Hysteresis through Endogenous Rigidity in Participation and Wages

Cynthia L. Doniger
Federal Reserve Board

David López-Salido
Federal Reserve Board

December 26, 2016

Abstract

A substantial body of empirical literature shows that high unemployment, falling wages, and reduced economic activity can have lasting consequences: hysteresis. We model hysteresis as resulting from a coordination failure among atomistic firms and workers in a frictional labor market that features random search, *ex-ante* wage commitments, and the possibility of worker non-participation. This coordination failure results in a continuum of possible equilibria with high-labor-demand equilibria welfare dominating low-labor-demand equilibria. We then introduce changes in labor productivity and specify a protocol for revelation of the new equilibria following shocks–*simultaneous best-response dynamics*–such that the model exhibits (1) periods of endogenous wage rigidity, (2) persistent changes in wages and labor force participation in response to transitory movements in labor productivity, and (3) sluggish recoveries including both a "jobless" phase and a "wageless" phase. Furthermore, expansions are insufficiently robust and in this sense recessions are scarring.

JEL Classification: E24, J42

Keywords: Hysteresis, Coordination Failure, Wage Rigidity, Labor Force Participation, "Job-less" recovery, "Wage-less" recovery.

Preliminary. This paper has previously been circulated as "Labor Demand Management in Search Equilibrium". We thank Susanto Basu, Russell Cooper, Marcus Hagedorn, Benjamin Johannsen, and Marianna Kudlyak for comments and Jason Sockin and Carter Bryson for excellent research assistance. We also thank participants in seminars at Cambridge University, Bank of Italy, EIEF Rome, ECB, Boston College, Philly Fed-Penn, the BIS, and FRB-NY. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or any other person associated with the Federal Reserve System. Comments and suggestions can be directed to `cynthia.l.doniger@frb.gov` and `david.lopez-salido@frb.gov`.

1 Introduction

A substantial body of economic literature shows that high unemployment, falling wages, and reduced economic activity can have lasting consequences. As such, an economic recession can lead to scarring–that is, a long-lasting damage of the economy. As Blanchard and Summers (1986a, 1987) suggested, hysteresis sets in. We examine these set of issues by building a model in which there is a coordination failure in wage setting and a corresponding multiplicity of equilibria. In our economy movements in labor productivity (stemming from TFP shocks, aggregate demand shocks, or other sources) can induce persistent changes in output, wages, labor force participation, and unemployment.

Key insights of our paper can be established using a stylized two-player game. One worker and one firm may match and produce. Before matching the worker will draw a flow value of leisure from a distribution known to both worker and firm. Conditional on the realization of the flow value of leisure, but before matching, the worker may decide to exit the game. The worker may take this option if she expects the wage offer to be lower than her flow value of leisure. We denote this decision as non-participation in the labor force. Further we will see that this behavior generates a "thick" side of the labor market and a "thin" side of the labor market. Also before matching the firm must commit to a posted wage with knowledge of the distribution of possible worker types but without knowledge of the worker's realization of the flow value. The firm's wage setting decision thus depends not only on the distribution of possible worker types but also on the firm's expectation for the worker's participation behavior. We show that this game exhibits positive spillovers and strategic complementarities.

Positive spillovers and strategic complementarities stem from a thick market externality denied to the firm through the worker's participation decision and a pecuniary externality denied to the worker through the firm's ex-ante wage posting. An increase in the posted wage–which is here the strategy of the firm–increases the payoff to the worker for all infra-marginal draws of the value of leisure through the pecuniary externality and increases the workers optimal participation threshold–the strategy of the worker. Conversely, an increase in the participation threshold–the strategy of worker–may increase the payoff to the firm via a thick-market externality and increases the optimal posted wage–the strategy of the firm. We specify an interval of wage and participation pairs on which a continuum of self-confirming rational expectations equilibria exist. These can be welfare ranked with high-wage and high-participation equilibria dominating low-wage, low-participation equilibria (Cooper and John, 1988).

The stylized assumptions of this two player game can be micro-founded by embedding the game in a frictional labor market in which a continuum of massless workers meet with a continuum of massless firms through a process of pair-wise random search. This supposition of a frictional labor market endows the firm with the studied market power in wage setting (monopsony) even for arbitrarily small market frictions (Diamond, 1971). Assuming massless agents and random search renders each worker (resp. firm) unable to affect the average participation threshold (resp. wage) through unilateral deviation. This supports modeling these aggregates as exogenous to each decision maker's problem even when expectations are informed by existing aggregate values as we shall consider. We close the model with the typical free-entry condition into vacancy posting. This provides a map from wages and participation levels to market tightness and unemployment rates.

Thus, in the full model, we are able to make assertions about the implications of moving between equilibria for tightness and unemployment in addition to the implications for output, participation, and wages already present in the simple two-player game. We follow a large part of the search and matching literature and assume the matching function exhibits constant returns to scale.[1] Still, match efficiency depends on the equilibrium selected, with more efficient equilibria leading to more efficient matching.

Finally, we subject this frictional labor market to stochastic shocks to labor productivity and impose the restriction that the status quo wage and participation pair persist unless

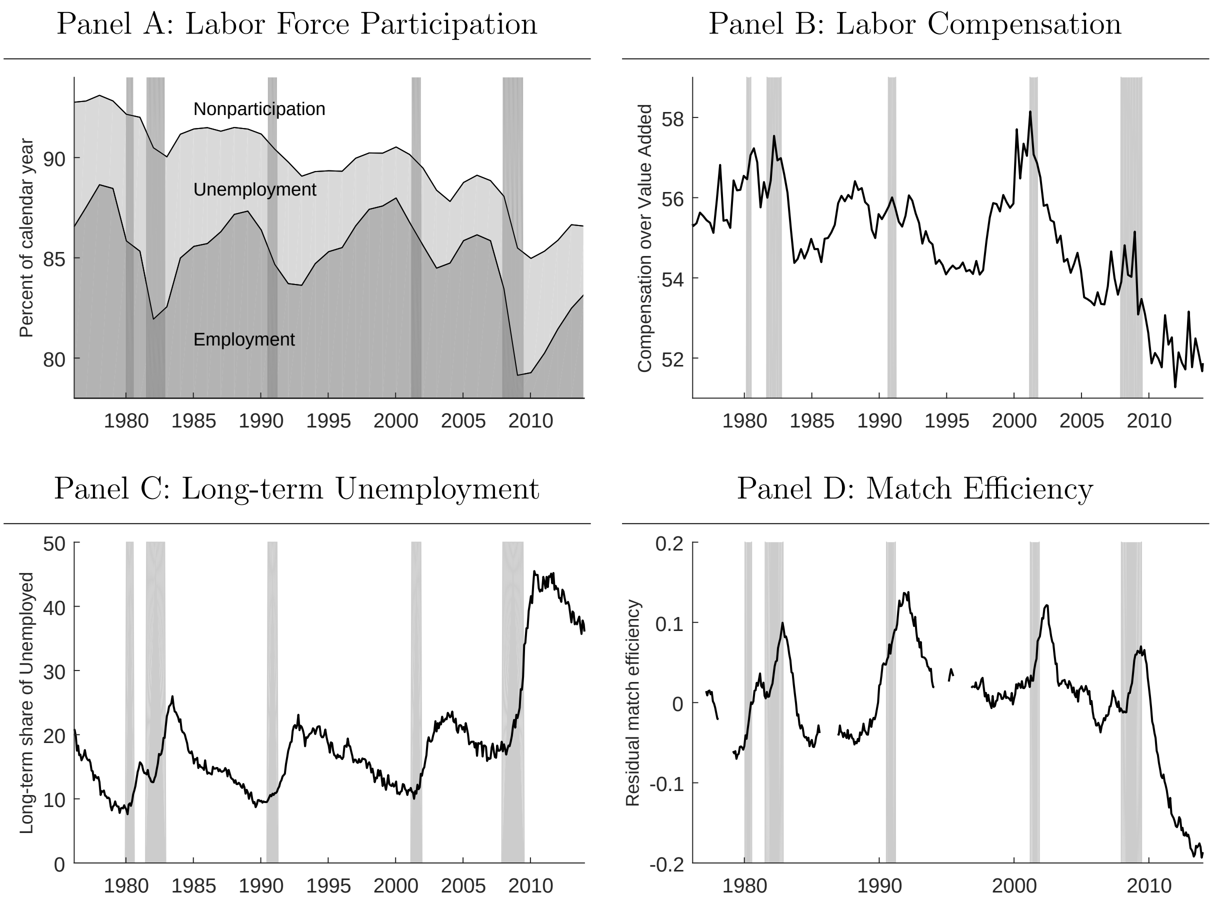
[1] It is important to note that for us this assumption is orthogonal to the assumptions necessary for multiplicity, unlike Diamond (1982).

deviation from this pair is the unilateral best-response of at least one player. Consistent with this assumption, movements between equilibria are governed by *simultaneous best-response dynamics* as in Vives (1990, 2005); Cooper (1994). This criteria imparts two usefull intuitions: first, equilibria are "sticky" unless at least one player has a unilateral incentive to deviate; second, the equilibrium selected following a deviation is "learnable" in the sense that any equilibrium is the conclusion of a rational and forward-looking algorithm on the part of every player.

We show that a corollary of the multiplicity of equilibria is that there are many productivity levels for which the same wage and participation threshold pair is an element of the equilibria. As a result, shocks within this range do not induce deviation from the status quo wage and participation threshold: an inaction range arrises in which wages and participation are rigid. The result has two important implications for the business cycle. First, contractions in the participation rate are associated with decreases in matching efficiency. This may lead to persistently elevated levels of unemployment. Second, wages exhibit endogenous stickiness. This may lead to a flatter relationship between the unemployment rate and wages when exiting a contraction than when entering a contraction. This enables us to generate both a "jobless-recovery" in the early phase and a "wageless-recovery" in the late phase of recovery from a severe economic contraction.

Further, recessions in our model are costly in terms of their "scarring" effects: following a severe contraction, economic activity in the post-recession economy is for a time less robust than in the pre-recession economy under identical fundamentals. This is because, in our model, the recent history of equilibria is also a determinant of equilibrium selection as a result of revealing equilibria through the *simultaneous best-response dynamics*: following any innovation in productivity each player considers her best deviation given the strategies all players played pre-shock. These may be summarized as the pre-shock wage and participation threshold. Further, during expansions firm's best deviation is to hire from the thin side of the labor market, which, we can show results in lower welfare than if the firm were

Figure 1: Medium-Run Stylized Facts
(Sample Period: 1975–2013)



Source A: Bureau of Labor Statistics, Current Population Survey, Annual Social and Economic Supplement following the methodology of Juhn et al. (1991, 2002); Murphy and Topel (1997) and Elsby and Shapiro (2012).
Source B: Bureau of Labor Statistics, Bureau of Economic Analysis, and author's calculations following the methodology of Elsby et al. (2013).
Source C: Bureau of Labor Statistics, Current Population Survey, Basic Monthly File.
Source D: Bureau of Labor Statistics, Current Population Survey; Conference Board, Help Wanted Index and Help Wanted Online Index; Bureau of Labor Statistics, Job Openings and Labor Turnover Survey. Vacancies are constructed from the Conference Board data as in Barnichon (2010) prior to 2001 and follow the Job Opening and Labor Turnover Survey thereafter.

not constrained by the differential in market thickness imposed by the worker's labor force participation decision. Thus, expansions are insufficiently robust regardless of their length or steepness and regardless of the severity of the contraction.[2]

Our theory accords well with four stylized facts in the medium-run data. First, over the

[2]Except in the trivial case that the distribution of worker types is bounded.

last fifty years, the U.S. economy has witnessed a persistent downward trend in participation in the labor force among prime-age males (e.g., Juhn et al. (1991), Murphy and Topel (1997), Juhn et al. (2002), and Elsby and Shapiro (2012)). Importantly for the present work, this trend takes the form of decline in participation during contractions without robust rebound in expansions. Second, over the same horizon, labor's share of output has eroded (Elsby et al., 2013; Karabarbounis and Neiman, 2014), reaching new lows during each expansion. Third, over the same horizon we have seen a shift toward long-term unemployment with the portion of newly unemployed likely to become long term unemployed spiking at recessions and not falling to pre-recession levels during expansions (e.g. Ahn and Hamilton (2016)). Forth, again over the same horizon, the aggregate matching function has become less efficient (e.g. Barnichon and Figura (2015); Diamond and Şahin (2015)) with a particularly precipitous drop following the financial crisis. Our theory offers declining labor force attachment as a complementary theory to labor market mismatch in accounting for this decline. Each of these is illustrated in Figure 1.[3]

Our paper also fits in an old and newly active literature regarding "fragile" labor market equilibria (Summers, 1992, pp. 332). These papers respond to the observation that seemingly small and/or seemingly transient macroeconomic shifts occasionally yield large and persistent shifts in the labor market. Stagnation of European labor markets in the 1980's and the "jobless recoveries" experienced by the United States after the 2001 and 2008 recessions are prime examples. These theories can be broken into two broad categories: (1) those that revolve around a bifurcation point: e.g. Diamond and Fudenberg (1989); Albrecht et al. (2013); Golosov and Menzio (2015) and Beaudry et al. (2015); and (2) those that posit criteria for switching between different equilibria: e.g. Blanchard and Summers (1986b); Eeckhout and Lindenlaub (2015); Kaplan and Menzio (2013) and Schaal and Taschereau-Dumouchel (2015).[4]

[3]Details regarding the construction of each data series and further observations about these medium-run facts are reserved for the Appendix.

