Search Models of the Business Cycle: Towards a New Synthesis*

Patrick J. Kehoe† Pierlauro Lopez‡ Virgiliu Midrigan§ Elena Pastorino¶

March 2026

Abstract

First-generation Diamond-Mortensen-Pissarides (DMP) models explain unemployment volatility through differential productivity between market and non-market activities—the idea that productivity shocks raise the relative payoff of market work. Once the opportunity cost of employment and non-productive activities moves cyclically, however, this mechanism collapses. Hall (2017) therefore proposed a second generation of DMP models in which discount-rate fluctuations, rather than differential productivity, amplify shocks to match observed labor market volatility. Despite its conceptual appeal, Hall's implementation faces three serious critiques: it requires implausibly volatile risk-free rates, generates counterfactually acyclical allocative wages, and relies on contracts that violate firm participation constraints. We show that discount-rate amplification can be disciplined to overcome these challenges. Replacing Hall's exogenous discount factor with a Campbell-Cochrane habit specification ties discount-rate movements to productivity-driven consumption risk, restores plausible asset pricing and wage cyclicality, and yields realistic unemployment volatility. This framework also clarifies the remaining quantitative constraint: absent additional match-surplus dynamics, matching unemployment volatility requires implausibly long negotiation horizons, implying extreme bargaining inefficiencies and unemployment-insurance elasticities far outside the documented empirical range. Introducing on-the-job human capital accumulation removes this dependence on bargaining delay. The resulting synthesis matches unemployment volatility, asset prices, wage cyclicality, and micro evidence on wage growth and unemployment duration elasticities.

Keywords: Shimer puzzle, User cost of labor, Alternating offer bargaining, Fundamental surplus, Hosios condition, Unemployment volatility.

JEL classification: E24, E32, J36, J64.

*The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

†Stanford University and Federal Reserve Bank of Minneapolis.

‡Federal Reserve Bank of Cleveland.

§New York University.

¶Hoover Institution and Stanford University.

In a masterful piece, Ljungqvist and Sargent (2017) synthesize the first generation of Diamond–Mortensen–Pissarides (DMP) models and provide a summary statistic that captures when these models can generate the observed volatility in unemployment. The *differential productivity* mechanism underlying them operates through changes in the payoff to market versus non-market activities as aggregate productivity fluctuates. As Shimer (2005, p. 25) puts it, "a*n increase in labor productivity relative to the value of nonmarket activity and to the cost of advertising a job vacancy makes unemployment relatively expensive and vacancies relatively cheap. The market substitutes toward vacancies.*" Intuitively, in expansions, higher market productivity draws workers into employment and induces firms to post more vacancies.

The differential productivity mechanism of these first-generation models have been challenged both empirically and conceptually. First, Chodorow-Reich and Karabarbounis (2016) showed that the flow opportunity cost of moving from unemployment to employment, primarily the forgone consumption value of nonworking time, is strongly procyclical. In contrast, a key feature of these first-generation models is that this flow opportunity cost is *acyclical*. Second, Shimer (2010) argued that since recruiting workers or bargaining over wages takes time away from production, the cost of doing so is proportional to the opportunity cost of a worker's time in production. We begin by showing that if we integrate these two ideas on the cyclicality of costs into either the canonical matching model or the Hall and Milgrom (2008) version of it with alternating-offer bargaining, as reviewed in Ljungqvist and Sargent (2017), then unemployment does not move *at all* in response to a change in productivity. This simple observation helps explain why these first-generation models fail to account for the observed fluctuations in unemployment.

The pioneering work of Hall (2017), henceforth *Hall*, launched a second generation of search models. Hall ensured that his model is consistent with both the insights of Chodorow-Reich and Karabarbounis (2016), by assuming that the flow payoff of unemployed workers is proportional to productivity, and of Shimer (2010), by assuming that vacancy-posting costs and the firm costs of bargaining are proportional to productivity. By doing so, Hall entirely eliminated the standard differential productivity mechanism and replaced it with a new one in which fluctuations in the stochastic discount factor lead to sizable fluctuations in unemployment. The motivation for doing so is the comovement between the unemployment rate and stock market prices over the cycle. In particular, unemployment is high when the stock market is low, so Hall's idea is that fluctuations in the stochastic discount factor can simultaneously lead to a drop in the stock market and an increase in unemployment. In Hall's framework, a newly hired worker generates a risky stream of firm cash flows analogous to dividends. Hence, when some exogenous event causes the

discount rates used to price risky streams to rise, the value of investing in job creation falls. Through this mechanism, the rise in the risk premium generates both a fall in the stock market and a fall in firm hiring and, hence, a rise in unemployment.

A second feature of this work is that Hall replaces Nash bargaining with alternating-offer bargaining introduced in Hall and Milgrom (2017). We are sympathetic to Hall and Milgrom (2017)'s view that the standard Nash bargaining protocol is not credible. These authors argue that "once a qualified worker meets an employer, a threat to walk away, permanently terminating the bargain, is not credible. The bargainers have a joint surplus, arising from search frictions, that glues them together." (Hall and Milgrom 2017, p. 1653) They argue that the alternating-offer bargaining protocol of Binmore, Rubinstein, and Wolinsky (1986) is built on more realistic threats during bargaining, namely, threats to extend bargaining rather than to terminate it.[1] With this bargaining protocol, the fundamental mechanism driving large increases in unemployment in downturns is that workers' bargaining power increases *greatly* and firms are reluctant to hire them. In particular, the key force, which drives almost all the fluctuations in unemployment, is that a worker's share of match surplus significantly increases during downturns. Intuitively, under alternating-offer bargaining, the cost to workers of turning down a job offer and continuing to bargain is that they lose the opportunity to search for a job and find a new match next period. When productivity is low, this opportunity cost is also low, and, hence, workers are more willing to reject offers. Hence, it is as if workers become more patient in bargaining during downturns and thus they obtain a greater share of the surplus.

Hall's seminal idea, however, has been critiqued in its execution on several grounds. One critique, the Borovička and Borovičková critique, is that the model's stochastic discount factor "*implies that most of the time-variation in discount rates is manifested as time-variation in the mean, as opposed to the dispersion, of the stochastic discount factor, and hence implies quantitatively implausible movements in the risk-free rate.*" (Borovička and Borovičková 2018, p.1) That is, Hall's model is inconsistent with the well-accepted view in the finance literature that asset price fluctuations are not driven by fluctuations in risk-free rates but rather by fluctuations in risk premia.

We extend this critique to demonstrate a fundamental problem with the entire class of discount factors of Hall's form. We show that, regardless of parameter values, Hall's discount factor features no permanent

[1]More generally, we share their view that Nash-bargaining is an axiomatic reduced-form description of wage outcomes, not intended to be a micro-founded noncooperative bargaining game that describes the actual process of bargaining. We also agree with them that since the alternating-offer bargaining setup offers an explicit microfoundation for how bargaining is conducted, it is a more useful environment to study policy interventions than the reduced-form Nash bargaining setup because it is easier to argue that the underlying parameters are structural and so invariant to standard policy interventions.

stochastic component. The key result in Alvarez and Jermann (2005), however, is that such a permanent component is necessary for an asset pricing model to generate reasonable properties for asset prices. Indeed, these authors show that without such a component, long-maturity risk-free bonds are riskier than any other asset in the economy—an implication that is starkly at odds with the data. Moreover, the empirical approach Hall used is related to that in Ross (2015) to recover pricing kernels from limited data, and is therefore subject to the criticism in Borovička, Hansen, and Scheinkman (2016). Briefly, the additional required restrictions imposed by Ross (2015) and, in this case, Hall, force the model to be inconsistent with key asset pricing facts.

A second critique is the Kudlyak critique of DMP models. Kudlyak (2014) builds on Becker's (1962) insight that only the present value of the wages paid by firms to workers over the course of an employment relationship is allocative for employment—an idea echoed by Barro (1977), Hall et al. (1980), and Kydland and Prescott (1980). Kudlyak establishes that for a large class of search models, the appropriate measure of rigidity of the allocative wage is the cyclicality of the user cost of labor, defined as the difference in the present values of wages between two firm-worker matches formed in two consecutive periods. As Kudlyak (2014) estimates, and Basu and House (2016) and Bils, Kudlyak, and Lins (2023) robustly confirm, the user cost of labor is highly procyclical: a 1 percentage point increase in the unemployment rate is associated with an approximately 5% decrease in the user cost of labor. When we compute this statistic for Hall's model, we find instead that a 1 percentage point increase in unemployment is actually associated with 0.1% *increase* in the user cost of labor. Using this metric, wages in Hall move about 1/50th as much as they do in the data and actually increase a little when unemployment increases rather than decrease.

A final critique is about Hall's quantitative implementation. Halls' constructed equilibrium violates the rationality condition on a firm's acceptance of a worker's offer. In Hall's model, the offer the worker makes to the firm implies negative profits for the firm and, hence, would not be accepted because the firm could always break off bargaining and earn zero profits. Once this rationality condition is imposed, the model produces less than a third of the volatility of unemployment observed in the data.

We resurrect Hall's key insight by replacing the problematic discount factor with a habit-based one along the lines of Campbell and Cochrane (1999). Our baseline model overcomes the three core critiques of Hall's model. First, it addresses the Borovička and Borovičková critique by using a discount factor consistent with Alvarez and Jermann's (2005) necessary conditions for empirically plausible asset pricing movements. Indeed, our model reproduces observed patterns of risk-free bonds and stocks. Second, it

passes Kudlyak's litmus test by generating unemployment as volatile as in the data, with user-cost-based wage flexibility that is even slightly greater than in the data. Third, we impose all the individual rationality constraints required by alternating-offer bargaining.

The baseline model, however, suffers from two important shortcomings. First, the mechanism is unappealing in that nearly all of the fluctuations in unemployment directly stem from inefficiencies in the bargaining process, which as we explain, are themselves driven by a parameter that is especially hard to pin down and, on the face of it, seems implausible. Second, Costain and Reiter (2008) argued that existing search models that rely on the differential productivity mechanism have the problem that the same forces that lead unemployment to move a lot for a given size shock over the business cycle invariably imply much greater responses to standard labor market policies than in the data. Even though the mechanism in the baseline model is very different from that in standard DMP models, it has the same problem for a well-studied intervention—increasing the replacement rate of unemployment benefits. Specifically, in the baseline model, the elasticity of the duration of unemployment to an increase is replacement rate is over 25 times as large as typical estimates from the data.

In the baseline model, a negative productivity shock increases the marginal utility of consumption and, hence, makes the risky streams of output produced by workers less attractive. As detailed earlier, such a negative shock makes the opportunity cost of a worker of turning down an offer decrease and, hence, their effective bargaining weight increase. By contrast, in an efficient equilibrium, the worker's share of the match surplus must be constant. By deriving for alternating-offer bargaining models the counterpart of the Hosios condition for Nash bargaining, we show that nearly all of the volatility in unemployment is driven by deviations from the efficient outcomes for the economy. This inefficiency is a type of production inefficiency at the firm level in that if firms were allowed to choose their bargaining protocol each period, then in a deep recession firms would pay almost 20% of their profits to switch bargaining protocols. In the efficient allocations, even for the same size shocks to the marginal utility of consumption that would generate hence, the discount factor, unemployment barely moves.

When we investigate the source of the inefficiency, we find that it stems from one key parameter δ: the probability of an exogenous breakdown in bargaining. Although in equilibrium, job offers are accepted immediately, this probability so low that, absent an agreement, the average duration of bargaining before a breakdown—namely, the *potential duration of bargaining*—is over six years for a job that is supposed to start immediately, which seems intuitively implausible. When the potential duration is long, the costs

of turning down an offer vary greatly over the cycle and workers' effective bargaining shares are very countercyclical. As we decrease the duration of potential bargaining this force weakens, workers' effective bargaining shares becomes less countercylical, and the volatility of unemployment decreases. In the limit in which bargaining breaks down immediately, firms and workers effectively make take-it-or-leave it offers and the model delivers the efficient Nash bargaining outcomes with very small movements in unemployment. Rather than pinning down δ by introspection, we can choose it so as to give an elasticity of unemployment duration as in the data. This exercise implies that δ is about 1 and at this level the model is at the efficient Nash limit and it generates about 1/50th of the volatility of the job-finding rate as in the data.

We show analytically that when δ is low, the discounting of future flows that determine the volatility of workers' bargaining weights is small. In this sense, these flows have long durations. With long duration valuations, a given persistent shock to the discount factor leads to large changes in the accompanying present values. Under this mechanism, the surplus itself is very stable and the volatility in unemployment rates arises from the volatility of the worker's share of the surplus. In the existing literature, an alternative approach to making unemployment volatile is to keep these shares close to constant, but instead make the surplus itself volatile. As shown by Ljungqvist and Sargent (1998), den Haan, Haefke, and Ramey (2005), Kehoe, Lopez, Midrigan, and Pastorino (2023), and Moscarini and Postle-Vinay (2024), a promising way to do so is to add human capital accumulation on the job and depreciation when unemployed. As the degree of human capital accumulation or decumulation increases, the duration of the surplus flows from the match also increases. With long durations, a given exogenous shock that increases the discount rate used to value the risky stream of surplus flows, leads to a large fall in the surplus and, hence, a large decline in the value of the investing in job creation. Hence, unemployment sharply increases. It turns out that a model which combines the forces in Hall (2017) with that of this human capital literature remedies the shortcomings of our baseline model.

We thus propose a *synthesis model* that incorporates the appealing features of Hall (2017) and Hall and Milgrom (2008) but dispenses with their problematic implications—the severe production inefficiencies and the Costain and Reiter (2008) critique. The key addition to our baseline model is with human capital accumulation on the job. We show that even with modest amounts of human capital—at the lower bound of conventional estimates—we can reduce the potential duration of bargaining to under a month, thereby reducing the inefficiencies to trivial amounts.

In our synthesis we produce a locus of pairs of potential durations and growth rates of human capital on

the job that produce the same amount of volatility in the job finding rate. To pin down the most reasonable pairs we use two statistics: the empirically relevant range of measured growth rates of human capital and the elasticity of the unemployment rate with respect to the replacement rate of unemployment benefits. Measured growth rates of human capital on the job range from .5 up to 4. The measured elasticity of the unemployment rate with respect to unemployment benefits range from .36 to .8. For our synthesis model to be consistent with both statistics, the growth rates of human capital must be in the range between 1.5 and 3.5. This synthesis seems a promising way forward for second-generation search models.

1 The Challenge to First Generation Search Models

The insights of Chodorow-Reich and Karabarbounis (2016) and Shimer (2010) imply that search models should satisfy the two key features that Hall (2017) imposed. First, the opportunity cost of employment should be procyclical, with an elasticity close to one. Second, the opportunity cost of a worker who takes time off from production to engage in other activities such as recruiting or haggling should be proportional to current productivity. We show that imposing these two requirements on a large class of models, those considered by Ljungqvist and Sargent (2017), implies that changes in productivity have *no effect* on the job-finding rate.[2] To illustrate how simple and persuasive this challenge is, we next consider the experiment conducted by Ljungqvist and Sargent (2017)—a steady-state increase in productivity—and contrast the resulting movement in unemployment with those under the traditional assumptions in the canonical DMP model and the Hall and Milgrom (2008) model.

1.1 Canonical DMP Model

Consider the *canonical DMP model* in Ljungqvist and Sargent (2017). Consumers are risk neutral with discount factor β. A consumer produces A units of output when employed and b units of output when unemployed. The cost of posting a vacancy is κ, the exogenous separation rate is σ, the worker's bargaining weight is γ, the elasticity of the matching function with respect to unemployment is η, and the job-filling rate for a firm is $\lambda_f(\theta)$ when market tightness is θ. Equation (12) in Ljungqvist and Sargent (2017, p. 2635) shows that the equilibrium value of market tightness is determined by the free-entry condition, which

[2]Note that these results are reminiscent of the result on the neutrality of productivity shocks by Shimer 2010. See also a related intuition by Ljungqvist and Sargent 2017 in footnote 28 of their paper, page 2664.

can be written in our notation as

$$(1) \quad \kappa = (1-\gamma)\lambda_f(\theta)\frac{\beta(A-b)}{1-\beta[1-\sigma-\gamma\theta\lambda_f(\theta)]}.$$

These authors then differentiate this equation to derive that the elasticity of market tightness θ to productivity is

$$\frac{d\log\theta}{d\log A} = \Upsilon\frac{A}{A-b} \text{ where } \Upsilon = 1/\left[\eta + \frac{\beta\gamma(1-\eta)\theta\lambda_f(\theta)}{1-\beta[1-\sigma-\gamma\theta\lambda_f(\theta)}\right]$$

and explain how for any given Υ, the *fundamental surplus fraction*, $(A-b)/A$ plays a critical role in determining the magnitude of this elasticity.

