A Macroeconomic Theory of Optimal Unemployment Insurance

Camille Landais, Pascal Michaillat, Emmanuel Saez *

November 20, 2014

ABSTRACT

This paper develops a theory of optimal unemployment insurance (UI) that accounts not only for workers' job-search behavior but also for firms' hiring behavior. The theoretical framework is a matching model of the labor market in which the production function, wage mechanism, matching function, and utility function are general. The labor market tightness plays a central role in the theory. The tightness is defined as the ratio of job vacancies to aggregate job-search effort and is determined in equilibrium to equalize labor demand and supply. We show that the optimal level of UI is the sum of a conventional Baily-Chetty term, which captures the trade-off between insurance and job-search incentives, and a correction term, which is positive if UI brings labor market tightness closer to its efficient level. Hence, optimal UI depends on whether tightness is inefficiently low or inefficiently high and whether UI raises or lowers tightness. For instance, if tightness is inefficiently low, optimal UI is more generous than the Baily-Chetty level if UI raises tightness and less generous if UI lowers tightness. Depending on the production function and wage mechanism, UI could raise tightness, for example by alleviating the rat race for jobs, or lower tightness, for example by increasing bargained wages. To determine whether UI raises or lowers tightness in practice, we develop an empirical criterion; the criterion involves comparing the microelasticity and macroelasticity of unemployment with respect to UI. Since tightness is subject to large fluctuations over the business cycle, our theory has direct implications for the cyclicality of optimal UI.

*Camille Landais: Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK; email: c.landais@lse.ac.uk. Pascal Michaillat: Department of Economics and Centre for Macroeconomics, London School of Economics, Houghton Street, London, WC2A 2AE, UK; email: p.michaillat@lse.ac.uk. Emmanuel Saez: Department of Economics, University of California, Berkeley, CA 94720, USA; email: saez@econ.berkeley.edu. This paper supersedes our earlier paper titled "Optimal Unemployment Insurance over the Business Cycle". We thank Daron Acemoglu, George Akerlof, Varanya Chaubey, Raj Chetty, Sanjay Chugh, Peter Diamond, Emmanuel Farhi, Jordi Galí, Mikhail Golosov, Yuriy Gorodnichenko, David Gray, Etienne Lehmann, Guido Lorenzoni, Emi Nakamura, Christopher Pissarides, Robert Shimer, Johannes Spinnewijn, Jón Steinsson, Aleh Tsyvinski, and numerous seminar and conference participants for helpful discussions and comments. This work was supported by the Center for Equitable Growth at the University of California Berkeley, the British Academy, the Economic and Social Research Council [grant number ES/K008641/1], and the Institute for New Economic Thinking.

1 Introduction

Unemployment insurance (UI) is a key component of social insurance because it provides relief to people who cannot find a job. The microeconomic theory of optimal UI is well understood: it is an insurance-incentive trade-off in the presence of moral hazard. UI helps workers smooth consumption when they are unemployed, but it also increases unemployment by discouraging job search. The Baily [1978]-Chetty [2006*a*] formula resolves this trade-off.

The macroeconomic theory of optimal UI, however, is less well understood.[1] This becomes problematic in slumps, when unemployment is unusually high, because adjusting UI could affect welfare through macroeconomic channels. For instance, some argue that UI should be decreased to exert downward pressure on wages, thereby encouraging job creation. Others argue that UI should be increased because in slumps jobs are simply not available to jobseekers, irrespective of their search effort, so UI does not influence unemployment much.

In this paper we propose a macroeconomic theory of optimal UI. The theory encompasses the microeconomic insurance-incentive tradeoff and the macroeconomic influence of UI on firms' hiring decisions. The theory provides a simple optimal UI formula, expressed with statistics that can be estimated empirically. Our theory provides guidance for the design of UI systems in which the generosity of UI is linked to the state of the labor market, such as in the US.[2]

To develop our theory, we embed the Baily-Chetty framework into a matching model with general production function, wage mechanism, and matching function.[3] The matching framework is well suited for our purpose because it captures both workers' job-search process and firms' hiring process. The labor market tightness, defined as the ratio of vacancies to aggregate job-search effort, is central to our theory. The tightness is determined in equilibrium to equalize labor demand and supply. UI generally affects the tightness because labor supply and demand respond to UI.

[1] A few papers study the optimal response of UI to shocks by simulating calibrated macroeconomic models, but by nature the results are somewhat specific to the calibration and structural assumptions about the wage mechanism or production function. See for instance Jung and Kuester [2014].

[2] US unemployment benefits have a duration of 26 weeks in normal times. The Extended Benefits program automatically extends duration by 13 weeks in states where the unemployment rate is above 6.5% and by 20 weeks in states where the unemployment rate is above 8%. Duration is often further extended by the government in severe recessions. For example, the Emergency Unemployment Compensation program enacted in 2008 extends durations by an additional 53 weeks in states where the unemployment rate is above 8.5%.

[3] In other contexts, matching models have been used to study UI. See for instance Cahuc and Lehmann [2000], Fredriksson and Holmlund [2001], and Lehmann and Van Der Linden [2007].

As in any matching model, the labor market equilibrium is not necessarily efficient. The labor market is efficient if tightness maximizes welfare keeping UI constant. The labor market is inefficient otherwise. The tightness is inefficiently high if reducing tightness with UI constant increases welfare; in that case unemployment is inefficiently low and firms devote too much labor to recruiting workers. The tightness is inefficiently low if raising tightness with UI constant increases welfare; in that case unemployment is inefficiently high and too few jobseekers find a job.

When the labor market is efficient, as one might expect, optimal UI satisfies the Baily-Chetty formula. The reason is that the marginal effect of UI on tightness has no first-order effect on welfare in this case. Hence, optimal UI is governed by the same principles as in the Baily-Chetty framework, in which tightness is fixed. Our theory uses the Baily-Chetty level as a baseline. This is convenient because the Baily-Chetty formula has been studied extensively so we have a good idea of the UI level implied by the formula [for example, Gruber, 1997].

But when the labor market is inefficient, optimal UI systematically departs from the Baily-Chetty level. This is our main result, and it has a simple intuition. Consider a labor market in which tightness is inefficiently low. If UI raises tightness, UI is desirable beyond the insurance-incentive trade-off, and optimal UI is higher than the Baily-Chetty level. Conversely, if UI lowers tightness, UI is not as desirable as what the insurance-incentive trade-off implies, and optimal UI is lower than the Baily-Chetty level. The same logic applies when tightness is inefficiently high.

Formally, we develop a formula that expresses the optimal UI replacement rate as the sum of the Baily-Chetty term plus a correction term. The correction term is equal to the effect of UI on tightness times the effect of tightness on welfare. The term is positive if UI brings tightness closer to efficiency, and negative otherwise. Hence, optimal UI is above the Baily-Chetty level if and only if increasing UI brings the labor market closer to efficiency.

In matching models, increasing UI could raise or lower tightness. UI can lower tightness through a *job-creation effect*: when UI rises, firms hire less because wages increase through bargaining, which reduces tightness. UI can also raise tightness through a *rat-race effect*: if the number of jobs available is somewhat limited, then by discouraging job search, UI alleviates the rat race for jobs and increases the job-finding rate and labor market tightness. When workers search less, they mechanically increase others' probability of finding one of the few jobs available. The overall effect of UI on tightness depends on which effect dominates. For example, in the standard model of Pissarides [2000] with wage bargaining and linear production function, only the job-creation effect operates and UI lowers tightness. But in the job-rationing model of Michaillat [2012] with rigid wage and concave production function, only the rat-race effect operates and UI raises tightness.

To improve the mapping of the theory with the data and make the theory more practical, we express the optimal UI formula in terms of estimable statistics, as in Chetty [2006*a*]. We develop a first empirical criterion to evaluate whether tightness is efficient, inefficiently low, or inefficiently high. A key statistic involved in this criterion is the share of labor devoted to recruiting. This share can be measured in data constructed by the Bureau of Labor Statistics (BLS). The criterion is that increasing tightness raises welfare if and only if the share of labor devoted to recruiting is low enough.

We develop a second empirical criterion to determine the effect of UI on tightness. The criterion involves a comparison of the microelasticity and macroelasticity of unemployment with respect to UI. The microelasticity measures the partial-equilibrium response of unemployment to UI, keeping tightness constant, whereas the macroelasticity measures its general-equilibrium response. The microelasticity accounts only for the response of job search to UI while the macroelasticity also accounts for the response of tightness. The criterion is that increasing UI raises tightness if and only if the microelasticity of unemployment with respect to UI is larger than its macroelasticity. This criterion is simple to understand. Imagine that UI raises tightness. Then UI increases the job-finding rate, which dampens the negative effect of UI on job search. Therefore, the macroelasticity is smaller than the microelasticity.

Combined with available empirical evidence, our theory has direct implications for the cyclicality of optimal UI. Using data from the Current Employment Statistics program and the National Employment Survey of the BLS, we construct a monthly time series for the share of labor devoted to recruiting. This share is subject to wide procyclical fluctuations in the US, suggesting that the tightness is inefficiently low in slumps and high in booms. Furthermore, recent empirical work suggests that the macroelasticity is smaller than the microelasticity, suggesting that UI raises tightness. This empirical evidence indicates that the correction term in our optimal UI formula is countercyclical; therefore, optimal UI is more countercyclical than the Baily-Chetty level of UI. Simulating the job-rationing model of Michaillat [2012], chosen because it generates both a macroelasticity smaller than the microelasticity and realistic fluctuations in the share of labor devoted to recruiting, we find that optimal replacement rate of UI is countercyclical, increasing by 20 percentage points when the unemployment rate rises from 5% to 12%.

2 A Matching Model of Unemployment Insurance

In this section we introduce the model on which we base our theory. The model embeds the framework of Baily [1978] and Chetty [2006*a*] into a matching model of the labor market.

The Labor Market. There is a measure 1 of identical workers and a measure 1 of identical firms. Initially, all workers are unemployed and search for a job with effort e. Each firm posts o vacancies to recruit workers. The number l of workers who find a job is given by a matching function taking as argument aggregate search effort and vacancies: $l = m(e,o)$. The function m has constant returns to scale, is differentiable and increasing in both arguments, and satisfies the restriction that $m(e,o) \leq 1$.

The labor market tightness is defined as the ratio of vacancies to aggregate search effort: $\theta = o/e$. Since the matching function has constant returns to scale, the labor market tightness determines the probabilities to find a job and fill a vacancy. A jobseeker finds a job at a rate $f(\theta) = m(e,o)/e = m(1,\theta)$ per unit of search effort; thus, a jobseeker searching with effort e finds a job with probability $e \cdot f(\theta)$. A vacancy is filled with probability $q(\theta) = m(e,o)/o = m(1/\theta,1) = f(\theta)/\theta$. The function f is increasing in θ and the function q is decreasing in θ. In other words, it is easier to find a job but harder to fill a vacancy when the labor market tightness is higher. We denote by $1-\eta$ and $-\eta$ the elasticities of f and q: $1-\eta = \theta \cdot f'(\theta)/f(\theta) > 0$ and $\eta = -\theta \cdot q'(\theta)/q(\theta) > 0$.

Firms. The representative firm hires l workers, paid a real wage w, to produce a consumption good. As in Michaillat and Saez [2013], we assume that the firm has two types of employees: some are engaged in production while others are engaged in recruiting. A number $n < l$ of workers are producing a quantity $y(n)$, where the production function y is differentiable, increasing, and concave. A number $l-n$ of workers are recruiting employees by posting vacancies. Posting a vacancy requires $r \in (0,1)$ recruiter, so the number of recruiters required to post o vacancies is $l-n=r\cdot o$. Since hiring l employees requires posting $l/q(\theta)$ vacancies, the number n of producers in a firm with l employees is limited to $n = l - r\cdot l/q(\theta)$. Hence, the numbers of employees and producers are related by

$$l = (1+\tau(\theta))\cdot n, \tag{1}$$

where $\tau(\theta) \equiv r/(q(\theta)-r)$ is the recruiter-producer ratio. The function τ is positive and increasing when $q(\theta) > r$, which holds in equilibrium. The elasticity of τ is $\theta\cdot\tau'(\theta)/\tau(\theta) = \eta\cdot(1+\tau(\theta))$.

The firm sells its output on a product market with perfect competition. Given θ and w, the firm chooses n to maximize profits $\pi = y(n) - (1+\tau(\theta))\cdot w\cdot n$. The optimal number of producers satisfies

$$y'(n) = (1+\tau(\theta))\cdot w. \tag{2}$$

At the optimum, the marginal revenue and marginal cost of hiring a producer are equal. The marginal revenue is the marginal product of labor, $y'(n)$. The marginal cost is the real wage, w, plus the marginal recruiting cost, $\tau(\theta)\cdot w$.

We implicitly define the labor demand $l^d(\theta,w)$ by

$$y'\left(\frac{l^d(\theta,w)}{1+\tau(\theta)}\right) = (1+\tau(\theta))\cdot w. \tag{3}$$

The labor demand gives the number of workers hired by firms when firms maximize profits.

The UI System. Search effort is not observable, so the receipt of UI cannot be contingent on search effort. Hence, UI provides all employed workers with consumption c^e and all unemployed workers with consumption $c^u < c^e$. We measure the generosity of UI in three different ways. UI is more generous if the consumption gain from work $\Delta c \equiv c^e - c^u$ decreases, the utility gain from work $\Delta v \equiv v(c^e) - v(c^u)$ decreases, or the replacement rate $R \equiv 1 - \Delta c/w$ increases. When a jobseeker finds work, she keeps a fraction $\Delta c/w = 1-R$ of the wage and gives up a fraction R as

UI benefits are lost. This is why we can interpret R as the replacement rate of the UI system.[4] The government must satisfy the budget constraint

$$y(n) = (1-l) \cdot c^u + l \cdot c^e. \tag{4}$$

If firms' profits π are equally distributed, the UI system can be implemented with a UI benefit b funded by a tax on wages t so that $(1-l) \cdot b = l \cdot t$ and $c^u = \pi + b$ and $c^e = \pi + w - t$. If profits are unequally distributed, a 100% tax on profits rebated lump sum implements the same allocation.

Workers. Workers cannot insure themselves against unemployment in any way, so they consume c^e if employed and c^u if unemployed. The utility from consumption is $v(c)$. The function v is differentiable, increasing, and concave. The disutility from job-search effort, e, is $k(e)$. The function k is differentiable, increasing, and convex. Given θ, c^e, and c^u, a representative worker chooses e to maximize expected utility

$$l \cdot v(c^e) + (1-l) \cdot v(c^u) - k(e) \tag{5}$$

subject to the matching constraint

$$l = e \cdot f(\theta), \tag{6}$$

where l is the probability to find a job and $1-l$ is the probability to remain unemployed. The optimal search effort satisfies

$$k'(e) = f(\theta) \cdot \Delta v. \tag{7}$$

At the optimum, the marginal utility cost and marginal utility gain of search are equal. The marginal utility cost is $k'(e)$. The marginal utility gain is the rate at which a unit of effort leads to a job, $f(\theta)$, times the utility gain from having a job, Δv. We implicitly define the effort supply

[4] Our definition of the replacement rate relates to the conventional definition as follows. Consider a UI system that provides a benefit b funded by a tax t so that $\Delta c = w - t - b$. Our replacement rate is defined as $R = (t+b)/w$. The conventional replacement rate is b/w; it ignores the tax t and is not the same as R. However, unemployment is small relative to employment so $t \ll b$ and $R \approx b/w$.

$e^s(f(\theta), \Delta v)$ as the solution of (7). The function e^s increases with $f(\theta)$ and Δv. Hence, search effort is higher when the labor market is tighter and when UI is less generous.

We define the labor supply by

$$l^s(\theta, \Delta v) = e^s(f(\theta), \Delta v) \cdot f(\theta). \tag{8}$$

The labor supply gives the number of workers who find a job when workers search optimally. The labor supply increases with θ and with Δv. Labor supply is higher when UI is less generous because search efforts are higher. The labor supply is higher when the labor market is tighter because the job-finding rate per unit of effort is higher and search efforts are higher.

Equilibrium. An equilibrium is a collection of variables $\{e, l, n, \theta, w, c^e, c^u, \Delta v\}$ such that workers maximize utility given tightness and UI, firms maximize profits given tightness and wage, and the government satisfies a budget constraint. These variables satisfy equations (1), (2), (4), (6),(7), and $\Delta v = v(c^e) - v(c^u)$.

Since there are eight variables but only six equations, two variables are indeterminate: one is the generosity of UI, Δv, and the other is the wage, w. In the rest of the paper, Δv is determined by the government to maximize welfare. As is well understood, the wage is indeterminate because of the matching frictions.[5] In Section 4, we assume that w follows a general wage schedule. In Section 5, we are more specific and study different wage mechanisms.