[4]These are certainly non-exhaustive lists.

In the first type of paper, periodic crises are a permanent feature of equilibria characterized as orbits rather than a single steady state point. As such, these periodic crises sustain higher welfare equilibria in the non-crisis periods than would otherwise be attainable. For example, the "purge" period in Kaplan and Menzio (2013) sustains high effort in all other periods. A drawback to such theories, which Beaudry et al. (2015) are working to remediate by integrating stochastic aggregate shocks, is the rigidly defined periodicity of the crisis.

In the second type of paper, multiplicity of equilibria yield regions of downward-sloping labor supply, as in the insider-outsider model of Blanchard and Summers (1986b), or upward-sloping labor demand, as in the thick-market externality of Diamond (1982). Evidence that wages have stagnated in the episodes in question suggests that a theory of upward-sloping labor demand is more in line with the evidence.

Our setup can also be used to generate "animal spirits" type business cycles by means of sun-spots that induce players to correlate expectations on periodically more or less efficient equilibria. The distinction between history dependence and the role of expectations in selecting equilibria has been studied in the macroeconomics literature, with a focus on rational expectations. See, for example, Krugman (1991). The indeterminacy highlighted could be resolved by refining equilibria via global games (Morris and Shin, 2000). However, such refinement yields a one-for-one mapping between fundamentals and equilibria. Refining instead only the path of equilibrium using approach of Vives (1990, 2005); Cooper (1994) provides for history dependence in equilibrium selection and allows our model to generate hysteresis.

Although this is not a focus of our paper, endogenous wage stickiness due to our coordination failure also speaks to the Shimer (2005) puzzle. As such, it relates to a larger literature that seeks to induce amplification through persistence in the wage level: for example Hall (2005), Gertler and Trigari (2009), and Kennan (2010). We differ from this literature in our focus on producing hysteresis as opposed to only amplification. In particular, a strict focus on amplification will yield "wage-less" but "job-full" recoveries following particularly severe contractions. Our model can generate both a "job-less" and "wage-less" recovery–that is a recovery in which unemployment remains persistently high even after labor productivity has recovered and a recovery in which unemployment falls to unprecedented lows before wages recover, respectively. An implication is that the social planner can improve total welfare by *raising* the wage.

The rest of the paper is structured as follows. Section 2 describes the benchmark model and establishes that each player's strategy imposes positive spillovers on and is a strategic complement to the other's. Section 3 houses the basic intuition of the two player game in a frictional labor market and argues that search friction microfounds the assumptions of the baseline model. Section 4 investigates how the model interacts with transitory labor productivity shocks to produce business cycles typified by swift contraction and sluggish recovery. Section 5 concludes and offers suggestions of further avenues of investigation.

2 A Stylized Two Player Game

We begin by considering a stylized one-worker-one-firm labor market. The worker's objective is to maximize income. The firm's objective is to maximize profit. Solving the two player game requires positing a fixed expectation for the firm's posted wage on the worker's behalf–we will call the worker's expectation of the wage w_0–and for the worker's participation strategy on the firm's behalf–we will see that we can summarize the strategy as a participation threshold and we will call the firm's expectation of the threshold r_0. Thus, w_0 (resp. r_0) is the *belief* that the worker (resp. firm) holds about the firm's (resp. worker's) wage offer (resp. participation threshold). In this section we focus on demonstrating that in this game there exist multiple self-confirming rational expectations equilibria. Later, when describing the equilibrium selection, we posit an expectation formation process that will depend on the history of aggregate shocks.

2.1 Worker

For the worker the game proceeds in two stages. In the first stage the worker draws a value of leisure, b, from a known distribution, $H(b)$, with density $h(b)$, defined on $[\underline{b}, \bar{b}]$.[5] In the second stage the worker may costlessly seek to contact the firm.[6] If she does not seek to contact the firm we call this nonparticipation. If she does seek to contact the firm she has a $(1-u)$ probability of making contact and being made a wage offer.[7] We will see that the decision to participate will depend on her expectation of the wage offer, w_0. If the worker accepts this wage offer we call her employed. If she rejects we call her voluntarily unemployed. Finally, in the case that there is no contact between worker and firm we call her involuntarily unemployed.

As is typical for such a problem, the worker will accept the wage offer if it is in a specific sense "high enough". There are also cases in which the workers value of leisure is in a specific sense "high enough" that nonparticipation may be desirable. Also as is typical for such problems the worker's strategy takes the form of threshold rule: a reservation wage and a threshold for labor force participation. Since search is costless the reservation wage is equal to the flow value of leisure.[8]

Let $V^W(r, w_0)$ be the expected payoff to a worker of choosing threshold r when the

[5] We do not rule out $\underline{b} = -\infty$ or $\bar{b} = \infty$.

[6] The assumption of costless search is not innocuous. Under this assumption the reservation wage and threshold participation level coincide. We will see when we turn to the firm's problem that this coincidence is essential for multiplicity. Costly search results in a unique equilibrium with less participation than the minimal participation consistent with the set of equilibria we recover under costless search. Costly non-participation results in a unique equilibrium with participation exceeding the maximal participation consistent with the set of equilibria we recover under costless search. One could consider situations in which the cost of participation fluctuates around zero at cyclical frequencies, inducing switching between these equilibria. We note that evidence on job search behavior does not support substantial cost–see Mukoyama et al. (2016)–and leave this for future work.

[7] In Section 3, the unemployment rate will be pinned down in equilibrium via the aggregate matching function and a free entry condition. Importantly, however, under the assumptions imposed there–which are standard for the search and matching literature–the unemployment rate is independent of each worker's reservation wage and participation threshold choice (since we will assume that each is massless) and thus, each player's choice of participation or wage is independent of the unemployment rate.

[8] Note, there is no option value to search since this is a one-shot game. In Section 3, we will show that even in the dynamic game there is no option value to search. This will follow from noting that, in equilibrium, the wage distribution is degenerate.

expected wage choice of the firm is w_0. Observe that there are two cases for the value function:

$$V^W(r,w_0) = \begin{cases} \overbrace{(1-u)w_0H(r)}^{\textit{employment}} + u\overbrace{\int_{\underline{b}}^{r} bh(b)db}^{\text{involuntary unemployment}} + \overbrace{\int_{r}^{\bar{b}} bh(b)db}^{\text{non-participation}}. & \text{if } r < w_0 \\ \underbrace{(1-u)w_0H(w_0)}_{\textit{employment}} + \underbrace{u\int_{\underline{b}}^{w_0} bh(b)db}_{\text{involuntary employment}} + \underbrace{\int_{w_0}^{r} bh(b)db}_{\text{voluntary unemployment}} + \underbrace{\int_{r}^{\bar{b}} bh(b)db}_{\text{non-participation}}. & \text{if } r \geq w_0 \end{cases} \tag{1}$$

In the case that $r < w_0$, the first term encodes the payoff should the game result in employment times the probability of that outcome. The payoff in this case will be w_0. For this to happen the worker must draw a value of leisure less than or equal to r, which occurs with probability $H(r)$, and she must receive a wage offer, which occurs with probability $(1-u)$. The second term encodes the expected payoff in the various cases in which the worker could be involuntarily unemployed. Involuntary unemployment occurs if the worker draws a value of leisure less than or equal to r, which occurs with probability $H(r)$, and does not receive a wage offer, which occurs with probability u. The expected payoff in these cases is the expected flow value of leisure conditional on participating: $\frac{\int_{\underline{b}}^{r} bh(b)db}{H(r)}$. So we have: $uH(r)\frac{\int_{\underline{b}}^{r} bh(b)db}{H(r)} = u\int_{\underline{b}}^{r} bh(b)db$. The third term encodes the payoff in the various cases in which the worker non-participates. Inactivity occurs if the worker draws a value of leisure greater than r, which occurs with probability $1-H(r)$. The expected payoff in these cases is the expected flow value of leisure conditional on non-participating: $\frac{\int_{r}^{\bar{b}} bh(b)db}{1-H(r)}$. So we have: $(1-H(r))\frac{\int_{r}^{\bar{b}} bh(b)db}{1-H(r)} = \int_{r}^{\bar{b}} bh(b)db$.

In the case that $r \geq w_0$, the first term similarly encodes the payoff should the game result in employment times the probability of that outcome. The payoff in this case will again be w_0. Now, for this to happen the worker must draw a value of leisure less than or equal to w_0 and thus accept the wage offer, which occurs with probability $H(w_0)$, and she must receive a wage offer, which occurs with probability $(1-u)$. The second term encodes the expected payoff in the various cases in which the worker could be involuntarily unemployed. Involuntary unemployment occurs if the worker draws a value of leisure less than or equal to w_0 and thus would accept the wage if offered, which occurs with probability $H(w_0)$, and does not receive a wage offer, which occurs with probability u. The expected payoff in these cases is the expected flow value of leisure conditional on participating and being willing to accept: $\frac{\int_{\underline{b}}^{w_0} bh(b)db}{H(w_0)}$. So we have: $uH(w_0)\frac{\int_{\underline{b}}^{w_0} bh(b)db}{H(w_0)} = u\int_{\underline{b}}^{w_0} bh(b)db$. The third term encodes the payoff in the various cases in which the worker is voluntarily unemployed. Voluntary unemployment occurs if the worker draws a value of leisure greater than w_0 and less than r, which occurs with probability $H(r) - H(w_0)$. The expected payoff in these cases is the expected flow value of leisure conditional on a draw in this range: $\frac{\int_{w_0}^{r} bh(b)db}{H(r)-H(w_0)}$. So we have: $(H(r) - H(w_0))\frac{\int_{w_0}^{r} bh(b)db}{H(r)-H(w_0)} = \int_{w_0}^{r} bh(b)db$. The fourth term encodes the payoff in the various cases in which the worker non-participates. Inactivity occurs if the worker draws a value of leisure greater than r, which occurs with probability $1 - H(r)$. The expected payoff in these cases is the expected flow value of leisure conditional on non-participating: $\frac{\int_{r}^{\bar{b}} bh(b)db}{1-H(r)}$. So we have: $(1 - H(r))\frac{\int_{r}^{\bar{b}} bh(b)db}{1-H(r)} = \int_{r}^{\bar{b}} bh(b)db$.[9]

Differentiating expression (1), it follows that $\frac{dV^W(r,w_0)}{dr} > 0$ whenever $r < w_0$. In this case, the worker can improve her payoff by increasing her participation threshold. That is, a low threshold causes the worker fail to seek work when there is a positive probability of receiving an acceptable employment offer. One can also see that $\frac{dV^W(r,w_0)}{dr} = 0$ whenever $r \geq w_0$: in this case the payoffs for participation and nonparticipation are the same. The worker obtains the same value–the flow value of leisure–in voluntary unemployment as she does in nonparticipation. Thus, in the region of indifference ($r \geq w_0$) we may posit any mixed strategy over the pure strategies "voluntary unemployment" and "nonparticipate".

[9]Note that the participation threshold appears in the value function as a limit of integration. In what follows it will be useful to recall Leibniz's rule:

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,t)dt\right) = f(x, b(x))b'(x) - f(x, a(x))a'(x) + \int_{a(x)}^{b(x)} \frac{d}{dx} f(x,t)dt.$$

Assumption 1. *When the worker expects to be indifferent between voluntary unemployment and nonparticipation she randomises between the two states with probability i placed on voluntary unemployment.*

We offer a *heuristic* justification for this assumption. Specifically, the worker holds lexicographic preferences. Whenever participation yields an increase in the worker's expected value of income, the worker strictly prefers to participate. However, when participation and nonparticipation yield the same expected income, the worker considers her "optimism" or "pessimism." With probability i she is "optimistic" and searches for work even when she is certain that the job offer will be unacceptable. With probability $(1-i)$ she is "pessimistic" and does not search for work when she anticipates that the offer will be unacceptable. Under this heuristic, i is pinned down by the probability that the worker is endowed with optimism.[10] We find this heuristic justification useful for two reasons: First, this gives a strict preference between participation and nonparticipation when the realization of the value of leisure is high. Second, under this heuristic we classify the worker as unemployed in the $i(1-u)$ case in which she participates and rejects a job offer. Thus, in the full model of a two-sided frictional labor market a worker endowed with optimism will persistently search. This gives us an opportunity to see a reflection of r_0 and i in data: differentials rates of long term unemployment at cyclical and secular frequencies.[11]

Under Assumption 1, the workers best response, $r^*(w_0)$, is to set a threshold:

$$r^*(w_0) = w_0, \tag{2}$$

such that she participates with probability one if $b < r^*(w_0)$ and with probability $i < 1$ if $b > r^*(w_0)$.

