1-\eta)\,\Omega_t^w - \eta\Omega_t^f \tag{A.12}$$

$$P_t\left(\varepsilon_t^e\right) - W_t\left(\varepsilon_t^e\right) = -\left[(1-\eta)\,\Omega_t^w - \eta\Omega_t^f\right] \tag{A.13}$$

Hence, if $(1-\eta)\,\Omega_t^w - \eta\Omega_t^f = 0$, then from (A.12) and (A.13), clearly $\varepsilon_t^w = \varepsilon_t^e = \varepsilon_t^f$. On the other hand, if $(1-\eta)\,\Omega_t^w - \eta\Omega_t^f > 0$, then from (A.12) and using the fact that $V_t\left(\varepsilon_t\right)$ and $W_t\left(\varepsilon_t\right)$ are strictly increasing, we have $\varepsilon_t^e > \varepsilon_t^w$. Also, from (A.13) and using the fact that $P_t\left(\varepsilon_t\right) - W_t\left(\varepsilon_t\right)$ is also strictly increasing in ε_t, we have $\varepsilon_t^f > \varepsilon_t^e$. Combining, we have $\varepsilon_t^w < \varepsilon_t^e < \varepsilon_t^f$. Similarly, if $(1-\eta)\,\Omega_t^w - \eta\Omega_t^f < 0$, then again from (A.12) and (A.13) respectively, we have $\varepsilon_t^e < \varepsilon_t^w$ and $\varepsilon_t^f < \varepsilon_t^e$.

A.4 Special case

Here I consider a special case where $\lambda = 1$ and $p_t\left(\varepsilon_t\right) = y_t\varepsilon_t$. Notice first that when $\lambda = 1$, the value of working V_t is independent of the current idiosyncratic shock. Putting $\varepsilon_t = \varepsilon_t^f$ into (2) and using the definition of ε_t^f, we have

$$W_t\left(\varepsilon_t^f\right) = P_t\left(\varepsilon_t^f\right) = p_t\left(\varepsilon_t^f\right) + \beta\left(1-s\right)\mathbb{E}_t\left(1-\chi_{t+1}\left(\varepsilon_{t+1}\right)\right)P_{t+1}\left(\varepsilon_{t+1}\right) \tag{A.14}$$

Also, putting $\varepsilon_t = \varepsilon_t^e$ into (2) and using the definition of ε_t^e, we have

$$U_t - V_t = P_t\left(\varepsilon_t^e\right) = p_t\left(\varepsilon_t^e\right) + \beta\left(1-s\right)\mathbb{E}_t\left(1-\chi_{t+1}\left(\varepsilon_{t+1}\right)\right)P_{t+1}\left(\varepsilon_{t+1}\right) \tag{A.15}$$

Combining (A.14) and (A.15) yields

$$V_t + W_t\left(\varepsilon_t^f\right) - U_t = p_t\left(\varepsilon_t^f\right) - p_t\left(\varepsilon_t^e\right)$$

Now using the bargaining solution (13) and rearranging, we have

$$p_t\left(\varepsilon_t^f\right) - p_t\left(\varepsilon_t^e\right) = \frac{1}{1-\eta}\left[(1-\eta)\,\Omega_t^w - \eta\Omega_t^f\right]$$

Hence, the result (i) follows by putting $p_t(\varepsilon_t) = y_t\varepsilon_t$. Similarly, by putting $\varepsilon_t = \varepsilon_t^w$ into (2) again, we have

$$P_t\left(\varepsilon_t^w\right) = p_t\left(\varepsilon_t^w\right) + \beta(1-s)\,\mathbb{E}_t\left(1-\chi_{t+1}\left(\varepsilon_{t+1}\right)\right)P_{t+1}\left(\varepsilon_{t+1}\right)$$

Now using (A.15) and the definition of ε_t^w, we have

$$\begin{aligned} W_t\left(\varepsilon_t^w\right) - P_t\left(\varepsilon_t^w\right) &= U_t - V_t - P_t\left(\varepsilon_t^w\right) \\ &= p_t\left(\varepsilon_t^e\right) - p_t\left(\varepsilon_t^w\right) \end{aligned}$$

Finally, using (13) yields

$$p_t\left(\varepsilon_t^e\right) - p_t\left(\varepsilon_t^w\right) = \frac{1}{\eta}\left[(1-\eta)\,\Omega_t^w - \eta\Omega_t^f\right]$$

Hence, the result (ii) follows by noting $p_t(\varepsilon_t) = y_t\varepsilon_t$.