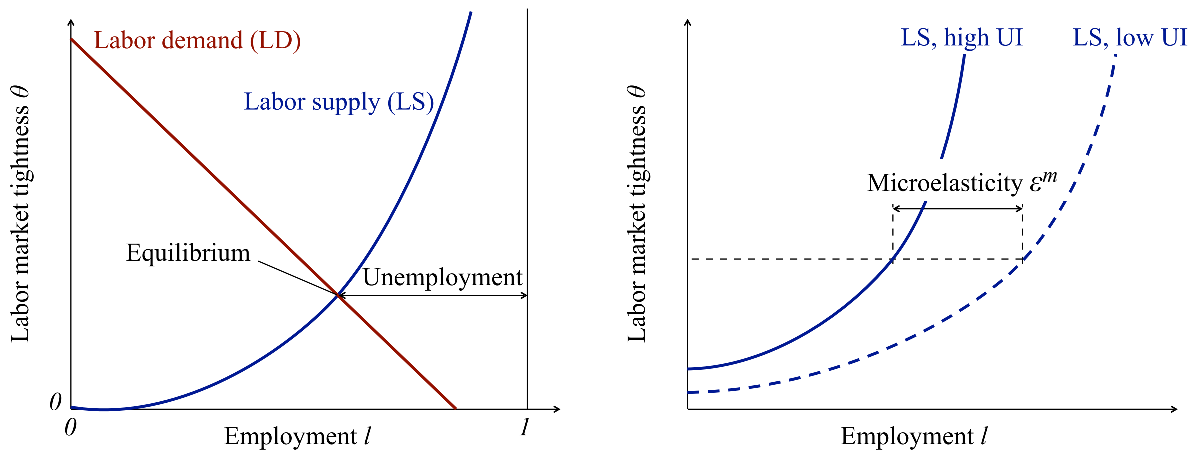
Given Δv and w, the key to solving the equilibrium is determining labor market tightness, θ. In equilibrium, θ equalizes labor supply and demand:[6]

$$l^s(\theta, \Delta v) = l^d(\theta, w). \tag{9}$$

Once θ is determined, l is determined from $l = l^s(\theta, \Delta v)$, e from $e = e^s(f(\theta), \Delta v)$, n from $n = l/(1+\tau(\theta))$, and c^e and c^u from the budget constraint (4) and $\Delta v = v(c^e) - v(c^u)$.

The equilibrium is represented in a (l, θ) plane in Figure 1(a). This diagram will be useful to understand the analysis.[7] The labor supply curve is upward sloping, and it shifts inward when UI

[5] See for instance the discussions in Howitt and McAfee [1987], Hall [2005*a*], and Michaillat and Saez [2013].
[6] Michaillat and Saez [2013] provide more details about the equilibrium concept.
[7] Michaillat [2014] and Michaillat and Saez [2014] show that the diagram is useful to study many other policies.



(a) Labor market equilibrium

(b) Microelasticity of unemployment

Figure 1: Equilibrium in the (l, θ) plane

increases. The labor demand curve may be horizontal or downward sloping, and it responds to UI when the wage responds to UI. The intersection of the labor supply and labor demand curve gives the equilibrium levels of labor market tightness, employment, and unemployment.

3 The Social Welfare Function

We express social welfare as a function of two arguments—UI and labor market tightness. We compute the derivatives of the social welfare function with respect to UI and tightness. These derivatives are key building blocks of the optimal UI formula derived in Section 4. Following Chetty [2006*a*], we express the derivatives in terms of statistics that can be estimated empirically. Section 6 will discuss estimates of these statistics.

We begin by defining the social welfare function. Consider an equilibrium parameterized by a utility gain from work, Δv, and a wage, w. For a given Δv, there is a direct relationship between w and the labor market tightness, θ. This relationship is given by (9). Hence, it is equivalent to parameterize the equilibrium by Δv and θ. Social welfare is thus a function of Δv and θ:

$$SW(\theta, \Delta v) = v(c^e(\theta, \Delta v)) - [1 - (e^s(\theta, \Delta v) \cdot f(\theta))] \cdot \Delta v - k(e^s(\theta, \Delta v)), \tag{10}$$

where $c^e(\theta, \Delta v)$ is the equilibrium level of consumption for employed workers. The consumption

$c^e(\theta,\Delta v)$ is implicitly defined by

$$y\left(\frac{l^s(\theta,\Delta v)}{1+\tau(\theta)}\right) = l^s(\theta,\Delta v)\cdot c^e(\theta,\Delta v) + (1-l^s(\theta,\Delta v))\cdot v^{-1}\left(v(c^e(\theta,\Delta v))-\Delta v\right). \tag{11}$$

This equation ensures that the government's budget constraint is satisfied when all the variables take their equilibrium values. Since $v(c^u) = v(c^e) - \Delta v$, the term $v^{-1}\left(v(c^e(\theta,\Delta v))-\Delta v\right)$ gives the equilibrium level of consumption for unemployed workers.

Next, we define two elasticities that measure how search effort responds to UI and labor market conditions. We will use them to analyze the social welfare function.

DEFINITION 1. *The* microelasticity of unemployment with respect to UI *is*

$$\varepsilon^m = \frac{\Delta v}{1-l}\cdot\left.\frac{\partial l^s}{\partial \Delta v}\right|_\theta.$$

The microelasticity measures the percentage increase in unemployment when the utility gain from work decreases by 1%, taking into account jobseekers' reduction in search effort but ignoring the equilibrium adjustment of labor market tightness. The microelasticity can be estimated by measuring the reduction in the job-finding probability of an unemployed worker whose unemployment benefits are increased, keeping the benefits of all other workers constant. In Figure 1(b), a change in UI leads to a change in search effort, which shifts the labor supply curve. The microelasticity measures this shift.

The empirical literature does not typically estimate ε^m. Instead, this literature estimates the microelasticity ε^m_R of unemployment with respect to the replacement rate, R. This is not an issue, however, because the two elasticities are closely related. Usually ε^m_R is estimated by changing benefits c^u while keeping c^e constant. As $\Delta c = (1-R)\cdot w$, $\Delta v = v(c^e) - v(c^e-(1-R)\cdot w))$ and $\left.\partial l^s/\partial R\right|_{\theta,c_e} = -w\cdot v'(c^u)\cdot\left(\left.\partial l^s/\partial \Delta v\right|_\theta\right)$. The empirical elasticity ε^m_R is thus related to ε^m by

$$\varepsilon^m_R \equiv \frac{R}{1-l}\cdot\left.\frac{\partial(1-l^s)}{\partial R}\right|_{\theta,c_e} = \frac{R\cdot w\cdot v'(c^u)}{\Delta v}\cdot\varepsilon^m. \tag{12}$$

DEFINITION 2. *The* discouraged-worker elasticity *is*

$$\varepsilon^f = \frac{f(\theta)}{e} \cdot \left.\frac{\partial e^s}{\partial f}\right|_{\Delta v}.$$

The discouraged-worker elasticity measures the percentage increase in search effort when the job-finding rate per unit of effort increases by 1%, keeping UI constant. In our model, workers search less when the job-finding rate decreases and $\varepsilon^f > 0$; hence, ε^f captures jobseekers' discouragement when labor market conditions deteriorate. The discouraged-worker elasticity determines the elasticity of labor supply with respect to labor market tightness:

LEMMA 1. *The elasticity of labor supply with respect to tightness is related to the discouraged-worker elasticity by*

$$\frac{\theta}{l} \cdot \left.\frac{\partial l^s}{\partial \theta}\right|_{\Delta v} = (1-\eta) \cdot (1+\varepsilon^f)$$

Proof. Obvious because $l^s(\theta, \Delta v) = e^s(f(\theta), \Delta v) \cdot f(\theta)$, ε^f is the elasticity of e^s with respect to f, and $1-\eta$ is the elasticity of f with respect to θ. □

Equipped with these elasticities, we can differentiate the social welfare function:

LEMMA 2. *The social welfare function admits the following derivatives:*

$$\left.\frac{\partial SW}{\partial \theta}\right|_{\Delta v} = \frac{l}{\theta} \cdot (1-\eta) \cdot \phi \cdot w \cdot \left[\frac{\Delta v}{\phi \cdot w} + R \cdot \left(1+\varepsilon^f\right) - \frac{\eta}{1-\eta} \cdot \tau(\theta)\right] \tag{13}$$

$$\left.\frac{\partial SW}{\partial \Delta v}\right|_{\theta} = (1-l) \cdot \frac{\phi \cdot w}{\Delta v} \cdot \varepsilon^m \cdot \left[R - \frac{l}{\varepsilon^m} \cdot \frac{\Delta v}{w} \cdot \left(\frac{1}{v'(c^e)} - \frac{1}{v'(c^u)}\right)\right], \tag{14}$$

where ϕ is an harmonic mean of workers' marginal utilities:

$$\frac{1}{\phi} = \frac{l}{v'(c^e)} + \frac{1-l}{v'(c^u)}. \tag{15}$$

Proof. We first derive (13). Since workers choose effort to maximize expected utility, a standard application of the envelope theorem says changes in effort, $e^s(\theta, \Delta v)$, resulting from changes in θ

have no impact on social welfare. The effect of θ on welfare therefore is

$$\frac{\partial SW}{\partial \theta} = \frac{l}{\theta} \cdot (1-\eta) \cdot \Delta v + v'(c^e) \cdot \frac{\partial c^e}{\partial \theta}. \tag{16}$$

The first term is the welfare gain from increasing employment by increasing tightness. It is obtained by noting that the elasticity of $f(\theta)$ is $1-\eta$ so $e \cdot f'(\theta) = (l/\theta) \cdot (1-\eta)$. This term accounts only for the change in employment resulting from a change in job-finding rate, and not for the change resulting from a change in effort. The second term is the welfare change arising from the consumption change following a change in θ.

To compute the consumption change $\partial c^e / \partial \theta$, we implicitly differentiate $c^e(\theta, \Delta v)$ with respect to θ in (11). A few preliminary results are helpful. First, (2) implies that $y'(n)/(1+\tau(\theta)) = w$. Second, Lemma 1 implies that $\partial l^s / \partial \theta = (l/\theta) \cdot (1-\eta) \cdot (1+\varepsilon^f)$. Third, $v^{-1}(v(c^e(\theta, \Delta v)) - \Delta v) = c^u$ so $c^e - v^{-1}(v(c^e(\theta, \Delta v)) - \Delta v) = \Delta c$. Fourth, the elasticity of $1+\tau(\theta)$ is $\eta \cdot \tau(\theta)$ so the derivative of $1/(1+\tau(\theta))$ with respect to θ is $-\eta \cdot \tau(\theta)/[\theta \cdot (1+\tau(\theta))]$. Fifth, the derivative of $v^{-1}(v(c^e) - \Delta v)$ with respect to c^e is $v'(c^e)/v'(c^u)$. The implicit differentiation therefore yields

$$\frac{l}{\theta} \cdot (1-\eta) \cdot (1+\varepsilon^f) \cdot (w - \Delta c) - \frac{l}{\theta} \cdot \eta \cdot \tau(\theta) \cdot w = \left(\frac{l}{v'(c^e)} + \frac{1-l}{v'(c^u)} \cdot \right) \cdot v'(c^e) \cdot \frac{\partial c^e}{\partial \theta}.$$

The first term on the left-hand side is the budgetary gain from the new jobs created. Each new job increases government revenue by $w - \Delta c$. The increase in employment results both from a higher job-finding rate and from a higher search effort. The term $(1+\varepsilon^f)$ captures the combination of the two effects. The second term is the loss in resources due to a higher tightness. A higher tightness forces firms to devote more labor to recruiting and less to producing. Introducing the variable ϕ defined by (15) and plugging the resulting expression for $\partial c^e / \partial \theta$ into (16) yields (13).

Next, we derive (14). Following the same logic as above, the effect of Δv on welfare is

$$\frac{\partial SW}{\partial \Delta v} = -(1-l) + v'(c^e) \cdot \frac{\partial c^e}{\partial \Delta v}. \tag{17}$$

The first term is the welfare loss suffered by unemployed workers after a reduction in UI benefits. The second term is the welfare change arising from the consumption change following a change in Δv. Next, we implicitly differentiate $c^e(\theta, \Delta v)$ with respect to Δv in (11). We need two preliminary results in addition to those above. First, the definition of the microelasticity implies that $\partial l^s/\partial\Delta v = [(1-l)/\Delta v]\cdot\varepsilon^m$. Second the derivative of $v^{-1}(v(c^e)-\Delta v)$ with respect to Δv is $-1/v'(c^u)$. The implicit differentiation yields

$$\frac{1-l}{\Delta v}\cdot\varepsilon^m\cdot(w-\Delta c)+\frac{1-l}{v'(c^u)}=\left(\frac{l}{v'(c^e)}+\frac{1-l}{v'(c^u)}\cdot\right)\cdot v'(c^e)\cdot\frac{\partial c^e}{\partial\Delta v}.$$

The first terms captures the budgetary gain from increasing employment by reducing the generosity of UI. This is a behavioral effect, coming from the response of job search to UI. The second term captures the budgetary gain from reducing the UI benefits paid to unemployed workers. This is a mechanical effect. As above, we plug the expression for $\partial c^e/\partial\Delta v$ into (17) and obtain

$$\frac{\partial SW}{\partial\Delta v}=(1-l)\cdot\phi\cdot\left(R\cdot\varepsilon^m\cdot\frac{w}{\Delta v}+\frac{1}{v'(c^u)}-\frac{1}{\phi}\right).$$

The definition of ϕ implies that $(1/v'(c^u))-(1/\phi)=-l\cdot[(1/v'(c^e))-(1/v'(c^u))]$. Combining this expression with the last equation yields (14). $\square$

4 The Optimal Unemployment Insurance Formula

In this section we derive the optimal UI formula, expressed in terms of estimable statistics.

It is well understood that in matching models with unemployment insurance, the equilibrium is usually inefficient in the sense that a marginal increase in labor market tightness keeping the utility gain from work constant has an effect on welfare. The equilibrium is generally inefficient in our model because, in any worker-firm pair, the wage is determined in a situation of bilateral monopoly.[8] Since the solution to the bilateral monopoly problem is indeterminate, there is no reason that the equilibrium wage ensures social efficiency.[9]

If increasing tightness while keeping the utility gain from work constant has no first-order

[8] See the discussions in Howitt and McAfee [1987] and Michaillat and Saez [2013].

[9] Our matching model assumes random search, in the sense that workers apply randomly to firms, irrespective of the wage that they may offer. With random search, there is clearly no reason for the equilibrium to be efficient [Pissarides, 2000]. In matching models with directed search, in the sense that workers apply to submarkets based on the wage-tightness compromise that they offer, the competitive search mechanism of Moen [1997] usually ensures efficiency. But in the presence of a UI system, even the competitive search mechanism cannot ensure efficiency because agents fail to internalize the effects of their actions on the government budget constraint [Lehmann and Van Der Linden, 2007]. Hence, inefficiency is generic in matching models with unemployment insurance.

effect on welfare, the tightness is efficient. But if the increase enhances welfare, the tightness is inefficiently low, and if the increase reduces welfare, the tightness is inefficiently high. The equilibrium wage is inefficiently high when the equilibrium tightness is inefficiently low, and it is inefficiently low when the equilibrium tightness is inefficiently high. We will show that the optimal level of UI depends on the effect of tightness on welfare. The following proposition is therefore important because it provides a simple condition to assess the effect of tightness on welfare.

DEFINITION 3. *The* efficiency term *is*

$$\frac{\Delta v}{\phi \cdot w} + R \cdot \left(1 + \varepsilon^f\right) - \frac{\eta}{1-\eta} \cdot \tau(\theta),$$

where ϕ is given by (15).

PROPOSITION 1. *Consider a marginal increase in labor market tightness, keeping the utility gain from work constant. The efficiency term is positive iff the increase raises welfare, zero iff the increase has no first-order effect on welfare, and negative iff the increase lowers welfare.*

Proof. The result directly follows from (13). □

This proposition is closely related to efficiency condition of Hosios [1990]. One difference is that the Hosios condition is a condition on the wage mechanism—it relates workers' bargaining power to the elasticity of the matching function—whereas our efficiency condition is a condition on estimable statistics. Another difference is that the Hosios condition applies to models with risk-neutral workers whereas our condition applies to models with risk-averse workers.

So far, we have parameterized each equilibrium by a pair $(\theta, \Delta v)$. We have used this preliminary analysis to characterize the social welfare function and introduce the efficiency term. However, to study optimal UI, we need to account for the equilibrium response of θ to Δv by specifying a wage mechanism. From now on, we assume that the wage is determined by a general wage schedule $w = w(\theta, \Delta v)$. This is the most general specification because all other endogenous variables are a function of $(\theta, \Delta v)$. With this wage mechanism, the equilibrium condition (9) implicitly defined equilibrium tightness θ as a function of Δv.