[10] For the purposes of this paper we take i to be exogenous.

[11] It is also possible that the worker adopts a more complex randomization strategy. In particular one might posit that the weight placed on "voluntary unemployment" might depend on the realization of b. As it will be clear, any randomization such that the probability of nonparticipation exceeds zero as the realized flow value of leisure approaches the expected wage level from above will yield qualitatively similar results.

Remark 1. *This game exhibits* ***positive spillovers*** *and* ***strategic complementarities*** *for the worker.*

An increase in the posted wage–which we will see is the strategy of the firm–increases the payoff for the worker regardless of worker's participation threshold strategy (***positive spillover***).[12] Also, an increase in the posted wage–which we will see is the strategy of the firm–increases the optimal participation threshold strategy of the worker (***strategic complementarity***).[13]

Remark 2. *The worker's threshold strategy is independent of the probability, i; the distribution of the value of leisure, $H(b)$; and the marginal product of labor.*

This last result will be useful when considering equilibrium selection in response to shocks to the fundamentals of the economy.

2.2 Firm

The firm posts a wage offer ex-ante, w, and may meet the worker with probability $(1-v)$.[14] If the firm successfully hires the worker it will earn rent $p-w$, where p is the marginal product of labor. If the firm and the worker do not meet or if the worker rejects the wage offer then the payoff to the firm is zero.[15]

[12]Formally, define ***positive spillover*** for player j as the case in which an increase in the other player's strategy increases the payoff to player j (Cooper and John (1988)). Now simply note that, from expression (1) it follows that $\frac{dV^W(r,w)}{dw} = (1-u)H(r) > 0$ if $r < w$ and $\frac{dV^W(r,w)}{dw} = (1-u)H(w) + uwh(w) > 0$ if $r > w$.

[13]Formally, we define ***strategic complementarities*** for player j as the case when an increase in the other player's strategy increases the best response of player j (Cooper and John (1988)). This trivially follows from expression (2), that is $\frac{dr^*(w)}{dw} = 1 > 0$.

[14]In Section 3, the vacancy rate will be pinned down in equilibrium via the aggregate matching function and a free entry condition. Importantly, however, under the assumptions imposed there–which are standard for the search and matching literature–the vacancy rate is independent of each firm's posted wage choice (since we will assume that each is massless) and thus, each player's choice of participation or wage is independent of the unemployment rate.

[15]In this simple two-player game the monopsonistic wage setting of the firm imposes inefficiency: some rent-producing matches are not formed as the firm restricts hiring in order to capture rents through the wage markdown. We will see, however, that in the full model including labor market frictions the rents that the monopsony expects to capture finance the cost of vacancy posting. Without monopsony rent no vacancies will be posted.

Let $V^F(w, r_0)$ be the payoff to the firm of posting wage w and let $w^*(r_0)$ be the best response of the firm when the firm's expectation of the worker's threshold for participation is r_0. We can write the payoff to the firm of selecting wage w when the expected participation threshold of the worker is r_0:

$$V^F(w, r_0) = \underbrace{(1-v)}_{\substack{\text{match}\\\text{probability}}} \underbrace{(p-w)}_{\substack{\text{revenue}\\\text{from}\\\text{successful}\\\text{hiring}}} \times \underbrace{\begin{cases} \frac{H(w)}{H(r_0)+i\bar{H}(r_0)} & \text{if } w < r_0 \\ \frac{H(r_0)+i[H(w)-H(r_0)]}{H(r_0)+i\bar{H}(r_0)} & \text{if } w \geq r_0 \end{cases}}_{\text{expected labor supply schedule}} \tag{3}$$

As noted above, the first term, $(1-v)$, and the second term, $(p-w)$, are the probability of meeting the worker and the payoff from successfully hiring, respectively.

The third term of expression (3) contains the core of the problem: this encodes the firm's expectation of the labor supply schedule it faces. In the present problem, labor supply takes the form of the probability that a wage offer of w will be accepted if the firm does indeed meet the worker. Notice that the participation threshold introduces a kink in the expected labor supply at the expected threshold. To the left of the threshold, where $w < r_0$ the labor market is "thick": the probability that the worker participates for draws of the value of leisure in this region is one. Thus, posting a wage less than or equal to the expected labor force participation threshold results in hiring a worker with probability $\frac{H(w)}{H(r_0))+i(1-H(r_0))}$.[16] To the right of the threshold the labor market is "thin": the probability that the worker participates for draws of the value of leisure in this region is only $i < 1$. Thus posting a wage greater than the expected labor force participation threshold results in hiring a worker with probability $\frac{H(r_0)+i[H(w)-H(r_0)]}{H(r_0))+i(1-H(r_0))}$.[17]

The firm's best response, i.e. its optimal ex-ante posted wage, for given a particular

[16] The denominator counts all of the cases in which the worker might participate and the numerator counts the fraction of these cases in which the worker accepts a wage offer equal to w.

[17] Again, the denominator counts all the cases in which the worker might participate and meet the firm and the numerator counts the fraction of these cases in which the worker accepts the wage offer.

expectation for the worker's strategy, $w^*(r_0)$, satisfies:

$$w^*(r_0) = \arg\max_w \left\{ V^F(w, r_0) \right\}.$$

As is usual for a wage-posting firm, the problem takes the form of monopsony wage setting. The firm understands and internalizes the impact of it's wage choice on labor supply. Our problem is slightly more complex due to the expectational element of the game. As noted, the firm's expectation over the worker's participation threshold introduces a kink in the expected labor supply schedule. As a result the best response function–the optimal wage posting strategy of the firm–is defined piecewise on intervals of the expectation for the worker's participation threshold. Our task is to determine these intervals and the optimal posted wage schedule within each of them. We distinguish three cases.

Thick market

A first case is that the firm expects that the participation threshold is high enough that it is non-binding ($w < r_0$), in other words the firm expects that it will be able to hire the marginal worker from the thick side of the market. Call the minimum non-binding threshold r^C, then $V^F(w, r_0 | r_0 \geq r^C)$ simplifies to $(1-v)(p-w)\frac{H(w)}{H(r_0))+i(1-H(r_0))}$ and, in this side of the market, the firm's best response satisfies:

$$w^C = \arg\max_w \left\{ V^F(w, r_0 | r_0 \geq r^C) \right\} = \arg\max_w \left\{ H(w)(p-w) \right\},$$

where the final equality follows from noting that, in expression (3), $(1-v)$ and $H(r_0)+i[1-H(r_0)]$ are exogenous to the firm. The first order condition can be written as follows:[18]

$$\underbrace{p}_{\substack{\text{marginal}\\\text{revenue}}} = \underbrace{w^C + \frac{H(w^C)}{h(w^C)}}_{\substack{\text{marginal}\\\text{cost}}} \tag{4}$$

The left hand side is the marginal revenue from an employee. The right hand side is the marginal cost of hiring the worker who will accept wages no less than w^C from the thick side of the market. A particular useful form of representing the firm's best response is that the marginal cost of labor is a markup over the wage, the markup is given by the elasticity of the (expected) labor supply faced by the firm. That is, expression (4) can be written as $\frac{p-w^C}{w^C} = \frac{H(w^C)}{w^C h(w^C)} = \frac{1}{\eta^{thick}}$, where η^{thick} is the elasticity of the expected labor supply curve in the thick side of the market. Notice that, $\eta^{thick} = \eta$, where $\eta = \frac{dH}{dw}\frac{w}{H}$ denote the elasticity of labor supply with respect to wages would pertain if the worker participates for all values of b. Thus, since as assumed the participation constraint is not binding in the thick market, this elasticity corresponds to the elasticity of the (unconstrained) firm's expected labor supply schedule.

The wage choice, w_C, is illustrated in the left panel of Figure 2. The x-axis plots the probability that a worker will participate. From the firm's perspective this is the labor supply schedule. The y-axis plots wages. To simplify the diagram we take $H(b)$ to be uniformly distributed on $[\underline{b}, \bar{b}]$. We plot both the underlying distribution of worker types and the labor supply schedule that the firm expects to face given the worker's participation threshold in gray hashed and solid respectively.[19] Note that the expected labor supply schedule is

[18]In this case we seek an interior solution on the thick side of the market so we search for a solution in which the first order condition holds with equality. Moreover, a sufficient condition such that w^C is the unique maximizer is given by: $\frac{1}{p-b} > \frac{d^2H(b)}{db^2}\frac{1}{2h(b)}$. This states that the distribution $H(b)$ is not more convex than the hyperbola defined by the firms iso-profit curves: $\frac{1}{p-b}$. This is trivially satisfied whenever the density is weakly decreasing – for example, the uniform and Pareto distributions as well as by the normal and logistic distributions when at least half of workers participate in the labor force. The remainder of this paper focuses on the case where the second order condition holds.

[19]Note that the underlying distribution of worker types is simply the cumulative distribution $H(b)$. Note also that both schedules are censored and become vertical at $H(\bar{b}) = 1$.

rotated counterclockwise relative to the underlying distribution of worker types around the probability that the worker is type r_0 or less, $H(r_0)$, creating a kink in the labor supply curve. This reflects that the market is segmented into a thick side to the left of the expected participation threshold and a thin side to the right.[20] The angle of rotation depends on the probability that the worker is optimistic, i.

We also plot the marginal revenue and marginal cost faced by the firm. Marginal revenue is exogenous, p. As in the typical monopsony problem, we find the optimal quantity of labor demanded at the intersection between marginal revenue and marginal cost. In the present case the firm is unconstrained by the worker's expected participation choice and the intersection of marginal cost and marginal revenue fall to the left of the kink in the expected labor supply curve. We will return to discussion of the kink presently. The quantity of labor demanded equals and the wage the markdown are determined as in a normal monopsony diagram. Note that $w_C < r_0$, we will return to this point when we consider equilibria defined as mutual best response.

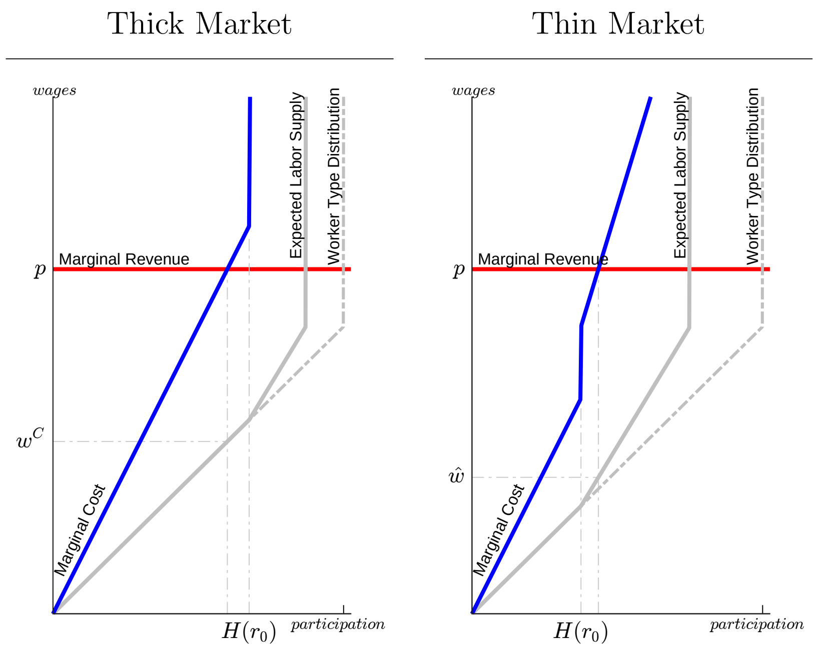
Thin market

A second case is that the firm expects that the participation threshold is low enough so it is binding and consequently the firm expects that will hire the marginal worker from the thin side of the market. In this case:

$$\hat{w} = \arg\max_{w} \left\{ V^F(w, r_0 | r_0 < r^C) \right\} = \arg\max_{w} \left\{ [H(r_0) - i[H(w) - H(r_0)]](p - w) \right\},$$

[20] Note that the angle of rotation will be higher as the fraction of worker choose voluntary unemployment over non-participation is closer to zero. For $i = 0$ it becomes vertical, and for $i = 1$ there is no rotation.

Figure 2: Optimal Wage Choice: Interior Solutions.