A.5 AOB model

Solving (28) and (29), the unique SPNE wage is given by

$$W_t(\varepsilon_t) = \frac{1}{2-\delta}\left[(1-\delta)\,P_t(\varepsilon_t) + U_t - V_t(\varepsilon_t) + \frac{(1-\delta)^2}{\delta}\gamma_t + \frac{1-\delta}{\delta}z_t\right]$$

Rearranging, we have

$$\frac{1-\delta}{2-\delta}\left[P_t(\varepsilon_t) - W_t(\varepsilon_t) - \left(-\frac{1-\delta}{\delta}\gamma_t\right)\right] = \frac{1}{2-\delta}\left[V_t(\varepsilon_t) + W_t(\varepsilon_t) - U_t - \frac{1-\delta}{\delta}z_t\right]$$

Hence, Proposition 2 follows by comparing with (13).

From Proposition 2 and the expression (16), we arrive at

$$\varepsilon_t^f - \varepsilon_t^e = \frac{1}{y_t}\left[\frac{1-\delta}{\delta} z_t + \frac{(1-\delta)^2}{\delta}\gamma_t\right]$$

Differentiating, we have

$$\begin{aligned}
\frac{\partial\left(\varepsilon_t^f - \varepsilon_t^e\right)}{\partial z_t} &= \frac{1}{y_t}\left[\frac{1-\delta}{\delta}\right] \\
\frac{\partial\left(\varepsilon_t^f - \varepsilon_t^e\right)}{\partial \gamma_t} &= \frac{1}{y_t}\left[\frac{(1-\delta)^2}{\delta}\right] \\
\frac{\partial\left(\varepsilon_t^f - \varepsilon_t^e\right)}{\partial \delta} &= -\frac{1}{y_t}\left[\frac{z_t + \left(1-\delta^2\right)\gamma_t}{\delta^2}\right]
\end{aligned}$$

A.6 Fundamental surplus in Ljungqvist and Sargent (2017)

Consider the steady state MP model with exogenous separation and homogeneous productivity in Ljungqvist and Sargent (2017). After setting $p(\varepsilon) = y$, $b = z$ and $\beta = (1+r)^{-1}$, the homogeneous version of the wage function (A.9) is

$$w = \eta(y + \theta c) + (1-\eta) z + \frac{r + s + \theta q(\theta)}{1+r}\Omega^{net} \tag{A.16}$$

Also, the homogeneous version of equation (A.11) is

$$\frac{c}{q(\theta)} = \frac{y-w}{r+s} \tag{A.17}$$

or

$$w = y - \frac{c(r+s)}{q(\theta)} \tag{A.18}$$

Equating (A.16) and (A.18) and rearranging, we have

$$\frac{1-\eta}{c}\left[y - z - \frac{\Omega^{net}(r+s)}{(1-\eta)(1+r)}\right] = \frac{r+s}{q(\theta)} + \eta\theta + \frac{\theta q(\theta)}{c(1+r)}\Omega^{net}$$

Hence, implicitly differentiating and rearranging, we get

$$\begin{aligned}\frac{d\theta}{dy} &= \frac{r+s+\eta\theta q(\theta)\left(1+\beta q(\theta)\frac{\Omega^{net}}{c\eta}\right)}{\alpha(r+s)+\eta\theta q(\theta)\left(1+(1-\alpha)\beta q(\theta)\frac{\Omega^{net}}{c\eta}\right)}\frac{\theta}{y-z-\frac{1-\beta(1-s)}{(1-\eta)}\Omega^{net}} \\ &= \Upsilon\left(\Omega^{net}\right)\frac{\theta}{y-z-\frac{1-\beta(1-s)}{(1-\eta)}\Omega^{net}}\end{aligned}$$

where $\alpha = \frac{-\theta q'(\theta)}{q(\theta)}$ is the elasticity of the matching function. Hence, the result follows by noting that $\epsilon_{\theta,y} = \frac{y}{\theta}\frac{d\theta}{dy}$.

B Business cycle statistics

In this appendix, I present the business cycle statistics in the data and in the model. I consider US labor market data from 1976 to 2016. The monthly unemployment rate series is taken from the official rates from the Labor Force Statistics of the Current Population Survey (CPS). The EU and UE transition rates are also taken from the same survey. For the period from January, 1976 to December, 2005, the separation rates and job finding rates are taken from Fujita and Ramey (2009). For the remaining years, I calculate the rates by $SR_{t+1} = \frac{EU_{t+1}}{e_t}$ and $FR_t = \frac{UE_{t+1}}{u_t}$, and adjust them for time aggregation error as in Fujita and Ramey (2009). Labor productivity p_t is defined by real GDP per workers, where the real GDP data is from Bureau of Economic Analysis and the number of employed workers is from CPS. Finally, the vacancy data is from the updated series by Barnichon (2010).