The response of the wage to UI has important implications for optimal UI because it determines how labor demand responds to UI. However, we do not need an explicit expression for the wage schedule. The only information needed is the response of employment to UI, measured by the following elasticity:

DEFINITION 4. *The* macroelasticity of unemployment with respect to UI *is*

$$\varepsilon^M = \frac{\Delta v}{1-l} \cdot \frac{dl}{d\Delta v}.$$

The macroelasticity measures the percentage increase in unemployment when the utility gain from work decreases by 1%, taking into account jobseekers' reduction in search effort and the equilibrium adjustment of labor market tightness. The macroelasticity can be estimated by measuring the increase in aggregate unemployment following a general increase in unemployment benefits. Of course, the macroelasticity is endogenous: it may respond to the labor market tightness or the generosity of UI.

We will show that the optimal level of UI depends on the effect of UI on tightness. The following proposition is therefore important because it shows that the effect of UI on tightness is measured by the wedge between the microelasticity and the macroelasticity of unemployment with respect to UI.

DEFINITION 5. *The* elasticity wedge *is* $1 - \varepsilon^M/\varepsilon^m$.

PROPOSITION 2. *The elasticity wedge measures the response of labor market tightness to UI:*

$$\frac{\Delta v}{\theta} \cdot \frac{d\theta}{d\Delta v} = -\frac{1-l}{l} \cdot \frac{1}{1-\eta} \cdot \frac{\varepsilon^m}{1+\varepsilon^f} \cdot \left(1 - \frac{\varepsilon^M}{\varepsilon^m}\right). \tag{18}$$

The elasticity wedge is positive if tightness increases with the generosity of UI, negative if tightness decreases with the generosity of UI, and zero if tightness does not depend on UI.

Proof. Since $l = l^s(\theta, \Delta v)$, we have:

$$\frac{\Delta v}{1-l} \cdot \frac{dl}{d\Delta v} = \left(\frac{\theta}{1-l} \cdot \frac{\partial l^s}{\partial \theta}\right) \cdot \left(\frac{\Delta v}{\theta} \cdot \frac{d\theta}{d\Delta v}\right) + \left(\frac{\Delta v}{1-l} \cdot \frac{\partial l^s}{\partial \Delta v}\right).$$

Using $(\theta/l)(\partial l^s/\partial \theta) = (1-\eta)(1+\varepsilon^f)$ from Lemma 1, we obtain

$$\varepsilon^M = \frac{l}{1-l} \cdot (1-\eta) \cdot \left(1+\varepsilon^f\right) \cdot \frac{\Delta v}{\theta} \cdot \frac{d\theta}{d\Delta v} + \varepsilon^m. \tag{19}$$

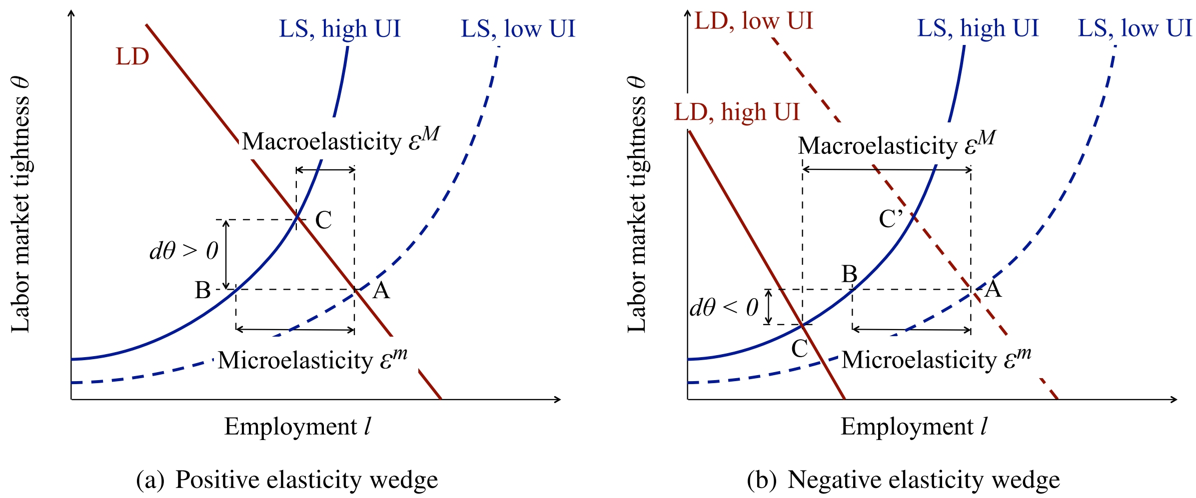


Figure 2: The effect of UI on tightness determines the sign of the elasticity wedge, $1 - \varepsilon^M / \varepsilon^m$

Notes: This figure illustrates the results of Proposition 2. Panel (a) considers a downward-sloping labor demand that does not respond to UI. Panel (b) considers a downward-sloping labor demand that shifts inward when UI increases.

Dividing this equation by ε^m and rearranging the terms yields the desired result. □

The proposition shows that a wedge appears between the microelasticity and macroelasticity when UI affects tightness, and that this elasticity wedge has the same sign as the effect of UI on tightness.

Figure 2 illustrates this result. In Figures 2(a) and 2(b), the horizontal distance A–B measures the microelasticity and the horizontal distance A–C measures the macroelasticity. In Figure 2(a), the labor demand curve is downward sloping, and it does not shift with a change in UI. After a reduction in UI, the labor supply curve shifts outward (A–B) and tightness increases along the new labor supply curve (B–C). Since tightness rises after the increase in UI, the macroelasticity is smaller than the microelasticity. In Figure 2(b), the labor demand also shifts inward with an increase in UI. Tightness falls along the new supply curve after the labor demand shift (C'–C). In equilibrium, tightness can rise or fall depending on the size of the labor demand shift. In Figure 2(b) tightness falls, so the macroelasticity is larger than the microelasticity. In Section 5, we will describe the mechanisms through which UI is able to affect tightness by looking at specific matching models.

Having analyzed the effects of tightness on welfare and those of UI on tightness, we are equipped to derive the optimal UI formula. The problem of the government is to choose the utility gain from work, Δv, to maximize social welfare, given by (10), subject to the equilibrium response of tightness, given by (9). The following proposition characterizes optimal UI:

PROPOSITION 3. *The optimal UI policy satisfies the formula*

$$R = \frac{l}{\varepsilon^m} \cdot \frac{\Delta v}{w} \cdot \left[\frac{1}{v'(c^e)} - \frac{1}{v'(c^u)}\right] + \frac{1}{1+\varepsilon^f}\left(1 - \frac{\varepsilon^M}{\varepsilon^m}\right)\left[\frac{\Delta v}{w \cdot \phi} + \left(1+\varepsilon^f\right) R - \frac{\eta}{1-\eta} \cdot \tau(\theta)\right], \quad (20)$$

where ϕ satisfies (15)*. The first term in the right-hand side is the* Baily-Chetty term*, and the second term is the* correction term.

Proof. We define welfare $SW(\theta, \Delta v)$ by (10). The derivative of the social welfare with respect to Δv is $dSW/d\Delta v = \partial SW/\partial \Delta v + (\partial SW/\partial \theta) \cdot (d\theta/d\Delta v)$. Therefore, the first-order condition $dSW/d\Delta v = 0$ in the current problem is a linear combination of the first-order conditions $\partial SW/\partial \theta = 0$ and $\partial SW/\partial \Delta v = 0$. Hence, the optimal UI formula is a linear combination of the efficiency condition and the Baily-Chetty formula. Moreover, the efficiency condition is multiplied by the wedge $1 - \varepsilon^M/\varepsilon^m$ because the factor $d\theta/d\Delta v$ is proportional to that wedge.

Equation (18) shows that the labor market tightness variation is given by

$$\frac{d\theta}{d\Delta v} = \frac{1-l}{l} \cdot \frac{1}{1-\eta} \cdot \frac{1}{1+\varepsilon^f} \cdot \frac{\theta}{\Delta v} \cdot \left(\varepsilon^M - \varepsilon^m\right).$$

We combine this equation with the derivatives provided by Lemma 2 to write the first-order condition $dSW/d\Delta v = 0$. Dividing the resulting equation by $(1-l) \cdot \phi \cdot w \cdot \varepsilon^m \cdot /\Delta v$ yields (20). □

COROLLARY 1. *If the labor market equilibrium is efficient, optimal UI satisfies the Baily-Chetty formula:*

$$R = \frac{l}{\varepsilon^m} \cdot \frac{\Delta v}{w} \cdot \left(\frac{1}{v'(c^e)} - \frac{1}{v'(c^u)}\right). \quad (21)$$

Proof. Equation (21) obtains from Propositions 1 and 3. It may not be immediately apparent that (21) is equivalent to the traditional Baily-Chetty formula. The equivalence becomes clear using (12), which allows us to rewrite formula (21) as

$$\varepsilon^m_R = l \cdot \left(\frac{v(c^u)}{v'(c^e)} - 1\right).$$

Table 1: Optimal replacement rate compared to Baily-Chetty replacement rate

		A. Comparison based on theoretical derivatives	
	$d\theta/d\Delta v > 0$	$d\theta/d\Delta v = 0$	$d\theta/d\Delta v < 0$
$\partial SW/\partial\theta\big\|_{\Delta v} > 0$	lower	same	higher
$\partial SW/\partial\theta\big\|_{\Delta v} = 0$	same	same	same
$\partial SW/\partial\theta\big\|_{\Delta v} < 0$	higher	same	lower
		B. Comparison based on estimable statistics	
	Elasticity wedge < 0	Elasticity wedge $= 0$	Elasticity wedge > 0
Efficiency term > 0	lower	same	higher
Efficiency term $= 0$	same	same	same
Efficiency term < 0	higher	same	lower

Notes: The replacement rate is $R = 1 - (c^e - c^u)/w$. The elasticity wedge is $1 - \varepsilon^M/\varepsilon^m$. The efficiency term is $(\Delta v + k(e))/(\phi \cdot w) + (1+\varepsilon^f) \cdot R - [\eta/(1-\eta)] \cdot \tau(\theta)$. The Baily-Chetty replacement rate is given by (21). The optimal replacement rate is higher than the Baily-Chetty rate if the correction term in (20) is positive, same as the Baily-Chetty rate if the correction term is zero, and lower than the Baily-Chetty rate if the correction term is negative.

This is the standard expression for the Baily-Chetty formula. □

Formula (20) shows that the optimal UI replacement rate, R, is the sum of the Baily-Chetty term and a correction term. The Baily-Chetty term captures the trade-off between the need for insurance, measured by $(1/v'(c^e)) - (1/v'(c^u))$, and the need for incentives to search, measured by ε^m, exactly as in the analysis of Baily [1978] and Chetty [2006*a*]. The correction term is the product of the effect of UI on tightness, measured by the elasticity wedge, and the effect of tightness on welfare, measured by the efficiency term. The correction term is positive if and only if increasing UI brings the labor market equilibrium toward efficiency.[10]

There are two situations where the correction term is zero and hence the optimal replacement rate is given by the Baily-Chetty formula. The first situation is when the labor market tightness is efficient such that the efficiency term is zero. This is the situation described by the corollary. When the labor market tightness is efficient, the marginal effect of UI on tightness has no first-order effect on welfare; hence, optimal UI is governed by the same principles as in the Baily-Chetty framework

[10] The response of tightness to UI can be interpreted as a pecuniary externality. The reason is that tightness can be interpreted as a price influenced by the search behavior of workers and influencing welfare when the labor market is inefficient. Under this interpretation, the additive structure of the formula—a standard term plus a correction term—is similar to the structure of many optimal taxation formulas obtained in the presence of externalities.

in which tightness is fixed. The second situation is when UI has no effect on tightness such that the elasticity wedge is zero. When tightness is fixed, our model is isomorphic to the Baily-Chetty framework, so optimal UI is guided by the same principles.

In all other situations, the correction term is nonzero and the optimal replacement rate departs from the Baily-Chetty rate. The main implication of our formula is that increasing UI above the Baily-Chetty rate is desirable if and only if increasing UI brings the labor market tightness closer to efficiency. UI brings the tightness closer to efficiency either if tightness is inefficiently low and UI raises tightness or if tightness is inefficiently high and UI lowers tightness. In terms of the estimable statistics, UI brings the tightness closer to efficiency if the efficiency term and elasticity wedge are both positive or if they are both negative. Table 1 summarizes all the possible situations.

As is standard in optimal tax formulas, the right-hand-side of (20) is endogenous to UI. Even though the formula characterizes optimal UI only implicitly, it is useful. First, it transparently shows the economic forces at play. Second, it gives general conditions for optimal UI to be above or below the Baily-Chetty level. These conditions apply to a broad range of matching models. Third, the right-hand-side term is expressed with statistics that are estimable. Hence, the formula could be combined with empirical estimates to assess a UI system. This assessment would be valid even if the right-hand-side term is endogenous to UI.

An important implication of our formula is that even in the presence of private provision of UI, the public provision of UI is justified. Indeed, small private insurers do not internalize tightness externalities and offer insurance according to the Baily-Chetty term. It is therefore optimal for the government to correct privately provided UI by a quantity equal to the correction term. The correction term is positive or negative depending on the state of the labor market and the sign of the elasticity wedge.

Our formula reveals some counterintuitive properties of optimal UI. First, even if UI has no adverse on unemployment and $\varepsilon^M = 0$, full insurance is not desirable. Consider a model in which the number of jobs is fixed. Increasing UI redistributes from employed workers to unemployed workers without destroying jobs, but, unlike what intuition suggests, the optimal replacement is strictly below 1. This can be seen by plugging $\varepsilon^M = 0$ and $\varepsilon^m > 0$ in (20). The reason is that increasing UI increases tightness and forces firms to allocate more workers to recruiting instead of producing, thus reducing output available to consumption. In fact, if the efficiency condition holds, UI is given by the standard Baily-Chetty formula and the magnitude of ε^M is irrelevant.

Second, even in the absence of any concern for insurance, some UI should be offered if UI brings the economy closer to efficiency.[11] Consider a model with risk-neutral workers and exogenous search effort ($\varepsilon^m = \varepsilon^f = 0$). Formula (20) boils down to $\tau(\theta) = (1-\eta)/\eta$, which is the condition on tightness to maximize output and hence restore efficiency.[12]

Third, even though wages may respond to UI, the response of wages does not appear directly in the formula. It does not matter whether wages change or not because wages do not enter in the government's budget constraint or workers' search decisions. The wage does appear in firms' decisions, but this effect is measured by the macroelasticity. In other words, the elasticity wedge is the sufficient statistics that captures the effects of UI on wages.

Finally, Formula (20) is quite robust: it would remain valid in a number of extensions. In the analysis, we assume that the utility of employed workers differs from that of unemployed worker only because employed workers consume more. In reality, they are many other differences between employed and unemployed workers that matter for their utility. For instance, it is well documented that the state of unemployment has high psychological and health costs [Clark and Oswald, 1994; Hawton and Platt, 2000; Sullivan and von Wachter, 2009]. To capture all the differences between employment and unemployment beyond consumption, we could assume that employed workers have utility $v_e(c^e)$ and unemployed workers have utility $v_u(c^u)$, where the functions v_e and v_u differ. In that case, formula (20) would carry over after adjusting the utility gain from work to $v_e(c^e) - v_u(c^u)$ and the marginal utilities to $v_u'(c^u)$ and $v_e'(c^e)$. In the analysis, we also assume that workers cannot insure themselves against unemployment. In reality, unemployed workers partially insure themselves [Gruber, 1997]. We could assume that workers self-insure partially against unemployment with home production.[13] In that case, formula (20) would carry over after adjusting the utility gain from work and the marginal utility of unemployed workers to account for home production.

[11] This result was noted by Rogerson, Shimer and Wright [2005].

[12] To see this, multiply formula (20) by $\varepsilon^m \cdot (1-R)$. With $\varepsilon^m = \varepsilon^f = 0$ and $\Delta v = \Delta c$ due to risk neutrality, we have $\phi = 1$ and $\Delta v/w = 1-R$. Therefore, $-\varepsilon^M \cdot [(1-R) + R - \eta \cdot \tau(\theta)/(1-\eta)] = 0$. If UI influences tightness, $\varepsilon^M > 0$ and we obtain $\tau(\theta) = (1-\eta)/\eta$.

[13] Home production is a reduced-form representation of all the means of self-insurance available to workers. In practice, workers self insure not only with home production but also with savings or spousal income. Analyzing savings or spousal income would be more complex.

Table 2: Effect of UI on labor market tightness in matching models

	Model			
	Standard	Rigid-wage	Job-rationing	Aggregate-demand
		A. Assumptions		
Production function	linear	linear	concave	linear
Wage mechanism	bargaining	rigid	rigid	rigid
		B. Properties		
Type of UI effect	job-creation	no effect	rat-race	rat-race
Effect of UI on tightness	$-$	0	$+$	$+$
Elasticity wedge	$-$	0	$+$	$+$

Notes: This table summarizes the results of Propositions 4, 5, 6, and 9. The job-creation and rat-race effects are depicted in Figure 3. The elasticity wedge is $1-\varepsilon^M/\varepsilon^m$.