Note: The left panel illustrates an expected threshold such that the firm's best response is in the thick side of the market. The right panel illustrates an expected threshold such that the firm's best response is in the thin side of the market. The location of the kink in the expected labor supply curve depends on the firm's expectation for the worker's participation threshold, r_0. The angle of rotation and the associated jump in the marginal cost curve depend on the probability that the worker searches when her value of leisure exceeds this threshold, i. Fore ease of illustration we take $H(b)$ to be uniformly distributed on $[\underline{b}, \overline{b}]$.

and the first order condition is given by:[21]

$$\underbrace{p}_{\substack{\text{marginal}\\ \text{revenue}}} = \underbrace{\hat{w}\left[1 + \frac{H(r_0) + i[H(\hat{w}) - H(r_0)]}{ih(\hat{w})\hat{w}}\right]}_{\substack{\text{marginal}\\ \text{cost}}} \tag{5}$$

Again, the left hand side is the marginal revenue from an employee. The right hand side is the marginal cost of hiring the worker who will accept wages no less than $\hat{w}$ from the thin side of the market. Manipulating equation (5) one can show that, hiring in this side of the market, the monopsony markup is equal to $\frac{p-\hat{w}}{\hat{w}} = \frac{H(r_0)+i[H(\hat{w})-H(r_0)]}{ih(\hat{w})\hat{w}} = \frac{1}{\eta^{thin}}$, where η^{thin}

[21] In this case we seek an interior solution on the thick side of the market so we search for a solution in which the first order condition holds with equality.

is the elasticity of the expected labor supply curve in the thick side of the market. Note that $\eta^{thin} = \eta \left[1 + \frac{H(r_0)}{H(w)} \frac{1-i}{i}\right]^{-1}$. Thus, the elasticity of the expected labor supply on the thin side of the market is strictly lower than the elasticity that would pertain if all workers participated (thick market) whenever $i < 1$ and $r_0 \geq \underline{b}$.[22] In other words, the markdown on the thin side of the market strictly exceeds the markdown that the firm would choose if it were not constrained by the expected participation threshold.

The wage choice, $\hat{w}$, is illustrated in the right panel of Figure 2. Now that we are considering a firm hiring from the thin side of the market and the kink induced by the rotation of the expected labor supply schedule relative to the underlying distribution of worker types is binding. Again, note that the expected labor supply schedule is rotated counterclockwise relative to the underlying distribution of worker types around the probability that the worker is type r_0 or less: $H(r_0)$. This reflects that the market is segmented into a thick side to the left of the expected participation threshold and a thin side to the right. The angle of rotation will be higher as the fraction of worker choose voluntary unemployment over non-participation is closer to zero. For $i = 0$ it becomes vertical, and for $i = 1$ there is no rotation.

We again plot the marginal revenue and marginal cost faced by the firm. Marginal revenue is exogenous, p. The slope of the marginal cost of hiring is lower in the thick market. At the quantity of labor supplied at the expected participation threshold the marginal cost jumps from marginal cost of hiring the $H(r_0) - \varepsilon$ worker to the marginal cost of the $H(r_0) + \varepsilon$ worker due to the kink in the expected labor supply curve. Again, as in the typical monopsony problem, we find the optimal quantity of labor demanded at the intersection between marginal revenue and marginal cost. The quantity of labor demanded equals the quantity of labor supply at the participation threshold and the wage the markdown are determined as in a normal monopsony diagram. Note that $\hat{w} > r_0$, we will return to this point when we consider equilibria defined as mutual best response.

[22] Note that η^{thin} is not well defined for $i = 0$ and $r_0 = \underline{b}$.

Corner solution

In this third case, the constraint imposed by the expected participation threshold is binding and it induces a corner solution (i.e., $w = r_0$). The marginal revenue strictly exceeds the marginal cost in the neighborhood of the participation threshold on the thick side of the market while the marginal cost strictly exceeds the marginal revenue on the thin side of the market:[23]

$$p > r_0^- + \frac{H(r_0^-)}{h(r_0^-)} \text{ and } p < r_0^+ + \frac{H(r_0^+)}{ih(r_0^+)}.$$

Notice also that, when $i < 1$, the elasticity of the expected labor supply with respect to the wage at the expected labor supply threshold is discontinuous: $\eta^{thin}(r_0^+) < \eta^{thick}(r_0^-)$. There is a wedge between marginal costs on the thick and thin side of the market stemming from that discontinuity.[24] Then, the optimal wage is a corner solution at $w^*(r_0) = r_0$.

What remains is to find the expected participation threshold such as the firm weakly prefer to hire the marginal worker from the thin side. In other words, the lowest participation threshold for which the corner solution is consistent. Call it r^L:

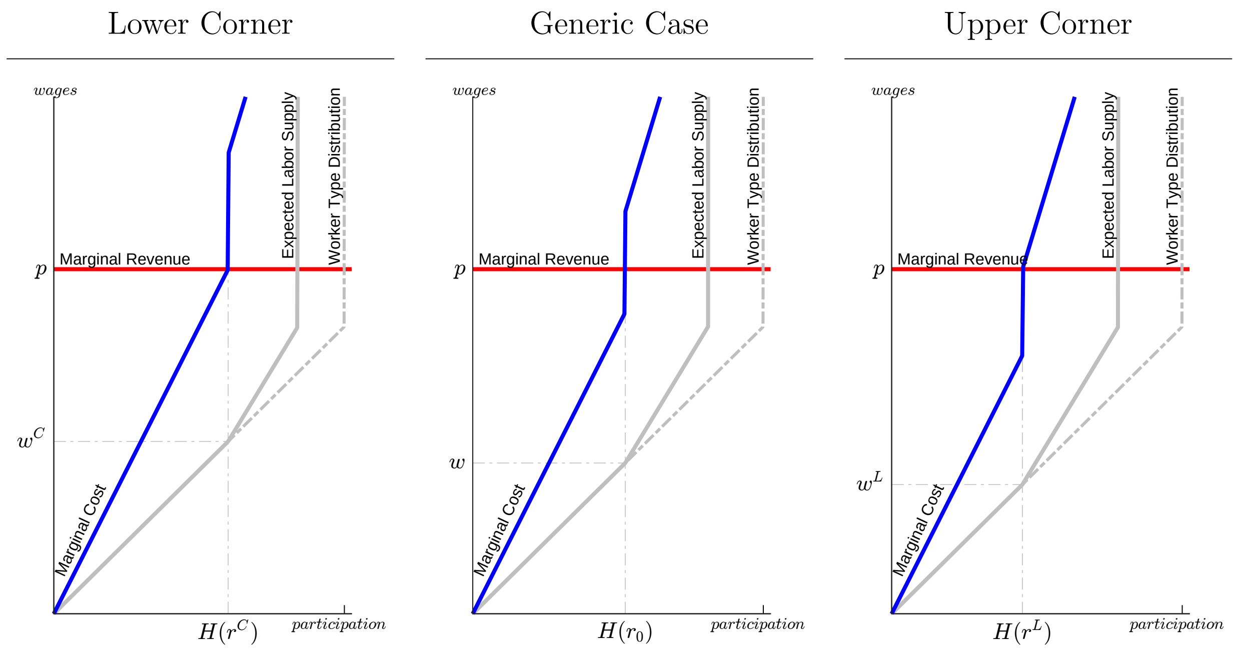
$$p = r^L + \frac{H(r^L)}{ih(r^L)}.$$

Figure 3 plots the range of participation thresholds for which the corner solution is optimal for the firm. The left panel plots the smallest threshold for which the firm optimally hires from the thick side of the market: $r_0 = r^C = w^C$. The right panel plots the largest threshold for which the firm optimally hires from the thick side of the market $r_0 = r^L = w^L$. The center panel plots a generic corner solution when $r^L < r_0 < r^H$ and the firm neither prefers to hire from the thick nor the thin side of the market. The wage is again found by

[23]Formally, taking the derivative from the left, $\lim_{w \nearrow r_0} \frac{dV^F(w,r_0)}{dw}$, and from the right, $\lim_{w \searrow r_0} \frac{dV^F(w,r_0)}{dw}$, respectively.

[24]For a given r_0 the size of the jump in the marginal cost depends on i, the of workers with $b \geq w_0$ who search for work: when all workers search then there is no difference in the thickness between both segments of the markets and the marginal cost is smooth; and when all workers with $b \geq w_0$ non-participate, the marginal cost of hiring from the thick side of the market is infinite.

Figure 3: Optimal Wage Choice: Corner Solutions.



Note: The left panel illustrates the smallest threshold for which the firm optimally hires from the thick side of the market. The right panel illustrates the largest threshold for which the firm optimally hires from the thin side of the market. The center panel illustrates a generic corner solution. The location of the kink in the expected labor supply curve depends on the firm's expectation for the worker's participation threshold, r_0. The angle of rotation and the associated jump in the marginal cost curve depend on the probability that the worker searches when her value of leisure exceeds this threshold, i. Fore ease of illustration we take $H(b)$ to be uniformly distributed on $[\underline{b}, \bar{b}]$.

evaluating the expected labor supply schedule at H_0 and we have $w \in (w^L, w^C)$.

We can now write the firm's wage best response (i.e., wage posting strategy) as follows:

$$w^*(w_0) = \begin{cases} \hat{w} & \text{if } r_0 < r^L \\ r_0 & \text{if } r_0 \in [r^L, r^C] \\ w^C & \text{if } r_0 > r^C. \end{cases} \tag{6}$$

Remark 3. *For all expected participation thresholds, r_0, in the interval $[r^L, r^C]$ the game exhibits **positive spillovers** and **strategic complementarities** for the firm.*

Whenever the marginal searching worker would accept the posted wage offer an increase in participation increases the pay off to the firm (***positive spillover***).[25] Moreover, whenever $r \in [r^L, r^C]$ an increase in the participation threshold increases the optimal posted wage strategy of the firm (***strategic complementarity***).[26]

Remark 4. *The firm's strategy depends on the probability that the worker chooses voluntary unemployment over non-participation, i; the distribution of the value of leisure, $H(b)$; and firm's productivity, p.*

These are determinants of r^L, r^C, and $\hat{w}$. This result will be useful when considering equilibrium selection in response to shocks to the fundamentals of the economy.

Remark 5. *Note that from expressions (4) and (5), $\hat{w} \geq r_0$ if $r_0 < r^L$ and $w^C \leq r_0$ if $r_0 > r^C$.*

This will be important when we consider below equilibria defined as mutual best response.

2.3 Self-Confirming Rational Expectations Equilibria

Definition 1. *A **self-confirming rational expectation equilibrium** of the two-player game is a pair–wage and participation threshold–such that wage and participation threshold are mutual best responses.*

In every equilibrium wage and participation threshold equal: $w^* = r^*$. This follows directly from the best response functions. In the remainder of the paper we will refer to equilibrium strategies, $\{w^*, r^*\}$, simply by the wage.

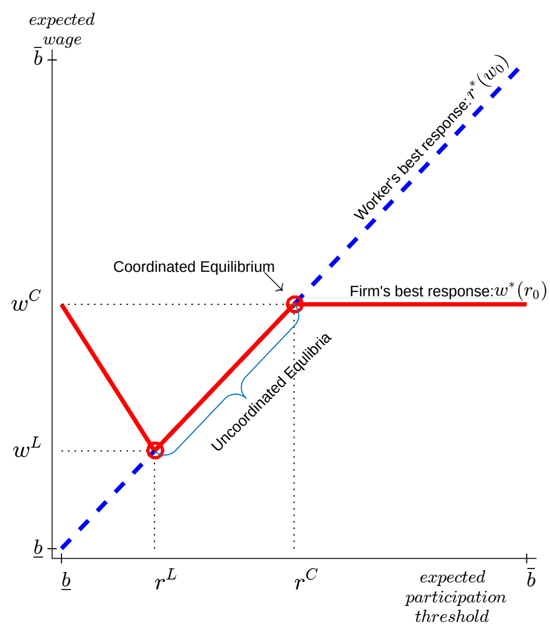
Proposition 1. *For $i < 1$, a continuum of equilibria exists on the interval $[w^L, w^C]$, with higher welfare for higher wage levels.*

Proof. Since we have already shown that both the firm and worker face positive spillovers and strategic complementarities on this interval the result is a straightforward application

[25]Formally, $\frac{dV^F(w,r)}{dr} = (1-v)(p-w)\frac{h(r)(1-i)(i(1-H(w)])}{[H(r)+i(1-H(r)]^2} > 0$ whenever $w \geq r_0$.
[26]Formally, $\frac{dw^*(r)}{dr} = 1 > 0$ whenever $r \in [r^L, r^C]$.

of Cooper and John (1988) Proposition 5. Note that for $i = 1$, w^C is the unique equilibrium wage. □

Figure 4: Mutual Best Responses.



Note: A single coordinated equilibrium corresponds to the wage and participation threshold that would be selected if the firm and worker could coordinate on the monopsony wage level. To the southwest a continuum of additional uncoordinated equilibria are supported by agents self-confirming beliefs about each other's strategies when coordination is not possible.

Figure 4 illustrates the best response function of the worker (hashed) and the best response of the firm (solid). These are mutual best responses on the 45 degree line from (r^L, w^L) to (r^C, w^C). In any equilibrium, randomly formed expectations must be self-confirming, that is $w^* = w_0$ and $r^* = r_0$. Consider the following: suppose the worker expects wages to be low. Her best response is to set a low participation threshold. The firm's best response is then to confirm the expectation by setting a low wage, as long as the original expectation is higher than w^L. Now consider the inverse of the problem: the firm expects the participation threshold to be low, as long as the original expectation is higher than r^L, its best response is to set a low wage. The worker's best response is then to confirm the expectation by setting a low threshold. Thus, the expectations that support equilibria are self-confirming while off equilibrium expectations fail to be confirmed.[27] As such, the expectations over wages and participation are self-confirming and rational in every equilibrium.