A summary table of the business cycle statistics of the US labor market is given in panel (a) of Table B.1. When calculating the business cycle statistics, all series are converted to quarterly data with simple averaging, as well as logged and HP filtered with smoothing parameter 1,600. In the table, σ_x^c denotes the cyclical volatility of the variable x, and $\epsilon_{x,p}$ denotes the elasticity of the variable x with respect to the labor productivity, as defined by

$$\epsilon_{x,p} = \frac{cov\left(x_t, p_t\right)}{var\left(p_t\right)} \tag{B.19}$$

It is worth noting that while the job finding rate is procyclical with an elasticity of 3.07 with respect to changes in labor productivity, the job separation rate is countercyclical with a comparable elasticity of -4.27. This shows that the separation margin of the labor market is indeed important at the business cycle frequency.

The business cycle statistics of the baseline calibration is given in panel (b) of Table B.1. Comparing with the data, we can see that the model matches well for the correlation matrix and the autocorrelation of the variables. A few significant differences can be observed. First, the volatility of job finding rate and the market tightness in the model are much lower than those in the data. This is related to Shimer (2005)'s finding that the MP model fails to generate enough volatility of the market tightness.[38] This problem can most likely be solved by relaxing the free entry condition, as shown in Coles and Moghaddasi Kelishomi (2018). Second, the model fails to produce the Beveridge curve (i.e. negative relationship between unemployment and vacancy). This is due to the fact that the volatility of separation rate in the model is too high compared with the data. As demonstrated in Fujita and Ramey

[38] It is well-noting, however, that the elasticity of job finding rate with respect to productivity in the model matches well with that in the data.

(a) Data

	p_t	u_t	FR_t	SR_t	UE_t	EU_t	v_t	θ_t
σ_x^c	0.008	0.109	0.085	0.064	0.044	0.057	0.128	0.233
$\epsilon_{x,p}$	1	-4.267	3.072	-4.266	-1.293	-4.059	7.884	12.130
$corr\,(x_t, x_{t-1})$	0.698	0.939	0.838	0.703	0.444	0.651	0.931	0.940

Correlation matrix

	p_t	u_t	FR_t	SR_t	UE_t	EU_t	v_t	θ_t
p_t	1	-0.295	0.272	-0.507	-0.221	-0.534	0.465	0.393
u_t		1	-0.921	0.750	0.646	0.663	-0.934	-0.981
FR_t			1	-0.612	-0.300	-0.524	0.887	0.918
SR_t				1	0.652	0.991	-0.779	-0.778
UE_t					1	0.617	-0.559	-0.609
EU_t						1	-0.708	-0.698
v_t							1	0.986
θ_t								1

(b) Simulated model

	p_t	u_t	FR_t	SR_t	UE_t	EU_t	v_t	θ_t
σ_x^c	0.009	0.109	0.027	0.103	0.084	0.097	0.069	0.044
$\epsilon_{x,p}$	1	-11.714	3.049	-10.865	-8.516	-10.167	-6.756	4.961
$corr\,(x_t, x_{t-1})$	0.812	0.863	0.824	0.703	0.861	0.681	0.852	0.823

Correlation matrix

	p_t	u_t	FR_t	SR_t	UE_t	EU_t	v_t	θ_t
p_t	1	-0.946	0.986	-0.933	-0.897	-0.922	-0.869	0.993
u_t		1	-0.959	0.887	0.990	0.870	0.981	-0.952
FR_t			1	-0.912	-0.914	-0.897	-0.887	0.998
SR_t				1	0.826	0.999	0.813	-0.929
UE_t					1	0.807	0.997	-0.902
EU_t						1	0.794	-0.916
v_t							1	-0.873
θ_t								1

Table B.1: Business cycle statistics of the labor market

Table B.2: Wage moments: data vs. model

Moment	Data	Model
Cyclical volatility	0.0105	0.0098
Autocorrelation	0.6891	0.8162
Elasticity w.r.t. productivity	0.1713	1.1083
Elasticity w.r.t. unemployment	0.0118	0.0895

(2012), this likely can be solved by introducing on-the-job search. Instead of matching all the summary statistics, I abstract from these complications by investigating how inefficient separations affect the labor market dynamics.

Table B.2 shows that the model does well in matching the overall size of wage fluctuations. The cyclical volatility of wages in the model is very close to that in the data. At the same time, wage fluctuations in the model are only slightly more persistent than in the data. However, the model makes wages too responsive to macroeconomic conditions. The wage elasticity with respect to productivity is much larger in the model than in the data, and the reported elasticity with respect to unemployment is also larger. In other words, the model captures the volatility of wages but overstates the systematic transmission of aggregate conditions into wages. It is well-documented that there exists some degree of wage rigidity in the data which is not captured by the standard model. Richer wage-setting environments of the kind emphasized by Gertler and Trigari (2009) could plausibly reduce the wage elasticities further, but introducing those additional sources of wage smoothing would move the paper away from its main goal of isolating the separation channel.