5 The Economics of the Elasticity Wedge

Section 4 showed that the optimal level of UI depends critically on the effect of UI on tightness, measured by the elasticity wedge. But Section 4 remained vague on the economic mechanisms through which UI affects tightness. In this section, we consider four matching models that differ by their wage mechanism and production function to describe possible mechanisms. Table 2 presents the models and summarizes their properties. In these models UI has two possible effects on tightness: a job-creation effect that lowers tightness, and a rat-race effect that raises tightness.

5.1 Elasticity Wedge in a Standard Model

The standard model shares the main features of the model developed by Pissarides [1985, 2000] and Shimer [2005]. The production function is linear: $y(n) = n$. When they are matched, worker and firm bargain over the wage. The outcome of this bargaining is that the match surplus is shared, with the worker keeping a fraction $\beta \in (0,1)$ of the surplus.[14] The parameter β is the worker's bargaining power.[15]

We need to derive the labor demand to analyze the model. We begin by determining the bar-

[14] In a seminal paper, Diamond [1982] also assumed a surplus-sharing solution to the bargaining problem. If workers and firms are risk neutral, the surplus-sharing solution coincides with the generalized Nash solution. Under risk aversion, these two solutions generally differ. We use the surplus-sharing solution for its simplicity.

[15] To obtain a positive wage, we impose that $\beta/(1-\beta) > \Delta v$.

gained wage. The worker's surplus from a match with a firm is $\mathscr{W} = \Delta v$. The firm's surplus from a match with a worker is $\mathscr{F} = 1 - w$ because once a worker is recruited, she produces 1 unit of good and receives a real wage w. Since worker and firm split the total surplus from the match, $\mathscr{W}/\beta = \mathscr{F}/(1-\beta)$. Hence, the bargained wage satisfies

$$w = 1 - \frac{1-\beta}{\beta} \cdot \Delta v.$$

Increasing UI raises the bargained wage. The reason is that the outside option of jobseekers increases after an increase in UI, so they are able to obtain a higher wage.

We combine the wage equation with (3) and $y'(n) = 1$ to obtain the labor demand:

$$\frac{\tau(\theta)}{1+\tau(\theta)} = \frac{1-\beta}{\beta} \cdot \Delta v. \tag{22}$$

This equation defines a perfectly elastic labor demand curve in a (l, θ) plane, as depicted in Figure 3(a). The labor demand shifts downward when UI increases. The reason is that when UI increases, wages rise so it becomes less profitable for firms to hire workers.

Having obtained the labor demand, we can describe the effect of UI on tightness:

PROPOSITION 4. *Increasing UI lowers tightness: $d\theta/d\Delta v > 0$. The elasticity wedge is*

$$1 - \frac{\varepsilon^M}{\varepsilon^m} = -\frac{l}{1-l} \cdot \frac{1-\eta}{\eta} \cdot \frac{1+\varepsilon^f}{\varepsilon^m} < 0.$$

Proof. We differentiate (22) with respect to Δv. Since the elasticities of $\tau(\theta)$ and $1+\tau(\theta)$ with respect to θ are $\eta \cdot (1+\tau(\theta))$ and $\eta \cdot \tau(\theta)$, we obtain $(\Delta v/\theta) \cdot (d\theta/d\Delta v) = 1/\eta$. It follows that $d\theta/d\Delta v > 0$. Using (18) then immediately yields the expression for the wedge. □

Figure 3(a) illustrates the results of the proposition. After an augmentation in UI, jobseekers search less, shifting the labor supply curve inward by a distance A–B, and bargained wages increase, shifting the labor demand downward and further reducing employment by the horizontal distance B–C. The total reduction in employment is given by the horizontal distance A–C. Since the labor demand is horizontal and shifts downward, tightness necessarily falls. Since A–C is larger than A–B, the macroelasticity is larger than the microelasticity.

The standard model nicely captures two effects of UI: the moral-hazard effect and the job creation effect. The moral-hazard effect is the reduction in employment caused by the reduction in search effort, which is not observable and thus a source of moral hazard. The distance A–B measures this effect. The job-creation effect is the reduction in employment caused by the reduction in hiring following the increase in wages. The distance B–C measures this effect. The job-creation effect is the reason why tightness falls when UI increases and why the macroelasticity is larger than the microelasticity.

The proposition implies that optimal UI is higher than the Baily-Chetty level when tightness is inefficiently high and lower when tightness is inefficiently low. For example, when tightness is inefficiently low, reducing UI lowers wages and increases tightness, which improves welfare beyond the insurance-incentive trade-off.

5.2 Elasticity Wedge in a Rigid-Wage Model

The rigid-wage model shares the main features of the model developed by Hall [2005*a*]. The production function is linear: $y(n) = n$. The wage is fixed: $w = \omega$, where $\omega \in (0,1)$. We combine the wage schedule with equation (3) to obtain the labor demand:

$$1 = \omega \cdot (1 + \tau(\theta)). \tag{23}$$

This equation defines a perfectly elastic labor demand in a (l, θ) plane, as depicted in Figure 3(b). The labor demand is unaffected by UI because the wage does not respond to UI.

Having obtained the labor demand, we can describe the effect of UI on tightness:

PROPOSITION 5. *Increasing UI has no effect on tightness. The elasticity wedge is 0.*

Proof. Equilibrium tightness is determined by (23). This equation is independent of Δv. □

Figure 3(b) illustrates the result. Since the labor demand is horizontal and independent of UI, UI has no effect on tightness. The only effect at play is the moral-hazard effect, as in the Baily-Chetty framework. The rigidity of wages with respect to UI eliminates the job-creation effect that was present in the standard model. The proposition implies that optimal UI is always given by the

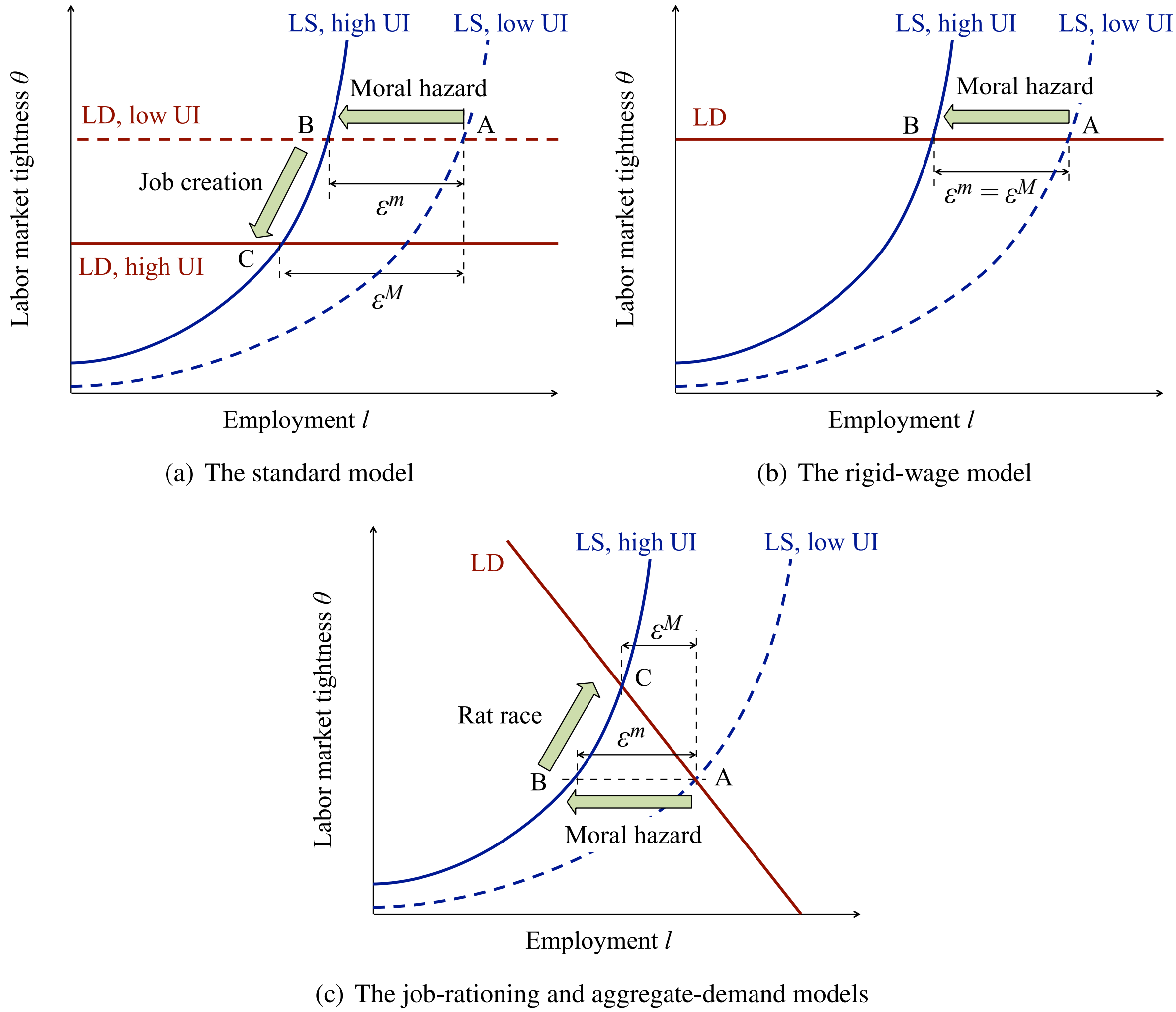


Figure 3: Effect of UI on labor market tightness and employment in matching models

Baily-Chetty formula even if the efficiency condition does not hold. Tightness may be inefficient but this inefficiency does not affect optimal UI because UI has no effect on tightness.

5.3 Elasticity Wedge in a Job-Rationing Model

The job-rationing model shares the main features of the model developed by Michaillat [2012]. The production function is concave: $y(n) = a \cdot n^{\alpha}$, where a is technology and $\alpha \in (0,1)$ parameterizes diminishing marginal returns to labor. The wage is independent of UI and partially rigid with respect to technology: $w = \omega \cdot a^{\gamma}$, where $\gamma \in [0,1)$ parameterizes the rigidity with respect to technology. If $\gamma = 0$, the wage is fixed: it does not respond to technology. If $\gamma = 1$, the wage is fully flexible: it is proportional to technology.

We combine the wage schedule with equation (3) to obtain the labor demand:

$$l^d(\theta, a) = \left(\frac{\omega}{\alpha} \cdot a^{\gamma-1}\right)^{-\frac{1}{1-\alpha}} \cdot (1+\tau(\theta))^{-\frac{\alpha}{1-\alpha}}. \tag{24}$$

The labor demand is unaffected by UI because the wage does not respond to UI. The labor demand is decreasing with θ. When the labor market is tighter, hiring workers is less profitable as it requires a higher share of recruiters, $\tau(\theta)$. Hence, firms choose a lower level of employment. The labor demand is also increasing in a. When technology is lower, the wage-technology ratio, $w/a = \omega \cdot a^{\gamma-1}$, is higher as wages are somewhat rigid, and hiring workers is less profitable. Hence, firms choose a lower level of employment. In the (l, θ) plane of Figure 3(c), the labor demand curve is downward sloping, and it shifts inward when technology falls.

Having characterized the labor demand, we describe the effect of UI on tightness:

PROPOSITION 6. *Increasing UI raises tightness: $d\theta/d\Delta v < 0$. The elasticity wedge is*

$$1 - \frac{\varepsilon^M}{\varepsilon^m} = \left(1 + \frac{\eta}{1-\eta} \cdot \frac{\alpha}{1-\alpha} \cdot \frac{1}{1+\varepsilon^f} \cdot \tau(\theta)\right)^{-1} > 0. \tag{25}$$

Proof. The elasticity of $1+\tau(\theta)$ with respect to θ is $\eta \cdot \tau(\theta)$. From (24), we infer that the elasticity of $l^d(\theta, a)$ with respect to θ is $-\eta \cdot \tau(\theta) \cdot \alpha/(1-\alpha)$. By definition, ε^M is $l/(1-l)$ times the elasticity of l with respect to Δv. Since $l = l^d(\theta, a)$ in equilibrium, we infer that

$$\varepsilon^M = -\frac{l}{1-l} \cdot \eta \cdot \frac{\alpha}{1-\alpha} \cdot \tau(\theta) \cdot \frac{\Delta v}{\theta} \cdot \frac{d\theta}{d\Delta v}$$

We plug the expression for $(\Delta v/\theta) \cdot (d\theta/d\Delta v)$ given by (18) into this equation and obtain

$$\varepsilon^M = \frac{\eta}{1-\eta} \cdot \frac{\alpha}{1-\alpha} \cdot \frac{1}{1+\varepsilon^f} \cdot \tau(\theta) \cdot \left(\varepsilon^m - \varepsilon^M\right).$$

Dividing this equation by ε^m and re-arranging yields (25). □

Figure 3(c) illustrates the results of the proposition. After an augmentation in UI, jobseekers search less, shifting the labor supply curve inward by a distance A–B. Since the labor demand is downward sloping, the initial reduction in employment is attenuated by a horizontal distance

B–C. The total reduction in employment is given by the horizontal distance A–C. Since the labor demand is downward sloping and does not respond to UI, tightness necessarily increases. Since A–C is smaller than A–B, the macroelasticity is smaller than the microelasticity.

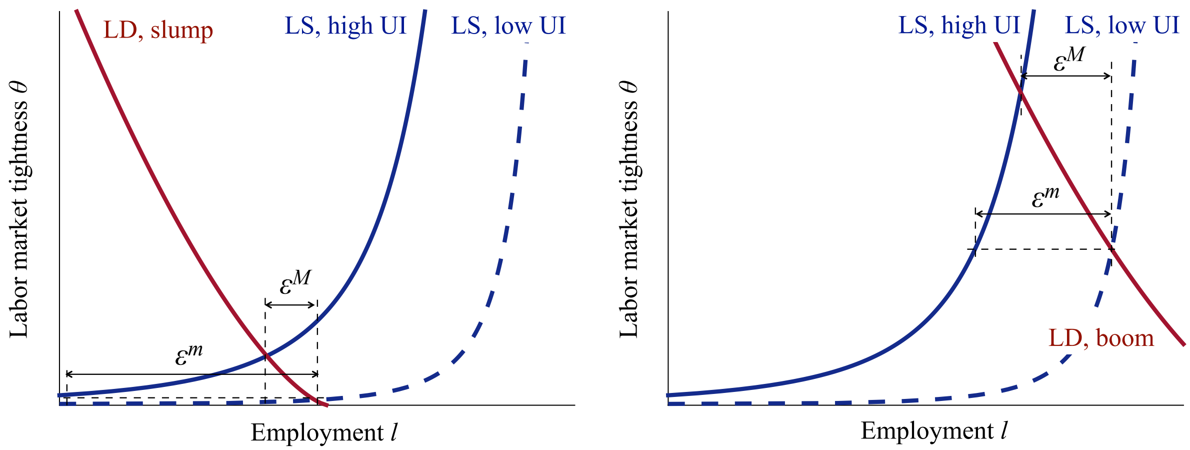
In addition to the moral-hazard effect, the job-rationing model features the rat-race effect, which is not present in the standard model. The rat-race effect is the increase in employment caused by the increase in tightness following the increase in UI. Intuitively, the number of jobs available is somewhat limited because of diminishing marginal returns to labor. Hence, when workers searching less, they reduce their own probability of finding a job but mechanically increases others' probability of finding one of the few jobs available. By discouraging job search, UI alleviates the rat race for jobs and increases the job-finding rate per unit of effort and labor market tightness.[16] The distance B–C measures employment gained through the rat-race effect. The rat-race effect is the reason why tightness rises when UI increases and why the macroelasticity is smaller than the microelasticity.

Proposition 6 implies that optimal UI is lower than the Baily-Chetty level when tightness is inefficiently high and higher than the Baily-Chetty level when tightness is inefficiently low. For instance, when tightness is inefficiently low, increasing UI raises tightness by alleviating the rat-race for jobs, which improves welfare beyond the insurance-incentive trade-off. The proposition also shows that the sign of the elasticity wedge does not depend at all on the rigidity of wages with respect to technology. Even if the wage were completely flexible with respect to technology ($\gamma = 1$), the job-rationing model would feature a positive elasticity wedge and thus a rat-race effect. In a way this is obvious because the response of tightness to UI is independent of the response of tightness to technology. What is critical to obtain a positive elasticity wedge is the rigidity of the wage with respect to UI.

The properties of the labor demand imply that equilibria with low technology are slumps, and equilibria with high technology are booms.

PROPOSITION 7. *For a given utility gain from work, an equilibrium with lower technology has*

[16] The formal argument is as follows. Consider an increase in UI. Imagine that tightness, θ, remained constant. Then the marginal recruiting cost, $\tau(\theta)$, would remain constant. As the wage, w, remains constant, the marginal cost of labor, $w \cdot (1 + \tau(\theta))$, would remain constant. Simultaneously, firms would employ fewer workers because workers search less. Hence, the marginal product of labor would be higher because of the diminishing marginal returns to labor. Firms would face the same marginal cost but a higher marginal product of labor, which would not be optimal. Thus, firms post more vacancies and the new equilibrium has higher labor market tightness.