Under Hall's assumptions, both home-produced output and the cost of posting a vacancy are proportional to productivity so that b and κ are replaced by bA and κA, respectively. Observe that scaling home production b by A is consistent with the findings in Chodorow-Reich and Karabarbounis (2016), as discussed above, and scaling κ by A is consistent with the view in Shimer (2010). Under these assumptions, the free-entry condition becomes

$$(2) \quad \kappa A = (1-\gamma)\lambda_f(\theta)\frac{\beta(1-b)A}{1-\beta[1-\sigma-\gamma\theta\lambda_f(\theta)]}.$$

Since A cancels out from both sides of this equation, market tightness θ does not vary with productivity in that $d\log(\theta)/d\log(A) = 0$ and so neither does the job-finding rate.

Proposition 1 (Zero Elasticity of Job-Finding Rate in Basic Matching Model). *In the basic matching model, if home-produced output and the cost of posting a vacancy are proportional to productivity, then the change in steady-state unemployment with respect to a change in steady-state productivity is zero regardless of all other parameters.*

Note that this result holds regardless of the size of the home production parameter b, which plays an important role in the debate between Shimer (2005) and Hagedorn and Manovskii (2008). More generally, this property holds independently of the size of the fundamental surplus, which, instead, is central to the analysis in Ljungqvist and Sargent (2017).

1.2 Alternating Offer Bargaining Model

A similar result applies to alternating-offer-bargaining models. Consider the exposition in Ljungqvist and Sargent (2017) of the model in Hall and Milgrom (2008). In this model, firms and workers make alternating offers and after each unsuccessful bargaining round, the firm incurs a haggling cost of ψ of making a new offer while the worker receives b. There is a probability δ that the job opportunity exogenously expires across bargaining rounds and the worker reenters unemployment. Ljungqvist and Sargent (2017) assume that $\delta = \sigma$ so the probability that a job opportunity expires equals the probability of an exogenous separation between a firm and a worker. Under this assumption, the free-entry condition (equation (36), p. 2648 of Ljungqvist and Sargent 2017) can be rearranged to obtain

$$\kappa = \frac{\beta\lambda_f(\theta)}{1-\beta(1-\sigma)}\left[A - \frac{b+\beta(1-\sigma)(A+\psi)}{1+\beta(1-\sigma)}\right] \tag{3}$$

so that the elasticity of market tightness with respect to productivity is

$$\frac{d\log\theta}{d\log A} = \Upsilon^{AOB}\frac{A}{A-b-\beta(1-\sigma)\psi}, \tag{4}$$

where Υ^{AOB} is a constant. Here the fundamental surplus fraction is the reciprocal of the fraction on the right side of (4), namely $[A-b-\beta(1-\sigma)\psi]/A$.

Now, suppose that we continue to scale b by A and extend Shimer (2010)'s idea that recruiting workers takes a fixed amount of a worker's time to the idea that each round of bargaining absorbs a fixed amount of a worker's time in haggling. Then it is natural to scale both κ and ψ by A, since both parameters reflect the foregone opportunity of producing goods for a worker engaged in either recruiting or bargaining. Hence, (3) becomes

$$\kappa A = \frac{\beta\lambda_f(\theta)}{1-\beta(1-\sigma)}\left[1 - \frac{b+\beta(1-\sigma)(1+\psi)}{1+\beta(1-\sigma)}\right]A. \tag{5}$$

Since A cancels out from both sides of this equation, θ is constant and so market tightness does not move as productivity moves, that is $d\log(\theta)/d\log(A) = 0$. Note that this same result holds even if δ does not equal σ because all value functions are proportional to A.

Proposition 2 (Zero Elasticity of Job-Finding Rate in Alternating Offer Bargaining Model).

In the alternating-offer-bargaining model, if home produced output, the cost of posting a vacancy, and the haggling cost are proportional to productivity, then the change in steady-state unemployment with respect to a change in steady-state productivity is zero regardless of all other parameters.

1.3 Implications

Here we have considered the standard differential productivity across sectors models surveyed by Ljungqvist and Sargent (2017) in their excellent synthesis of the work on the unemployment volatility puzzle. We have shown that once we impose Hall's key assumptions, they lead to exactly zero movements in unemployment in response to a permanent change in productivity. It is easy to show that stochastic versions of these models with empirically relevant specifications of productivity yield very small movements in unemployment.

Given these results, the challenge to the DMP literature is as follows. Construct a model that generates the observed volatility of unemployment in U.S. data for shocks of plausible magnitude that has three features. First, the model is consistent with the observed high correlation between stock prices and unemployment. Second, the model is consistent with the Chodorow-Reich and Karabarbounis (2016) finding that the opportunity cost of employment in the data is procyclical with an elasticity of one. Third, the model is consistent with the Shimer (2010) view that the opportunity cost of workers taking time away from production to spend time on other tasks in the firm, such as vacancy posting (or, more generally, recruiting) or bargaining with workers is proportional to the foregone output that the worker would have produced during the time.

As we have seen existing DMP models do not meet his challenge. Hence, we agree with Hall that a new generation of search models is needed in which a force different from the standard differential productivity mechanism drives variation in unemployment over the cycle. The key idea, explored in Hall (2017, p.306) is that an investment in hiring workers is similar to an investment in physical capital. We know that investment falls greatly when risk rises. Hall's idea is that "*the flow of benefits from a newly hired worker has financial risk comparable to corporate earnings, so the dramatic increase in the market discount applied to earnings that occurred in the crisis implied similarly high discounting of benefit flows accruing to firms from adding workers*". Then this same increase in the market discount will lead firms to decrease their investment in hiring workers. We turn now to examining Hall's economy and the set of issues his quantification implies.

2 Analysis of Hall's Model

In the Appendix we lay out Hall's model in our notation to make it comparable to our baseline model which augments Hall's setup by explicitly modeling consumers and, thus, obtaining a standard endogenous discount factor. Here we briefly discuss how Hall's model works. Basically, the AOB bargaining protocol makes the share of surplus that workers receive rise greatly in recessions. Hence, firms shy away from hiring them in recessions and, thus, unemployment rises substantially. To demonstrate this intutition we rely on a result, shown formally later, that the allocations under the AOB protocol can be decentralized under a standard Nash-bargaining scheme with exogenously varying bargaining weights denoted by γ_t^{AOB} and refer to them as *effective bargaining weights*. In Figure 1 we graph these weights against the unemployment rate in Hall's binned data. We delve into this feature in great detail later in our analysis of the baseline model.

We turn to demonstrating three issues with Hall's analysis. The first is that the model's implications for fluctuations in asset prices are starkly at odds with the data. The second is that the model generates wages that are much stickier than those in the data as measured by the cyclicality of the user cost of labor, discussed below. The third is that the computed contract violates a rationality condition for firms. We then ask is there an easy fix to the model's asset pricing problems and we argue that not within the class of pricing kernels that Hall posits.

2.1 The Three Critiques

We now demonstrate three main issues for Hall's model.

2.1.1 Implications for Asset Prices: the Borovicka and Borovickova Critique

We first show that, as Borovicka and Borovickova (2018) claim in their critique of Hall, the pricing kernel that Hall uses implies that fluctuations in the expected asset returns in the model are almost entirely driven by fluctuations in (real) risk-free rates and almost none by fluctuations in excess returns, whereas in the data the opposite is true. This property of the pricing kernel implies that almost all of the movements in unemployment in the model are driven by counterfactual movements in risk-free rates and essentially none are driven by risk. Moreover, these movements in risk-free rates are an order of magnitude larger than those observed in the data. Finally, we show that if we make the pricing kernel consistent with the observed fluctuations in risk-free rates then the model generates fluctuations in the unemployment rate

less than one-tenth as large as in the data.

In Figures 2a and 2b, we compare the time series for risk-free rates and realized equity premium from the model to the appropriately binned versions from Hall's dataset. Clearly, relative to the data, the fluctuations in risk-free rates are an order of magnitude too large and those of the equity premium are an order of magnitude too small. To quantify this point, we construct a variance decomposition of expected returns on stocks $\mathbb{E}_t R_{t+1}$ using the identity $\mathbb{E}_t R_{t+1} = \mathbb{E}_t(R_{t+1} - R_{ft}) + R_{ft}$, where $\mathbb{E}_t(R_{t+1} - R_{ft})$ is the equity premium and R_{ft} is the risk-free rate. This identity implies a decomposition of the source of variance of movements in the expected return on stocks. In particular, since $var(\mathbb{E}_t R_{t+1}) = cov(\mathbb{E}_t R_{t+1}, \mathbb{E}_t(R_{t+1} - R_{ft})) + cov(\mathbb{E}_t R_{t+1}, R_{ft})$ we have that

$$\frac{cov(\mathbb{E}_t R_{t+1}, \mathbb{E}_t(R_{t+1} - R_{ft}))}{var(\mathbb{E}_t R_{t+1})} + \frac{cov(\mathbb{E}_t R_{t+1}, R_{ft})}{var(\mathbb{E}_t R_{t+1})} = 1. \tag{6}$$

In Table 1 we see that, in the data, the first term in (6), which is the fraction of the variance in the expected return on stocks due to the comovements with the equity premium, accounts for over 93% of the variance in expected stock returns, and the second term, which is the fraction of this variance that due to comovements with the risk-free rate, accounts for less than 7%. In contrast, in Hall's model, comovements with the equity premium, account for essentially zero of theessentially 100% of the variance of expected stock returns is due to the comovements with the risk-free rate . This property is starkly at odds with the data.

Table 1: Key Moments in Hall (2017) and Variants

	Data	Hall (2017)	Risk neutral	Risk neutral	Pure discount	Alternative barg. durations		
		$\delta = .013$	SDF I	SDF II	shocks	$\delta = \sigma = .035$	$\delta = .23$	$\delta = 1$
Potential barg. duration (months)		78	78	78	78	28	3.3	0
$E(R_{ft})$ (% p.a)	1.48	7.68	7.68	1.48	7.71	7.68	7.68	7.68
$\sigma(R_{ft})$ (% p.a)	0.84	5.31	5.31	0.84	5.28	5.31	5.31	5.31
$E(R_{t+1} - R_{ft})$ (% p.a)	6.64	0.41	0.00	0.00	0.37	0.41	0.41	0.41
$\sigma(R_{t+1} - R_{ft})$ (% p.a)	4.95	0.22	0.00	0.00	0.23	0.22	0.22	0.22
$\sigma(\Delta a_t)$ (%)	1.08	0.20	0.20	0.20	0.00	0.20	0.20	0.20
$\sigma(\lambda_{wt})$ (%)	10.2	11.1	11.5	0.96	10.9	2.92	0.49	0.14
$\sigma(u_t)$ (%)	0.84	0.97	1.02	0.09	0.93	0.25	0.04	0.01
Semi-elast. UC labor to unempl.	-5.32	0.10	0.11	-1.09	4.90	-0.73	-5.32	-116
S.d. effective bargaining power $\sigma(\gamma_t)$		5.49	5.69	4.80	5.46	1.40	0.19	0.00

Note: The potential bargaining duration is defined as $\frac{1}{\delta} - 1$. The "Data" column refers to Hall's discretized data, except the semi-elasticity UC labor to unemployment that comes from Bils et al. (2023). Throughout the alternative models, the haggling cost parameter is adjusted to keep the mean unemployment rate constant. The "Pure discount shocks" variant turns off the effect of productivity shocks everywhere but on discount rates.

We turn to showing that nearly all the movements in unemployment in the model are driven by the fluctuations in the risk-free rate. To do so, we consider a risk-neutral pricing kernel, which shuts down all movements in the equity premium, make it consistent with Hall's risk-free rates, and recompute Hall's model with such a kernel. Specifically, we replaces Hall's discount factor $\beta m(s')/m(s)$ what that from a risk neutral agent with a time-varying discount factor $\beta(s)$. We choose $(\beta(s))$ so that its risk free rates, $R_f(s) = 1/\beta(s)$ coincide with the risk free rates $R_t^{Hall}(s)$ in Hall's baseline model[3]

$$(7) \quad \frac{1}{R_f^{Hall}(s)} = \sum_{s'} \beta \pi(s, s') \frac{m(s')}{m(s)},$$

As the column labeled *Risk neutral SDF I* in Table 1 shows, the standard deviation of unemployment in the risk neutral model is 1.02%, which is a bit larger than the 0.97% standard deviation for unemployment in the baseline. The key message is that a risk neutral model with fluctuations in risk-free interest rates set to equal the counterfactually large risk-free rates in Hall's model, gives essentially identical fluctuations to those in Hall.

So far we have set the risk-free rates equal to those in Hall's model. We now consider what happens if instead we set the risk-free rates, $R_f\ (s) = 1/\beta(s)$, in our risk neutral model equal to those in Hall's data, that is, equal to the binned data on risk-free rates in Figure 2a. In Figure 2a we plot the realized time series for the unemployment rate in Hall's model against the unemployment rate implied by a risk-neutral version of it when the primitive shocks are the actual risk-free rates. As the column labeled *Risk neutral SDF II* in Table 1 shows, the resulting standard deviation of unemployment is about 10% of that in Hall's model.

In sum, Table 1 and Figure 2 confirm the critique of Borovicka and Borovickova (2018), namely that the fluctuations in unemployment in Hall's model are driven not by risk but rather by counterfactually large movements in risk-free rates.

2.1.2 Implications for the Stickiness of Wages: the Kudlyak Critique

We turn now to studying the implications of Hall's model for the stickiness of wages. Motivated by the work of Becker (1962), who emphasized that the allocative for employment is the present value of the flow wages paid to a worker over the course of an employment relationship. Kudlyak (2014) proved that for

[3]We keep all parameters fixed at their values in Hall's baseline, except for the haggling cost ψ, which we increase from 0.567 to 0.568 to ensure that the mean unemployment rate is unchanged.

Figure 1: Implied Nash bargaining weights vs unemployment in Hall (2017)

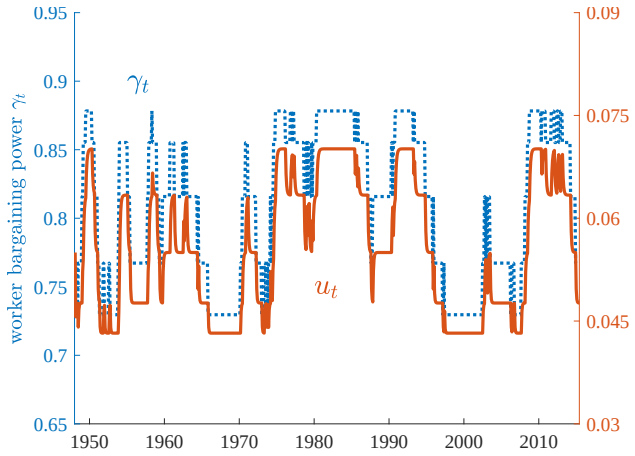
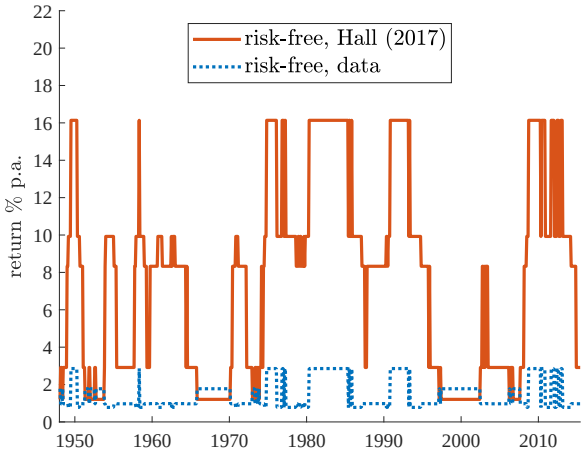
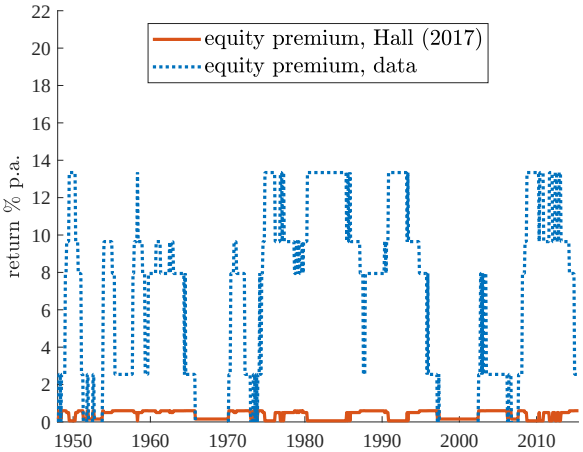


Figure 2: Asset prices in Hall (2017)

(a) *Risk-free rate*



(b) *Equity premium*



a large class of search models, the appropriate allocative wage is the difference in the present value of wages between two matches that start in two consecutive periods, as captured by the user cost of labor. Intuitively, in a search model, hiring a worker is akin to acquiring a long-term input subject to adjustment costs. Thus, by measuring the rental price of the services of a worker potentially employed for several years, the user cost of labor is a more relevant measure of the cost of hiring a worker than the flow wage. Specifically, the *user cost of labor* is defined as the difference between the present value of the cost of hiring a worker at some date t and that of hiring a worker at $t+1$ over the states in which the worker has not separated, and is given by

$$UC_t = W_t - (1-\sigma)\mathbb{E}_t Q_{t+1} W_{t+1}.$$

Intuitively, the user cost of labor measures the shadow wage that would make a firm indifferent between hiring a worker today at that wage, thus creating a match that survives to tomorrow with probability $1-\sigma$, and hiring a worker tomorrow at the present value of wages prevailing then. Importantly, the user cost of labor in t captures not just the flow wage of a new hire in t but also the difference in the present value of wages from $t+1$ on between a worker hired in t, who remains employed in $t+1$, and a worker hired in $t+1$. Hence, the user cost measure incorporates any potential extra cost or benefit of committing in t to a (possibly state-contingent) sequence of wage payments from $t+1$ on, relative to waiting and hiring an identical worker in $t+1$ at the relevant present value of wages in $t+1$. In particular, if recessions are times of scarring in that relative to workers hired in booms, workers hired in recessions not only obtain a lower wage when hired but also in any subsequent period, then the user cost of labor is more cyclical than the flow wage.