C Computation strategy

First, the stochastic processes of the aggregate and idiosyncratic productivity are discretized as follows. The AR(1) stochastic process (42) is approximated by a finite state Markov chain with N states and a transition probability matrix $[\pi_{i,j}]$ by using the Tauchen (1986) procedure. Also, I approximate the truncated log-normal distribution with K points of support $\{\varepsilon_1, ..., \varepsilon_K\}$, with $G(\varepsilon_K) = 1$. Let $g(\cdot)$ be the probability mass function of the discretized distribution. In the computation exercise, I set $N = 10$ and $K = 400$.

Next, I solve the model by using a fixed point algorithm. Since only the net bargaining wedge matters, I assume $\Omega^w = -\Omega^f \equiv \Omega$. In this case, we have $\Omega^{net} = \Omega$. Then I use a standard iteration procedure to find a fixed point. Specifically, given initial guess of $\{P_0, W_0, V_0, U_0, \theta_0, \chi_0\}$, I update the value functions and the market tightness as follows (derived from equations (1) - (5) and (14)).

$$
\begin{aligned}
\theta_{n+1}(y_i) &= \left\{ \begin{array}{c} \left(\frac{\beta A}{c}\right) \sum_{j=1}^{N} \sum_{k=1}^{K} g(\varepsilon_k) \pi_{i,j} \\ \left[1 - \chi_n(y_j, \varepsilon_k)\right] \left[P_n(y_j, \varepsilon_k) - W_n(y_j, \varepsilon_k) - k\right] \end{array} \right\}^{\frac{1}{\alpha}} \\
P_{n+1}(y_i, \varepsilon_l) &= y_i \varepsilon_l + \beta (1-s) \sum_{j=1}^{N} \pi_{i,j} \left[\begin{array}{c} \lambda \sum_{k=1}^{K} g(\varepsilon_k) \left[1 - \chi_n(y_j, \varepsilon_k)\right] P_n(y_j, \varepsilon_k) \\ + (1-\lambda) \left[1 - \chi_n(y_j, \varepsilon_l)\right] P_n(y_j, \varepsilon_l) \end{array} \right] \\
V_{n+1}(y_i, \varepsilon_l) &= \beta \sum_{j=1}^{N} \pi_{i,j} \left\{ \begin{array}{c} \left[s + (1-s) \left[\begin{array}{c} \lambda \sum_{k=1}^{K} g(\varepsilon_k) \chi_n(y_j, \varepsilon_k) \\ + (1-\lambda) \chi_n(y_j, \varepsilon_l) \end{array} \right] \right] U_n(y_j) \\ + (1-s) \left[\begin{array}{c} \lambda \sum_{k=1}^{K} g(\varepsilon_k) \left[1 - \chi_n(y_j, \varepsilon_k)\right] V_n(y_j, \varepsilon_k) \\ + (1-\lambda) \left[1 - \chi_n(y_j, \varepsilon_l)\right] V_n(y_j, \varepsilon_l) \end{array} \right] \end{array} \right\} \\
U_{n+1}(y_i) &= b + \beta \sum_{j=1}^{N} \sum_{k=1}^{K} g(x_k) \pi_{i,j} \\
& \left[\begin{array}{c} A\theta_n(y_i)^{1-\alpha} \left[1 - \chi_n(y_j, \varepsilon_k)\right] \left[W_n(y_j, \varepsilon_k) + V_n(y_j, \varepsilon_k)\right] \\ + \left(1 - A\theta_n(y_i)^{1-\alpha} + A\theta_n(y_i)^{1-\alpha} \chi_n(y_j, \varepsilon_k)\right) U_n(y_j) \end{array} \right] \\
W_{n+1}(y_i, \varepsilon_l) &= \eta P_{n+1}(y_i, \varepsilon_l) + (1-\eta) \left(U_{n+1}(y_i) - V_{n+1}(y_i, \varepsilon_l)\right) + \Omega \\
\chi_{n+1}(y_i, \varepsilon_l) &= \begin{cases} 1 & \begin{array}{c} P_{n+1}(y_i, \varepsilon_l) \leq W_{n+1}(y_i, \varepsilon_l) \text{ or} \\ W_{n+1}(y_i, \varepsilon_l) + V_{n+1}(y_i, \varepsilon_l) \leq U_{n+1}(y_i) \end{array} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

for all $i = 1, ..., N$ and $l = 1, ..., K$. The error of iteration is defined by

$$
\max\left\{\|P_{n+1} - P_n\|, \|V_{n+1} - V_n\|, \|U_{n+1} - U_n\|, \|\theta_{n+1} - \theta_n\|\right\}
$$

where $\|\cdot\|$ is the entrywise matrix norm. The iteration continues until the error is less than the tolerance level. In the computation exercise, I set the tolerance level to be 10^{-6}. From the solution of $\{P, W, V, U\}$ from above, I then find the cutoff productivities $\{\varepsilon^w, \varepsilon^f, \varepsilon^e\}$ from the definitions (10) - (9).