Coordination. As labeled in Figure 4, we can categorize equilibria into two sets: coordinated and uncoordinated equilibria. We have considered expectations on the posted wage, w_0, and the participation threshold, r_0, that are formed in some sense randomly. Now consider the following sense of coordination: augment the game to allow the worker and the firm to meet prior to playing the game in order to form their expectations. Given that our game has both positive spillovers and strategic complementarities there exists a subset of equilibria that welfare dominate. Thus, in the pre-meeting the worker and firm gain from coordinating their expectations on wage and participation pairs consistent with the dominating equilibrium. We label this as the coordinated equilibrium, and it corresponds to the standard monopsony in which the firm's hire from the thick side of the market. The remaining equilibria exist only as equilibria of the game in which the pre-meeting is not allowed and expectations are formed randomly and subsequently self-confirmed.[28]

At this point, one might be tempted to impose an equilibrium refinement with a "trembling hand" flavor (Selten, 1975). This temptation arises from noting that if there is non-zero probability that wage offer exceeds the worker's expectation then the threshold must be raised to exceed the support of the distribution of possible wage offers. Thus, if there is any chance that the firm's hand "trembles" and it may make a mistake in some off equilibrium path and post a high wage such a consideration could be used to rule out all equilibria other than the constrained efficient. Such a consideration is a valid criticism of our two-player game. However, we have devised this simply to illustrate clearly the mechanism at work driving multiplicity in our model. In the full labor market that we want to consider, and which

[27]It is easy to construct an analogous heuristic example illustrating that (off equilibrium) expectations above w^C or below w^L are not confirmed.

[28]Note that coordination requires only strategic complementarity–consider the game "which side of the road to drive on"–however with positive spill-over one can institute coordination by allowing the worker and firm to agree to *share the surplus* in the pre-meeting.

we will present in Section 3 one can see that such a refitment is less plausible as it would require not just that a single firm's hand "trembles" but that a discrete mass of firms all simultaneously mistakenly play an off-equilibrium wage. We suggest that such a coordinated mistaken deviation from equilibrium is too implausible to impose as a refinement.

We now define a mechanism for equilibrium revelation in the two player game.

Definition 2. *A* Cournot tâtonnement *is a sequence indexed by* $k = 1, 2, ...$ *such that* $\{w_0, r_0\} \in [\underline{b}, \bar{b}]^2$*, and* $\{w_k, r_k\} = \{w^*(r_{k-1}), r^*(w_{k-1})\}$ *is the* simultaneous best-response *of the firm and worker at iteration k to the* $k-1$ *value of the wage and participation threshold respectively.*

In words, a *Cournot tâtonnement* in the two player game is a sequence of doubles, $\{w_k, r_k\}$, such that w_k (respectively r_k) is the best response of the firm (respectively worker) to the past-round realization of r_{k-1} (respectively w_{k-1}). This equilibrium revelation mechanism has the desirable property that, if the *Cournot tâtonnement* starting at the pair w_0, r_0 converges, a rational and forward looking agent can "learn" the equilibrium that will arise after observing w_0, r_0 by thinking through her and her opponent's *simultaneous best-response dynamics*.

Proposition 2. *Starting at beliefs,* $a_0 = \{w_0, r_0\}$*, such that the worker's best response to the initial wage is the initial threshold,* $r^*(w_0) = r_0$*, then* simultaneous best-response dynamics

1. *converge monotonically to* w^C *whenever* $w_0 \geq w^C$*.*
2. *converge to* $w \in (w^L, w^C]$ *whenever* $w_0 \in (\underline{b}, w^C)$*.*[29]

The proof is in the appendix, here we provide a heuristic sketch. First, without loss of generality, we can make all the arguments from the perspective of the worker, but of course similar arguments hold from the firm's perspective. Second, if at iteration k of the *simultaneous best-response dynamics* the worker's best response (to her expectation of the

[29] with monotone convergence whenever the second order condition $\frac{2h(b)}{p-b} > \frac{d^2H(b)}{db^2}$ holds. Note that this second order condition is the same as is required for a unique solution to the monopsonist's problem.

wage) confirms the firm's expectation of the participation threshold (i.e., $r_k = r^*(w_k)$), then for every $(k+n)$ iteration, with n even, the same holds. Third, at iteration $(k+1)$ the firm confirms the expectation of the worker, that is $w_{k+1} = w^*(r_{k+1})$. Let assume that k is even, then for all the even iterations the worker confirms the firm's expectation, and for all the odd iterations the firm confirms the worker expectations. Four, then the task is to show that there exists a n such that, $r_{k+n+1} = r^*(w_{k+n+1})$ also holds, thus for every iteration after $(k+n+1)$, r^* and w^* are mutual best responses. The appendix shows that the conditions for (iterative) monotone convergence are the same as the sufficient conditions guaranteeing an unique optimal wage for the unconstrained firm. Finally, since the workers strategy only depends on her expectations. Then we will be able to obtain comparative statics and dynamics by focusing on the case in which the *simultaneous best-response dynamics* starts at a point where there is at least one player (the firm) whose expectations are confirmed in the first iteration.

3 A Two Sided Frictional Labor Market

We embed this stylized game in a frictional labor market that is endowed with appropriate features to justify the assumptions of our two player game: *atomistic* workers and *atomistic* firms meet via a process of *sequential random matching*; and there exists free entry into vacancy creation. From these assumptions we show that firms' and workers' strategies in the multi-player game follow those of the two player game. We close the model with a standard matching function and free entry into vacancy creation to pin down the (equilibrium) labor market tightness. Thus, equilibria of the full model are characterized by triplets: wage, participation threshold, and labor market tightness.

In the two-player game we assumed that the probability of a worker (firm) meeting a firm (worker) is exogenous to that player's strategy. Here, this assumption is micro-founded by the assumption that every agent is atomistic. Thus, first, the expectation of the strategy to be played by any given firm (resp. worker) is equivalent to the average wage (resp. participation threshold); and second, the equilibrium job finding and filling hazards are exogenous to each worker's and each firm's strategy. In the two-player game we also assumed that the firm holds monopsony power and posts wages ex-ante. Here, as Diamond (1971) shows, sequential random matching endows the firm with this monopsony power and the monopsony wage level prevails even in the limit as search friction fades.

We close the model by positing a standard matching function as in the Diamond-Mortensen-Pissarides (DMP) model. Firms post vacancies at flow cost, c and workers engage in search at zero cost. Both discount the future at rate ρ. As in the baseline DMP model the flow of new matches is determined by the matching function, denoted as $m(U, V)$, where U is the mass of unemployed workers and V is the mass of vacancies.[30] Imposing Inada conditions and constant returns to scale, the job-finding rate of unemployed workers, $f(\theta) \equiv \frac{m}{U} = m(1, \theta)$, is increasing and concave in the market tightness defined as the ratio of vacancies to the unemployed, $\theta = \frac{V}{U}$. Analogously, the rate at which vacancies meet unemployed workers, $q(\theta) \equiv \frac{m}{V} = \frac{f(\theta)}{\theta}$, is a positive and decreasing function of market tightness.[31]

Note that since we have assumed random matching, the matching rate of a worker with a high flow value of leisure and a low flow value of leisure are the same whenever they participate in the labor market. Thus, the mass of workers who search for work but will reject wage offers of w^*, $i(1 - H(w^*))$, are just as likely to meet a firm as the $H(w^*)$ mass of workers who search and will accept wage offer w^*.[32] Since the $i(1 - H(w^*))$ mass of workers

[30] U and V are the mass of job seekers and vacancies. u and v are the exogenous odds of meeting a trading partner in the two player game. We will endogenize u and v here using the matching function.

[31] It is important to note here that our model differs in an important way from the classic model of upward sloping labor demand: Diamond (1982). In that model multiplicity derives from a thick market externality generated by increasing returns to scale in the matching function. In our model multiplicity derives from a pair of externalities, a thick market externality derived from workers' participation decision and a pecuniary externality derived from firms' wage posting decision. We follow the main stream DMP literature and impose constant returns to scale on our matching function. We appeal to the empirical results summarized in Pissarides and Petrongolo (2001) to justify this assumption.

[32] Note, under the assumptions defining the frictional labor market, the equilibrium wage coincides with the equilibrium wage in the two-player game.

reject the wage offer whenever they meet a firm the vacancy filling rate is:

$$g(\theta) = \Lambda(w^*)q(\theta),$$

where $\Lambda(w^*) \equiv \frac{H(w^*)}{H(w^*)+i(1-H(w^*))}$ is a measure of the severity of the congestion imposed by workers with leisure value above the equilibrium wage level searching for work while at the same time rejecting all offers. $\Lambda = 1$ occurs when workers with flow value of leisure above the expected wage offer never search and is the case when congestion is minimized. For $i > 0$ and $w^* < \bar{b}$, $\Lambda < 1$ and congestion drives a wedge between the rate at which vacancies meet employees and the rate at which jobs fill.

Job destruction is an exogenous shock that arrives with Poisson arrival δ. At the steady state, the flow into and the flow out of unemployment amongst workers with value of leisure below the participation threshold are equal, $(1-u)\delta = f(\theta)u\Lambda(w^*)$, where u is the unemployment rate and $\Lambda(w^*)$ accounts for the fraction of workers that will accept the offer. This gives the steady state unemployed rate:

$$u = \frac{\delta}{\delta + f(\theta)\Lambda(w^*)}. \tag{7}$$

Again, notice that the DMP model corresponds to the minimum value of congestion, $\Lambda = 1$, that is when workers with flow value of leisure above the expected wage offer never search. For $i > 0$ and $w^* < \bar{b}$, $\Lambda < 1$ and thus congestion increases the steady state unemployment.

In the Appendix we present the Bellman equations associated with each worker's and each firm's decision problem and establish the equivalence between the strategies in the frictional two-sided labor market with those in the two-player game. These yield identical participation and wage posting strategies as in the two-player game: the best response of workers is a participation threshold strategy as in equation (2) and the best response of firms is a wage posting strategy as in equation (6).

Given the wage, free entry into vacancy creation pins down the labor market tightness as the value of a vacancy is driven to zero. Thus equilibrium tightness satisfies:

$$\frac{c}{q(\theta(w^*))\Lambda(w^*)} = \frac{p - w^*}{\rho + \delta}. \tag{8}$$

Notice that the slope of the job creation condition depends on the participation threshold through the congestion effect. For a given wage, more congestion implies a steeper job creation condition and therefore a looser labor market.

Definition 3. *A symmetric* ***self-confirming rational expectations equilibrium*** *of the frictional labor market is a triple – wage level, participation threshold, labor market tightness – such that:*

1. *the wage level and participation threshold are mutual best responses.*

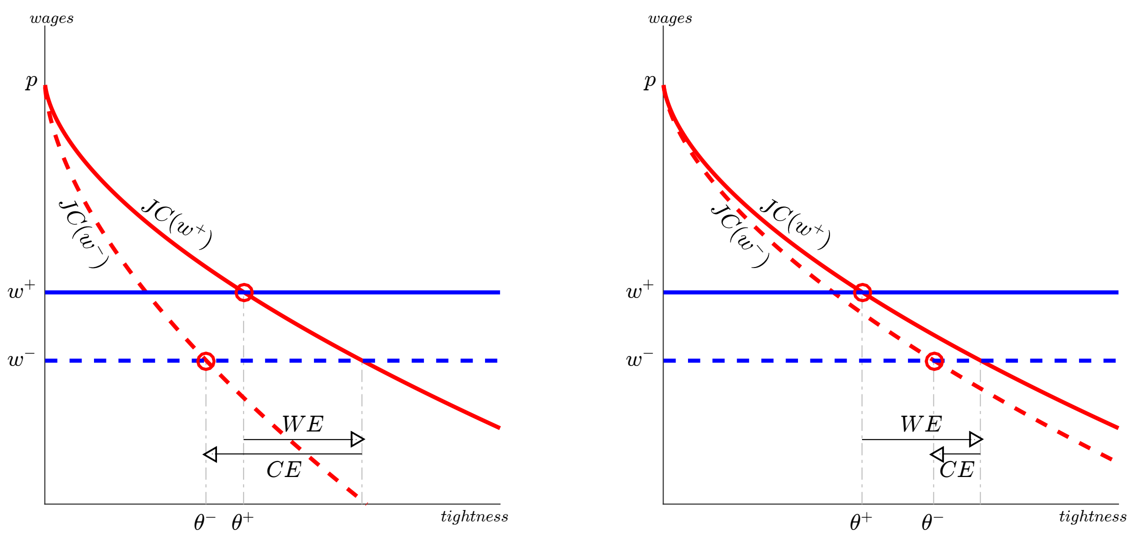
2. *labor market tightness satisfies the free entry condition, equation (8).*

Note that all firms (resp. workers) face the same objective function and we have restricted our attention to equilibria that are symmetric in the sense that every firm (resp. worker) play the same strategy as every other firm (resp. worker). Thus our equilibria and our convergence results from the two player game naturally extend to the two sided frictional labor market.