Kudlyak (2014), Basu and House (2016), and Bils et al. (2023) construct a measure of the user cost of labor using NLSY data and measure its cyclicality as the semi-elasticity β_K of the user cost of labor with respect to the unemployment rate using the following regression

$$\log UC_t = \beta_0 + \beta_K u_t + x_t'\beta_x + \varepsilon_t, \tag{8}$$

where u_t is the unemployment rate, x_t is a vector of controls, and we refer to β_K as the *Kudlyak statistic*. Kudlyak (2014) estimated $\beta_K = -5.2\%$, Basu and House (2016) refine this estimate to $\beta_K = -5.8\%$, and Bils et al. (2023), after conducting a battery of checks showing that the estimates are robust, ended up

focusing on $\beta_K = -5.3\%$ in their Table 8. For instance, this last estimate implies that a one percentage point increase in the unemployment rate is associated with a 5.3% decrease in the user cost of labor. In this specific sense, wages are not very sticky. The second column of Table 1 shows that the Kudlyak statistic is equal to 0.1 in Hall's baseline, which is less then 1/50th of the size measured by Kudlyak (2014), Basu and House (2016), and Bils et al. (2023) (as well as the opposite sign of that in the data). In this sense, Hall's model has extremely sticky wages relative to those in the data.

It turns out that the stickiness of wages is closely connected to the exogenous breakdown rate of bargaining δ. When this breakdown rate is very low, wages and much stickier than the Kudlyak statistic indicates. If we increase that breakdown rate to be consistent with the Kudlyak statistic, the model no longer produces volatile unemployment. That is, even though in equilibrium firms and workers come to an agreement immediately, the probability that bargaining would have continued if they did not, δ, plays a critical role in determining both the stickiness of wages, as well as the implied volatility in unemployment. To that end, in Table 1 we consider four levels for this breakdown probability δ, namely, Hall's breakdown probability of $\delta = .013$ as well as $\delta = .028, 0.23$,and 1. We find it intuitive to associate the level of δ with its implied mean duration in months, namely $1/\delta$, and refer to this measure as the *potential bargaining duration.*

In Table 1 we see that at Hall's baseline setting with $\delta = .013$, the potential bargaining duration is 6.4 years (77 months) and the corresponding wages are extremely sticky. Next, we note that Hall and Milgrom (2008 p 1664) argue that "Our knowledge of δ is limited to the view that the probability cannot be smaller than the probability σ that a job itself loses its productivity". So in the next column we consider $\delta = \sigma = .028$ and refer to this value as the *minimal breakdown probability.* Moving to this minimal level from Hall's baseline level causes the volatility of unemployment to fall by a factor of 4. Here wages are procylical, as in the data, but the Kudlyak statistic indicates that wages are still over 7 times as sticky as in the data $(-5.3/-.7)$.

We next increase δ to a level that allows the model to reproduce the Kudlyak statistic in the data. Specifically, we set $\delta = .23$, which implies a potential bargaining duration of 4.3 months. Now that the model matches the observed level of wage stickiness, it predicts a volatility of unemployment that is less than 1/20th of that in the data (.04/.84). Finally, when $\delta = 1$, the volatility of unemployment is only 1/100th of that in the data and wages are over 20 times as flexible as those in the data. It is useful to note that, regardless of the breakdown probability, the asset pricing statistics reported in the first 6 rows

in the table are unchanged. In this sense, asset prices by themselves shed no light on the appropriate level of the breakdown probability.

2.1.3 Critiques on Implementation

The first issue with implementation is that the constructed productivity series used in the model is not very connected to actual productivity. Specifically, Hall does not feed the raw productivity data directly into the model. Instead, he constructs a binned version of productivity growth, where the bins are based on a latent state derived from non-productivity indicators, which combines price-dividend ratios and labor market tightness. Because this binning is based on a state that is only weakly correlated with productivity, the resulting transformation removes most of the original volatility. Intuitively, it acts similarly to a regression of productivity growth on the latent state. Specifically, the backed-out productivity growth series using Hall's procedure has a correlation of only 16.7% with actual productivity growth and its volatility is less than one-fifth as large as actual productivity growth. We bring this up because in our exercise below we use actual labor productivity.

The second issue with implementation is that it neglects to impose an important rationality constraint. Specifically, the model is solved without imposing the rationality condition requiring that the wage offered by the worker W_{wt} for a match be less than the present value of output Y_t made by that worker during that match. That is, instead of satisfying $W_{wt} \leq Y_t$, for all five states of s_i, $W_{wt} > Y_t$, which implies that if a firm accepted the offer in any such state it would earn strictly *negative* profits.[4] Clearly this outcome cannot be an equilibrium because the firm has the option of quitting bargaining and earning zero profits. In Figure 3, we plot the unemployment rate in Hall's original model and the unemployment rate once we solved for equilibrium imposing the correct constraint. In this case, as this figure makes clear, the standard deviation of unemployment is much lower—it drops from 0.97% to 0.28%, which is less than a third of the observed volatility of unemployment as in the data.[5]

[4]Mechanically, this violation occured because the equations in the code used to solve the model dropped the maximum in equation describing the worker's offer in the bargaining equilibrium, equation (24) and instead replaced it by

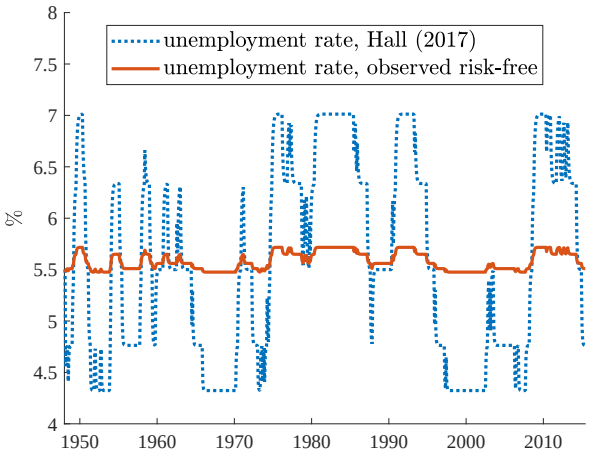
$$Y_t - W_{wt} = (1-\delta)\left[-\psi A_t + \mathbb{E}_t Q_{t+1}(Y_{t+1} - W_{ft+1})\right], \tag{9}$$

which permitted solutions with $Y_t - W_{wt} < 0$. .

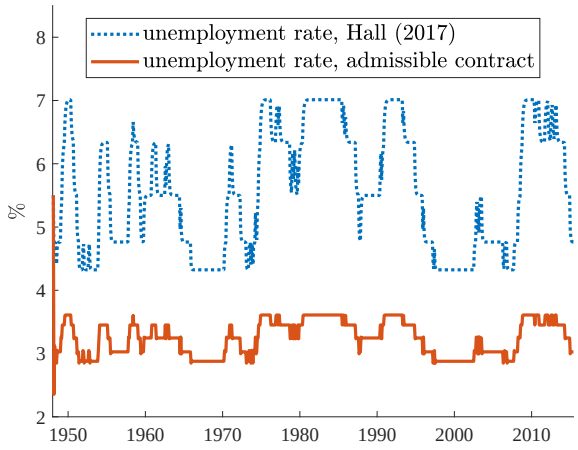
[5]We had a fruitful discussion with Jarda Borovicka on this point.

Figure 3: Unemployment rates in alternative versions of Hall (2017)

(a) *Model matching observed risk-free rate*



(b) *Unemployment rate under admissible contract*



2.2 Is There an Easy Fix to the Asset Pricing Implications?

Hall proposed a class of discount factors and showed an easy way to recover the associated parameters by using data on stock market returns. At first blush, one might think if there are issues with the implications of the discount factor constructed by this method, there are easy ways to extend his approach, say by adding more states and adjusting some parameters, so that the extended model fits the basic features of real rates and stock market returns in the data well. Here we emphasize that doing so is impossible in the class of discount factor models that Hall uses. Specifically, we apply the key proposition of Alvarez and Jermann (2005) to Hall's stochastic discount factor specified and obtain the following result.

Proposition 3. (No Permanent Component in Hall's Discount Factor) *The stochastic discount factor in Hall's economy has no permanent component. Consequently, the limiting long-term bond delivers a higher expected log return than any other asset in the economy.*

This proposition implies that, no matter what are the parameters used, Hall's specification of the stochastic discount factor necessarily implies that the excess return on long maturity bonds is greater than the excess return on any other asset in the economy—including the excess return on stocks, namely the equity premium. This implication that is starkly at odds with the data.

More precisely, Hall's family of stochastic discount factor has the form $Q_{t+1} = \beta m_{t+1}/m_t$ with $m_t = m(s_t)$ and s_t follows a Markov chain such that $\lim_{n\to\infty} E_t(m_{t+n}) = m$ is a constant. To see this, note that the return $R_{\infty,t+1}$ from holding an infinite-duration bond for one period is the inverse of the stochastic

discount factor because

$$R_{\infty,t+1} = \lim_{n\to\infty} \frac{E_{t+1}Q_{t+1,t+n}}{E_tQ_{t,t+n}} = \lim_{n\to\infty} \frac{m_t}{\beta m_{t+1}} \frac{E_{t+1}(m_{t+n})}{E_t(m_{t+n})} = \frac{1}{Q_{t+1}}$$

Therefore, this return is lowest when the marginal utility of consumption is the highest. Specifically, it has the lowest possible correlation with the stochastic discount factor since $corr_t(\log(Q_{t+1}), \log(R_{\infty,t+1})) = -1$. Thus, the long horizon bond is the riskiest asset in the economy and, hence, commands the largest risk premium.

Interestingly, this implication follows from the same assumptions that ensure that the stochastic discount factor can be recovered from observed stock prices, which is a key assumption in Hall's empirical implementation strategy. It turns out that Hall's strategy to recover the discount factor is closely related to that used by Ross (2013) and is thus subject to the critique of Borovicka, Hansen, and Scheinkman (2013) of that strategy, namely that it necessarily recovers one without a permanent component, and, hence, by construction cannot fit the asset pricing data well.

Motivated by that conclusion, we try a more conventional route. We next consider an economy that is virtually identical to Hall's economy except that we replace the problematic reduced-form discount factor with a standard endogenous discount factor derived from explicit preferences which has been shown to have desirable asset pricing properties and explore Hall's insights in it.

3 An Economy with an Endogenous Discount Factor

Here we consider an economy that is nearly identical to Hall's economy except that consumers have a version of Campbell and Cochrane (1999) habit-formation preferences together with a standard specification of productivity shocks. We show that this *baseline model* can overcome all of the problems of Hall's economy that we have discussed. First, it can generate realistic patterns for asset prices. Second, it generates a processes for wages that is consistent with the measures of the cyclicality of the user cost of labor in Kudlyak (2014) and Basu and House (2016). Third, even with actual productitivity growth and admissible contracts, it can generate the observed fluctuations in unemployment rates. In this sense, our baseline model resurrects Hall's seminal idea.

We then explore why our baseline model is a quantitative success. Our main result is that the model's ability to generate volatile unemployment rates is intimately connected to the high degree of inefficiency

of the allocations generated by the alternating-offer-bargaining protocol when there is a very low chance of an exogenous breakdown in bargaining. We establish this finding in several steps. First, we establish an efficiency result for alternating-offer-bargaining: if the probability p that the worker makes the first offer equals the elasticity η of the matching function with respect to unemployment then, as the probability of an exogenous breakdown δ approaches 1, the allocations under alternating-offer-bargaining converge to the efficient outcomes. Second, we show that under our baseline parameters, the efficient allocations have a volatility of unemployment that is 1/50th of that in the data.

3.1 Consumers and Firms

We consider an economy with a large but finite number of families N and firms F. Each family has a measure m_w of workers so that the measure of workers per firm $m_f = Nm_w/F$. We focus on an economy in which m_w and m_f are fixed positive numbers and N and F are large enough so that no individual family or firm views itself as having market power. See Burdett and Judd (1985) for similar assumptions.

We replace Hall's reduced-form discount factor by one generated by a version of the Campbell and Cochrane (1999) economy with their external habits replaced by exogenous habits. We do so in order to eliminate the consumption externality generated by external habits discussed in Ljungqvist and Uhlig (2015) while retaining the desirable asset pricing properties of their specification. With the exogenous habit X_t, a family's utility is given by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{(C_t - X_t)^{1-\alpha}}{1-\alpha}. \tag{10}$$

In a symmetric equilibrium, the consumption of the representative family C_t equals aggregate consumption and we can define the *aggregate surplus consumption ratio* as $S_t = (C_t - X_t)/C_t$ so that aggregate marginal utility is $\beta^t(C_t - X_t)^{-\alpha} = \beta^t C_t^{-\alpha} S_t^{-\alpha}$. As in Campbell and Cochrane (1999), we specify the law of motion for the exogenous habit X_t indirectly by specifying a law of motion for the aggregate surplus consumption ratio S_t. Specifically, we assume that productivity $a_t = \log(A_t)$ follows a random walk with drift g_a, namely,

$$a_{t+1} = g_a + a_t + \varepsilon_{t+1}, \tag{11}$$

and that $s_t = \log(S_t)$ follows an autoregressive process given by

$$(12) \quad s_{t+1} = (1 - \rho_s)\, s + \rho_s s_t + \lambda_a(s_t)\left(\Delta a_{t+1} - \mathbb{E}_t(\Delta a_{t+1})\right),$$

where s denotes the mean of s_t and $\Delta a_{t+1} = a_{t+1} - a_t = \varepsilon_t$. The s_t process acts like an exogenous shock to the discount factor and we refer to it as the *discount factor shock*. The *sensitivity function* $\lambda_a(s_t)$ in (12) is defined as

$$(13) \quad \lambda_a(s_t) = \frac{1}{S}\left[1 - 2\left(s_t - s\right)\right]^{1/2} - 1,$$

when the right side of (13) is nonnegative and otherwise $\lambda_a(s_t) = 0$. As in Campbell and Cochrane (1999), the function $\lambda_a(s_t)$ is chosen so that in a downturn following a productivity shock, risk aversion rises sharply but the risk-free rate does not. The *pricing kernel* for the economy is

$$(14) \quad Q_{t,t+1} = \beta\left(\frac{S_{t+1}}{S_t}\frac{C_{t+1}}{C_t}\right)^{-\alpha}.$$

This kernel determines the intertemporal price of consumption goods and is the discount factor used by individual consumers and firms where we note that S_t is simply an exogenous shock defined by (12), so that, differently from Campbell and Cochrane (1999), there are no externalities. We let $Q_{t,r} = \beta^{r-t}[S_r C_r/(S_t C_t)]^{-\alpha}$ denote the discount factor for period $r \geq t+1$ in units of the period-t consumption good.

The aggregate resource constraint is

$$(15) \quad C_t + \kappa A_t v_t = A_t(1 - u_t) + bA_t u_t,$$

$\kappa A_t v_t$ is the aggregate resources use for posting vacancies, $A_t(1-u_t)$ is aggregate production by the measure $1-u_t$ employed consumers, and $bA_t u_t$ is aggregate home production, where aggregate vacancies $v_t = \theta_t u_{t-1}$ which follows from $\theta_t = v_t/u_{t-1}$.

3.2 Search

The timing of the model is as follows. At the beginning of period t, current productivity A_t is realized. At that time, the unemployed from the end of period $t-1$, namely the measure u_{t-1}, search for new matches and firms post a measure v_t of vacancies. New matches with consumers are formed and each of these

matches immediately begins to produce A_t units of output. At the end of the period t, a fraction $1-\sigma$ of workers keep their jobs and a fraction σ separate and enter the unemployment pool in period $t+1$. During period $t+1$, these unemployed workers produce bA_{t+1} in home production and then search for a job that can only start next period, $t+2$. A worker who was employed in period t and separates at the end of the period must therefore spend period $t+1$ in unemployment.