(a) High elasticity wedge in a slump (b) Low elasticity wedge in a boom

Figure 4: Countercyclicality of the elasticity wedge, $1-\varepsilon^M/\varepsilon^m$, in the job-rationing model

Notes: This figure illustrates the results of Proposition 8. A slump corresponds to an equilibrium with low technology. A boom corresponds to an equilibrium with high technology.

lower tightness and lower employment: $\partial\theta/\partial a\big|_{\Delta v} > 0$ *and* $\partial l/\partial a\big|_{\Delta v} > 0$.

Proof. The equilibrium condition is $l^d(\theta,a) = l^s(\theta,\Delta v)$, where l^d is given by (24) and l^s by (8). Implicit differentiation of the equilibrium condition yields $\partial\theta/\partial a = \left(\partial l^d/\partial a\right)\cdot\left(\partial l^s/\partial\theta - \partial l^d/\partial\theta\right)^{-1}$. We have seen that $\partial l^d/\partial a > 0$, $\partial l^s/\partial\theta > 0$, and $\partial l^d/\partial\theta < 0$. Thus $\partial\theta/\partial a > 0$. The other result follows since $l = l^s(\theta,\Delta v)$ and $\partial l^s/\partial\theta > 0$. □

The proposition says that when technology is low, tightness and employment are low, as in a slump. Conversely, tightness and employment are high when technology is high, as in a boom. The mechanism is simple. When technology is low, the wage-technology ratio is high by wage rigidity, which depresses labor demand and therefore tightness and employment. Figure 4(a) plots the labor demand curve for a low technology, which represents a slump, and Figure 4(b) plots it for a high technology, which represents a boom.

Furthermore, the properties of the labor demand imply that jobs are rationed in slumps. When technology is low enough ($a < (\alpha/\omega)^{\frac{1}{\gamma-1}}$), then $l^d(\theta = 0,a) < 1$ and jobs are rationed: firms would not hire all the workers even if workers searched infinitely hard and tightness was zero. In Figure 3(c), the labor demand crosses the x-axis at $l < 1$.

To conclude, we discuss how that the rat-race effect varies over the business cycle. The following proposition establishes that the elasticity wedge is higher in slumps than in booms:

ASSUMPTION 1. *The matching function and the marginal disutility of search effort are isoelastic: $m(e,v) = \omega_h \cdot e^{\eta} \cdot v^{1-\eta}$ and $k'(e) = \omega_k \cdot e^{\kappa}$ for $\eta \in (0,1)$, $\kappa > 0$, $\omega_h > 0$, and $\omega_k > 0$.*

PROPOSITION 8. *Suppose that Assumption 1 holds. An equilibrium with lower technology has a higher elasticity wedge: $\partial \left[1 - \varepsilon^M / \varepsilon^m\right] / \partial a\big|_{\Delta v} < 0$.*

Proof. This result follows from combining Proposition 7 with (25), the fact that $\tau(\theta)$ is increasing with θ, and the fact that $\varepsilon^f = 1/\kappa$ and η are constant under Assumption 1. □

The proposition shows that the elasticity wedge is higher in slumps than in booms. This implies the rat-race effect is stronger in slumps than in booms. This result is illustrated by comparing the boom in Figure 4(b) to the slump in Figure 4(a). The wedge between ε^M and ε^m is driven by the slope of the labor supply relative to that of the labor demand. In a boom, the labor supply is steep at the equilibrium point because the matching process is congested by the large number of vacancies. Hence, ε^M is close to ε^m. Conversely, in a slump, the labor supply is flat at the equilibrium point because the matching process is congested by search efforts. Hence, ε^M is much lower than ε^m. Formally, let $\varepsilon^{ls} \equiv (\theta/l) \cdot (\partial l^s / \partial \theta)$ and $\varepsilon^{ld} \equiv -(\theta/l) \cdot \left(\partial l^d / \partial \theta\right)$ be the elasticities of labor supply and labor demand with respect to tightness (ε^{ld} is normalized to be positive). We could rewrite the elasticity wedge as $1 - \varepsilon^M / \varepsilon^m = 1 / \left[1 + (\varepsilon^{ld} / \varepsilon^{ls})\right]$. The elasticity wedge is countercyclical because $\varepsilon^{ld} / \varepsilon^{ls}$ is procyclical.[17]

5.4 Elasticity Wedge in an Aggregate-Demand Model

The aggregate-demand model does not appear elsewhere in the literature, but it is useful to establish the robustness of the rat-race effect. This model shows that diminishing marginal returns to labor are not required to obtain a rat-race effect. Here the rat-race effect is present even though the marginal returns to labor are constant. We will see that this effect is present because the general-equilibrium labor demand is downward sloping in a (l, θ) plane, such that the number of jobs is limited for a given tightness. This model also shows that technology shocks combined with

[17] The cyclicality of the elasticity wedge is closely connected to the cyclicality of the public-employment multiplier in Michaillat [2014]. Both rely on the cyclicality of the ratio $\varepsilon^{ld} / \varepsilon^{ls}$. The difference is that the elasticity wedge describes the response to a shift in labor supply whereas the multiplier describes the response to a shift in labor demand.

real wage rigidity is not the only mechanism that can generate slumps and booms. In this model, slumps and booms are generated by money-supply shocks and nominal wage rigidity.

We make the following assumptions on the production function and wage schedule. The production function is linear: $y(n) = n$. The nominal wage is independent of UI and partially rigid with respect to the price level P: $W = \mu \cdot P^{\zeta}$, where $\zeta \in [0,1)$ parameterizes the rigidity of the nominal wage with respect to the price level. The real wage is $w = W/P = \mu \cdot P^{\zeta-1}$.

Because of nominal wage rigidity, it is necessary to define the price-setting mechanism. As in Mankiw and Weinzierl [2011], we assume that workers are required to hold money to purchase consumption goods and that the money market is described by a quantity equation: $M = P \cdot y$. The parameter $M > 0$ is the money supply. The quantity equation says that nominal consumer spending is equal to the money supply.[18] Since $y = n$, the quantity equation implies that $P = M/n$. A high number of producers implies high output and, for a given money supply, a low price. In general equilibrium, the real wage is therefore related to the number of producers by

$$w = \mu \cdot \left(\frac{n}{M}\right)^{1-\zeta}.$$

When money supply falls or the number of producers rises, the price falls and the real wage rises.

The product market is perfectly competitive; hence, firms take the price as given and (3) remains valid. We combine the wage equation with (3) and $l = (1+\tau(\theta)) \cdot n$ to obtain the general-equilibrium labor demand:

$$l^d(\theta, M) = M \cdot \mu^{-\frac{1}{1-\zeta}} \cdot (1+\tau(\theta))^{-\frac{\zeta}{1-\zeta}}.$$

This is a general-equilibrium demand because it takes into account the quantity equation describing the money market. Labor demand decreases with θ, as in the job-rationing model. But the mechanism is different. Higher employment implies more production, lower prices in the goods market, and higher real wages by nominal wage rigidity. Firms are willing to hire more workers only if tightness is lower, which reduces recruiting costs and compensates for the higher real wage. Moreover, the labor demand increases with M: After a negative money-supply shock, prices fall. Nominal wage rigidity combined with a lower price level leads to a higher real wage and a higher

[18] We implicitly assume that the velocity of money is constant and normalized to 1.

marginal cost of labor, which leads to lower hiring and higher unemployment. The labor demand slopes downward in the (l,θ) plane, as depicted in Figure 3(c). The labor demand shifts inward when the money supply decreases, but the labor demand does not shift after a change in UI. Jobs are also rationed in slumps. If money supply is low enough ($M < \mu^{1/(1-\zeta)}$), then $l^d(\theta = 0, M) < 1$ and jobs are rationed: firms would not hire all the workers even if workers searched infinitely hard.

It is clear by now that the aggregate-demand model has exactly the same properties as the job-rationing model, except that money-supply shocks and not technology shocks generate fluctuations in tightness and employment. To conclude, we list all the properties of the aggregate-demand model. The interpretations and proofs are the same as in the job-rationing model.

PROPOSITION 9. *Increasing UI raises tightness: $d\theta/d\Delta v < 0$. The elasticity wedge is*

$$1 - \frac{\varepsilon^M}{\varepsilon^m} = \left(1 + \frac{\eta}{1-\eta} \cdot \frac{\zeta}{1-\zeta} \cdot \frac{1}{1+\varepsilon^f} \cdot \tau(\theta)\right)^{-1} > 0.$$

PROPOSITION 10. *For a given utility gain from work, an equilibrium with lower money supply has lower tightness and lower employment: $\partial\theta/\partial M\big|_{\Delta v} > 0$ and $\partial l/\partial M\big|_{\Delta v} > 0$.*

PROPOSITION 11. *Suppose that Assumption 1 holds. An equilibrium with lower money supply has higher elasticity wedge: $\partial\left[1 - \varepsilon^M/\varepsilon^m\right]/\partial M\big|_{\Delta v} < 0$.*

6 Implications of the Theory for Business Cycles

A common view is that unemployment may be efficient on average, but because of macroeconomic shocks, unemployment is inefficiently high in slumps and inefficiently low in booms. This view would suggest that the efficiency term in our optimal UI formula systematically changes sign over the business cycle. If the elasticity wedge is nonzero, this view therefore implies that optimal UI fluctuates over the business cycle.

In this section we explore this view. We combine our optimal UI formula with empirical evidence to draw the implications of our theory for business cycles. We provide empirical evidence that the efficiency term is positive in slumps but negative in booms, and that the elasticity wedge is positive. Our theory therefore implies that the optimal replacement rate of UI is countercylical. We also find that the countercylical fluctuations of the optimal replacement rate are sizable.

6.1 The Optimal Unemployment Insurance Formula in a Dynamic Model

To offer a better mapping between the theory and the data, we first embed the static model into a dynamic environment. We work in continuous time.

At time t, the number of employed workers is $l(t)$ and the number of unemployed workers is $u(t) = 1 - l(t)$. Labor market tightness is $\theta_t = o_t/(e_t \cdot u_t)$. Jobs are destroyed at rate $s > 0$. Unemployed workers find a job at rate $e(t) \cdot f(\theta(t))$. Thus, the law of motion of employment is

$$\dot{l}(t) = e(t) \cdot f(\theta(t)) \cdot (1 - l(t)) - s \cdot l(t). \tag{26}$$

In steady state $\dot{l}(t) = 0$. Hence, employment, effort, and tightness are related by

$$l = \frac{e \cdot f(\theta)}{s + e \cdot f(\theta)}. \tag{27}$$

Let $L(x) = x/(s+x)$. The elasticity of L with respect to x is $1 - L$. It is because $l = L(e \cdot f(\theta))$ in the dynamic model instead of $l = e \cdot f(\theta)$ in the static model that the factor $1 - l$ appears in many formulas of the dynamic model.

Firms employ $n(t)$ producers and $l(t) - n(t)$ recruiters. Each recruiter handles $1/r$ vacancies so the law of motion of the number of employees is

$$\dot{l}(t) = -s \cdot l(t) + \frac{l(t) - n(t)}{r} \cdot q(\theta(t)). \tag{28}$$

In steady state $\dot{l}(t) = 0$ and the number of employees is proportional to the number of producers: $l = (1 + \tau(\theta)) \cdot n$, where $\tau(\theta) = (s \cdot r)/[q(\theta) - (s \cdot r)]$.[19]

We focus on the steady state of the model with no time discounting. Firms, workers, and government maximize the flow value of profits, utility, and social welfare subject to the steady-state constraints. Given w and θ, the firm chooses n to maximize $y(n) - w \cdot (1 + \tau(\theta)) \cdot n$. The optimal employment level satisfies (2). Given θ, c^e, and c^u, the representative unemployed worker

[19] The wedge $\tau(\theta)$ is a different function of the parameters in the static and dynamic environments, but $\tau(\theta)$ should be considered as a sufficient statistic defined as $1 + \tau(\theta) \equiv l/n$. That is, $\tau(\theta)$ is the recruiter-producer ratio.

chooses e to maximize

$$l \cdot v(c^e) + (1-l) \cdot v(c^u) - (1-l) \cdot k(e) \tag{29}$$

with l given by (27). Routine calculations show that the optimal search effort e satisfies

$$k'(e) = \frac{l}{e} \cdot (\Delta v + k(e)). \tag{30}$$

Finally, the government chooses c^e and c^u to maximize (29) subject to (1), (2), (27), (30), and (4).

Without discounting, the static results are barely modified in the dynamic model. Following the same steps as in the static model, we can show that the formula of Lemma 1 becomes

$$\frac{\theta}{l} \cdot \frac{\partial l^s}{\partial \theta}\bigg|_{\Delta v} = (1-l) \cdot (1+\varepsilon^f) \cdot (1-\eta)$$

The only difference with the original formula is the extra factor $1-l$. In the dynamic environment, equation (19), which links micro- and macroelasticity becomes

$$\varepsilon^M = \varepsilon^m + l \cdot (1-\eta) \cdot \left(1+\varepsilon^f\right) \cdot \frac{\Delta v}{\theta} \cdot \frac{d\theta}{d\Delta v}.$$

As expected, the only difference with the original formula is that a factor l replaces the factor $l/(1-l)$. Accordingly, in the dynamic model, the wedge becomes

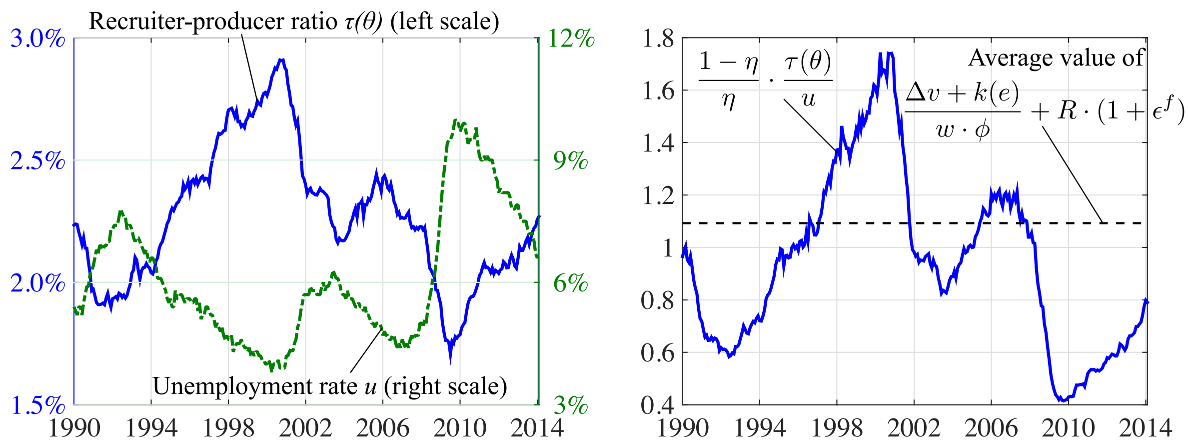
$$1 - \frac{\varepsilon^M}{\varepsilon^m} = -l \cdot (1-\eta) \cdot \frac{1+\varepsilon^f}{\varepsilon^m} \cdot \frac{\Delta v}{\theta} \cdot \frac{d\theta}{d\Delta v}.$$

Here again, the only difference with the original wedge is that a factor l replaces the factor $l/(1-l)$. Accordingly, in the dynamic model, the formula of Proposition 2 becomes

$$R = \frac{l}{\varepsilon^m} \frac{\Delta v}{w} \left[\frac{1}{v'(c^e)} - \frac{1}{v'(c^u)}\right] + \frac{1}{1+\varepsilon^f} \left[1 - \frac{\varepsilon^M}{\varepsilon^m}\right] \left[\frac{\Delta v + k(e)}{w \cdot \phi} + R\left(1+\varepsilon^f\right) - \frac{\eta}{1-\eta} \frac{\tau(\theta)}{1-l}\right]. \tag{31}$$

where ϕ satisfies equation (15). Two differences appear in the correction term: Δv is replaced by $\Delta v + k(e)$ and $\tau(\theta)$ is replaced by $\tau(\theta)/(1-l)$.