To compute the unemployment volatility, I simulate a time series of unemployment by using the flow equations (18) - (20). I choose the sample size to be $12,000$ weeks, which is equivalent to 1000 quarters. The weekly series is transformed to quarterly frequency by calculating the simple average in a quarter. Finally, the log deviation of the quarterly series computed by using a hp-filter with a coefficient of 1600, and the cyclical part is used to measure the unemployment volatility.

D Additional quantitative results

D.1 Alternative aggregate productivity shock process

Here I estimate the aggregate productivity shock process by using the cyclical component of the quarterly labor productivity time series $\{\tilde{y}_t\}$. I first estimate an AR(1) process:

$$\tilde{y}_{t+1} = \rho \tilde{y}_t + \varepsilon_t$$

where

$$\varepsilon_t \sim N\left(0, \sigma^2\right)$$

Then I convert the quarterly process into weekly process as follows:

$$\begin{aligned} \hat{\alpha} &= \hat{\rho}^{\frac{1}{13}} \\ \hat{\gamma} &= \frac{\hat{\sigma}^2}{\sum_{i=0}^{12} \hat{\alpha}^{2i}} \end{aligned}$$

where $\hat{\alpha}$ and $\hat{\gamma}$ are the estimates for the persistence and standard deviation of the weekly process. I got $\hat{\alpha} = 0.9889$ and $\hat{\gamma} = 0.00397$, which is slightly more volatile than the baseline calibration. The quantitative results are shown in Figure D.1. As a result of the more volatile productivity shocks, the unemployment volatility is higher for each value of Ω^{net}. Therefore, the value of Ω^{net} to produce a realistic level of unemployment volatility is a bit lower at about 0.17. The contribution of inefficient unemployment is almost the same at 29% (vs. 29.1% for the baseline). I conclude that neither qualitative nor quantitative results have changed due to the more volatile process.

D.2 Alternative distribution for the idiosyncratic shocks

In the baseline calibration, the new drawing of the idiosyncratic shocks follows a truncated log-normal distribution. Here I consider an alternative specification where it is gamma distributed. Specifically, I calibrate the shape parameter κ_1 and the scale parameter κ_2 to target the average productivity and average job separation rate. The results are shown in Figure D.2.[39] While we also observe an increasing unemployment volatility as Ω^{net} increases as in the baseline, the curvature is different. In fact, for large values of Ω^{net}, unemployment volatility seems to be less sensitive to the bargaining friction. It is likely because the

[39] The estimated parameters are $\hat{\kappa}_1 = 50$ and $\hat{\kappa}_2 = 0.02$.