At the beginning of each period t firms post a measure of vacancies v_t which together with the measure u_{t-1} of unemployed consumers from $t-1$ form a measure $m_t = m(u_{t-1}, v_t)$ of matches, where $m(u,v) = Bu^{\eta}v^{1-\eta}$ with $\eta \in (0,1)$ is a Cobb-Douglas matching function. All matched workers then produce. After matches have occurred there are new entrants into the pool of unemployed for period t namely, the $\sigma(1-u_{t-1})$ employed workers from $t-1$ who separated their firms at the end of the period. Hence, the transition law for unemployed workers is

$$(16)\quad u_t = \sigma(1-u_{t-1}) + (1-\lambda_w(\theta_t))u_{t-1},$$

where $\lambda_w(\theta_t) = m_t/u_{t-1}$ is the job-finding rate of an unemployed worker and $\theta_t = v_t/u_{t-1}$ is market tightness. Given the Cobb-Douglas matching function, the job-finding rate is $\lambda_w(\theta_t) = B\theta_t^{1-\eta}$ and is an increasing function of θ_t.

After a consumer and a firm match, the pair engage in an alternating-offer-bargaining game, which we describe later. The present value of output over the course of a match that begins at t in exogenous state (A_t, s_t) can be written recursively as $Y(A_t, s_t) = A_t + (1-\sigma)\sum_{s_{t+1}} \pi(s_t, s_{t+1})Q(s_t, s_{t+1})Y(g(s_t, s_{t+1})A_t, s_{t+1}))$ which encodes that the worker produces A_t in period t and that with probability $1-\sigma$ the match continues and the state transits to s_{t+1} in period $t+1$ and the worker produces output $A_{t+1} = g(s_t, s_{t+1})A_t$ and with probability σ the match dissolves and no future output is produced. From now on we let subscript t denote dependence of variables on the aggregate state (u_t, A_t, s_t) and use the notation

$$(17)\quad Y_t = A_t + (1-\sigma)\mathbb{E}_t Q_{t+1} Y_{t+1}$$

for the expression in $Y(A_t, s_t)$ and similar expressions.

Wage bargaining results in a wage contract with a present value W_{t+1} for the measure of workers that constitute a family. After a given worker leaves the current match the worker produces at home whenever the worker is subsequently unemployed and earns wages whenever the worker is subsequently employed.

We refer to the expected present value of income after the current match dissolves as the subsequent career value or *post-match value* and denote it by

$$(18) \quad P_t = \mathbb{E}_t Q_{t+1} \left[\sigma U_{t+1} + (1-\sigma) P_{t+1} \right]$$

where the value of unemployment is[6]

$$(19) \quad U_t = bA_t + \mathbb{E}_t Q_{t+1} \left[\lambda_w(\theta_{t+1})(W_{t+1} + P_{t+1}) + (1 - \lambda_w(\theta_{t+1}))U_{t+1} \right]$$

Note that the post-match value in (18) is simply the present discounted value of income in all future dates and states after the current match has finished. Next, note that the value of unemployment in (19) has three terms: current home production when unemployed bA_t; the present value of income after finding a job tomorrow, which occurs with probability $\lambda_w(\theta_{t+1})$, times the sum of the present value of wages in that match, W_{t+1} and the career value P_{t+1} following such a match; and the present value of income if the worker remains unemployed tomorrow, U_{t+1}, times the probability of that event, $1 - \lambda_w(\theta_{t+1})$.

Using the definition of the present value of production and wages, the expected profit for a firm that has just hired a worker is

$$(20) \quad J_t = Y_t - W_t,$$

and free entry for employers implies that the expected profit from posting a vacancy, with vacancy posting cost κA_t is zero, so that

$$(21) \quad \kappa A_t = \lambda_f(\theta_t) J_t$$

where $\lambda_f(\theta_t) = m_t/v_t$ is the job-filling rate of a firm with an open vacancy. With our Cobb-Douglas matching function $\lambda_f(\theta) = B\theta^{-\eta}$, a decreasing function of market tightness θ_t.

[6]There is a typo in Hall's equation for unemployment, which leads to an inconsistency in timing. Specifically, rather than writing U_t as in (19), Hall writes $U_t = zA_t + \mathbb{E}_t Q_{t+1} \left[\lambda_w(\theta_t)(W_{t+1} + P_{t+1}) + (1 - \lambda_w(\theta_t))U_{t+1} \right]$. But this version incorrectly uses the current-period market tightness, θ_t, to compute the probability of finding a job in the next period.

3.3 Alternating Offer Bargaining

We now turn to describing the alternating-offer-bargaining game that occurs upon the matching of a worker and a firm. The worker makes the first wage offer with probability p and the firm makes the first wage offer with probability $1-p$. In each subsequent period, if bargaining has not broken down, firms and workers deterministically alternate making offers each period until an offer is accepted. If period t is one in which the firm makes the wage offer this wage offer is denoted W_{ft}, whereas if period t is one in which the consumer makes the offer this wage offer is denoted W_{wt}.

In each period, the current offer by an agent can be accepted by the other agent, the other agent can abandon negotiations altogether, or the other agent can reject the current offer. If the current offer is accepted, the worker produces output immediately. If the other agent abandons negotiations, the worker immediately enters unemployment and the firm returns to the market without a worker. If the other agent rejects the offer but wants to continue negotiations then with probability δ bargaining exogenously breaks down, in which case next period the firm returns to the market with an unfilled vacancy and the consumer enters unemployment, and with probability $1-\delta$ bargaining continues and the agent that rejected the offer makes an offer tomorrow. The firm faces a *haggling cost* ψA_t that the firm must pay in order to make a counteroffer to the worker at t. The worker instead faces only an opportunity cost in continuing to bargain, which we discuss below. Thus, the three parameters that characterize this bargaining scheme are (p, δ, ψ).

As explained in Hall and Milgrom (2008), standard recursive logic implies that when it is the firm's turn to make an offer in period t, the firm will make the best possible offer from its perspective, such that the worker will (weakly) prefer to accept it immediately rather than choose one of the two alternatives: abandon negotiations and get U_t immediately or reject the current offer and, if there is not an exogenous breakdown, make a counteroffer at $t+1$. Thus, the firm's offer W_{ft} satisfies

$$(22)\quad W_{ft} + P_t = \max\left\{U_t, \delta U_t + (1-\delta)\left[bA_t + \mathbb{E}_t Q_{t+1}(W_{wt+1} + P_{t+1})\right]\right\},$$

where the maximum ensures that the worker weakly prefers to accept the firm's offer and get $W_{ft}+P_t$ today rather than the better of two options: abandoning negotiations and getting the value of unemployment today U_t and rejecting the current offer.[7] If the worker rejects the offer, with probability δ there is an

[7]Of course, even though (22) holds with equality, meaning the worker is indifferent between accepting the offer and not,

exogenous breakdown in which case the worker receives the value of unemployment U_t and with probability $1-\delta$ there is no breakdown, in which case the worker produces bA_t in home production and makes a counteroffer W_{wt+1} to the firm at $t+1$. Clearly, its offer W_{ft} must be smaller than the discounted value of output from the match with the worker, Y_t, so that the firm makes nonnegative profits,

$$(23)\quad Y_t - W_{ft} \geq 0$$

which is the individual rationality constraint for the firm.

Next, when it is the worker's turn to make an offer in period t, the worker will make the best possible offer from its perspective such that the firm will prefer to accept it immediately rather than choose one of two alternatives: abandon negotiations and get 0 or reject the current offer. Therefore, the worker's offer W_{wt} satisfies

$$(24)\quad Y_t - W_{wt} = \max\{0, \delta \cdot 0 + (1-\delta)\left[-\psi A_t + \mathbb{E}_t Q_{t+1}(Y_{t+1} - W_{ft+1})\right]\},$$

where the maximum ensures that the firm weakly prefers to accept the offer and get $Y_t - W_{wt}$ today rather than the better of two options: abandon negotiations and obtain 0 and rejecting the current offer. Here, if the firm rejects the offer, then with probability δ bargaining breaks down and the firm gets 0 and with probability $1-\delta$ bargaining does not break down, the firm pays $-\psi A_t$ in haggling costs and makes a counteroffer of W_{ft+1} in the next period. Clearly, workers will only make offers such that if it is taken by the firm, workers prefer employment to unemployment. That is,

$$(25)\quad W_{wt} + P_t \geq U_t$$

which is the individual rationality constraint for the worker.

In each period t a measure of family members are matched with a firm and each of them independently draws the option to make the first offer in bargaining and each shares equally in the family income. Thus, the law of large numbers implies that, for a given realization of the aggregate shock, the value to a family is deterministic and equals the mean of the wages of all its consumers who are bargaining at t which is

we should think this equality as arising from the limit of a sequence of monotonically decreasing offers $\left\{W_{ft}^n\right\}$ which converge to W_{ft} in which for all n, $W_{ft}^n + P_t$ is strictly greater than the right side of (22). Hence, along all points in this sequence the worker strictly prefers to accept the offer.

given by

(26) $W_t = pW_{wt} + (1-p)W_{ft}$.

Similarly, since production is linear in employment, we can think of the firm as dealing with a measure of workers in each period. Hence, the present value of wages to both workers and firms that result from bargaining is deterministic at W_t.

Given a discount factor $Q(s_t, s_{t+1})$, a *credible bargaining equilibrium* is a collection $\{W_{wt}, W_{ft}, W_t, P_t, U_t, J_t, Y_t, \theta_t$ that satisfies (16)–(26). In equilibrium, the receiving party always accepts the first offer. Here the alternating-offer structure affects the wages that are set in the contract because it provides incentives on how to set wages for the first offer. The aggregate state at the beginning of period t is $x_t = (u_{t-1}, a_t, s_t)$, namely, the mass of unemployed agents and the current levels of productivity and the discount factor shock. The rest of the economy is the same as before. In particular, equations (16)-(21) hold and wages are set by the same alternating-offer-bargaining scheme as before. Here a collection of state dependent functions

(27) $\{C_t, Q_{t,t+1}, W_{ft}, W_{wt}, W_t, P_t, U_t, Y_t, J_t, \theta_t, u_t, v_t\}_{t=0}^{\infty}$

is a *credible bargaining equilibrium* if for each t, *i)* the wage offers W_{ft} and W_{wt} satisfy (22)-(26), *ii)* the values Y_t, P_t, U_t, and J_t, satisfy (17)-(20), *iii)* the law of motion for unemployment and satisfies (16), *iv)* the free-entry condition (21) holds, *v)* the resource constraint (15) holds, and *vi)* the pricing kernel $\{Q_{t,t+1}\}$ satisfies (14). For ease of notation we let the subscript t indicate the dependence on the state x_t. In period 0, the economy inherits initial conditions (u_{-1}, a_0, s_0).

When we analyze the properties of the AOB equilibrium it will be convenient to exploit its equivalence to a Nash-bargaining game with appropriately chosen time-varying bargaining weights. Given an AOB equilibrium, namely the collection of functions in (27) with a superscript A added, $u_t^A = \sigma(1 - u_{t-1}^A) + \left[1 - \lambda_w(\theta_t^A)\right] u_{t-1}^A$, recursively defines the evolution of the endogenous component u_{t-1}^A of the state $x_t^A = (u_{t-1}^A, a_t, s_t)$ while (11) and (12) recursively define the evolution of the exogenous shocks a_t and s_t. For a given AOB equilibrium, denote firm surplus by $J_t^A = Y_t^A - W_t^A$, worker surplus by $W_t^A + P_t^A - U_t^A$ and

total surplus by $V_t^A = Y_t^A + P_t^A - U_t^A$. Next, define the (effective) worker bargaining weight as

$$(28) \quad \gamma_t^A \equiv \frac{W_t^A + P_t^A - U_t^A}{V_t^A}$$

which is a function of x_t^A. When doing so, it is useful to note that we can recursively solve for the state x_t^A as a function of the given initial conditions (u_{-1}^A, a_0^A, s_0^A) of the AOB equilibrium and the history of the realizations of innovations ε_t from 0 to t, so that x_t^A is simply an exogenous stochastic process and, hence, so is γ_t^A.

Consider the equilibrium of an economy identical to the one above in which we replace the AOB wage setting mechanism with that of a Nash bargaining game in which the worker bargaining weight is given by the exogenous stochastic process γ_t^A. An equilibrium of the economy with Nash bargaining with the exogenous stochastic process for weights $\{\gamma_t^A\}$ is $\left\{C_t^N, Q_{t,t+1}^N, W_t^N, P_t^N, U_t^N, Y_t^N, J_t^N, \theta_t^N, u_t^N, v_t^N\right\}_{t=0}^{\infty}$that satisfy (16)–(21) and (11)-(15), but with the wage-setting equations (22)-(26) replaced by the Nash bargaining conditions

$$(29) \quad W_t^N + P_t^N - U_t^N = \gamma_t^A V_t^A \text{ and } Y_t^N - W_t^N = \left(1 - \gamma_t^A\right) V_t^A.$$

Inspection of the conditions in the two equilibria makes the following proposition immediate.

Proposition 4. (Equivalence of equilibria under AOB and Nash bargaining with time varying weights) *The equilibrium under AOB bargaining coincides with that under Nash bargaining when the worker bargaining weights are given by γ_t^A in (28).*

To understand this proposition, note that in the standard Nash bargaining problem with constant weights, all these weights do is determine how the surplus is split between firms and workers. Here instead the AOB equilibrium has determined the surpluses as a function of the state and we have constructed the effective bargaining weight function $\gamma_t^A(x_t^A)$ so that the Nash bargaining equilibrium with these weights implements the effective surplus-splitting rule from the AOB equilibrium. The connection will prove useful later when we analyze the mechanism underlying our results. Note that while we have shown this equivalence result for AOB bargaining, a similar result will hold for many bargaining protocols.

3.4 Intuition for Why AOB Weights Fluctuate

Now we turn to developing intuition for why and how the effective AOB weights for the worker's share of the surplus, γ_t^A, fluctuates. To do so we derive a functional equation for the AOB weights γ_t^A. We begin by noting that we substituting for Y_t^A, P_t^A, and U_t^A into $V_t^A = Y_t^A + P_t^A - U_t^A$ from (17), (18), and (19) gives that the surplus satisfies the recursive formula

$$V_t^A = 1 - bA_t + E_t Q_{t+1}^A (1 - \sigma - \lambda_{wt+1}^A \gamma_{t+1}^A) V_{t+1}^A$$

where, in contrast to the standard Nash bargaining problem, the worker's share γ_t^A is endogenous. To that end, let γ_{wt}^A and γ_{ft}^A denote the effective bargaining weights–surplus shares— the worker receives when the worker and the firm make the first offer, respectively, defined by

$$(30) \quad \gamma_{ft}^A \equiv \frac{W_{ft}^A + P_t^A - U_t^A}{V_t^A} \text{ and } \gamma_{wt}^A \equiv \frac{W_{wt}^A + P_t^A - U_t^A}{V_t^A}.$$

Since by definition the worker's surplus share is $\gamma_t^A = W_t^A + P_t^A - U_t^A$, then $W_t^A = pW_{wt}^A + (1-p)W_{ft}^A$ from (26) together (30) imply $\gamma_t = p\gamma_{wt} + (1-p)\gamma_{ft}$. Now substituting $Y_t^A - W_{wt}^A = (1 - \gamma_{wt}^A)V_{wt}^A$ and $Y_{t+1}^A - W_{ft+1}^A = (1 - \gamma_{ft+1}^A)V_{t+1}^A$ into (24), after some manipulation we have

$$(31) \quad \gamma_{wt}^A V_t^A = V_t^A - (1-\delta) \max\left(0,\, -\psi A_t + \mathbb{E}_t Q_{t+1}^A \left(1 - \gamma_{ft+1}^A\right) V_{t+1}^A\right).$$

Next, subtracting U_t^A from the left side and $\delta U_t^A + (1-\delta)U_t^A$ from the right side of (22) and using (30) after some manipulation gives

$$(32) \quad \gamma_{ft}^A V_t^A = (1-\delta) \max\left(0,\, -\left(U_t^A - bA_t - \mathbb{E}_t Q_{t+1}^A U_{t+1t}^A\right) + \mathbb{E}_t Q_{t+1}^A \gamma_{wt+1}^A V_{t+1}^A\right).$$

In (31), the expected value for the firm of making an offer tomorrow is the last term in (31), namely $\mathbb{E}_t Q_{t+1}^A \left(1 - \gamma_{ft+1}^A\right) V_{t+1}^A$, but in order to do so the firm must pay the cost ψA_t today. Similarly, in (32), the expected value for the worker of making an offer tomorrow is the last term in (31), namely $\mathbb{E}_t Q_{t+1}^A \gamma_{wt+1}^A V_{t+1}^A$, but in order to do so the worker incurs the opportunity cost $U_t^A - \mathbb{E}_t Q_{t+1}^A U_{t+1}^A - bA_t$ today. To interpret

the worker's cost, use (19) to write the worker's cost of delay as

$$(33)\quad U_t^A - \mathbb{E}_t Q_{t+1}^A U_{t+1}^A - bA_t = \mathbb{E}_t Q_{t+1}\left[\lambda_{wt+1}(W_{t+1} + P_{t+1} - U_{t+1})\right] = \mathbb{E}_t Q_{t+1}\left[\lambda_{wt+1}(\gamma_{t+1}^A V_{t+1}^A)\right].$$

That is, turning down the offer means that the worker gives up the ability to search tomorrow, namely the probability λ_{wt+1} of getting a 'bonus' of $W_{t+1} + P_{t+1} - U_{t+1}$ over unemployment.