Although it seems that the formula is modified in the dynamic model, the two formulas are in



(a) Recruiter-producer ratio and unemployment

(b) Fluctuations of $[\eta/(1-\eta)] \cdot \tau(\theta_t)/u_t$

Figure 5: A method to measure the fluctuations of the efficiency term

Notes: The time period is January 1990–February 2014. In panel (a) the unemployment rate, u_t, is the seasonally adjusted monthly unemployment rate constructed by the BLS from the CPS. The recruiter-producer ratio is constructed as $\tau_t = \sigma \cdot rec_t/(l_t - \sigma \cdot rec_t)$, where rec_t is the seasonally adjusted monthly number of workers in the recruiting industry (NAICS 56131) computed by the BLS from CES data, l_t is the seasonally adjusted monthly number of workers in all private industries computed by the BLS from CES data, and $\sigma = 8.4$ is a scaling factor ensuring that the level of the recruiter-producer ratio matches the evidence in Villena Roldan [2010]. In panel (b) the solid blue line is $[\eta/(1-\eta)] \cdot \tau_t/u_t$ with $\eta = 0.7$. The dashed black line is the average value of $(\Delta v + k(e))/(\phi \cdot w) + (1+\varepsilon^f) \cdot R$, calibrated in Appendix A. This value is obtained by setting $R = 58\%$, $\varepsilon^f = 0.021$, and $(\Delta v + k(e))/(\phi \cdot w) = 0.51$.

fact identical once expressed with the correct statistics. The derivation of the formula shows that the Δv in the correction term stands for the utility gap between employed and unemployed workers. This gap is Δv in the static model where both employed and unemployed workers search, whereas it is $\Delta v + k(e)$ in the dynamic model where only unemployed workers search. The derivation also shows that the $\tau(\theta)$ in the correction term is divided by the elasticity of the function L, defined by $l^s(\theta, \Delta v) = L(e^s(\theta, \Delta v) \cdot f(\theta))$. This elasticity is 1 in the static model but $1-l$ in the dynamic model.

6.2 An Estimate of the Efficiency Term

To determine empirically whether the labor market tightness is inefficiently high or low, we measure the efficiency term in US data. In the dynamic environment, the efficiency term is

$$\frac{\Delta v_t + k(e_t)}{w_t \cdot \phi_t} + \left(1+\varepsilon^f\right) \cdot R_t - \frac{\eta}{1-\eta} \cdot \frac{\tau_t}{u_t},$$

where ϕ_t satisfies equation (15).

We start by measuring $[\eta/(1-\eta)] \cdot \tau_t/u_t$. We set $\eta = 0.7$, in line with empirical evidence [Petrongolo and Pissarides, 2001]. We measure the unemployment rate u_t with the seasonally adjusted monthly unemployment rate constructed by the Bureau of Labor Statistics (BLS) from the Current Population Survey (CPS). The challenging part is to measure the recruiter-producer ratio τ_t. We measure τ_t in three steps.

The first step is to measure the number rec_t of employees in the recruiting industry. We set rec_t to the seasonally adjusted monthly number of workers in the recruiting industry computed by the BLS from Current Employment Statistics (CES) data. The recruiting industry is the industry with North American Industry Classification System (NAICS) code 56131. Its official name is "employment placement agencies and executive search services". It comprises firms primarily engaged in listing employment vacancies and referring or placing applicants for employment, and firms providing executive search, recruitment, and placement services. The series is available from January 1990 to February 2014. This industry comprises 279,800 workers on average.

Of course, the employees in the recruiting industry are only a small share of all the workers devoted to recruiting. Many workers in firms outside of the recruiting industry spend a lot of time and effort to recruit workers for their own firm. To account for these workers and capture the total amount of labor devoted to recruiting in the economy, we scale up our measure based on the recruiting industry by a factor 8.4. This scaling factor ensures that the average share of labor devoted to recruiting in 1997 is 2.5%. We obtain this amount from the National Employer Survey (NES) conducted in 1997 by the Census Bureau. The survey gathered employer data on employment practices, especially recruiting. In the 1997 survey, 4500 establishments answered detailed questions about the methods used to recruit applicants. Villena Roldan [2010] analyzes this survey and finds that firms spend 2.5% of their total labor cost in recruiting activities. In other words, 2.5% of the workforce is devoted to recruiting.[20]

The last step is to construct τ_t as $\tau_t = 8.4 \cdot rec_t/(l_t - 8.4 \cdot rec_t)$, where l_t is the seasonally adjusted monthly number of workers in all private industries computed by the BLS from CES data. Figure 5(a) displays the recruiter-producer ratio τ_t and the unemployment rate u_t. The series τ_t is very procyclical. Figure 5(b) displays the term $[\eta/(1-\eta)] \cdot \tau_t/u_t$. This term is extremely

[20] In monetary terms in 1997, firms spent on average $4200 per recruited worker.

procyclical: it is more than four times higher at the peak of the expansion in 2001 (when $\tau_t = 1.74$) than at the bottom of the recession in 2009 (when $\tau_t = 0.41$).

To estimate the efficiency term, we also need to measure $(\Delta v_t + k(e_t))/(\phi_t \cdot w_t) + (1+\varepsilon^f) \cdot R_t$. It is difficult to obtain a precise value for this expression. Nevertheless, we argue that this expression varies much less than $[\eta/(1-\eta)] \cdot \tau_t/u_t$ over the business cycle; therefore, it is not critical to obtain a precise value for this expression to know broadly when the efficiency term is positive and negative. With a log utility $\Delta v = \ln(c^e/c^u)$. Chodorow-Reich and Karabarbounis [2014] show that the ratio c^e/c^u is around 0.7–0.8 in the US and acyclical over the period 1983–2012 (Figure 2). Shimer [2004] shows that search effort among the unemployed does not vary systematically over the business cycle suggesting that $k(e)$ is relatively acyclical as well. Since $l \approx 1$ and using log utility, we find that $1/(\phi \cdot w) \approx 1/(v'(c^e) \cdot w) \approx c^e/w$. The ratio c^e/w varies relatively little over the cycle and certainly not by a factor four. Hence, $(\Delta v + k(e))/(w \cdot \phi)$ is overall relatively acyclical. Finally, $(1+\varepsilon^f) \cdot R$ is somewhat countercyclical in the United States as the generosity of UI increases in recessions, but its fluctuations are relatively small. Therefore, in net, the term $(\Delta v_t + k(e_t))/(\phi_t \cdot w_t) + (1+\varepsilon^f) \cdot R_t$ is fairly stable over the business cycle, perhaps slightly countercyclical.

Since the term $(\Delta v_t + k(e_t))/(\phi_t \cdot w_t) + (1+\varepsilon^f) \cdot R_t$ varies little over the business cycle, and in any case varies much less than the term $[\eta/(1-\eta)] \cdot \tau_t/u_t$, we only display its average value in Figure 5(b). This average value is calibrated in Appendix A. When $[(1-\eta)/\eta] \cdot \tau(\theta)/u$ is above $(\Delta v + k(e))/(\phi \cdot w) + (1+\varepsilon^f) \cdot R$, the efficiency term is negative and the tightness is inefficiently high, and when it is below, the efficiency term is positive and the tightness is inefficiently low. We are not able to delimit precisely periods when the efficiency term is exactly zero, but given the amplitude of the fluctuations in $[\eta/(1-\eta)] \cdot \tau_t/u_t$, it is likely that the efficiency term was negative in the 1991–1994 and 2009–2013 periods and positive in the 1998–2001 period.

More empirical research is needed to obtain a precise value of the efficiency term. For instance, if unemployed workers are particularly miserable and suffer large psychological costs from unemployment, the term $\Delta v + k(e)$ may be much higher than in our calibration. In that case, the efficiency term would be higher than in our calibration, and the efficient tightness higher. Conversely, if unemployed workers are particularly happy and enjoy utility from leisure, the term $\Delta v + k(e)$ may be much lower than in our calibration. In that case, the efficiency term would be lower than in our calibration, and the efficient tightness lower. In any case, our optimal UI formula would apply, and our methodology could be helpful to measure the efficiency term.

A limitation of our measure of τ_t is that it is only valid as long as the share of recruitment done through recruiting firms is stable over the business cycle. In Appendix B, we propose an alternative measure of τ_t that does not rely on this assumption. The alternative measure relies instead on the assumption that the cost of posting a vacancy is constant over the business cycle. This alternative measure is available from December 2000 to February 2014. Over this period, we find that the two measures are quite similar (see appendix Figure 7).

6.3 Estimates of the Elasticity Wedge

To assess empirically whether increasing UI raises or lowers labor market tightness, we need an estimate of the elasticity wedge. The ideal experiment to estimate the elasticity wedge is a design with double randomization : (i) some randomly selected areas are treated and some are not, and (ii) within treated areas, all but a randomly selected and small subset of jobseekers are treated. The treatment is to offer higher or longer UI benefits. The elasticity wedge can be estimated by comparing the unemployment durations of non-treated jobseekers in non-treated areas to that of non-treated jobseekers in treated areas. We discuss two recent studies that estimate this wedge.

Lalive, Landais and Zweimüller [2013] use a natural experiment that offers the desired design: the Regional Extended Benefit Program (REBP) implemented in Austria in 1988–1993. The treatment was an increase in benefit duration from 52 to 209 weeks for eligible unemployed workers in a subset of regions. They estimate a positive elasticity wedge, $1-\varepsilon^M/\varepsilon^m = 0.3$.

Marinescu [2014] follows another route to assess the sign and magnitude of the elasticity wedge: she directly estimates the effect of a change in UI on labor market tightness. The changes in UI that she considers are the long UI extensions implemented in the US from 2009 to 2013. Using detailed information on vacancies and job applications from CareerBuilder.com, the largest American online job board, she computes the effect of a change in UI on aggregate search effort, measured by the number of job applications sent, and on vacancies. At the state level, she finds that an increase in UI has a negative effect on job applications but no effect on vacancies. Since labor market tightness is the ratio of vacancies to aggregate search effort, these results imply that an increase in UI raises tightness and thus that the elasticity wedge is positive.

The two previous studies find a positive elasticity wedge when the rat-race and job-creation effects are accounted for. Several papers study these effects in isolation. They find evidence of rat-race effect but not of job-creation effect, which is further evidence of a positive elasticity wedge.

Papers that aim to measure the rat-race effect include Burgess and Profit [2001], Ferracci, Jolivet and van den Berg [2010], Gautier et al. [2012], and especially Crepon et al. [2013]. All these papers find that an increase in the search effort of some jobseekers, induced for example by job training programs, has a negative effect on the job-finding probability of other jobseekers. These findings are consistent with rat-race effects.

The best way to measure the job-creation effect is to measure the effect of an increase in UI on wages at the micro level. A number of studies have investigated whether more generous UI benefits affect the re-employment wage. Most studies find no effect on wages or even slightly negative effects [for example, Card, Chetty and Weber, 2007]. Since the job-creation effect operates when an increase in UI raises wages, the absence of effect on wages suggests that the job-creation effect is likely to be small.[21]

In sum, the available evidence suggests that the elasticity wedge is positive, with best estimate $1-\varepsilon^M/\varepsilon^m = 0.3$. We will use this estimate to quantity the cyclical fluctuations of optimal UI.

6.4 Quantifying the Fluctuations of Optimal Unemployment Insurance

Small Reform Evaluation. Our formula is useful to assess the desirability of a small reform around the current system. The statistics in the Baily-Chetty term—microelasticity and coefficient of risk aversion—are well measured and are commonly used to estimate the Baily-Chetty level of UI [for example, Gruber, 1997]. The statistics in the correction term—elasticity wedge and efficiency term—can be estimated following the methodology discussed above. The right-hand-side term in formula (31) can therefore be estimated at the current UI system. If the current

[21] More generous benefits induce longer unemployment durations that may have a negative effect on wages if, for instance, the duration of unemployment spells affects the productivity of unemployed workers or is interpreted by employers as a negative signal of productivity. It is difficult to disentangle this negative effect from the positive effect of UI on wages through bargaining, which is the relevant effect for our analysis. In German data, Schmieder, von Wachter and Bender [2013] attempt such a decomposition by controlling for the duration of the unemployment spell. They find a negative effect of UI on wages through longer unemployment durations but zero effect through wage bargaining. Lalive, Landais and Zweimüller [2013] use the same methodology and find a positive but small effect of UI through wage bargaining in Austria.

replacement rate is less than the right-hand-side term in formula (31), increasing UI increases welfare, and conversely, if the current replacement rate is more than the right-hand-side term, decreasing UI increases welfare.

In fact, since the actual UI replacement rate in the US is very close to the Baily-Chetty rate, we can assume that the evidence provided above about the signs of the elasticity wedge and efficiency term are valid with a UI system at the Baily-Chetty level.[22] This evidence indicates that the elasticity wedge is positive, and that the efficiency term is positive in slumps but negative in booms. Using the argument in Table 1, we infer from formula (31) that the optimal UI replacement rate is more countercyclical than the Baily-Chetty rate.

We can also use the formula to get a sense of the amplitude of the cyclical fluctuations of optimal UI. With standard estimates of the microelasticity and risk aversion, a replacement rate of $R = 50\%$ roughly satisfies the Baily-Chetty formula [Chetty, 2006*a*; Gruber, 1997]. So the Baily-Chetty term is roughly 0.5. Available evidence indicates that the elasticity wedge is positive, maybe around 0.3. Our measure of the efficiency term in Figure 5 indicates that the between a boom and a slump, the efficiency term increases by 1.4. Hence, if the labor market tightness is efficient on average, the efficiency term increases by $1.4/2 = 0.7$ in slumps and falls by $1.4/2 = 0.7$ in booms. This means that the correction term maybe increases by $0.3 \times 0.7 = 0.2$ in slumps and falls by 0.2 in booms. Compared to a Baily-Chetty term of 0.5, these fluctuations in the correction term are quantitatively significant, and will lead to marked departure of the optimal UI replacement rate from the Baily-Chetty rate.

Calibration and Simulation. To assess more precisely the amplitude of the fluctuations of optimal UI over the business cycle, we also simulate a dynamic and calibrated version of the job-rationing model of Section 5—chosen because it generates realistic fluctuations in the efficiency term and a positive elasticity wedge. We use formula (31) to determine the optimal UI replacement over the business cycle.

The elasticity wedge and efficiency term, which are the two critical statistics determining the cyclicality of optimal UI, are outcomes of the simulations. This does not mean that the empirical estimates of the elasticity wedge and efficiency term are irrelevant; these estimates are fundamental

[22] With the calibration in Table 3, the current US replacement rate of $R = 58\%$ exactly satisfies the Baily-Chetty formula. We establish this result in Appendix A.

to conduct relevant simulations. The sign of the elasticity wedge estimated in empirical work allows researchers to choose the right model of the labor market: a negative wedge advocates for using the standard model whereas a positive wedge advocates for using the job-rationing or aggregate-demand model. Here we simulate the job-rationing model because empirical evidence on the elasticity wedge suggests that it is positive, and because this model is more conventional than the aggregate-demand model.

We represent the business cycle as a succession of steady states with different values for technology. Formally, the simulations provide a comparative steady-state analysis. This analysis provides a good approximation to a dynamic simulation because matching models of the labor market reach their steady state quickly.[23] We compute a collection of steady states spanning all the stages of the business cycle, from slumps with high unemployment to booms with low unemployment. We perform two types of simulations. We first simulate a collection of steady states in which the replacement rate remains constant at its average value of 58%. We then simulate steady states for the same values of the underlying parameter but with the optimal replacement rate given by (31).

We calibrate the model to US data, as summarized in Table 3. The calibration procedure is standard so we relegate it to Appendix A. Here we only explain how we calibrate the model to match empirical evidence on the elasticity wedge and microelasticity.

A large body of work has estimated the microelasticity ε^m.[24] The ideal experiment to estimate ε^m is to offer higher or longer UI benefits to a randomly selected and small subset of jobseekers within a labor market and compare unemployment durations between treated and non-treated jobseekers. In practice, ε^m is estimated by comparing individuals with different benefits in the same labor market at a given time, while controlling for individual characteristics. Most empirical studies estimate the elasticity with respect to the replacement rate, which we denote ε^m_R. In US administrative data from the 1980s, the classic study of Meyer [1990] finds an elasticity $\varepsilon^m_R = 0.6$ using state fixed-effects (Table VI, columns (6) to (9)). In a larger US administrative dataset for the same early 1980s years, and using a regression kink design to identify the elasticity, Landais [2014]

[23]Shimer [2005] and Pissarides [2009] argue that in a standard matching model, the steady-state equilibrium with technology a approximates well the equilibrium in a stochastic environment when the realization of technology is a. They have two reasons. First, after a shock the labor market rapidly converges to a situation where inflows to and outflows from employment are balanced because rates of inflow to and outflow from unemployment are large [Hall, 2005*b*]. Second, technology is very persistent. Appendix A5 in Michaillat [2012] validates this approximation with numerical simulations.

[24]See Krueger and Meyer [2002] for a comprehensive survey.