A comment is in order: the nature of the coordination failure is deeper in the multi-player economy. It is no longer sufficient for a single worker and single firm to meet and *pre-arrange* their expectations before entering the search market and engaging in random matching. Since each is massless the probability of meeting again is null. It must be the case that each worker can *coordinate* her expectation with every firm and visa versa. Instead we can think of a pre-meeting in which *all firms* can coordinate in order to provide a single wage signal on which workers can then coordinate their expectations of the wage level (and visa versa). In addition, note that the equilibrium prediction of the multi-player game is stronger than that of the two-player game in the sense that a unilateral deviation from the

Figure 5: Free Entry into Vacancy Creation:
Wage Level Effect (WE) and Congestion Effect (CE).

Congestion Effect Dominates | Wage Level Effect Dominates



Note: equilibrium labor market tightness in a high wage (solid) and low wage (hashed) equilibrium and when the congestion effect is large relative to the wage rigidity effect (left) and the converse (right).

equilibrium by any player, even if observable to the other players, triggers no response since the odds of encountering the particular player through random matching are null. In the following sections we refer to an uncoordinated firm or uncoordinated worker as one who recognizes that her unilateral deviation from a given set of expectations does not impact the expectation of any other player.

3.1 Comparative Statics: Output, Wages, and Unemployment

Figure 5 illustrates two equilibria for the labor market tightness conditional on a high wage level, w^+, (depicted as the intersection of the solid lines in each panel) and on a low wage level, w^-, (depicted as the intersection of the dashed lines in each panel) under two assumptions on the magnitude of congestion. As is typical, the job creation condition is downward sloping. Meanwhile the wage schedule is flat. Thus we are guaranteed a unique labor market tightness for every aggregate wage level and firm productivity pair.[33] Heavy congestion is illustrated in the left panel and light congestion in the right.[34] The slope of the job creation condition depends on the participation threshold through the congestion effect. Fixing wages, more congestion implies a steeper job creation condition and therefore a looser labor market at any wage level. This accounts for the difference in slope between the dashed job creation conditions in the two panels. Fixing congestion, a lower wage implies a tighter labor market.

If the increase in congestion triggered by moving from the high wage to the low wage equilibrium is large enough then unemployment will be higher in the low-output equilibrium. This is illustrated in the left panel: the congestion effect (labeled CE) is larger than the wage level effect (labeled WE). If the increase in congestion is mild unemployment *falls*. In this case the change in unemployment and the change in the output gap are in opposition. This is illustrated in the right panel: the congestion effect (labeled CE) is smaller than the wage level effect (labeled WE). Thus, without knowledge of the magnitude of the congestion effect the relation between unemployment fluctuations and fluctuations in output slack is ambiguous and we fail to generate satisfying business cycle regularities, such as counter-cyclical unemployment, simply from "animal spirits."[35] This, seemingly undesirable ambiguity will enable to generate instances of "job-lessness" and also instances of "wage-lessness" recoveries. In particular, we posit an expectation formation process through which congestion and wages evolve endogenously. This allows us to generate business cycles which feature an initial phase of "jobless-ness" followed by phase of "wageless-ness." We turn to this analysis in the next section.

[33] As in the standard DMP model, tightness and unemployment move inversely according to equation (7).

[34] Note that in both cases, wages are lower than w^C, and as the equilibria can be welfare ranked, total output and total employment are larger in the higher wage equilibrium.

[35] The model could be used to create economic fluctuations *via* coordinated changes in expectations (i.e., animal spirits). For such fluctuations to match a basic business cycles fact–countercyclical unemployment–we would need to impose that congestion is large enough.

4 Macroeconomic Implications

In this section we focus on fluctuations driven by exogenous changes in labor productivity. In particular, we consider productivity as a stochastic process and we show how the history of productivity shocks might influence (expectations of) the steady-state aggregate wage level. Since workers may move freely between participation and non-participation and since free entry into vacancies guarantees that the value of a vacancy is zero in all periods, worker and firm value functions are independent of the shock process so long as shocks are not large enough to destroy existing matches.[36] Following the search literature (e.g. Shimer (2012)) we appeal to the short half-life of unemployment and vacancies to note that convergence to steady state is rapid and thus consider serial steady-states and abstract from dynamics between steady-states in the following analysis.

Assumption 2. *The equilibrium wage and participation threshold pair change only when unilateral deviation is the best response for some player. Any deviation triggers* simultaneous best-response dynamics*.*

This assumption imposes that the equilibrium wage-labor force participation pair changes only when productivity evolves in such a way that the existing pair is no longer in the set of possible equilibria. The new equilibrium, if it exists, is then selected from the new feasible set via a *simultaneous best-response dynamics* as in Vives (1990, 2005); Cooper (1994).[37]

Proposition 3. *Given any shock to labor productivity, p,* simultaneous best-response dynamics *converge to an equilibrium. Further, a positive (negative) shock yields a weakly positive (negative) revision to the wage and to the participation threshold.*

Proof. This follows directly from Remarks 2 and 4 Proposition 2. Importantly, revisions to

[36] Note that matches formed when wages are lower will be more robust to shocks, since the workers involved are selected on having lower flow value of leisure. This effect will rotate the job creation condition counter clockwise. This effect is outside the scope of this research and the results presented in the following are not qualitatively affected.

[37] Note that whenever the existing equilibrium remains in set of possible equilibria then it is also the point of convergence of the *simultaneous best-response dynamics* starting from that point.

labor productivity leave workers' best response *unaffected*. Meanwhile, positive (negative) revisions to productivity weakly increase (decrease) firms' best response. □

4.1 Endogenous Rigidity in Wages and in Participation

The model predicts that there is a range of values both for labor productivity and the unemployment rate over which wages are endogenously ridged.

Proposition 4. *Given an initial wage and participation threshold, $\{w_0, r_0\}$, these values are* ***endogenously rigid*** *for a range of labor productivity, $[p^L, p^H]$, and associated unemployment rates, where*

$$p^L = w_0 + \frac{H(w_0)}{h(w_0)} \qquad\qquad p^H = w_0 + \frac{H(w_0)}{ih(w_0)}.$$

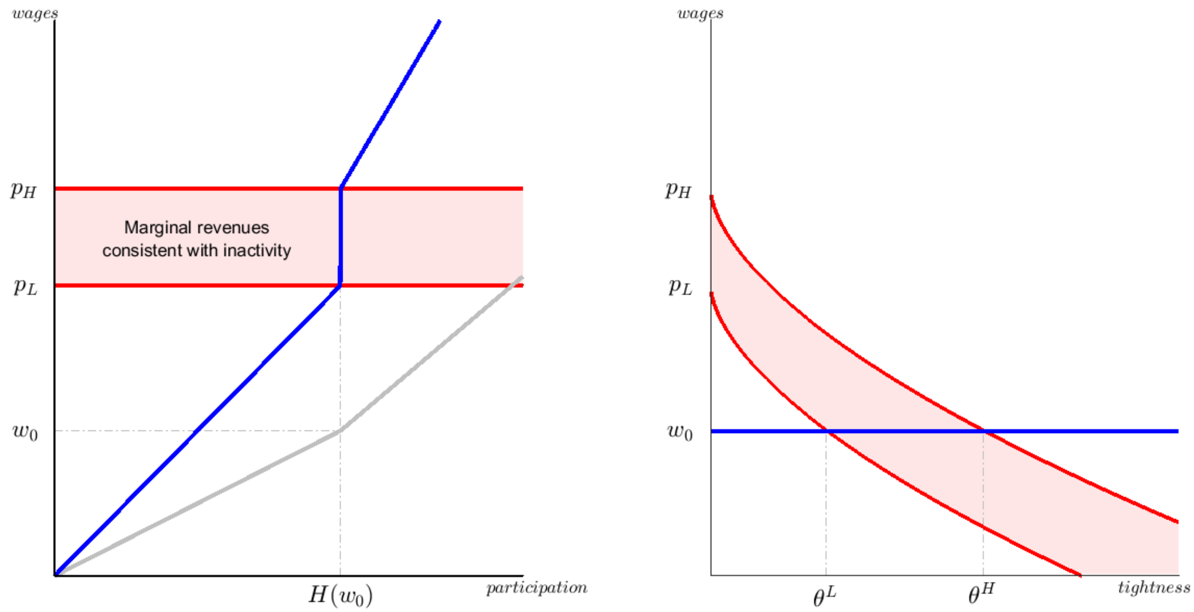
and the unemployment rate is pinned down by the labor market tightness that satisfies the free entry condition, equation (8), for each $p \in [p^L, p^H]$.

Proof. The proof follows from considering the first order conditions of the problem of a firm. This is essentially a corollary to the existence of the interval of equilibria on $[w^L, w^C]$ for any given productivity level. See proposition **??**. □

For $i \in (0, 1)$ this range of inactivity is depicted in Figure 6. The left panel presents the firm's monopsony problem constrained by the participation threshold as in Figures 2 and 3. For productivity in $[p^L, p^H]$ we see that the intersection between marginal cost and marginal revenue is at $H(w_0)$ and consistent with wage level w_0. The right panel presents the range of labor market tightness consistent with wage level w_0. Thus, we have ranges of values both for labor productivity and the unemployment rate over which wages are endogenously rigid.

Figure 7 illustrates firms' and workers' best response in the face of a positive and a negative productivity shock and the associated *simultaneous best-response dynamics* from each firm's perspective, as in Figures 2 and 3. Figure 12, located in the appendix, illustrates the same from the perspective of mutual best-response as in Figure 4.

Figure 6: Endogenous rigidity in wage and participation.



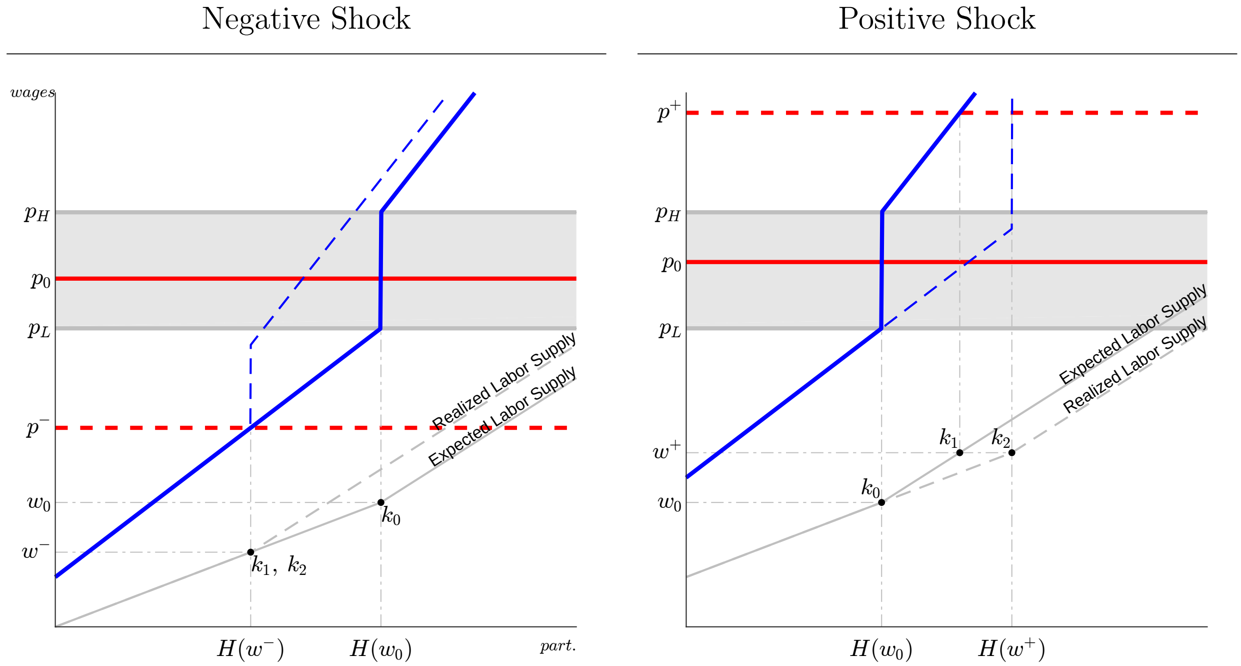
Note: The left panel presents the firm's monopsony problem constrained by the participation threshold as in Figures 2 and 3. The right panel presents the range of labor market tightness consistent with wage level w_0.