This difference in how the costs of delaying vary over the cycle for the firm and the worker is at the heart of why the worker's effective bargaining share γ_t^A rises in downturns. Scaled by the state, the firm's cost of delay ψ is simply a constant. Loosely, the firm's impatience to strike an offer is constant over the upturns and downturns. In contrast, the worker is losing a time-varying bonus by delaying. Since the probability of finding a job in the next period, λ_{wt+1}, is low when the state today is bad, the worker does not mind the delaying bargaining in downturns. Thus, the worker is relatively more patient in downturns to striking a bargain and so in downturns the worker gets a larger share of the surplus than in upturns. (Note that this feature is analogous to the result that in classic bargaining games the more patient player gets a larger surplus share.) Because firms understands that entering into bargaining in downturns means that they will get a smaller share of the surplus, from the free entry condition it follows that firms post fewer vacancies downturns than in upturns. Hence, the job-finding rate must fall in downturns and unemployment rise.

We can go beyond this general intuition if we make more assumptions. Assume that it is not optimal for either firms or workers to quit negotations during bargaining—so we can drop the max operators in (22) and (24)—and that the individual rationality constraints for firms and workers, (23) and (25) never bind. Dropping the max operators we can write (31) and (32) as

$$(34)\quad \gamma_{wt}^A V_t^A = V_t^A - (1-\delta)\left[-\psi A_t + \mathbb{E}_t Q_{t+1}^A \left(1 - \gamma_{ft+1}^A\right) V_{t+1}^A\right]$$

and

$$(35)\quad \gamma_{ft}^A V_t^A = (1-\delta)\left[\mathbb{E}_t Q_{t+1}^A (\gamma_{wt+1}^A - \lambda_{wt+1}\gamma_{t+1}^A) V_{t+1}^A\right]$$

where we substituted from (33) to obtain (35).

For simplicity, we assume $p = 1/2$ and note that $\eta = 1/2$ is the elasticity of the matching function with respect to the measure of the unemployed. In this case, multiplying (34) and (35) by $1/2$ gives, after

simple manipulation, that the AOB weight γ_t^A solves the recursive expression

$$(36)\quad \left(\gamma_t^A - \frac{1}{2}\right) V_t^A = \frac{1}{2}(1-\delta)\left(\psi A_t - \mathbb{E}_t Q_{t+1}\gamma_{t+1}^A \lambda_{wt+1}^A V_{t+1}^A\right) + (1-\delta)\,\mathbb{E}_t Q_{t+1}^A \left(\gamma_{t+1}^A - \frac{1}{2}\right) V_{t+1}^A.$$

Letting $\omega_t = \psi A_t - \mathbb{E}_t Q_{t+1}^A \lambda_{wt+1}\gamma_{t+1}^A$ be the haggling cost of firms net of the haggling cost of worker in t. The lower is the cost of haggling for workers—as it is in recessions—the higher will be the worker's effective bargaining weight. Solving this equation forward gives that

$$\left(\gamma_t^A - \frac{1}{2}\right) V_t^A = \frac{1-\delta}{2}\left\{\omega_t + (1-\delta)\mathbb{E}_t Q_{t+1}^A V_{t+1}^A \omega_{t+1} + (1-\delta)^2 \mathbb{E}_t Q_{t+2}^A V_{t+2}^A \omega_{t+2} + ...\right\}$$

where we used that for $\delta < 1$, $\lim_{s\to\infty}(1-\delta)^s \mathbb{E}_t Q_{t+s}^A V_{t+s}^A \omega_{t+s} = 0$. For $\delta = 1$, the effective bargaining weight is constant at $\gamma_t^A = 1/2$.

Hence, the deviation of the worker's AOB weight from $p = 1/2$ equals the present value of the difference between the cost of haggling for firms, ψA_t, and the cost of haggling for workers, $\mathbb{E}_t Q_{t+1}^A \gamma_{t+1}^A \lambda_{wt+1}^A V_{t+1}^A$. These costs are scaled by the probability that bargaining continues, $1-\delta$. When δ is equal to 1, the bargaining weight is equal to $1/2$, which also the value of the elasticity η. In contrast, when $\delta < 1$, the bargaining weight fluctuates over time in response to aggregate shocks. In particular, in periods of high discount rates, in which λ_{wt+1} is low, the worker's opportunity cost of haggling is low, and the AOB bargaining weight γ_t^A is high.

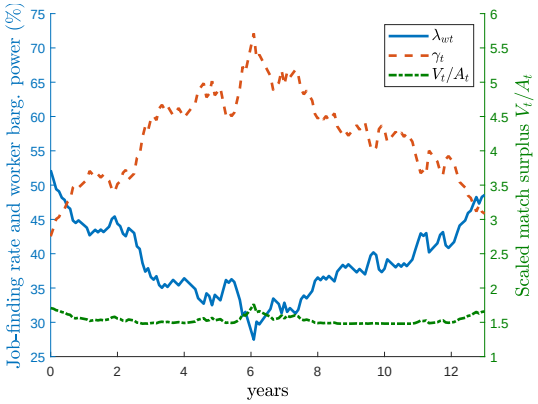
As we will see, under our quantification this weight will fluctuate as it does in Figure 4. This expression also illustrates the critical role played by the probability of a breakdown δ in generating fluctuations in the AOB bargaining weights. When δ is close to 1, the flows in (36) are heavily discounted and the AOB bargaining weights are close to $1/2$ and unemployment has little volatility. In contrast, when δ is low, the flows in (36) are less heavily discounted and the AOB bargaining weights fluctuate significantly.

4 Quantification and Results

We turn now to describing how we choose parameters for our quantitative analysis and then discuss the results.

Figure 4: Job-finding rate, worker's bargaining share, and scaled surplus

(a) *Baseline model*



(b) *Model with* $\delta = 1$

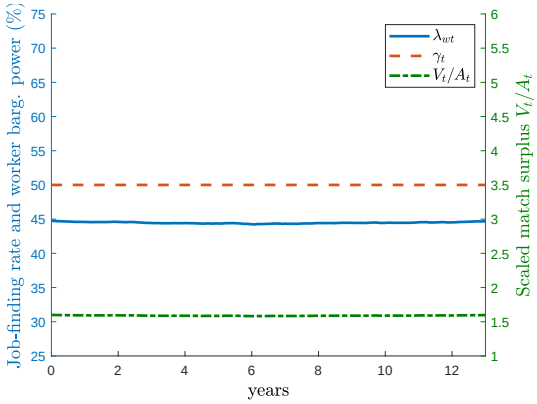


Table 2: Baseline Model

Panel A: Parameters		Panel B: Moments		
Endogenously chosen parameters		*Targeted moments*	Data	Model
g_a, mean productivity growth (% p.a.)	2.28	Mean productivity growth (% p.a.)	2.28	2.28
σ_a, s.d. productivity growth (% p.a.)	1.84	S.d. productivity growth (% p.a.)	1.84	1.84
β, time preference factor	0.999	Mean risk-free rate (% p.a.)	0.74	0.74
S, mean of state S_t	0.194	S.d. risk-free rate (% p.a.)	2.35	2.35
α, inverse EIS	5.00	Sharpe ratio of market excess return (p.a.)	0.45	0.45
ρ_s^{12}, annualized persistence of state	0.935	Autocorrelation log price-dividend ratio (annual)	0.94	0.94
κ, vacancy posting cost	0.37	Mean unemployment rate (%)	5.9	5.9
ψ, haggling cost	0.353	Mean job-finding rate (%)	45.5	45.5
Assigned parameters		*Results*		
b, home production parameter	0.6	S.d. job-finding rate (%)	6.67	6.77
η, matching function elasticity	0.5	S.d. unemployment rate (%)	0.97	1.12
σ, probability of separation	0.028	Correlation unemployment and job-finding rate	-0.96	-0.95
B, efficiency of matching technology	0.455	Correlation unemployment and productivity growth	0.04	-0.03
δ, bargaining breakup probability	0.013	Autocorrelation job-finding rate	0.94	0.99
p, probability worker moves first	0.5	Autocorrelation unemployment rate	0.97	0.99
		Mean excess return (% p.a.)	7.05	7.13
		S.d. excess return (% p.a.)	15.6	14.9
		Mean log price-dividend ratio	3.43	3.17
		S.d. log price-dividend ratio	0.45	0.37
		Semi-elasticity UC labor to unemployment	-5.32	-6.40

4.1 Quantification

The model is monthly and its parameters are listed in Table 2: six parameters, $\{b, \eta, \sigma, B, \delta, p\}$, are assigned and the remaining eight, $\{g_a, \sigma_a, \beta, S, \alpha, \rho_s, \kappa, \psi\}$, are endogenously chosen to match nine moments from the data. For the assigned parameters, following Ljungqvist and Sargent (2017), we set the home production parameter b to 0.6 and we set the matching function elasticity η to be $1/2$. We choose the separation rate σ to match the Abowd-Zellner corrected estimate of a separation rate of 2.8% by Krusell et al. (2017) based on CPS data. For the match efficiency parameter B in the Cobb-Douglas matching function $m = Bu^{\eta}v^{1-\eta}$ we proceed as follows. We consider a steady state and, as in Shimer (2005), normalize the market tightness $\theta_{ss} = 1$, so that the steady-state job-finding rate $\lambda_w^{ss} = B$ is 45.5%. Finally, we set the bargaining breakup probability $\delta = 0.013$ per quarter and the probability that the worker makes the first offer $p = .5$ to be the same as in Hall (2017).

Consider next the endogenously chosen parameters. Here, even though we jointly choose eight parameters to match the eight targeted moments in Table 2, it is intuitive to discuss each parameter together with the moment most sensitive to it. We choose the parameters g_a and σ_a of the exogenous productivity process to match the mean and standard deviation of labor productivity growth from the BLS for the period between January 1947 and December 2007.

Consider next the preference parameters, $\{\beta, S, \alpha\}$. We choose the rate of time preference β and the mean S of the state S_t to match the mean and the standard deviation of the real risk-free rate r_{ft} constructed as $i_t - \mathbb{E}_t\pi_{t+1}$, where i_t is the one-month Treasury bill rate.[8] To see how the mean S of the process governing the state S_t can be chosen to generate only a modest volatility in the risk-free rate, note that when consumption is conditionally lognormally distributed, the real risk-free rate approximately satisfies

$$r_{ft} = -\log(\beta) + \alpha\mathbb{E}_t\Delta c_{t+1} + \alpha\mathbb{E}_t\Delta s_{t+1} - \frac{\alpha^2[1+\lambda_a(s_t)]^2}{2}\sigma_t^2\left(\varepsilon_{ct+1}\right), \tag{37}$$

which holds since the conditional standard deviation of the innovation to consumption growth $\sigma_t(\varepsilon_{ct+1}) \cong \sigma_t(\varepsilon_{t+1})$. Thus, the impact of $\sigma_t(\varepsilon_{ct+1})$ on r_{ft} is affected by the level of S through $\lambda_a(s_t)$ by (13).

In comparing equity returns in the model and the data, we follow Mehra and Prescott (1985), Campbell

[8] To compute $\mathbb{E}_t\pi_{t+1}$, we use an updated version of the Fama and French's (1993) data available from Kenneth French's website and proxy $\mathbb{E}_t\pi_{t+1}$ by the projection of monthly CPI inflation on twelve of its lags. Fitting a univariate AR(1) specification for $\mathbb{E}_t\pi_{t+1}$ with lags up to one year is standard. See, for instance, Hur, Kondo, and Perri (2019).

and Cochrane (1999), and Wachter (2006) and interpret dividends as claims to aggregate consumption in the model and as claims to aggregate dividends from CRSP in the data. We then choose the inverse elasticity of intertemporal substitution α in the model to match the mean of the maximum Sharpe ratio of the aggregate stock market return measured from the CRSP value-weighted stock index, which covers all firms continuously listed on NYSE, AMEX, and NASDAQ.[9] Specifically, in the model we compute the Sharpe ratio of the stock market excess return $\mathbb{E}(R_{t+1}-R_{ft})/\sigma(R_{t+1}-R_{ft})$ using the return on consumption claims. This strategy is similar to that used by Campbell and Cochrane (1999) and Wachter (2006) in a very related context.

To pin down the persistence of the discount factor shock, we choose ρ_s to match the observed autocorrelation of price-dividend ratios. When we do so, the standard deviation of the price-consumption ratio is 82% of the standard deviation of the price-dividend ratio in the CRSP data.Finally, we choose the vacancy posting cost parameter κ and the haggling cost parameter ψ to match the mean unemployment rate of 5.9% and the mean job-finding rate of 45.5% from BLS data between January 1948 and December 2007.

4.2 Results

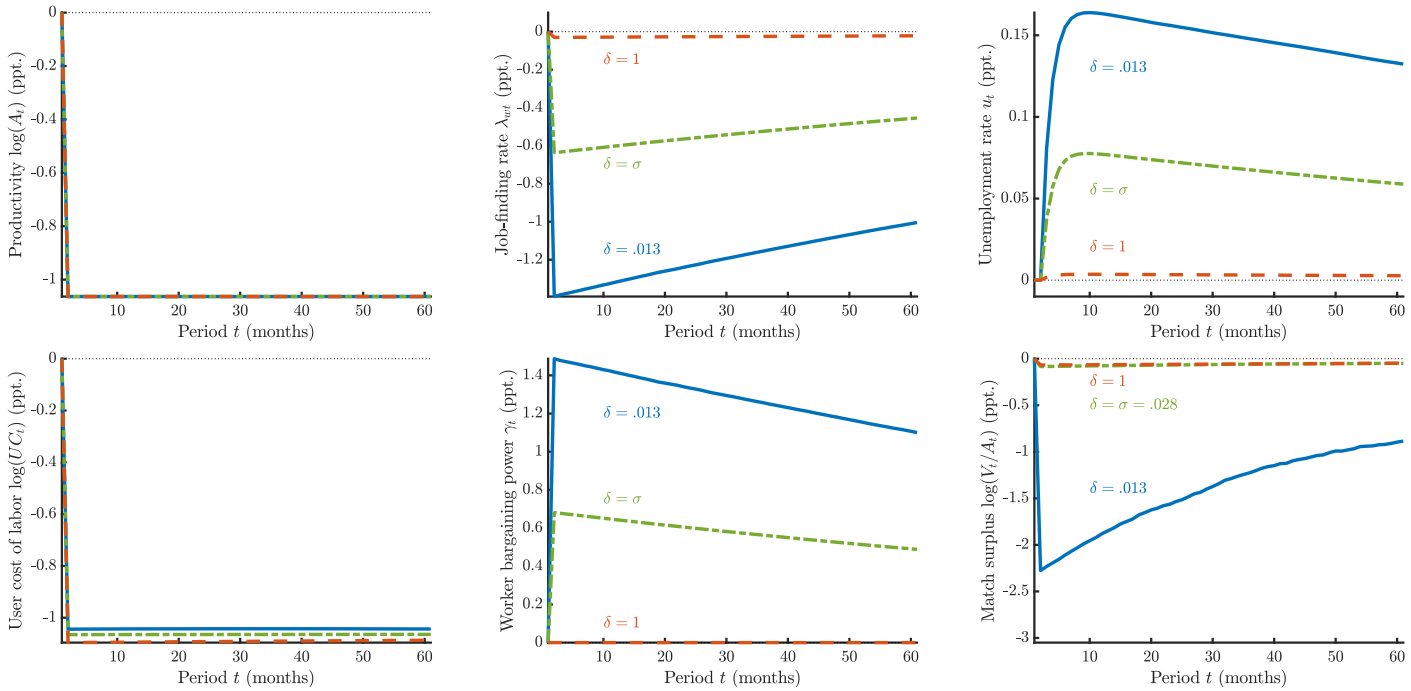
We now show that the baseline model can generate the observed volatility in job-finding rates and unemployment as well as overcome the Borovicka and Borovickova critique on asset prices, the Kudlyak critique on wage stickiness and the two critiques on implementation for Hall. We then turn to examining the impulse responses to shocks to give a better feel for the mechanism behind these results.

4.2.1 Solving the Unemployment Volatility Puzzle

As Table 2 shows, our baseline model produces a volatility for the job-finding rate (6.77) very similar to that in the data (6.67). It also produces a value for the volatility of unemployment (1.12) even higher than that in the data (0.97). The autocorrelation of the unemployment rate in the model (0.99) is also close to that in the data (0.97). This table shows that our model also reproduces well the highly negative correlation between job-finding and unemployment rates, namely -0.95 in the model and -0.96 in the data. This result is consistent with Shimer's (2005) emphasis that unemployment rises in recessions because the job-finding rate falls due to a decline in vacancy creation.