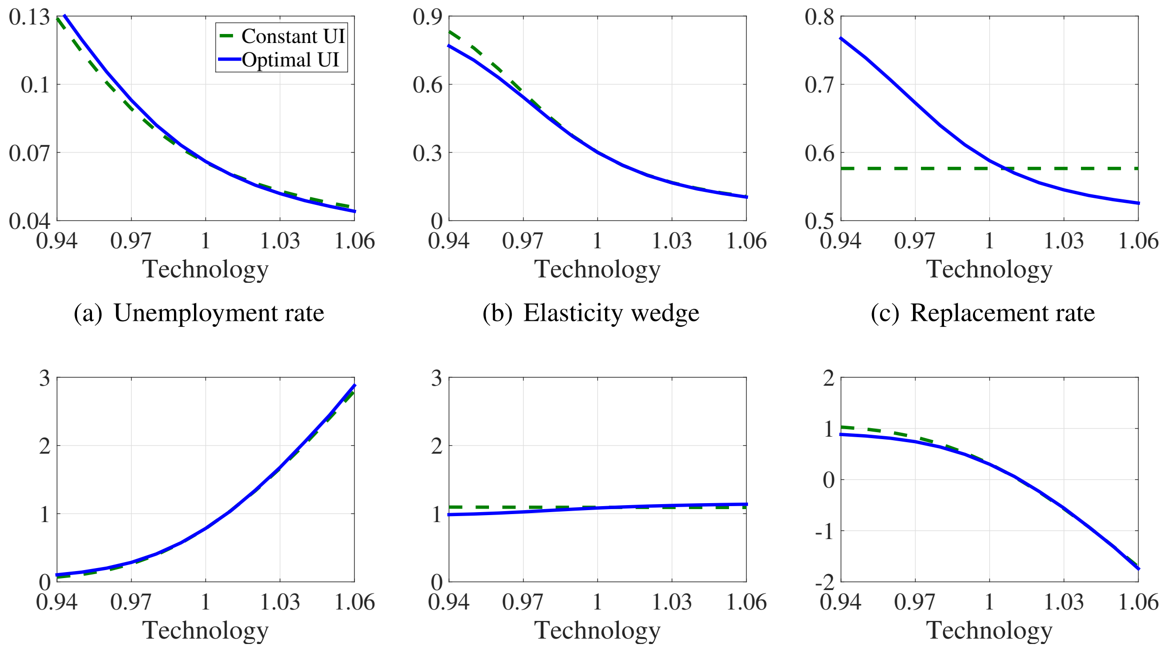
Table 3: Parameter values in the simulations of the job-rationing model

	Description	Value	Source
	A. Average values used for calibration		
u	unemployment rate	6.6%	CPS, 2000–2014
θ	labor market tightness	0.37	JOLTS and CPS, 2000–2014
$\tau(\theta)$	recruiter-producer ratio	2.2%	Villena Roldan [2010] and CES, 2000–2014
ε^m	microelasticity	0.3	literature
e	search effort	1	normalization
R	replacement rate	58%	Chetty [2008]
	B. Calibrated parameters		
η	unemployment-elasticity of matching	0.7	Petrongolo and Pissarides [2001]
ρ	relative risk aversion	1	Chetty [2006*b*]
γ	technology-elasticity of real wage	0.5	Haefke, Sonntag and van Rens [2008]
s	monthly job-destruction rate	3.5%	JOLTS, 2000–2014
κ	convexity of disutility of search	3.1	matches $\varepsilon^m = 0.3$
α	marginal returns to labor	0.75	matches $1-\varepsilon^M/\varepsilon^m = 0.3$
ω_m	matching efficacy	0.67	matches average values
r	recruiting cost	0.84	matches average values
ω_k	level of disutility of search	0.38	matches average values
ω	level of real wage	0.75	matches average values for $a=1$

finds an elasticity $\varepsilon^m_R = 0.3$ (Table 4, column (1)). Overall, $\varepsilon^m_R = 0.5$ seems a reasonable estimate. Using (12), this estimate $\varepsilon^m_R = 0.5$ for the microelasticity of unemployment with respect to the replacement rate is translated into an estimate $\varepsilon^m = 0.3$ for the microelasticity of unemployment with respect to Δv.

To obtain a search behavior consistent with the empirical evidence on the microelasticity, we assume a disutility of effort $k(e) = \omega_k \cdot e^{\kappa+1}/(\kappa+1) - \omega_k/(\kappa+1)$, and we set $\kappa = 3.1$ to match the microelasticity of $\varepsilon^m = 0.3$. We do not have direct evidence on the discouraged-worker elasticity, ε^f, but ε^f is closely related to ε^m. We find that our estimate $\varepsilon^m = 0.3$ implies an estimate $\varepsilon^f = 0.021$. Since $\varepsilon^m = 0.3$, our calibration implies a strong response of search effort to UI. But since $\varepsilon^f = 0.021$, our calibration is also consistent with the evidence that the response of search effort to the job-finding rate is weak [Shimer, 2004].

To match the estimate of the elasticity wedge $1-\varepsilon^M/\varepsilon^m = 0.3$, we calibrate the production-function parameter α. This parameter determines the magnitude of the elasticity wedge and of the



(a) Unemployment rate (b) Elasticity wedge (c) Replacement rate

(d) $[\eta/(1-\eta)]\,\tau(\theta)/u$ (e) $(\Delta v + k(e))/(w\phi) + (1+\varepsilon^f)R$ (f) Efficiency term = (e)−(d)

Figure 6: Steady states under constant UI and optimal UI in the job-rationing model

rat-race effect. We calibrate α to match the estimate of the elasticity wedge provided by Lalive, Landais and Zweimüller [2013]: $1-\varepsilon^M/\varepsilon^m = 0.3$. Extending (25) to the dynamic environment without discounting yields

$$1-\frac{\varepsilon^M}{\varepsilon^m} = \left(1+\frac{\eta}{1-\eta}\cdot\frac{\alpha}{1-\alpha}\cdot\frac{1}{1+\varepsilon^f}\cdot\frac{\tau(\theta)}{u}\right)^{-1}.$$

We set $1-\varepsilon^M/\varepsilon^m = 0.3$, $\eta = 0.7$, $\varepsilon^f = 0.021$, $\tau(\theta) = 2.2\%$, and $u = 6.6\%$ and obtain $\alpha = 0.75$.

Figure 6 displays the simulations of the job-rationing model. Each steady state is indexed by technology, $a \in [0.94, 1.06]$. Because of wage rigidity, the steady states with low a have a high wage-technology ratio and therefore high unemployment: they represent slumps. Conversely, the steady states with high a have a low wage-technology ratio and low unemployment: they represent booms. As showed in Figure 6(a), unemployment falls from 12.9% to 4.6% when a increases from 0.94 to 1.06 and UI remains constant. This numerical results implies that even a modest amount of wage rigidity, in line with the empirical findings of Haefke, Sonntag and van Rens [2008], generates realistic fluctuations in unemployment. Indeed, the elasticity of unemployment with respect to technology implied by the simulations is 9.5, larger than the elasticity of 4.2 measured in US data.[25]

It is not the unemployment rate that matters for optimal UI, but the efficiency term, depicted in Figure 6(f). In a slump, the efficiency term is positive. The efficient term is 0 for $a = 1.02$ and an unemployment rate of 6.0%. In booms, the efficiency term is negative. In fact, the efficiency term is constructed as the difference between $(\Delta v + k(e))/(w \cdot \phi) + (1 + \varepsilon^f) \cdot R$, depicted in Figure 6(e), and $[\eta/(1-\eta)] \cdot \tau(\theta)/u$, depicted in Figure 6(d). The former term is broadly constant at 1.1 for any value of technology; the latter term increases steeply from 0.07 to 2.80 when technology increases from $a = 0.94$ to $a = 1.06$; hence the efficiency term decreases from 1.03 to -1.70 when technology increases from $a = 0.94$ to $a = 1.06$.

Figure 6(b) displays the elasticity wedge. The wedge is positive, which implies that optimal UI should be below the Baily-Chetty level in booms and above it in slumps. The wedge is also countercyclical, consistent with the results of Proposition 8. In an average situation ($a = 1$), the wedge is 0.3. The average wedge matches exactly the estimate from Lalive, Landais and Zweimüller [2013] due to our calibration of the production-function parameter, α. From a slump ($a = 0.94$) to a boom ($a = 1.06$), the wedge falls from 0.83 to 0.11.

Figure 6(c) shows that the optimal replacement rate and the technology parameter driving business cycles are negatively correlated: the optimal replacement rate falls from 77% when $a = 0.94$) to 53% when $a = 1.06$. Thus, the optimal replacement rate is sharply countercyclical.

Next, the unemployment rate responds to the adjustment of the replacement rate from its original level to its optimal level. In slumps, the optimal replacement rate is higher than its original level so the unemployment rate increases above its original level, but not by much. At $a = 0.94$ the unemployment rate increases by 0.6 percentage point from 12.9% to 13.5%. UI has little influence on unemployment in a slump because the macroelasticity is very low, as suggested by the high elasticity wedge in slumps: for $a = 0.94$, $\varepsilon^M = 0.05$ whereas $\varepsilon^m = 0.28$. In booms, the optimal replacement rate is below its original level so the unemployment rate falls below its original level. At $a = 1.06$ the unemployment rate falls by 0.2 percentage point from 4.6% to 4.4%. The macroelasticity is much larger in a boom, as suggested by the low elasticity wedge in booms: for $a = 1.06$,

[25] The elasticity is obtained by looking at a small change in technology around the average. When $a = 1$, $u = 6.56\%$ and when $a = 0.99$, $u = 7.18\%$ so the elasticity is $(1/6.56) \cdot (7.18 - 6.56)/(1 - 0.99) = 9.5$. Michaillat [2012] shows that the elasticity of unemployment with respect to technology over the 1964–2009 period in the US is 4.2.

$\varepsilon^M = 0.27$ whereas $\varepsilon^m = 0.31$.

To conclude, we would like to repeat two points. First, diminishing marginal returns to labor are not necessary to obtain a countercyclical optimal UI. The necessary ingredients to obtain this result are a downward-sloping labor demand and wages that do not respond much to UI. Second, technology shocks are not necessary to obtain large unemployment fluctuations. In fact, when we simulate the aggregate-demand model calibrated in the same fashion as the job-rationing model, we find the same quantitive results as those displayed in Figure 6. In the aggregate-demand model, marginal returns to labor are constant and the business-cycle shocks are monetary shocks; yet, unemployment is subject to large fluctuations and optimal UI is sharply countercyclical.

7 Conclusion

In this paper we propose a macroeconomic theory of optimal UI. We show that the optimal replacement rate of UI is the sum of a conventional Baily-Chetty term, which captures the trade-off between insurance and job-search incentives, and a correction term, which is positive if UI brings the labor market tightness closer to efficiency. We develop two empirical criteria: one to determine whether the labor market tightness is inefficiently low or inefficiently high, and another to determine whether UI raises or lowers tightness. Recent empirical evidence suggests that UI raises tightness. We also provide evidence suggesting that in the US, the labor market tightness is inefficiently low in slumps and inefficiently high in booms.

In our model optimal UI varies over the business cycle because the labor market tightness is subject to inefficient fluctuations over the business cycle and UI influences tightness. Additional mechanisms may also justify adjusting UI over the business cycle. For example, unemployed workers may be more likely to exhaust their savings or be less able to borrow in slumps; in that case, it would be desirable to provide even more UI in slumps, when the consumption-smoothing value of UI is higher. In our formula, the higher consumption-smoothing value of UI would appear as a higher Baily-Chetty term. This mechanism and others could be incorporated into our model to quantify precisely the fluctuations of optimal UI over the business cycle.

In principle, the methodology developed in this paper could be applied to the optimal design of other public policies. We conjecture that a policy maximizing welfare in an economy with inefficient tightness obeys the same general rule as the one derived in this paper for UI. The optimal policy is the sum of the optimal policy when the economy is efficient plus a correction term when the economy is inefficient. If a marginal increase of the policy increases tightness, the correction is positive when tightness is inefficiently low and negative when tightness is inefficiently high. The methodology could be applied to the provision of public good. In the model of Michaillat and Saez [2013], public good provision stimulates aggregate demand and increases tightness. As a result, the government should provide more public good than in the Samuelson [1954] rule if and only if tightness is inefficiently low. The methodology could also be applied to income taxation. In the model of Michaillat and Saez [2013], if high-income earners have a lower propensity to consume than low-income earners, transfers from high-income earners to low-income earners stimulate aggregate demand and increase tightness. As a result, the top income tax rate should be higher than in the Mirrlees [1971] optimal top income tax formula if and only if tightness is inefficiently low. Our methodology could maybe help bridge the gap between the analysis of tax, social-insurance, and public-good policies in public economics and the analysis of stabilization policies in macroeconomics.[26]

References

Baily, Martin N. 1978. "Some Aspects of Optimal Unemployment Insurance." *Journal of Public Economics*, 10(3): 379–402.

Blanchard, Olivier J., and Peter Diamond. 1989. "The Beveridge Curve." *Brookings Papers on Economic Activity*, 20(1): 1–76.

Burgess, Simon, and Stefan Profit. 2001. "Externalities in the Matching of Workers and Firms in Britain." *Labour Economics*, 8(3): 313–333.

Cahuc, Pierre, and Etienne Lehmann. 2000. "Should Unemployment Benefits Decrease with the Unemployment Spell?" *Journal of Public Economics*, 77(1): 135–153.

Card, David, Raj Chetty, and Andrea Weber. 2007. "Cash-On-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market." *Quarterly Journal of Economics*, 122(4): 1511–1560.

Chetty, Raj. 2006*a*. "A General Formula for the Optimal Level of Social Insurance." *Journal of Public Economics*, 90(10-11): 1879–1901.

Chetty, Raj. 2006*b*. "A New Method of Estimating Risk Aversion." *American Economic Review*, 96(5): 1821–1834.

[26]This agenda is closely related in spirit to the conceptual framework proposed by Farhi and Werning [2013] to study optimal macroeconomic policies in the presence of price rigidities. They obtain the same decomposition of optimal policies into standard policies plus a correction term. However, their correction term arises not because of matching frictions but because of price rigidities.

Chetty, Raj. 2008. "Moral Hazard versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy*, 116(2): 173–234.

Chodorow-Reich, Gabriel, and Loukas Karabarbounis. 2014. "The Cyclicality of the Opportunity Cost of Employment." http://scholar.harvard.edu/files/chodorow-reich/files/opportunity_cost_employment.pdf.

Clark, Andrew E., and Andrew J. Oswald. 1994. "Unhappiness and Unemployment." *The Economic Journal*, 104(424): 648–659.

Crepon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do Labor Market Policies Have Displacement Effect? Evidence from a Clustered Randomized Experiment." *Quarterly Journal of Economics*, 128(2): 531–580.

Diamond, Peter A. 1982. "Wage Determination and Efficiency in Search Equilibrium." *Review of Economic Studies*, 49(2): 217–227.

Farhi, Emmanuel, and Iván Werning. 2013. "A Theory of Macroprudential Policies in the Presence of Nominal Rigidities." National Bureau of Economic Research (NBER) Working Paper 19313.

Ferracci, Marc, Grégory Jolivet, and Gérard J. van den Berg. 2010. "Treatment Evaluation in the Case of Interactions within Markets." Institute for the Study of Labor (IZA) Discussion Paper 4700.

Fredriksson, Peter, and Bertil Holmlund. 2001. "Optimal Unemployment Insurance in Search Equilibrium." *Journal of Labor Economics*, 19(2): 370–399.

Gautier, Pieter, Paul Muller, Bas van der Klaauw, Michael Rosholm, and Michael Svarer. 2012. "Estimating Equilibrium Effects of Job Search Assistance." Institute for the Study of Labor (IZA) Discussion Paper 6748.

Gertler, Mark, and Antonella Trigari. 2009. "Unemployment Fluctuations with Staggered Nash Wage Bargaining." *Journal of Political Economy*, 117(1): 38–86.

Gruber, Jonathan. 1997. "The Consumption Smoothing Benefits of Unemployment Insurance." *American Economic Review*, 87(1): 192–205.

Haefke, Christian, Marcus Sonntag, and Thijs van Rens. 2008. "Wage Rigidity and Job Creation." Institute for the Study of Labor (IZA) Discussion Paper 3714.

Hall, Robert E. 2005*a*. " Employment Fluctuations with Equilibrium Wage Stickiness." *American Economic Review*, 95(1): 50–65.

Hall, Robert E. 2005*b*. "Employment Efficiency and Sticky Wages: Evidence from Flows in the Labor Market." *Review of Economics and Statistics*, 87(3): 397–407.

Hawton, Keith, and Stephen Platt. 2000. "Suicidal Behaviour and the Labour Market." In *The International Handbook of Suicide and Attempted Suicide*. , ed. Keith Hawton and Kees van Heeringen, 309–384. West Sussex, England:John Wiley & Sons, Ltd.

Hosios, Arthur J. 1990. "On the Efficiency of Matching and Related Models of Search and Unemployment." *Review of Economic Studies*, 57(2): 279–298.

Howitt, Peter, and R. Preston McAfee. 1987. "Costly Search and Recruiting." *International Economic Review*, 28(1): 89–107.

Jung, Philip, and Keith Kuester. 2014. "Optimal Labor-Market Policy In Recessions." *American Economic Journal: Macroeconomics*.

Krueger, Alan B., and Bruce D. Meyer. 2002. "Labor Supply Effects of Social Insurance." In *Handbook of Public Economics*. Vol. 4, , ed. Alan J. Auerbach and Martin Feldstein, 2327–2392. Elsevier.

Lalive, Rafael, Camille Landais, and Josef Zweimüller. 2013. "Market Externalities of Large Unemployment Insurance Extension Programs." econ.lse.ac.uk/staff/clandais/cgi-bin/Articles/Austria.pdf.