If reduction in productivity renders $p^- < p^L$, then from proposition 2 it follows that $w^- = w^C(p^-)$. In words, the contraction is severe enough to render the initial wage expectations outside the equilibrium set. The each firm's constrained monopsony problem in this event is illustrated in the left panel of figure 7. The gray region indicates the interval of labor productivities consistent with a pre-shock equilibrium at wage w_0. The red line indicates a productivity preceding the shock. The red dashed line represents the productivity following the shock, p^-. The gray solid lines indicate the expected labor supply schedule on impact and the labor supply schedule that arises in the equilibrium following the shock. The blue solid and dashed lines plot the marginal cost curve on impact and in the post-shock equilibrium respectively. Prior to impact of the shock the economy is in equilibrium at wage level w_0. We label this as k_0 indicating the initial conditions from whence *simultaneous best-response dynamics* will commence. Notice that on impact of the shock each firm is unconstrained by the existing expected participation threshold, r_0. Each selects a new wage such that the first order condition of an unconstrained monopsonist is satisfied: $h(w^-)(p^- - w^-) = H(w^-)$.

Meanwhile the worker best responds to w_0 by maintaining $r^*(w_0) = w_0 = r_0$. The first iteration of the *simultaneous best-response dynamics* suggests k_1 as a candidate equilibrium. A second iteration confirms, now workers best respond with $r^*(w^-) = w^-$ and the labor supply curve shifts to the realized labor supply curve. Firms best respond with $w^*(r_0) = w^-$. At this point the two are mutual best response and the economy has converged to a new equilibrium, we label this k_2.

If increase in productivity renders $p^+ < p^H$, then from proposition 2 it follows that $w^+ \in \{w^L(p^+), w^C(p^+)\}$. In words, the expansion is sufficient to render the initial wage expectations outside the equilibrium set. The each firm's constrained monopsony problem in this event is illustrated in the right panel of figure 7. Again, the gray region indicates the interval of labor productivities consistent with a pre-shock equilibrium at wage w_0. The red line indicates a productivity preceding the shock. The red dashed line represents the productivity following the shock, p^+. The gray solid lines indicate the expected labor supply schedule on impact and the labor supply schedule that arises in the equilibrium following the shock. The blue solid and dashed lines plot the marginal cost curve on impact and in the post-shock equilibrium respectively. Prior to impact of the shock the economy is in equilibrium at wage level w_0. We label this as k_0 indicating the initial conditions from whence *simultaneous best-response dynamics* will commence. Notice that on impact of the shock each firm *is constrained* by the existing expected participation threshold, r_0. Each selects a new wage such that the first order condition of a constrained monopsonist is satisfied: $ih(w^+)(p^+ - w^+) = H(r_0) + i[H(w^+) - H(r_0)]$. Meanwhile the worker best responds to w_0 by maintaining $r^*(w_0) = w_0 = r_0$. The first iteration of the *simultaneous best-response dynamics* suggests k_1 as a candidate equilibrium. In a second iteration workers respond to the new proposed wage level with $r^*(w^+) = w^+$ and the labor supply curve shifts to the realized labor supply curve. Firms best respond with $w^*(r_0) = w^+$ since is wage level is now in $\{w^L(p^+), w^C(p^+)\}$. At this point the two are mutual best response and the economy has converged to a new equilibrium, we label this k_2.

Figure 7: Large Productivity Shocks.



Note: Equilibrium rehalation following large productivity shocks. The left panel illustrates a negative productivity shock and the right panel a postive productivity shock. Hashed lines indicate post-shock schedules. Iterations of *simultatnous best-response dynamics* are indicated as the points labeled k_0, k_1, and k_2. As before, fore ease of illustration we take $H(b)$ to be uniformly distributed on $[\underline{b}, \bar{b}]$.

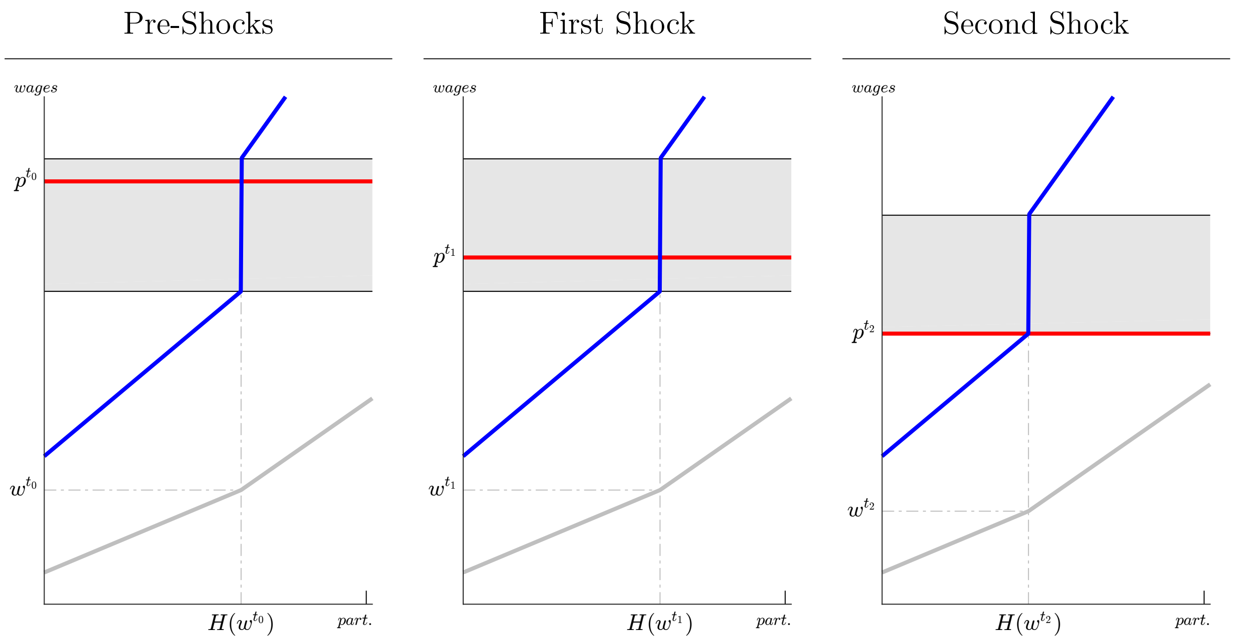
We now turn to some further implications.

Remark 6. *The economy reacts to productivity innovations sluggishly: Productivity may grow (fall) for some time before yielding a rise (decline) in wages and participation.*

Figure 8 illustrates an example. The left panel depicts a steady state at time t_0 in which productivity is within some inaction range.[38] Wages and participation are at levels w^{t_0} and $H(w^{t_0})$ respectively. At time t_1 a shock arrives such that productivity at t_1 is inside the time t_0 inaction rage. As a result wages and participation remain the same. At time t_2 a second and same sized productivity arrives with the result that productivity at time t_2 is outside and below the time t_0 inaction range. As a result wages and participation fall to $w^{t_1} < w^{t_0}$ and $H(w^{t_1}) < H(w^{t_0})$ respectively.

[38] The following holds with strict inequality whenever $p^{t_0} < p^H$.

Figure 8: Sluggish Response to Productivity Innovations



Note: Productivity may grow (fall) for some time before yielding a rise (decline) in wages and participation.

4.2 Hysteresis

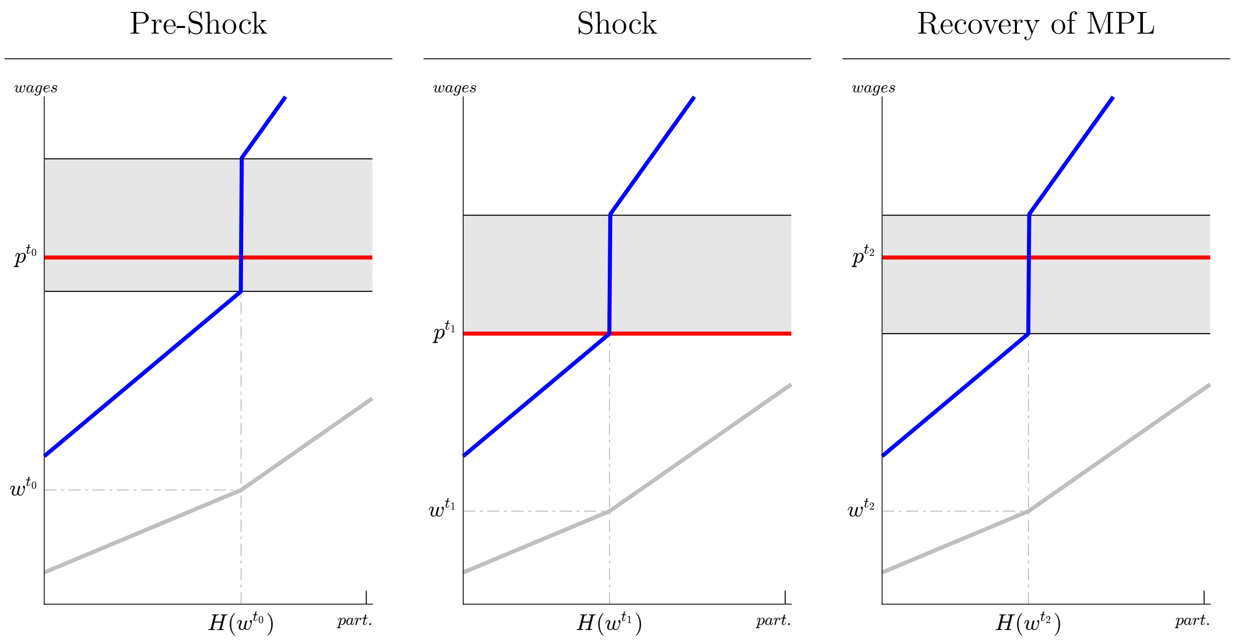
Since there are a set of equilibria consistent with a range of productivity levels, transition from one wage level to another may require a large shock or an accumulation of small shocks that put productivity outside that range. As a result aggregate variables, such as total output, may exhibit asymmetric responses to positive and negative shocks of equal magnitude. Also as a result recessions may result in scaring.

Remark 7. *A transient productivity shock may lead to persistent changes in wages, participation, unemployment, and output.*

Figure 9 illustrates an example. The left panel depicts an equilibrium at time t_0 in which productivity is within some inaction range.[39] Wages and participation are at levels w^{t_0} and $H(w^{t_0})$ respectively. At time t_1 a shock arrives such that productivity at t_1 is

[39]The following holds with strict inequality whenever $p^{t_0} < p^H$.

Figure 9: Persistent Response to Transitory Productivity Shock.



Note: A transitory negative productivity shock leads to a persistent decrease in wages, participation, and output.

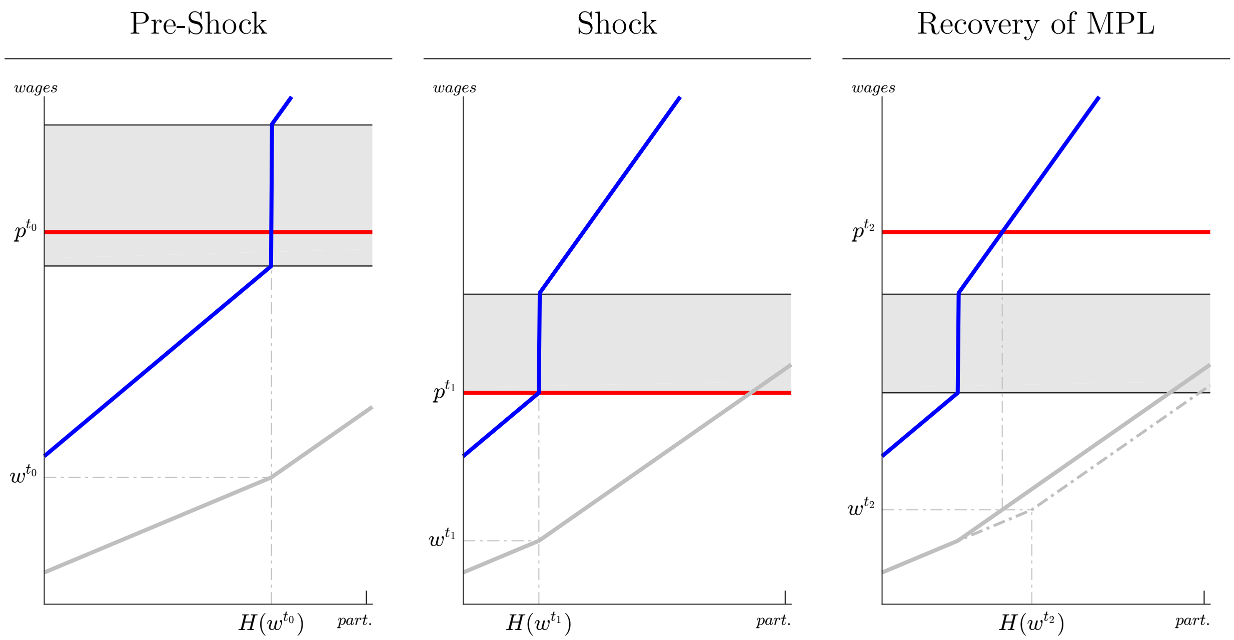
outside and below the time t_0 inaction rage. As a result wages and participation fall to $w^{t_1} < w^{t_0}$ and $H(w^{t_1}) < H(w^{t_0})$ respectively. A new inaction range is established following the shock that partially over-laps and partially falls below the pre-shock inaction range.[40] At time t_2 productivity recovers to the pre-shock level: $p^{t_2} = p^{t_0}$. However, even after the recovery productivity remains within the shock-state inaction range. As a result wages and participation *do not* respond to the recovery of productivity. Applying Proposition 1 we can welfare rank the time t_0 and time t_2 equilibria. In addition to wages and participation, total output and total welfare are lower at time t_2 than at time t_0.