[9]Note that we would have obtained similar results by using data from the Flow of Funds, since, as shown by Larrain and Yogo (2008), the returns measured from CRSP are highly correlated with the returns on the aggregate stock market measured from the Flow of Funds. In our sample, this correlation is of 0.97.

Figure 5: Varying δ: Responses to a negative two-standard deviation productivity shock in baseline



4.2.2 Overcoming the Borovicka and Borovickova Critique on Asset Prices

We next show that the baseline model generates reasonable patterns for risk-free rates and stock market returns. In Table 2, we see that the model's predictions for the mean and standard deviation of the (real) risk-free rate, 0.74% per year and 2.35% per year, are the same as in the data. The model's implications for the risk-return trade-off, given by a Sharpe ratio on market excess returns of 0.45, is identical to that in the data. Also, standard measures of riskiness, such as the equity premium, namely, the mean excess return on equity, and its standard deviation, are similar in the model (7.13% and 14.9%) and the data (7.05% and 15.6%). Finally, the log price-dividend ratio has a similar mean, standard deviation, and autocorrelation in the model (3.43, 0.45, and 0.94) as in the data (3.20, 0.35, and 0.94). In sum, our baseline model has similar fluctuations in asset prices as in the data. Thus, it successfully overcomes the Borovicka and Borivickova (2018) critique of the source of movements in discount rates in Hall's model.

4.2.3 Overcoming the Kudlyak Critique on Wage Stickiness

As discussed earlier, the Kudlyak statistics estimated by Kudlyak (2014), Basu and House (2016), and Bils et al. (2023) are $\beta_K = -5.2\%$, -5.8%, and -5.3% respectively, so a one percentage point increase in unemployment results in more than a 5% drop in the user cost of labor. As the last entry in Panel B of Table 2 shows, our baseline model implies that a one percentage point increase in unemployment is

associated with a 6.4% decrease in wages, which even a bit larger than the statistics from the data. In this sense, the baseline model can generate the observed volatility of unemployment with an even greater procyclicality of wages than in the data. Hence, in contrast to Hall's model, the mechanism of the baseline model is consistent with the cyclicality of wages observed in the data.

To get some intuition about this mechanism, Figure 5 plots the impulse responses to a two-standard deviation drop in productivity of nearly 1%. Consider the responses for the baseline model with $\delta = .013$. As the top middle panel shows, this drop in productivity leads to a drop in the job-finding rate of about 1.4 percentage points on impact that slowly reverts back to zero. The top right panel shows that this movement in the job-finding rate is associated with a hump-shaped increase in the unemployment rate. The left panel in the second row shows that the associated user cost of labor tracks productivity: it permanently falls by a little over 1%. A rough way to get a feel the magnitude of the Kudlyak statistic is to take the fall in the user cost, about 1 percentage point and divide it by the approximate rise in the unemployment rate, about 0.15 percentage points to get $\Delta UC/\Delta u = -1/.15 = -6.6\%$, which is close to the actual Kudylak statistic of -6.4% in Table 2. The bottom middle panel plots the impulse response of the implied Nash bargaining weight of a worker to this negative shock to productivity and we see that this bargaining weight increases as productivity falls. In the bottom right panel we see that the scaled surplus of the match, V_t/A_t, barely moves.

We next argue that our model reproduces the Kudlyak statistic because of the high correlation between productivity shocks and discount-rate movements in our model, which contrasts with the weak correlation between productivity shocks and discount-rate movements in Hall's model. To illustrate the need for this high correlation, we construct a variant of our model with *independent discount rate shocks*, in which discount-rate shocks are independent of productivity innovations. Specifically, we keep the random walk process on productivity, namely, $a_{t+1} = g_a + a_t + \varepsilon_{t+1}$ but for discount rate shocks assume $s_{t+1} = (1 - \rho_s)\, s + \rho_s s_t + \lambda_a(s_t)\varepsilon_{st+1}$, where the innovations ε_{st+1} to surplus consumption, are independent of the innovations ε_{t+1} to productivity, rather than being perfectly correlated as we had in our baseline in (12). Table 3 shows the results of this variant for some key statistics. This variant is reparameterized to match the same empirical moments as our baseline and is nearly observationally equivalent in all respects—except for the behavior of the user cost. In this case, the Kudlyak statistic collapses to near zero (0.27)—less than 1/40th of that in the data and the opposite sign. This pattern mirrors the result in Hall's model in which the user cost of labor barely falls when productivity falls. These findings highlight that a high correlation

Table 3: Key Moments Across Variants of Baseline Model

	Data	Baseline	Pure discount	Alternative bargaining durations		
		$\delta = .013$	shocks	$\delta = .012$	$\delta = \sigma = .028$	$\delta = 1$
Potential bargaining duration $\frac{1}{\delta} - 1$ (months)		78	78	85	35	0
$E(R_{ft})$ (% p.a)	0.74	0.74	0.74	0.74	0.74	0.74
$\sigma(R_{ft})$ (% p.a)	2.35	2.35	2.35	2.35	2.35	2.35
$E(R_{t+1} - R_{ft})$ (% p.a)	7.05	7.13	6.64	7.07	6.88	6.88
$\sigma(R_{t+1} - R_{ft})$ (% p.a)	15.6	14.9	15.0	14.8	14.5	14.4
$\sigma(\Delta a_t)$ (%)	1.84	1.84	1.84	1.84	1.84	1.84
$\sigma(\lambda_{wt})$ (%)	6.67	6.77	6.95	7.86	2.97	0.15
$\sigma(u_t)$ (%)	0.97	1.12	1.26	1.30	0.39	0.02
Semi-elast. UC labor to unemployment	-5.32	-6.40	0.15	-5.32	-13.8	-305
S.d. effective bargaining power $\sigma(\gamma_t)$		7.80	8.04	9.11	3.24	0.00
Average elasticity of unemployment duration w.r.t. b	(0.36, 0.80)	20.8	20.8	21.2	19.0	0.88

Note: Throughout the alternative models, the haggling cost parameter is adjusted to keep the mean job-finding rate. The "Pure discount shocks" variant turns off the effect of productivity shocks everywhere but on discount rates.

between productivity shocks and discount-rate movements is essential for the model to generate realistic cyclicality in the user cost of labor. Next, we consider a case with discount rate shocks only and find that the Kudylak statistic collapses even more to 0.15.

The key takeaway from these last two experiment is that it is complete misreading of the second generation literature to claim that discount rate shocks alone produce reasonable comovements of wages and unemployment. Instead, successful versions of these models have discount rate shocks that are *driven* by the productivity shocks, rather than by some independent force.

4.2.4 Dependence on Extremely Low Exogenous Breakdown Rates

To understand the mechanism in the model, it is critical to understand that the baseline model generates the observed volatility in unemployment and the observed degree cyclicality of wages only because the exogenous breakdown rate of bargaining δ is very low—so that the potential duration of bargaining is very long. In Table 3 we investigate what happens as we increase the breakdown rate from .013 in our baseline to what Hall and Milgrom consider to be the *minimum breakdown probability*, namely equal to the separation rate, $\delta = \sigma = .028$, to the *maximum breakdown probability* $\delta = 1$. As we do so, the volatility of unemployment falls from 1.12 to .39 to .02 whereas the Kudlyak statistic increases from -6.4 to -13.8 to -305. That is, as the breakdown rate increases, a one percent increase in unemployment is associated with a 6.4%, 13.8%, and 305% fall in wages, respectively.

To get a feel for the patterns over time differ over time as the level of the breakdown rate is changed, in Figure 4, we consider the two extreme levels of the breakdown probability. For the same sequence of

productivity shocks chosen to produce a typical looking recessions, the left panel plots the job finding rate, the scaled surplus, and the effective bargaining rate as the baseline breakdown probability, $\delta = .013$ and the right panel plots the same series for $\delta = 1$. Consistently with Table 3, we see that as the breakdown rate gets large—so the potential bargaining duration becomes smaller, the job-finding rate and the effective bargaining weight become much less volatile.

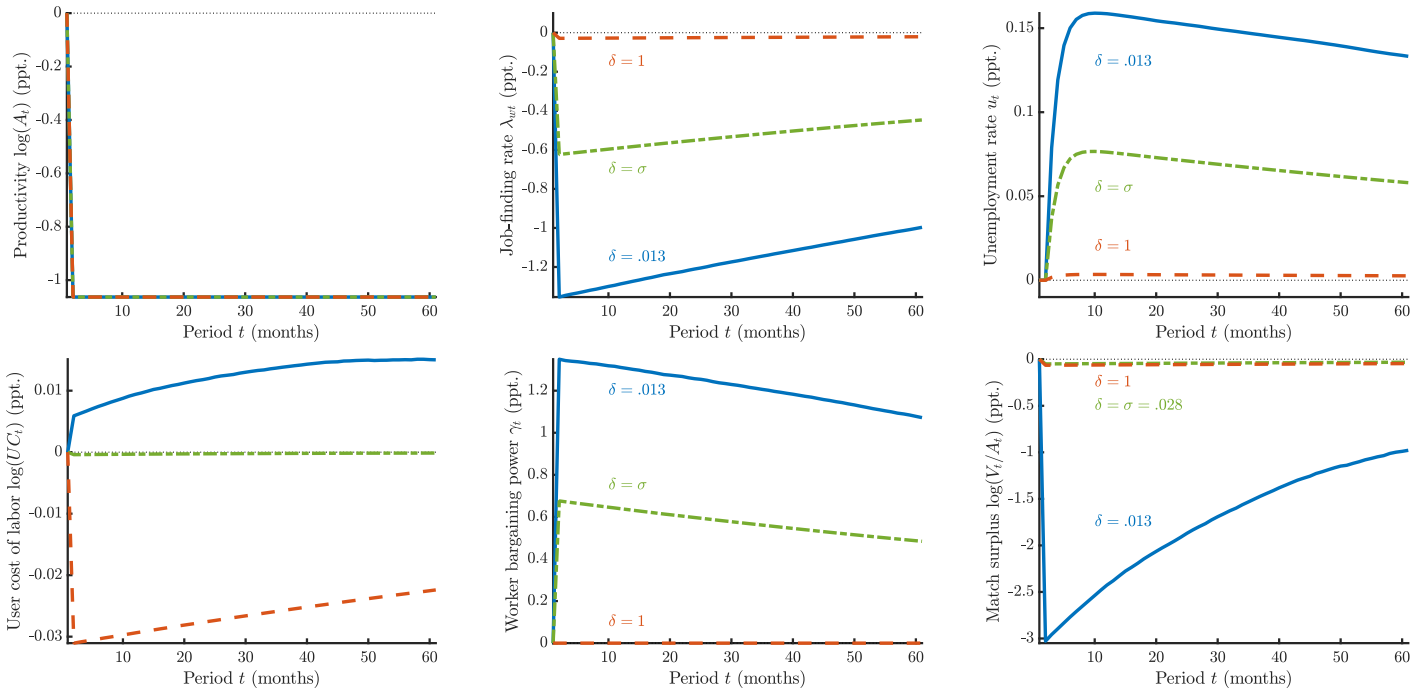
Figure 5 helps provide some intuition for these results. The bottom left panel of this figure shows that, regardless of the breakdown rate, when productivity permanently falls by 1 percent, so does the user cost. In contrast, as the top middle and right panel show, the lower is the breakdown rate, the larger is the fall in the job-finding rate and, hence, the larger is the hump-shaped rise in the unemployment rates. Here also we can roughly calculate the Kudlyak statistic by eye-balling the movements in $\Delta UC/\Delta u$ from the figure in the three cases to be $-1/.15 = -6.6\%$, $-1/.07 = -14\%$, and $-1/.005 = 200\%$. In short, regardless of the breakdown rate, the user cost falls one for one with productivity, but lower breakdown rates are associated with much greater increases in the unemployment rate. So as the breakdown rate falls, the semi-elasticity $\Delta UC/\Delta u$ falls by much less and, hence, the user cost is much less cyclical. Since the user cost tracks productivity nearly one-for-one in all three cases, the key difference across these three experiments is how much unemployment rises for this given fall in productivity, not how much the user cost moves overall.

In Figure 6 we see what happens in the model with pure discount shocks as we consider these same three levels of the breakdown probability. Here, regardless of the breakdown probability, a contractionary discount rate shock leads to essentially no change in the user cost. Note from Table 1 that a broadly similar pattern happens in Hall's model: as δ increases the Kudlyak statistic increases from a small positive number to a very large negative number (-116 when $\delta = 1$). In this sense, we think that the variant with shocks to the discount factor that are independent of productivity are useful in helping to understand what is going on in Hall.

5 Some Unappealing Implications of the Baseline Model

Even though the baseline model overcomes the criticisms of Hall's original work, once we dig deeper into its mechanisms we find some unappealing features. The first two features arise because low breakdown probabilities mean the equilibrium is very inefficient in two senses. First, it is not Pareto efficient. Second, it is not even production efficient. That is, in an extended version of the model, regardless of whatever

Figure 6: Varying δ: Responses to a positive two-standard deviation productivity shock in the variant that turns off the effect of productivity shocks everywhere but on discount rates



protocol the other firms are using, each individual firm in each period would prefer to choose a different protocol than that in the baseline model. Indeed in a deep recession firms would pay almost 20% of their profits to be allowed to adopt an efficient bargaining protocol. Third, the critique of Costain and Reiter (2008) of standard search models applies to the baseline model: if such models generate unemployment as volatile as it is the data then they generate much too large a response to policy changes. Concretely, in the baseline model the elasticity of the response of unemployment duration to an increase in the replacment rate of unemployment benefits is an order of magnitude larger than it is in the data.

5.1 Efficiency and Inefficiency in AOB Models

We begin by establishing a set of conditions on the parameters of the AOB bargaining protocol such that it is efficient—in the sense that equilibrium allocations maximize the ex-ante welfare of each family subject to the search frictions. These conditions are the analog of the Hosios condition for efficiency under Nash bargaining. The key condition is that duration of potential bargaining $1/\delta - 1$ converges to zero. Moreover, under this condition the effective bargaining weights that decentralize the AOB equilibrium are constant. Guided by this result, consider the last row of Table 3. As the duration of potential bargaining falls from 85 months, to 35 months, to zero months, the volatility of the effective bargaining weights falls monotonically—indicating greater efficiency—from being very volatile to being constant. But as this volatility falls, so does

the volatility of job-finding rates and unemployment. Next, Figure 1 for Hall's model and Figure 4 we see that the extra volatility generated by alternating-offer-bargaining relative to efficient Nash bargaining manifests itself by having the workers' effective bargaining power inefficiently rise in recessions relative the constant Hosios level of bargaining power and inefficiently fall in booms relative to this same constant level of bargaining power. In Figure 4 we see that when $\delta = 1$, fluctuations in employment are essentially zero. In this sense, inefficiency drives nearly all of the fluctuations in job-finding rates.

The following proposition links this key parameter δ to the efficiency of the resulting allocations: when the duration is short, allocations are close to efficient but when the duration is long allocations are inefficient.