Landais, Camille. 2014. "Assessing the Welfare Effects of Unemployment Benefits Using the Regression

Kink Design." *American Economic Journal: Economic Policy*.

Lehmann, Etienne, and Bruno Van Der Linden. 2007. "On the Optimality of Search Matching Equilibrium When Workers Are Risk Averse." *Journal of Public Economic Theory*, 9(5): 867–884.

Mankiw, N. Gregory, and Matthew Weinzierl. 2011. "An Exploration of Optimal Stabilization Policy." *Brookings Papers on Economic Activity*, 42(1): 209–272.

Marinescu, Ioana. 2014. "The General Equilibrium Impacts of Unemployment Insurance: Evidence from a Large Online Job Board." http://www.marinescu.eu/Marinescu_UI_2014.pdf.

Meyer, Bruce. 1990. "Unemployment Insurance and Unemployment Spells." *Econometrica*, 58(4): 757–782.

Michaillat, Pascal. 2012. "Do Matching Frictions Explain Unemployment? Not in Bad Times." *American Economic Review*, 102(4): 1721–1750.

Michaillat, Pascal. 2014. "A Theory of Countercyclical Government Multiplier." *American Economic Journal: Macroeconomics*, 6(1): 190–217.

Michaillat, Pascal, and Emmanuel Saez. 2013. "Aggregate Demand, Idle Time, and Unemployment." National Bureau of Economic Research (NBER) Working Paper 18826.

Michaillat, Pascal, and Emmanuel Saez. 2014. "An Economical Business-Cycle Model." National Bureau of Economic Research (NBER) Working Paper 19777.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, 38(2): 175–208.

Moen, Espen R. 1997. "Competitive Search Equilibrium." *Journal of Political Economy*, 105(2): 385–411.

Pavoni, Nicola, and Giovanni L. Violante. 2007. "Optimal Welfare-To-Work Programs." *Review of Economic Studies*, 74(1): 283–318.

Petrongolo, Barbara, and Christopher A. Pissarides. 2001. "Looking into the Black Box: A Survey of the Matching Function." *Journal of Economic Literature*, 39(2): 390–431.

Pissarides, Christopher A. 1985. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages." *American Economic Review*, 75(4): 676–690.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory.* . 2nd ed., Cambridge, MA:MIT Press.

Pissarides, Christopher A. 2009. "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica*, 77(5): 1339–1369.

Rogerson, Richard, Robert Shimer, and Randall Wright. 2005. "Search-Theoretic Models of the Labor Market: A Survey." *Journal of Economic Literature*, 43(4): 959–988.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *Review of Economics and Statistics*, 36(4): 387–389.

Schmieder, Johannes F., Till M. von Wachter, and Stefan Bender. 2013. "The Causal Effect of Unemployment Duration on Wages: Evidence from Unemployment Insurance Extensions." National Bureau of Economic Research (NBER) Working Paper 19772.

Shimer, Robert. 2004. "Search Intensity." http://sites.google.com/site/robertshimer/intensity.pdf.

Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*, 95(1): 25–49.

Sullivan, Daniel G., and Till M. von Wachter. 2009. "Job Displacement and Mortality: An Analysis using Administrative Data." *Quarterly Journal of Economics*, 124(3): 1265–1306.

Villena Roldan, Benjamin. 2010. "Aggregate Implications of Employer Search and Recruiting Selection." Center for Applied Economics, University of Chile Working Paper 271.

Appendix A: Calibration

This appendix provides a detailed description of the calibration of the job-rationing model simulated in Section 6.4. Note that all the derivations pertain to the steady state of the dynamic model without discounting. We first present the average values that we target in the calibration—some are normalizations and some are based on empirical evidence for the US. We then explain how we obtain the calibrated values for the parameters.

Normalizations. We normalize the average technology to $a = 1$. We normalize the average search effort to $e = 1$ and the disutility of effort such that $k(e = 1) = 0$. This normalization implies that on average, the costs of search while unemployed are of same magnitude as the costs of work while employed (which are not modeled).

Average Values of Labor Market Variables. We focus on the December 2000–February 2014 period. We compute the average value of key labor market variables in the data constructed by the BLS from the Job Openings and Labor Turnover Survey (JOLTS) and from the Current Population Survey (CPS). We take the average of the seasonally adjusted monthly total separation rate in all nonfarm industries from the JOLTS; we find that this average job-destruction rate is $s = 3.5\%$. We take the average of the seasonally adjusted monthly unemployment rate from the CPS; we find that this average unemployment rate is $u = 6.6\%$. We take the average of the seasonally adjusted monthly vacancy level in all nonfarm industries from the JOLTS, and divide it by the average of the seasonally adjusted monthly unemployment level from the CPS; we find that this average labor market tightness is $\theta = 0.37$ (the average effort is normalized to 1).

Average Values of the Recruiter-Producer Ratio. We take the average over the December 2000–February 2014 period of the recruiter-producer ratio series constructed in Section 6.2 and plotted in Figure 5(a). We find that the average recruiter-producer ratio is $\tau(\theta) = 2.2\%$.

Average Value of the Replacement Rate. UI benefits replace between 50% and 70% of the pre-tax earnings of a worker [Pavoni and Violante, 2007]. Following Chetty [2008] we set the benefit rate to 50%. Since earnings are subject to a 7.65% payroll tax, we set the average replacement rate to $R = 0.5 + 0.0765 = 58\%$.

Calibrating Utility of Consumption. We assume that the utility of consumption is log: $v(c) = \ln(c)$. This assumption implies a coefficient of relative risk aversion $\rho = 1$, consistent with labor supply behavior [Chetty, 2006*b*]. This coefficient of relative risk aversion is maybe on the low side of available estimates, but using log utility simplifies other part of the calibration. Naturally, the higher risk aversion, the more generous optimal UI.

Measuring ε^m. Using (12), we can translate the estimate $\varepsilon^m_R = 0.5$ obtained in the literature into an estimate of ε^m. With log utility, and since $c^e - c^u = (1-R) \cdot w$, we have $\Delta v / (v'(c^u) \cdot R \cdot w) = (1-R) \cdot \ln(c^e/c^u) / [R \cdot (c^e/c^u - 1)]$.

The next step is to express c^e/c^u as a function of known quantities. The production function is $y(n) = n^\alpha$ so the firm's optimality condition, given by (2), implies that $y(n) = w \cdot l/\alpha$. We can therefore rewrite the government's budget constraint, given by (4), as $c^e - u \cdot (c^e - c^u) = w \cdot (1 - u)/\alpha$, where $u = 1 - l$ is the unemployment rate. We divide this budget constraint by $\Delta c = c^e - c^u$. We get $(c^e/c^u)/(c^e/c^u - 1) = u + (1-u)/[(1-R)\cdot\alpha]$. After some algebra we obtain

$$\frac{c^e}{c^u} = \frac{1 + \alpha \cdot (1-R) \cdot u/(1-u)}{1 - \alpha \cdot (1-R)}, \tag{A1}$$

On average, $u = 6.6\%$ and $R = 58\%$. But α is unknown for the moment. However, for a broad range of α, the value for ε^m implied by $\varepsilon^m_R = 0.5$ does not change much. With $\alpha = 1$, $c^e/c^u = 1.79$, and $\varepsilon^m = 0.27$. With $\alpha = 0.66$, $c^e/c^u = 1.42$, and $\varepsilon^m = 0.31$. We therefore take $\varepsilon^m = 0.30$ as a calibration target. This is the value of ε^m implied by $\varepsilon^m_R = 0.5$ and $\alpha = 0.75$ which is the value we use in our simulation.

Linking ε^f to ε^m. Let $\varepsilon^e_\Delta \equiv (\Delta v/e)\cdot(\partial e^s/\partial \Delta v)$, $\kappa \equiv (e/k'(e))\cdot k''(e)$, and $L(x) \equiv x/(s+x)$. The elasticity of $L(x)$ is $1 - L(x)$. The effort supply $e^s(f, \Delta v)$ satisfies $k'(e^s) = (L(e^s \cdot f)/e^s)\cdot(\Delta v + k(e^s))$. Differentiating this condition with respect to Δv yields

$$\kappa \cdot \varepsilon^e_\Delta = (1-l)\cdot\varepsilon^e_\Delta - \varepsilon^e_\Delta + \frac{\Delta v}{\Delta v + k(e)} + \varepsilon^e_\Delta \cdot \frac{e \cdot k'(e)}{\Delta v + k(e)}.$$

In equilibrium, $(e \cdot k'(e))/(\Delta v + k(e)) = l$. Therefore, $\varepsilon^e_\Delta = (1/\kappa)\cdot \Delta v/(\Delta v + k(e))$. Since the labor supply satisfies $l^s(f, \Delta v) = L(e^s(f,\Delta v)\cdot f)$, the elasticity of $l^s(\theta, \Delta v)$ with respect to Δv is $(1-l)\cdot\varepsilon^e_\Delta$. By definition, ε^m is $l/(1-l)$ times the elasticity of $l^s(\theta,\Delta v)$ with respect to Δv. Thus,

$$\varepsilon^m = \frac{1-u}{\kappa}\cdot\frac{\Delta v}{\Delta v + k(e)}. \tag{A2}$$

Similarly, differentiating the effort supply condition with respect to f yields

$$\kappa \cdot \varepsilon^f = (1-l)\cdot(\varepsilon^f + 1) - \varepsilon^f + \varepsilon^f \cdot \frac{e \cdot k'(e)}{\Delta v + k(e)},$$

which implies that

$$\varepsilon^f = \frac{u}{\kappa}.$$

Combining this equation with (A2), we find that when $e = 1$ and thus $k(e) = 0$,

$$\varepsilon^f = \frac{u}{1-u}\cdot\varepsilon^m.$$

With $\varepsilon^m = 0.3$ and $u = 6.6\%$, we get $\varepsilon^f = 0.021$.

Calibrating Matching Parameters. We use a Cobb-Douglas matching function $m(e \cdot u, o) = \omega_m \cdot (e \cdot u)^\eta \cdot o^{1-\eta}$. We set $\eta = 0.7$, in the range of available estimates reported by Petrongolo and Pissarides [2001]. To calibrate ω_m, we exploit the steady-state relationship $u \cdot e \cdot f(\theta) = s \cdot (1-u)$, which implies $\omega_m = s \cdot \theta^{\eta - 1}\cdot(1-u)/(u \cdot e)$. With $s = 3.5\%$, $u = 6.6\%$, $e = 1$, and $\theta = 0.37$, we get $\omega_m = 0.67$.

To calibrate r we exploit the steady-state relationship $\tau(\theta) = r \cdot s/[o_m \cdot \theta^{-\eta} - r \cdot s]$, which implies $r = o_m \cdot \theta^{-\eta} \cdot \tau(\theta)/[s \cdot (1+\tau(\theta))]$. With $\omega_m = 0.67$, $s = 3.5\%$, $\theta = 0.37$, and $\tau(\theta) = 2.2\%$, we obtain $r = 0.84$.

Calibrating α and Related Quantities. As explained in Section 6.4, we set $\alpha = 0.75$. Using (A1), $u = 6.6\%$, $R = 58\%$ and $\alpha = 0.75$, we obtain $c^e/c^u = 1.50$ and $\Delta v = \ln(c^e/c^u) = 0.41$. With log utility and $k(e) = 0$, we have $(\Delta v + k(e))/(\phi \cdot w) = (1-R) \cdot \ln(c^e/c^u) \cdot (u + (1-u) \cdot c^e/c^u)/(c^e/c^u - 1)$. With $R = 58\%$, $u = 6.6\%$, and $c^e/c^u = 1.50$, we obtain $(\Delta v + k(e))/(\phi \cdot w) = 0.51$.

Calibrating Disutility of Search. We use a disutility of search $k(e) = \omega_k \cdot e^{\kappa+1}/(\kappa+1) - \omega_k/(\kappa+1)$. To calibrate κ, we use (A2). On average, $u = 6.6\%$, $k(e) = 0$, and ε^m=0.3 so $\kappa = 3.1$. To calibrate ω_k, we exploit the steady-state relationship $k'(e) = [(1-u)/e] \cdot (\Delta v + k(e))$. This implies $\omega_k = (1-u) \cdot \Delta v$ when $e = 1$. With $u = 6.6\%$ and $\Delta v = 0.41$, we obtain $\omega_k = 0.38$.

Calibrating the Wage Schedule. To calibrate ω, we exploit the steady-state relationship $\alpha \cdot n^{\alpha-1} = \omega \cdot a^{\gamma-1} \cdot (1+\tau(\theta))$. With $a = 1$, $\tau(\theta) = 2.2\%$, $\alpha = 0.75$, and $n = (1-u)/(1+\tau(\theta)) = 0.915$, we obtain $\omega = 0.75$.

We calibrate γ from microeconometric estimates of the elasticity for wages in newly created jobs—the elasticity that matters for job creation [Pissarides, 2009]. In panel data following production and supervisory workers from 1984 to 2006, Haefke, Sonntag and van Rens [2008] find that the elasticity of new hires' earnings with respect to productivity is 0.7 (Table 6, panel A, column 4). If the composition of the jobs accepted by workers improves in expansion, 0.7 is an upper bound on the elasticity of wages in newly created jobs [Gertler and Trigari, 2009]. A lower bound is the elasticity of wages in existing jobs, estimated between 0.1 and 0.45 [Pissarides, 2009]. Hence we set $\gamma = 0.5$, in the range of plausible values.

Solving the Baily-Chetty Formula. With log utility the Baily-Chetty formula, given by (21), can be written as

$$\frac{R}{1-R} = (1-u) \cdot \frac{\Delta v(R)}{\varepsilon^m}$$

where

$$\Delta v(R) = \ln\left(\frac{c^e}{c^u}\right) = \ln\left(\frac{1+\alpha \cdot (1-R) \cdot u/(1-u)}{1-\alpha \cdot (1-R)}\right).$$

The expression from $\Delta v(R)$ comes from (A1). Setting $\varepsilon^m = 0.3$, $\alpha = 0.75$, and $u = 6.6\%$ and solving the Baily-Chetty formula for R, we obtain $R = 58\%$. Therefore, the average replacement rate in the US exactly satisfies the Baily-Chetty formula.

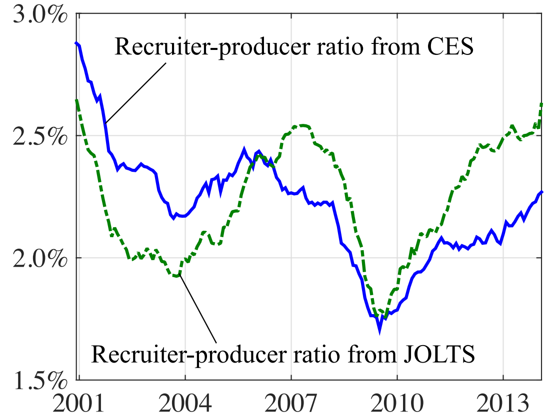


Figure 7: The recruiter-producer ratios constructed from CES data and from JOLTS

Notes: The time period is December 2000–February 2014. The recruiter-producer ratio depicted by the solid blue line is constructed from the number of workers in the recruiting industry computed by the BLS from CES data. The construction of this ratio is described in Section 6.2. The recruiter-producer ratio depicted by the dashed green line is constructed from the numbers of vacancies and new hires computed by the BLS from JOLTS. The construction of this ratio is described in Appendix B. Both time series are scaled to match the evidence on recruiting costs provided by Villena Roldan [2010]. This latter time series is smoothed using a 5-point moving average to remove monthly volatility.

Appendix B: Another Measure of the Recruiter-Producer Ratio

A limitation of the measure of the recruiter-producer ratio proposed in Section 6.2 and plotted in Figure 5(a) is that it is only valid as long as the share of recruitment done through recruiting firms is stable over the business cycle. In this appendix, we propose an alternative measure of the recruiter-producer ratio that does not rely on this assumption. This alternative measure is available from December 2000 to February 2014.

Given the nature of the matching process in a dynamic environment and assuming that the cost of posting a vacancy is constant over time, the recruiter-producer ratio satisfies

$$\tau_t = \frac{s \cdot r}{q_t - s \cdot r},$$

where q_t is the rate at which vacancies are filled, s is the job-destruction rate, and r is the flow cost of posting a vacancy. This expression is derived in Section 6.1. We measure q_t in the data constructed by the BLS from the JOLTS. Our measure is $q_t = h_t/v_t$, where v_t is the seasonally adjusted monthly vacancy level in all nonfarm industries, and h_t is the seasonally adjusted monthly number of hires in all nonfarm industries. Next, we set $s = 3.5\%$ as explained in appendix A. Last, we set $r = 0.81$ such that the average value of τ_t is 2.2%, the same as the average value between December 2000 and February 2014 of the original recruiter-producer ratio in Section 6.2.

Figure 7 plots the two measures of the recruiter-producer ratio from December 2000 to February 2014. Over this period, we find that the two measures behave very similarly. The correlation between the two series is 0.43.