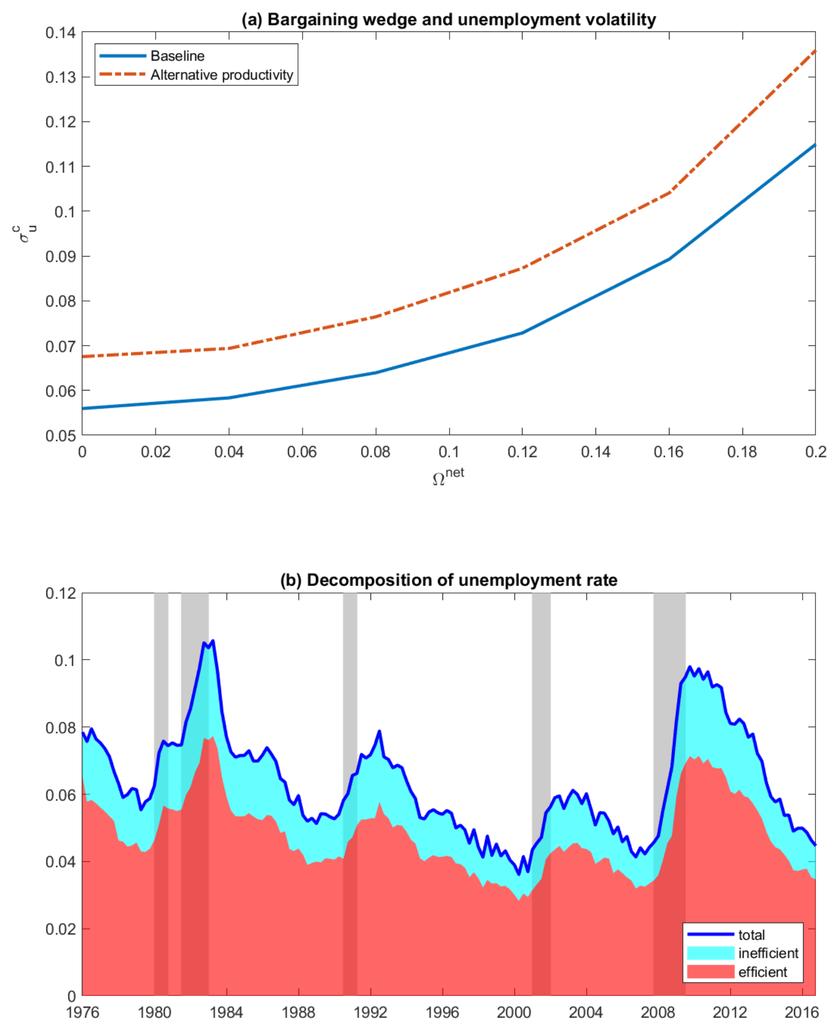

Figure D.1: Robustness: alternative aggregate productivity process

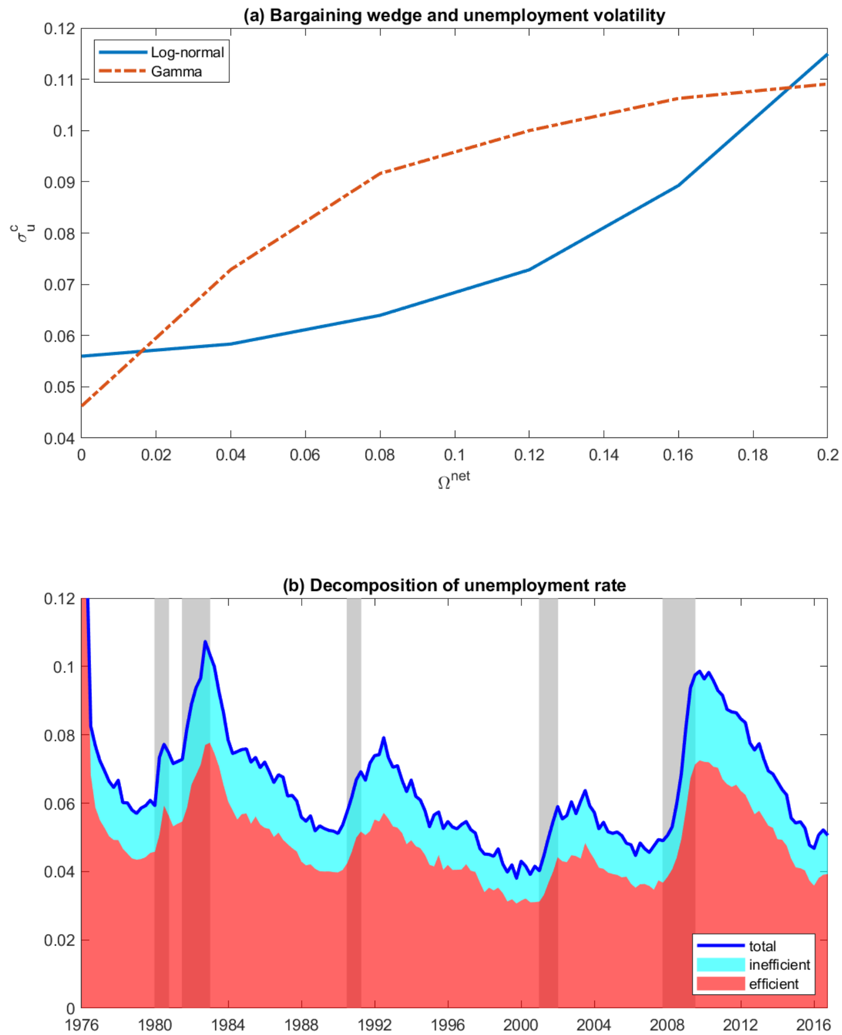

Figure D.2: Robustness: alternative idiosyncratic shocks

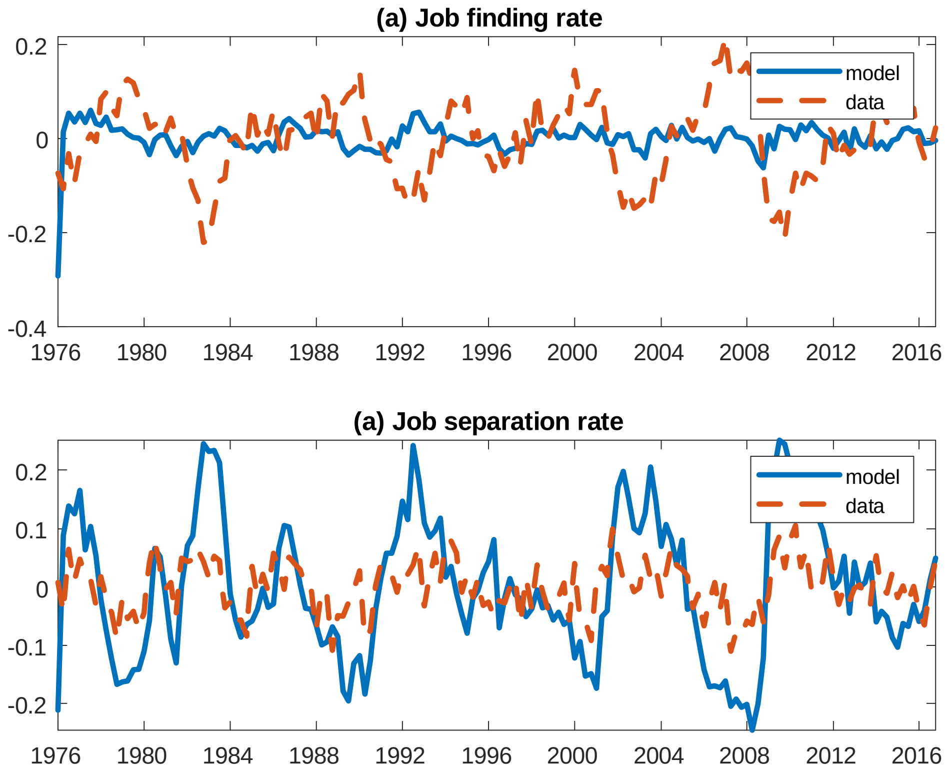

Figure D.3: Job finding and separation rates in the model vs. data

calibrated gamma distribution has a smaller right tail, hence the separations in the model, which mostly happen on the right tail, is less sensitive to changes in the specification of wages. The resulting estimated Ω^{net} is higher at around 0.20, and the importance of inefficient unemployment is slightly higher at 30%. Overall, neither qualitative nor quantitative results have changed for this alternative distribution.

D.3 Job finding and separations rates