Proposition 5. (Efficiency under AOB) *When the probability p that the worker makes the first offer equals the elasticity η of the matching function with respect to the measure of unemployed workers, then the allocations in a sequence of bargaining games indexed by the breakdown probabilities $\{\delta_n\}_{n=1}^{\infty}$ converge to the efficient allocations as δ_n converges to one.*

Proof. We use the well-known result that if workers and firms receive shares $1-\eta$ and η of the surplus $V_t = Y_t + P_t - U_t$ so $W_t + P_t - U_t = \eta V_t$ and $Y_t - W_t = (1-\eta)V_t$ then the Hosios condition is satisfied and the equilibrium is efficient. For each δ_n, $W_{ft}(n)$ and $W_{wt}(n)$ satisfy

$$W_{ft}(n) + P_t(n) = \max\left\{U_t, \delta(n)U_t + (1-\delta(n))\left[bA_t + \mathbb{E}_t Q_{t+1}(W_{wt+1} + P_{t+1})\right]\right\}, \tag{38}$$

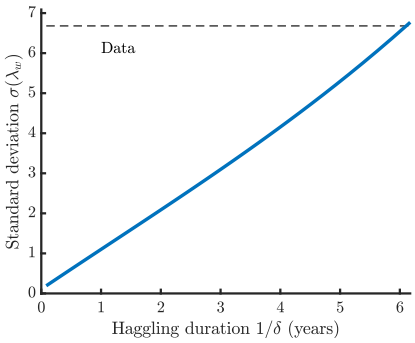
$$Y_t - W_{wt}(n) = \max\{0, (1-\delta(n))\left[-\psi A_t + \mathbb{E}_t Q_{t+1}(Y_{t+1} - W_{ft+1})\right]\}. \tag{39}$$

Taking the limit of both sides of these equations in n yields $W_{ft} + P_t = U_t$ and $Y_t - W_{wt} = 0$. By continuity, the participation constraints $W_{ft} \leq Y_t$ and $W_{wt} + P_t \geq U_t$ clearly hold. Using these limiting values to substitute for W_{wt} and W_{ft} from (38) and (39) into $W = pW_{wt} + (1-p)W_{ft}$ and using that $p = \eta$, we obtain $W_t = \eta Y_t + (1-\eta)(U_t - P_t)$. Adding $P_t - U_t$ to both sides of this equation and collecting terms gives that $W_t + P_t - U_t = \eta(Y_t + P_t - U_t) = \eta V_t$ and $Y_t - W_t = (1-\eta)V_t$. Thus, workers receives a share η of the surplus and firm receive a share $1-\eta$. Hence, the Hosios condition holds and the equilibrium is efficient. *Q.E.D.*

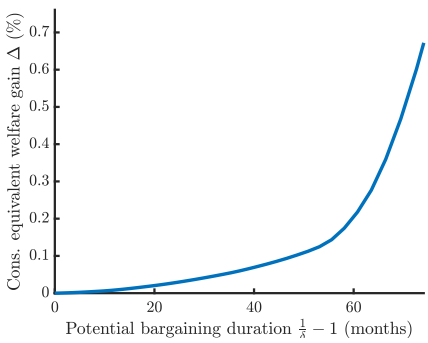
The intuition for efficiency when $\delta(n)$ converges to 1 and $p = \eta$, is as follows. In the limit, with probability η the worker goes first and makes a take-it-or-leave-it offer in which the worker gets all of the output during the current match and the firm gets 0, so the worker wages $W_{wt} = Y_t$ and their surplus is

Figure 7: Volatility of job-finding rate and average consumption equivalent welfare gain as a function of bargaining duration

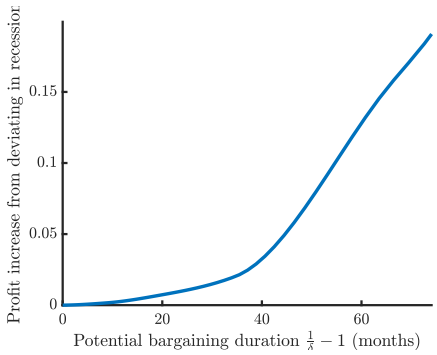
(a) *Volatility of job-finding rate*



(b) *Average consumption equivalent welfare gain*



(c) *Average profit increase from deviating in recession*



Note: The consumption equivalent welfare gain (Δ) measures the proportional change in lifetime consumption that makes a consumer indifferent between the baseline model and the efficient allocation. It is defined by $(1+\Delta)^{1-\alpha}V^{aob} = V^{\text{eff.}}$, where V^{aob} is welfare under the baseline and $V^{\text{eff.}}$ is welfare under the efficient allocation. A positive Δ implies that the efficient allocation yields a welfare improvement equivalent to a $\Delta\%$ permanent increase in consumption.

$W_{wt} + P_t = Y_t + P_t$. With probability $1-\eta$ the firm goes first and makes a take-it-or-leave-it offer in which the firm gets largest surplus it can subject to the constraint that the worker gets in total $W_{ft} + P_t = U_t$. Since each family has a measure of workers, each of which gets to make the first offer with probability p and each firm makes a measure of offers and for each offer it gets to make the first offer with probability $1-p$ then by the law of large numbers this economy is equivalent to one in which workers and firms bargain over wages W_t—equal to $pW_{wt} + (1-p)W_{ft}$—in a standard Nash bargaining model with the bargaining weight equal to the elasticity of the matching function η.

Figure 7 panels *a* and *b* show how increasing the potential bargaining duration increases both the volatility of the job-rate and the consumption equivalent welfare gain from switching to the efficient AOB of Proposition 5. Here this consumption gain is the percentage change in lifetime consumption that a consumer would need to have to be indifferent between being in the equilibrium of the baseline model rather than the equilibrium with efficient AOB bargaining. We see that for short potential durations, say less than a couple of months, this welfare gain is miniscule but it grows rapidly with the duration of potential bargaining.

5.2 Production Efficiency in an Extended Search Model

We turn to different perspective on this inefficiency. We argue that the failure of efficiency is essentially a failure of each individual firm choose the technology for bargaining that maximizes its profits. In the standard AOB setup this inefficiency is hidden because the firm has no choice of the technology is uses to attract workers. Here we augment our setup so that in each period, the firm can choose both the bargaining protocol it desires and the measure of vacancies to post.

The production technology for output has two parts. In the first stage, at the beginning of each period, firms can post their vacancies and bargaining protocol—namely the rules and the parameters that define it—that determine the wages workers will receive upon matching with the firm. Posting a protocol means that the firm is commited to using that protocol in the current period. In the second stage, workers direct their search to a group of firms with the same protocol and then randomly search among them. So for example, if some firms choose the baseline AOB protocol and others choose, say, efficient Nash, then it is as if the firms with the AOB protocol are located on the *AOB island* and the firms with the efficient Nash protocol are located on the *efficient Nash island*. Workers can then choose which island to search on. Once workers arrive on an island they randomly match with the firms there. This setup keeps the spirit of random search conditional on a protocol, but the first stage allows firms to compete with each other over protocols. Intuitively, it is this head-to-head competition among protocols at the island-choosing stage that allows better protocols to drive out worse ones in the market.

To make the analysis simple, we consider an economy with no productivity growth ($g_a = 0$) and no shocks that is in a steady state in which productivity is constant and normalized at 1. Next, recall that our baseline model has a large but finite number of firms F and a measure of workers Nm_w with $m_f = Nm_w/F$ strictly positive, so that in a symmetric equilibrium, each firm hires a finite measure of workers.[10] Since the following result holds for a general constant-returns-to-scale matching function $m(u, v)$, we set up notation for that case. Note that defining the elasticity of the matching function with respect to unemployment as $\eta(\frac{v}{u}) = d\log m(u, v)/d\log u$ and using that $\theta = v/u$ and $\lambda_w(\theta) = \theta\lambda_f(\theta)$, simple

[10]We do so because our matching function is defined over measures of workers and vacancies, and we want to consider a deviation by a single firm. Since we want the associated matching probabilities to be well-defined for such a deviation, we need that each firm posts a measure of vacancies and attracts a measure of workers. For simplicity, we also want each firm to be small in a strategic sense relative to other $F-1$ firms. To do so, we consider a sequence of economies $(F, m_w(F))$ such that the measure of workers per firm, $m_f(F)/F$ is a constant $m_f > 0$, for all F, even as F becomes arbitrarily large. (See Burdett and Judd (1983) for similar assumptions in a different context.) We focus on the case in which F is sufficiently large that deviations by a single firm will not change the equilibrium.

algebra gives $\eta(\theta) = -\theta\frac{\lambda_f'(\theta)}{\lambda_f(\theta)}$ and $1-\eta(\theta) = \theta\frac{\lambda_w'(\theta)}{\lambda_w(\theta)}$.Turning to the proof, suppose, by way of contradiction, that all firms choose to post the same vacancies v^A and the same AOB protocol. We suppose that a single firm, the deviant firm, posts the same vacancies v^A as the others but deviates to a Nash protocol with continuous recontracting and a worker bargaining weight γ. We will derive the bargaining weight that the deviant firm will choose when it maximizes profits. Recall that Proposition 2 established the equivalence of any AOB protocol with particular parameters (p, ψ, δ) and the Nash bargaining protocol with a particular choiceof weights, namely, $\gamma = \gamma^A$ where $\gamma^A \equiv (W^A + P^A - U^A)/V^A$ is the worker's surplus under the AOB protocol. We now show that the profit maximizing γ chosen by the deviant firm does not equal γ^A.

In so doing, note from the viewpoint of the deviant firm, this change in protocol will lead workers to flow to it, but at the level of the economy, given we are focusing on the limit as the number of firms F grows large, it will have no affect on aggregate allocations, which will remain in the AOB steady state. In particular, at the choice-of-island stage, workers will flow to this deviant firm's Nash island until the value of searching on it, $\mathcal{W}_t^N$, equals the value of searching on the AOB island, $\bar{\mathcal{W}}^A$, that is, until

$$(40)\quad \mathcal{W}_t^N \equiv \lambda_w(\theta_t^N)\left(W_t^N + P^A\right) + [1-\lambda_w(\theta_t^N)]U^A = \bar{\mathcal{W}}^A \equiv \lambda_w(\theta^A)\left(W^A + P^A\right) + [1-\lambda_w(\theta^A)]U^A,$$

where $\theta_t^N = u_t^N/v^A$ and $\theta^A = u^A/v^A$. Now, after this one-shot deviation, all firms use AOB, so the *post-match value* at t, P^A, is that from the original AOB steady state and is given recursively by $P^A = \sigma\beta U^A + (1-\sigma)\beta P^A$, where the value of unemployment for all workers, regardless of which type of firm they choose, is $U^A = b + \beta\{\lambda_w(\theta^A)[W^A + P^A] + [1-\lambda_w(\theta^A)]U^A\}$. Note that the total surplus of the match $\tilde{V} = (\tilde{W} + P^A - U^A) + (Y - W^N) = Y + P^A - U^A = V^A$ is unaffected by the one-shot deviation. We then have

Proposition 6. (AOB is not production efficient) *In our endogenous protocol setup, a firm can increase profits by deviating from an inefficient AOB protocol to an efficient protocol.*

Proof. Consider the profit-maximization problem of a firm at the stage in which it posts a protocol and vacancies, which is equivalent to posting a protocol and market tightness. From Proposition 2, posting a protocol is equivalent to posting a worker bargaining share γ under Nash bargaining. The profits of the deviant firm at t is given by the product of the probability of finding a worker, $\lambda_f(\theta)$, and the share $1-\gamma$

of the surplus $\tilde{V}$ that accrues to the firm, namely,

$$(41) \quad \max_{\gamma,\theta} \lambda_f(\theta)(1-\gamma)\tilde{V},$$

where $\theta = u_t^N/v^A$, subject to the constraint that workers flow to this firm so as to satisfy (40), namely,

$$(42) \quad \lambda_w(\theta)\left(\tilde{W} + P^A\right) + [1-\lambda_w(\theta)]\,U^A \geq \bar{\mathcal{W}}^A.$$

Subtracting U^A from both sides of this constraint, dividing both sides by $\tilde{V}$ gives

$$\lambda_w(\theta)\frac{\tilde{W} + P^A - U^A}{\tilde{V}} \geq \frac{\bar{\mathcal{W}}^A - U^A}{\tilde{V}} \equiv K,$$

where K is a constant. Since the share of the surplus of workers who flow to this firm $\gamma\tilde{V}$, we can rewrite it as $\lambda_w(\theta)\gamma \geq K$. The first-order conditions for γ and θ for maximizing (41), subject to $\lambda_w(\theta)\gamma \geq K$ are

$$\gamma : \lambda_f(\theta)\tilde{V} = \chi\lambda_w(\theta) \text{ and } \theta : (1-\gamma)\lambda_f'(\theta)\tilde{V} = -\chi\left[\gamma\lambda_w'(\theta)\right]$$

where χ is the multiplier on (40). Dividing the first-order condition for θ by that for γ gives

$$(43) \quad (1-\gamma)\frac{\lambda_f'(\theta)}{\lambda_f(\theta)} = -\gamma\frac{\lambda_w'(\theta)}{\lambda_w(\theta)}.$$

Multiplying (43) by $-\theta$ and using $\eta(\theta) = -\theta\frac{\lambda_f'(\theta)}{\lambda_f(\theta)}$ and $1-\eta(\theta) = \theta\frac{\lambda_w'(\theta)}{\lambda_w(\theta)}$ implies $(1-\gamma)\eta(\theta) = \gamma\left[1-\eta(\theta)\right]$ or $\gamma = \eta(\theta)$. So the profit maximizing protocol for the deviant firm is Nash bargaining under the Hosios condition, namely to choose the worker's bargaining share γ to equal the elasticity of the matching function $\eta(\theta)$. *Q.E.D.*

This same logic implies that profit maximizing firms in our endogenous protocol game—for which the analog of Proposition 2 applies—will never choose an inefficient protocol because any such inefficient protocol has an effective Nash bargaining weight of $\tilde{\gamma} \neq \eta$ and the same proof applies.

5.3 The Costain-Reiter Critique

Costain and Reiter (2008) argue that in a large class of search models it is difficult to construct a model that is simultaneously consistent with business cycle facts and with the effects of labor market policies. Their

Table 4: Key Moments Across Variants of Model with Human Capital Accumulation and AOB

	Data	Alternative models					
On-the-job human capital growth rate g_e (% p.a.)	(1.50, 3.50)	0.00	0.50	1.50	2.50	3.50	4.00
Bargaining breakup probability δ		0.01	0.13	0.47	0.73	0.90	1.00
Potential bargaining duration $\frac{1}{\delta} - 1$ (months)		78.0	7.00	1.13	0.38	0.11	0.00
$E(R_{ft})$ (% p.a)	0.74	0.74	0.74	0.74	0.74	0.74	0.74
$\sigma(R_{ft})$ (% p.a)	2.35	2.35	2.35	2.35	2.35	2.35	2.35
$E(R_{t+1} - R_{ft})$ (% p.a)	7.05	7.13	6.90	6.88	6.87	6.87	6.86
$\sigma(R_{t+1} - R_{ft})$ (% p.a)	15.6	14.9	14.8	14.8	14.8	14.7	14.7
$\sigma(\Delta a_t)$ (%)	1.84	1.84	1.84	1.84	1.84	1.84	1.84
$\sigma(\lambda_{wt})$ (%)	6.67	6.67	6.67	6.67	6.67	6.67	6.67
Semi-elast. UC labor to unemployment	-5.32	-6.40	-5.50	-5.44	-5.35	-6.05	-6.30
S.d. effective bargaining power $\sigma(\gamma_t)$		7.80	4.89	2.66	1.72	0.46	0.00
$cov(\log(1-\gamma_t), \log(\lambda_{wt}))/var(\log(\lambda_{wt}))$ (%)		93.6	52.3	31.7	20.8	6.92	0.00
$cov(\log(V_t/A_t), \log(\lambda_{wt}))/var(\log(\lambda_{wt}))$ (%)		6.40	47.7	68.3	79.2	93.1	100
Consumption equivalent welfare gain (%)		0.69	0.02	0.01	0.01	0.00	0.00
Mean profit increase for deviating to efficient bargaining in recession (%)		16.6	0.94	0.86	0.79	0.00	0.00
Average elasticity of unemployment duration w.r.t. b	(0.36, 0.80)	20.8	2.20	0.72	0.46	0.29	0.25

basic idea is that once forces in the model are such that the model generates the substantial volatility of unemployment seen in the data, the model necessarily generates much too large a response to policies that address unemployment. We focus on the responses to one such policy that has been well studied, namely how much the duration of unemployment increases as the replacement rate of unemployment benefits increases.

The survey by Schmieder and von Wachter (2016) provides a comprehensive review of estimates of the elasticity of the duration of unemployment with respect to the replacement rate of unemployment benefits. We focus on their summary of studies for the United States, conducted on a large number of states as summarized in their Table 2, namely, studies by Moffitt (1985), Katz and Meyer (1990), Chetty (2008), and Kroft and Notowidigdo (2016). The estimates from these studies range from .36 to .80 with a mean of .58. For the baseline model this elasticity is 20.8, see first model column in Table 4, so that a five percentage point increase in the replacement rate of unemployment benefits leads to an increase in the duration of unemployment two years ($5 \cdot 20.8$ weeks).

6 A New Synthesis

We are sympathetic to the Hall and Milgrom (2008) and Hall (2017) view that since their alternating offer bargaining setup offers an explicit microfoundation for how bargaining is conducted, it is a more useful environment to study policy interventions than the reduced-form Nash bargaining setup because it is easier to argue that the underlying parameters describing the process are structural. To our taste, however, a major drawback of our quantitatively successful implementation of Hall's idea is that it relies too heavily

on inefficiency in order to generate volatile unemployment which in turn requires what seems like extremely long potential bargaining. Apriori one might think that a reasonable length of time for potential bargaining duration is a small fraction of the potential duration of the job. Instead, in the baseline parameterization of Hall (2017), in which jobs lasts about 2.5 years, the potential bargaining duration is over 6 years.