Figure D.3 shows the job finding and separation rates in the model and in the data, respectively. As seen in the comparison of business cycle statistics, while the cyclicality of the transition rates in the model tend to follow that in the data, the model produces a less volatile job finding rate and more volatile job separation rate than in the data.

D.4 Constant bargaining wedge

Here I consider another specification of the matching process by assuming that the bargaining wedge is constant over the business cycle. As mentioned before, in the data, the average

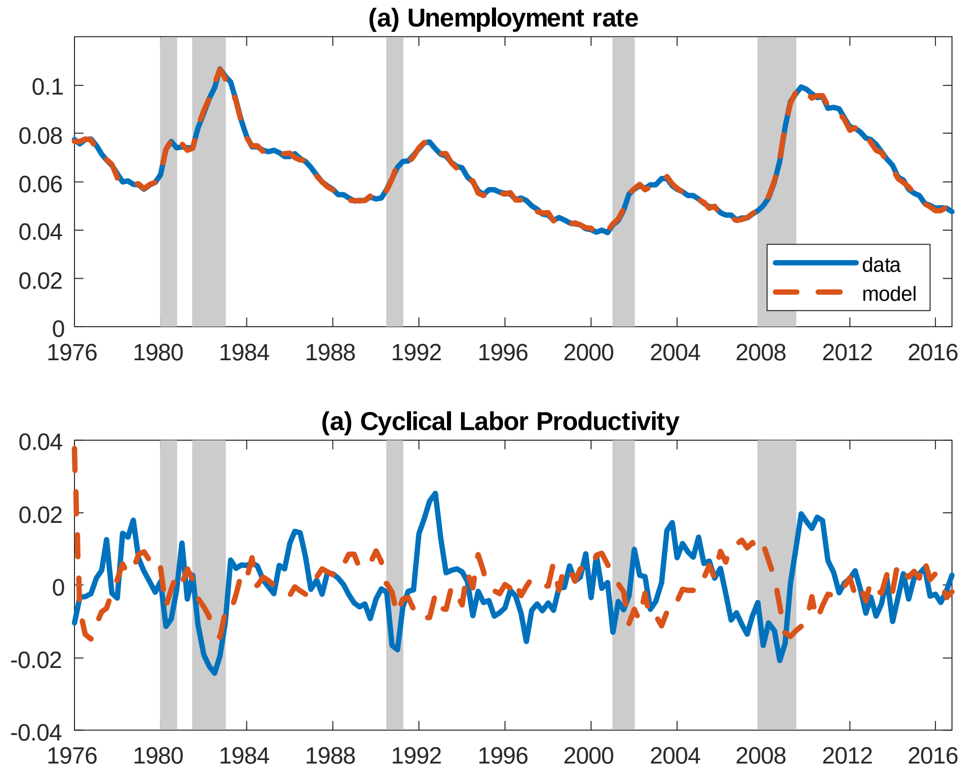

Figure D.4:
Unemployment rate and cyclical productivity (contant bargaining wedge)

unemployment volatility is about 10.9%, entailing a bargaining wedge of 0.188. I then perform the same procedure to decompose the unemployment rate. Figure D.4 and D.5 shows the results.