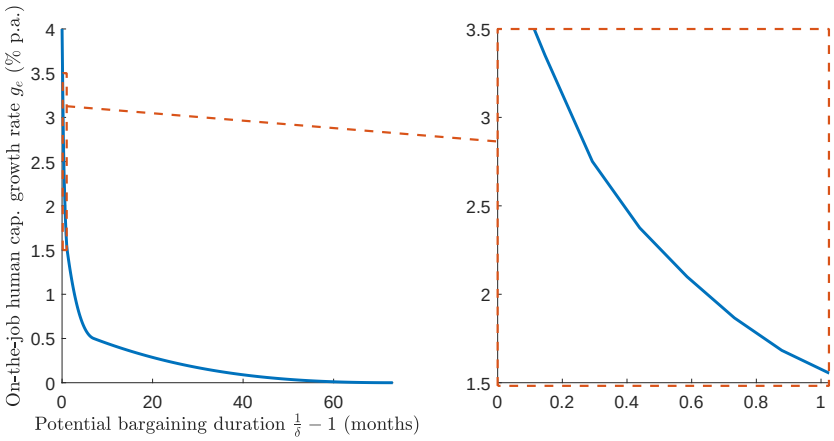
Motivated by these considerations, we integrate the best features of Hall (2017), the alternating offer setup, with another feature observed in the data, human capital accumulation. Our previous work, Kehoe, Midrigan, and Pastorino (2019) and Kehoe, Lopez, Midrigan, and Pastorino (2023), demonstrates that adding quantitatively reasonable amounts of human capital accumulation on the job can help match the volatility to the unemployment rate. Here we do exactly that and argue that the *synthesis model* is a more desirable than either the baseline model or our previous work which yield wages equivalent to those from axiomatic Nash bargaining under the Hosios' condition.

Our main results are that with even a modest amount of human capital accumulation this model can generate all the key moments that the baseline model does, but the level of inefficiency in terms of both the cost to average welfare and to the cost in firm profits is minimal. Moreover, it can overcome the Costain-Reiter critique: the model can generate both a volatile of unemployment in line with the data as well as responses to experiments such as an increase in unemployment benefits that are also in line with the data.

The only significant change in the integrated model is that it adds human capital accumulation on the job. In order to make the model amenable to matching statistics on life cycle wages we introduce a life-cycle in a tractable way. As before, each consumer belongs to one of a large number of families that insure their members against idiosyncratic risk. Each consumer survives from one period to the next with probability ϕ. A measure $1-\phi$ of new consumers is born each period so that the measure of consumers in the economy is constant over time and equal to one. Consumers' human capital accumulates during employment and depreciates during unemployment. Firms post vacancies to hire consumers with any desired level of human capital.

Specifically, each consumer is indexed by a state variable, z_t—*human capital*—that summarizes their ability to produce output. A consumer with human capital z_t produces $A_t z_t$ units of output when employed and $bA_t z_t$ units of output when unemployed with $0 < b < 1$ and the cost of posting a vacancy for a worker with human capital z_t is $\kappa A_t z_t$. Newly born consumers draw their initial human capital from a distribution $\nu(z)$ with mean one and enter the labor market unemployed. After entry, when a consumer is employed,

Figure 8: Locus of human capital and AOB parameters corresponding to baseline job-finding rate volatility



Note: The locus depicts the pairs (δ, g_e) that yield the same job-finding rate volatility as in the baseline model. The variants with positive human capital accumulation require death and rebirth; to preserve stability we assume a survival probability $\phi = 1/(1 + g_e)$. Across models, the cost of vacancy parameter is adjusted to keep the mean unemployment rate fixed.

human capital increases on the job at rate g_e so that according to $z_{t+1} = (1 + g_e)z_t$,where are constant rates of human capital accumulation on the job and depreciation off the job. Markets are segmented by the human capital level of workers z_t in that market—we think of z_t as indexing an island. Hence, each firm posts measures of vacancies of the form $v_t(z_t)$ to whatever market z_t it chooses. After all firms have posted vacancies, workers of type z_t randomly search on island z_t for the jobs from the firms that have posted there. The rest of the model is the same as our baseline model.

The only new parameters are the rate of growth of human capital g_e and the probability of survival, ϕ. We consider a range of growth rates for human capital growth rates g_e from 0 percent per year to 4% per year based on the existing range of empirical estimates.[11] We find it convenient to set ϕ so that $\phi(1 + g_e) = 1$. This way when $g_e = 0$ we nest the baseline model and this normalization leads to nearly identical numbers to when we set ϕ to be the same constant across models.

The synthesis model has two ways of generating volatile job-finding rates, a low δ as in Hall or a positive g_e or any combination thereof. To help us think about what these different combinations entail, we find a locus of pairs (δ, g_e) which generate the identical volatility of job-finding rates and then investigate what other statistics vary.[12] Figure 8 plots the locus of these pairs and shows the clear inverse relationship

[11] Specifically, Altonji and Shakotko (1987) estimate an annual wage growth of the job of 2% to 5% whereas Rubinstein and Weiss (2006 Table 2b) in their survey report estimates of average annual growth rates of nearly 6%. Subtracting an annual growth of productivity of about 2% gives a range for g_e of between 0% and 4%.

[12] Across this locus of models, the cost κ of posting a vacancy is adjusted so as to keep the mean unemployment rate fixed.

between the potential duration ($\frac{1}{\delta} - 1$) and the growth rates. Table 4 presents the key moments. First, note that across all of these models the volatility of the job-finding rate is equal to that in the data, 6.67%. The first three rows of the table give some feel for the how the magnitudes of (δ, g_e) compare along this locus. At one extreme, the baseline model produces this volatility with $g_e = 0$ and a potential duration of 78 months, while at the other if $g_e = 4$ then the potential duration is 0 months, so that all offers are take-it-or-leave it. Next notice that the asset pricing statistics are essentially identical in all of them. Moreover, in all of them the Kudlyak statistic is similar. So neither asset prices nor the Kudlyak statistic are of help for choosing between points on the locus.

Let us turn to features of the models that differ along this locus. Clearly, the standard deviation of a worker's bargaining power goes from very high with no human capital accumulation namely 7.80 and monotonically decreases to zero as g_e is increases from 0 to 4.0 percent per year. Along this locus we can decompose the source of the volatility in the job-finding rate into that coming from the volatility of the bargaining weight, $1 - \gamma_t$, and that coming from the volatility of the scaled surplus, V_t/A_t. To construct this decomposition we start from the free entry condition, written as

$$(44) \quad \kappa A_t = \lambda_f(\theta_t)(1 - \gamma_t)V_t$$

and use that with our matching function $\lambda_f = B\theta^{-\eta}$ and $\lambda_w = B\theta^{1-\eta}$. Setting $\eta = 1/2$ as in our quantification and performing some standard manipulations gives

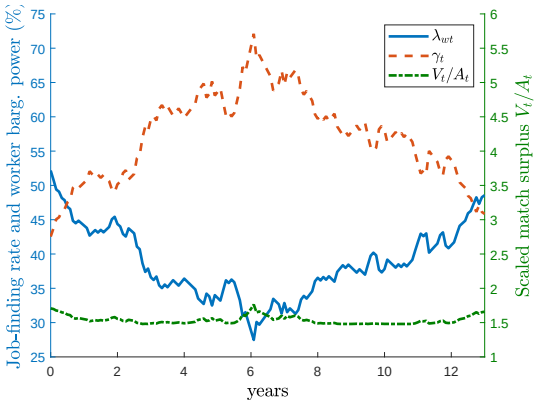
$$-\frac{cov(\log(1-\gamma_t), \log \lambda_{wt})}{var(\log \lambda_{wt})} - \frac{cov(\log \frac{V_t}{A_t}, \log \lambda_{wt})}{var(\log \lambda_{wt})} = 100\%$$

which is a decomposition of the fluctuations into the part due to the covariance of the firm bargaining weight $1 - \gamma_t$ with the job-finding rate, λ_{wt} and the covariance of the scaled surplus V_t/A_t with the job-finding rate. From Table 4 we see that in the case of no human capital accumulation ($g_e = 0$) only 6.4% come from fluctuations in the scaled surplus so that nearly all of the fluctuations the job-finding rate come from the fluctuations in the bargaining weight, 93.6%. As g_e is increased the fluctuations in the surplus account for a progressively larger share of total fluctuations.

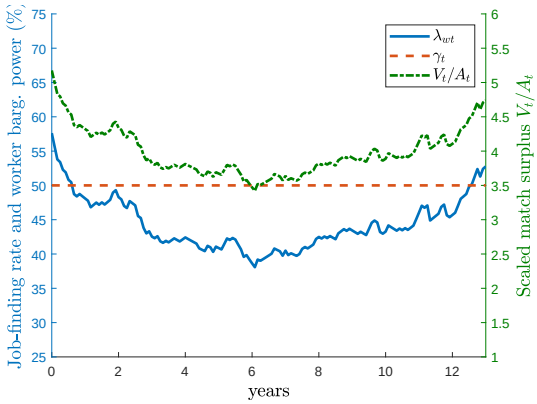
Figure 9 simulates a sequence of shocks that gives rise to a recession and shows the associated movements in the two determinants of the job-finding rate: worker bargaining power and the scaled matched surplus. In panels 9a and 9b we see the two extremes. With no human capital growth, the scaled surplus is

Figure 9: Job-finding rate in baseline model and version of synthesis model

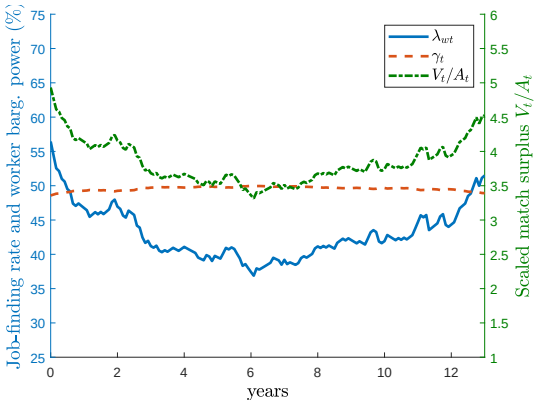
(a) *Implied determinants:* $g_e = 0$



(b) *Implied determinants:* $g_e = 4.0$



(c) *Implied determinants:* $g_e = 1.5$



(d) *Implied determinants:* $g_e = 3.5$

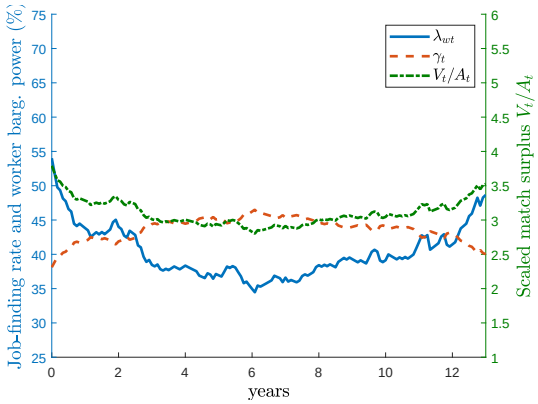
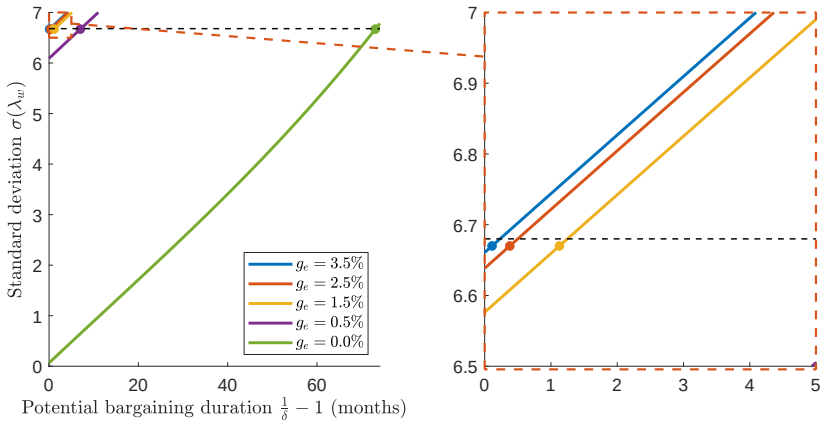


Figure 10: Volatility of job-finding rate as a function of potential bargaining duration and rate of human capital growth



Note: The circles indicate points on the locus of Figure 8.

essentially flat whereas the worker bargaining weight increases enormously from 0.42 in normal times to 0.72 during the recession. In contrast, in panel 9b with very fast human capital accumulation, the worker bargaining share is constant whereas the scaled surplus is very volatile, and tracks closely the job-finding rate. Panels 9c and 9d show intermediate cases.

Figure 10 shows for any given rate of human capital accumulation g_e, what the potential bargaining duration, $1/\delta - 1$, must be to generate the same standard deviation in the job-finding rate. For $g_e = 0.0$, as in Hall's baseline, this duration must be more than 6 years. As the right panel of this figure shows, for a rate of human capital accumulation in the empirically plausible range between 1.5 and 3.5, the needed duration falls to 1 month or less.

For the new synthesis we focus on the conservative range of estimates for g_e between 0.5 and 3.5. First notice from Table 4 that, in that range, the potential bargaining duration ranges from 7 months to about 3 days. Next, notice that the reasonable range for the Costain-Reiter elasticity of unemployment duration with respect to unemployment benefits is from 0.36 to 0.80. For the synthesis model to be consistent with that range, we need g_e between 1.5 and 3.5. So to be consistent with *both* the range of estimates for g_e and the Costain-Reiter statistic we need g_e to be in the range 1.5 to 3.5. In that range inefficiency is minimal.

7 Conclusion

The first generation DMP models built the theoretical foundations for search models of the business cycle on which all future work is based. They struggled, however, to produce volatile unemployment while being consistent with arguments in Chodorow-Reich and Karabarbounis (2016), Shimer (2010), and Kudlyak (2014). The pioneering work of Hall (2017) launched a promising second generation of search models but was subjected to several critiques. We have shown that the synthesis model which combines the most desirable aspects of Hall with an attractive stochastic discount factor and empirically modest amounts of human capital and is able to overcome not only these critiques but also the unappealing features of the baseline model. In this sense, the synthesis model helps to carry out Hall's original vision of second generation business cycle models.

References

Alvarez, Fernando, and Urban J. Jermann. 2005. Using asset prices to measure the persistence of the marginal utility of wealth. *Econometrica* 73.6: 1977-2016.

Barro, Robert J. 1977. Long-term contracting, sticky prices, and monetary policy. *Journal of Monetary Economics* 3.3: 305-316.

Basu, Susanto, and Christopher L. House. 2016. Allocative and remitted wages: New facts and challenges for Keynesian models. *Handbook of Macroeconomics*. Vol. 2. Elsevier: 297-354.

Becker, Gary S. 1962. Investment in human capital: A theoretical analysis. *Journal of Political Economy* 70.5, Part 2: 9-49.

Bils, Mark, Marianna Kudlyak, and Paulo Lins. 20123. The quality-adjusted cyclical price of labor. *Journal of Labor Economics* 41.S1: S13-S59.

Borovička,, Jaroslav, and Katarına Borovičková. 2018. Risk premia and unemployment fluctuations. Manuscript, New York University.

Borovička, Jaroslav, Lars Peter Hansen, and José A. Scheinkman. 2016. Misspecified recovery. *Journal of Finance* 71.6: 2493-2544.

Burdett, Kenneth, and Kenneth L. Judd. 1983. Equilibrium price dispersion. *Econometrica*, 955-969.

Campbell, John Y., and John H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107.2: 205-251.

Chodorow-Reich, Gabriel, and Loukas Karabarbounis. 2016. The Cyclicality of the Opportunity Cost of Employment". *Journal of Political Economy* 124 (6): 1563–1618.

Hagedorn, Marcus, and Iourii Manovskii. 2008. The cyclical behavior of equilibrium unemployment and vacancies revisited. *American Economic Review* 98.4: 1692-1706.

Hall, Robert E., Martin Neil Baily, Lawrence H. Summers, James Duesenberry, Franco Modigliani, William Nordhaus, James Tobin et al. 1980. Employment fluctuations and wage rigidity. *Brookings Papers on Economic Activity* no. 1: 91-141.

Hall, Robert E. 2017. High discounts and high unemployment. *American Economic Review* 107.2: 305-330.

Hall, Robert E., and Paul R. Milgrom. 2008. The limited influence of unemployment on the wage bargain. *American Economic Review* 98.4: 1653-1674.

Kydland, Finn and Edward C. Prescott. 1980. A Competitive Theory of Fluctuations and the Feasibility and Desirability of Stabilization Policy. In Rational Expectations and Economic Policy, edited by StanelyFischer, 169–98. University of Chicago Press.

Krusell, Per, Toshihiko Mukoyama, Richard Rogerson, and Ayşegül Şahin.2017. Gross worker flows over the business cycle. *American Economic Review* 107.11: 3447-3476.

Ljungqvist, Lars, and Thomas J. Sargent. 2017. The fundamental surplus. *American Economic Review* 107.9: 2630-2665.

Ljungqvist, Lars, and Harald Uhlig. 2015. Comment on the Campbell-Cochrane habit model. *Journal of Political Economy* 123.5: 1201-1213.

Mehra, Rajnish, and Edward C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15.2: 145-161.

Pissarides, Christopher A. 2009. The unemployment volatility puzzle: Is wage stickiness the answer? *Econometrica* 77.5: 1339-1369.

Ross, Steve. 2015. The recovery theorem. *Journal of Finance* 70.2: 615-648.

Shimer, Robert. 2005. The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review* 95.1: 25-49.

Wachter, Jessica. A. 2013. Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, 68(3), 987-1035.