We can see a similar decomposition with the case of cyclical bargaining wedge. In this case, we have $\gamma = 0.22$. Hence, an acyclical bargaining wedge alone can explain about 22% of the unemployment volatility in the data. Adding cyclicality to the bargaining wedge increases the explanatory power by another 7%.

D.5 Sensitivity analysis

Sensitivity analysis for various parameters is shown in Table D.1. Panel (a) varies the flow value of unemployment b. As b increases, cyclical unemployment volatility rises. The economic intuition is consistent the small-surplus literature. At the same time, the share of volatility explained by inefficient unemployment remains fairly stable. Thus, changing b mainly alters the overall amount of amplification rather than overturning the importance of inefficient separations.

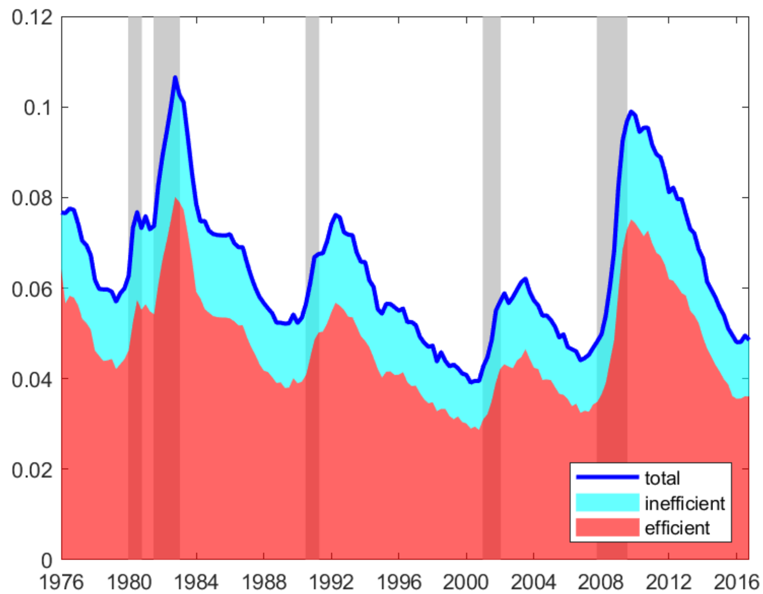

Figure D.5: Decomposition of unemployment rate (constant bargaining wedge)

Table D.1: Sensitivity analysis

Case	Cyclical unemployment volatility, σ_u^c	Contribution of inefficient unemployment, γ
Baseline	0.109	0.291
(a) Varying b		
$b = 0.60$	0.063	0.278
$b = 0.65$	0.090	0.304
$b = 0.75$	0.131	0.279
$b = 0.80$	0.202	0.256
(b) Varying η		
$\eta = 0.40$	0.102	0.321
$\eta = 0.45$	0.107	0.310
$\eta = 0.55$	0.113	0.274
$\eta = 0.60$	0.211	0.254
(c) Varying λ		
$\lambda = 0.08$	0.074	0.280
$\lambda = 0.09$	0.085	0.289
$\lambda = 0.11$	0.112	0.289
$\lambda = 0.12$	0.112	0.226

Notes. The baseline calibration sets $b = 0.71$, $\eta = 0.5$, and $\lambda = 0.104$. I vary one parameter at a time while holding the baseline net bargaining wedge fixed at $\Omega^{net} = 0.1881$ and all other parameters at their baseline values. The column σ_u^c reports the standard deviation of quarterly cyclical unemployment, and $\gamma \equiv \text{cov}(u_t, u_t^i)/\text{var}(u_t)$ is the fraction of unemployment volatility explained by inefficient unemployment.

Panel (b) varies the worker's bargaining power η. Moving η away from the baseline changes both wage determination and the size of the effective separation distortion. A larger η shifts the wage bargain further in favor of the worker and tends to compress the firm's surplus, which weakens job creation and raises unemployment volatility. This effect is modest as η increases. The contribution of inefficient unemployment, however, declines from 0.321 at $\eta = 0.40$ to 0.254 at $\eta = 0.60$.

Panel (c) varies the switching rate of match-specific productivity λ. Raising λ increases unemployment volatility from 0.074 to 0.112. Intuitively, more frequent redraws make both efficient and inefficient separations more responsive to aggregate conditions. The contribution of inefficient unemployment is remarkably stable around $0.28 - 0.29$ for $\lambda \in [0.08, 0.11]$, which suggests that the decomposition is quite robust over economically plausible values.

Overall, Table D.1 shows that the main quantitative conclusions are robust. The exact level of unemployment volatility depends on the calibration, as expected, but inefficient unemployment remains a quantitatively meaningful source of business-cycle fluctuations across all parameter values considered here.