Remark 8. *Contractions lead to scarring: following a sufficiently severe shock wages, participation, and output that run below pre-shock levels for any given labor productivity for some time.*

Figure 10 illustrates an example. The left panel depicts an equilibrium at time t_0 in

[40]Note: the two inaction ranges need not overlap.

Figure 10: Scaring Effects of Contractions



Note: Following a severe contraction, wages, participation, and output run below pre-shock levels associated with the same labor productivity.

which productivity is within some inaction range.[41] Wages and participation are at levels w^{t_0} and $H(w^{t_0})$ respectively. At time t_1 a shock arrives such that productivity at t_1 is outside and below the time t_0 inaction rage. As a result wages and participation fall to $w^{t_1} < w^{t_0}$ and $H(w^{t_1}) < H(w^{t_0})$ respectively. A new inaction range is established following the shock that partially over-laps and partially falls below the pre-shock inaction range.[42] This is the all the same as in the previous illustration with the exception that the shock is of greater magnitude. At time t_2 productivity recovers to the pre-shock level: $p^{t_2} = p^{t_0}$. However, after the recovery firms are constrained by the shock-level participation threshold. As a result wages and participation do not *fully* respond to the recovery of productivity. Again, applying Proposition 1 we can welfare rank the time t_0 and time t_2 equilibria. In addition to wages and participation, total output and total welfare are lower at time t_2 than at time t_0.

[41]The following holds with strict inequality whenever $p^{t_0} < p^H$.
[42]Note: the two inaction ranges need not overlap.

As the portion, i, of workers who search regardless of the relation between the expected wage and their flow value of leisure rise to unity, which eliminates the difference in market thickness above or below the expected participation threshold, the upper bound of the range of inaction, p^H, falls to the lower bound, p^L. In this case, the unique equilibrium coincides with the constrained efficient equilibrium. In such an economy, responses to shocks will also always be symmetric which eliminates hysteresis. As i fall to zero and there are no searchers in the thin side of the market, the upper bound, p^H, rises without bound and the economy converges to a strictly asymmetric case in which the only possible revisions to wages and participation are downward. In the intermediate case we have a moderated degree of asymmetry: responses to positive and negative revisions to labor productivity *of the same absolute magnitude* may lead to revisions in output of *differing absolute magnitude*.

4.3 "Job-less" and "Wage-less" Recoveries

Closing the model with free entry into vacancies, as in Section 3, can give rise to both "job-less" and "wage-less" phases of recovery from productivity contractions. Job-lessness operates through the increase in congestion induced by the fall in the participation threshold during a severe contraction. Wage-lessness operates through an increase in the wage rigidity effect induced by persistence of the lower severe-contraction wage level beyond the duration of the disturbance. When the increase in congestion is not particularly severe, wage-lessness may dominate throughout a recovery. However, whenever the drop in the participation threshold is severe enough to trigger a job-less recovery, job-lessness will occur during the early part of the recovery and will be followed by a period during which unemployment falls to unprecedented lows while wages continue to lag: a "wage-less" phase.

Figure 11 illustrates. The left two panels illustrate a steady state at time 0. The firm's wage setting problem is illustrated in the top panel and the job creation condition in the bottom. At this pre-shock steady state, wages, participation, and labor market tightness are w^0, H^0, and θ^0 respectively. The economy is then hit by a shock that lowers productivity to

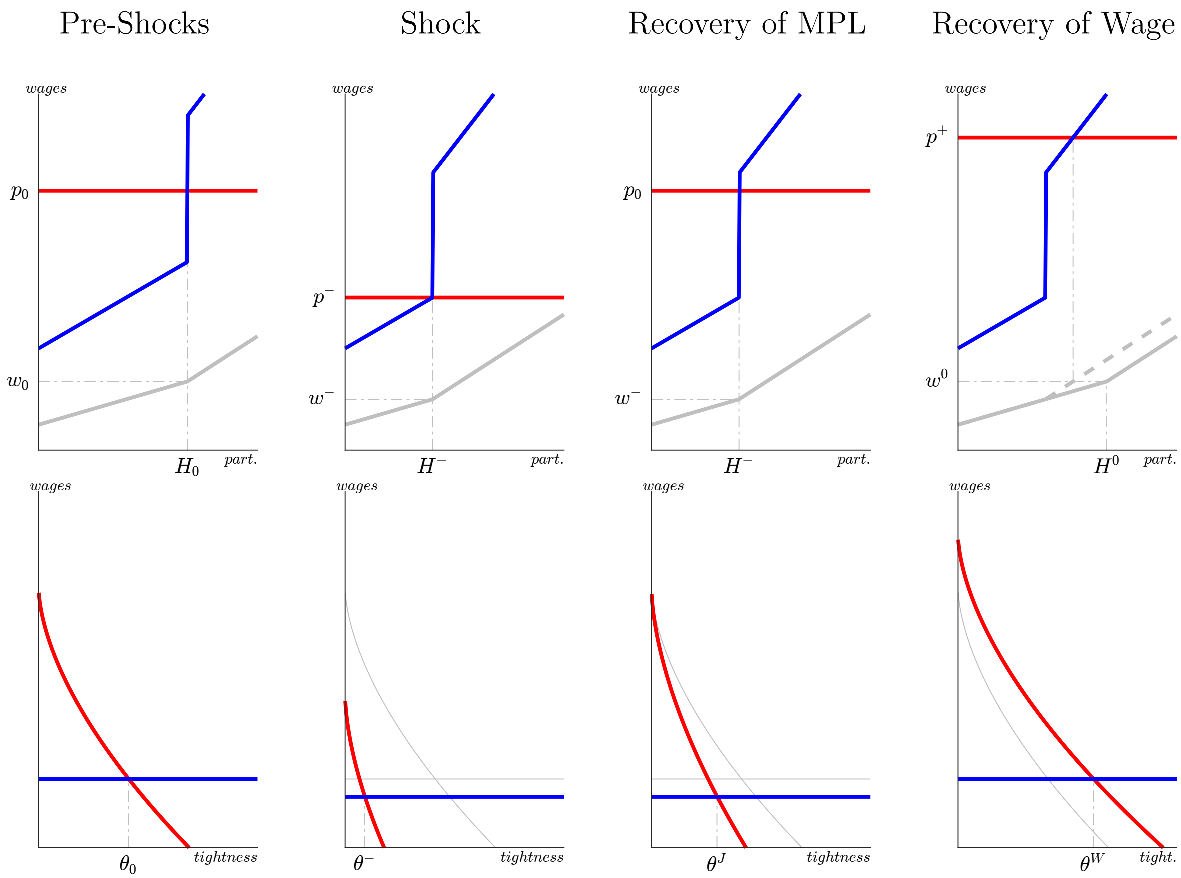
p^{-}, a value below the pre-shock inaction range. This is illustrated in the second set of panels. As a result of the shock, wages and participation fall to w^{-} and H^{-} as firms lower the wage level due to their lower productivity (e.g. w^0 is now above the unconstrained wage choice in the shock state: $w^0 > w^C(p^-)$). Labor market tightness also falls. The drop in labor market tightness is the the result of three forces: the drop in productivity, an increase in congestion due to the drop in the participation threshold, and a drop in wages. The first two clearly push tightness down. That these dominate the third is a result of firms' optimization. Note that a fall in labor market tightness implies an increase in unemployment.

The third set of panels illustrates wages, participation, and labor market tightness if the economy is subsequently hit by a productivity shock that returns productivity to the pre-shock level. Since pre-shock productivity is in the inaction range determined in the shock state, wages and participation do not respond.[43] As a result labor market tightness rises. Now tightness is a result of two of the three forces: the increased in congestion due to the lower participation threshold and the lower wage level. In the case illustrated the congestion effect dominates and tightness falls short of the pre-shock level. Note that a shortfall in tightness implies a higher level of unemployment: we have a job-less recovery.

The final set of panels illustrates the productivity shock required to return the economy to the pre-shock wage level following the contraction. Due to the asymmetry induced by the inaction range the productivity level required to return the pre-shock wage level exceeds the pre-shock productivity level. Thus, labor market tightness in the period in which wages and participation recover *must* exceed the tightness experienced just before the impact of the first shock. Noting that congestion reduces as productivity approaches the level required to return wages to the pre-shock wage, we see that the wage rigidity effect must dominate the congestion effect for an interval of productivity levels before wages and participation recover to the pre-recession levels. Thus we have that at least the end phase of each recovery from a severe contraction must be wage-less. Thus, every job-less recovery must be followed by a

[43] Alternately, as in Remark 8 they may paartially recover.

Figure 11: “Job-less” and “Wage-less” Recoveries.



Note: every severe contraction is followed by a period of “wage-less” recovery. Some severe contractions also feature a “job-less” phase early in recovery which is always followed by a “wage-less” phase.

wage-less recovery.

5 Conclusion

We contribute a theoretical model of hysteresis in the labor market to complement a substantial body of empirical literature that shows that high unemployment, falling wages, and reduced economic activity can have lasting consequences. We model hysteresis as resulting from a coordination failure among atomistic firms and atomistic workers in a frictional labor market that features random search, *ex-ante* wage commitments, and the possibility of worker non-participation. This coordination failure results in a continuum of possible equilibria with high-wage, high-participation equilibria welfare dominating low-wage, low-participation equilibria. We then introduce changes in labor productivity and show that if the transition among equilibrium is determined by a *simultaneous best-response dynamics* then endogenous rigidity in wages and participation arises in response to transitory movements in labor productivity. The model is consistent with a host of other medium-run and business cycle facts following a severe recession: outward shifts in the Beveridge curve, jobless and wage-less recoveries, and a reduction in the sensitivity of wages to unemployment. Furthermore, expansions are insufficiently robust and in this sense recessions are scarring.

We offer here some further thoughts in conclusion.

Wage rigidity and the Barro (1977) critique

Our objective is to construct a Keynesian model in the sprit of the discussion of Howitt (1986). As such we strive for wage rigidity that is *evidence of* market failure while not directly causing market failure itself. The goal is to *end up* with endogenously rigid wages *because of* the market failure–as Keynes calls for–and *not* the other way around.

It is useful to dissect rigidity in our mechanism into ex-ante rigidity (before meeting any particular worker) and ex-post rigidity (after having met a particular worker). Firms in our model have ex-ante total wage flexibility; however, equilibrium posted wages are at times non-responsive to fundamentals due to coordination failure. Thus, ex-ante rigidity imposed by coordination failure is impervious to the Barro (1977) critique.

On the other hand, firms in our model experience ex-post total wage rigidity. We model wages this way, as does much of the wage posting literature, in order to capture an informational asymmetry between worker and firm: the firm does not know and can not learn the worker's flow value of leisure. A typical critique of wage posting is that in the sub-game in which a rent-producing match is ruled out due to wage posting the informed party should reveal her type and bargaining should commence.[44]

[44] This criticism is valid for any model of monopolistic pricing: given the flexibility the monopolist would always prefer to price discriminate and price discrimination is more efficient than monopoly pricing.

In the limit as $i \searrow 0$, this suspect sub-game never occurs: only workers who will accept the posted wage search and all matches are consummated. Although ex-post wage rigidity never binds and ex-ante wages are flexible, wages are endogenously rigid on a vast region of values for the fundamental (labor productivity). For all realizations of productivity above p_L wages remain unchanged. Thus, for shocks that return p in this region wages are endogenously totally rigid. This case leads to the *largest* range of fundamentals for which wages are rigid in our model.

This result is too strong. Subjected to stochastic shocks, an economy in which all workers choose non-participation when their flow value of leisure exceeds the expected wage will eventually converge to zero participation, zero employment, and zero output. We take an intermediate value of i, allowing $i \in [0, 1]$, in order to admit the possibility of macroeconomic recovery following contractions. When we have $i > 0$ we are *partially* subject to Barro's criticism in the sense that ex-post some rent producing matches are ruled out.

Of course, if negotiation is allowed in this sub-game, there will be cases when a mutually beneficial deal can be struck. Further, if such negotiation is permitted this will be known by workers and, as a result, beliefs consistent with inactivity will be difficult to support (at least at the threshold for participation predicted in the baseline model). One possible solution to this sub-game that preserves our results is to assume that workers can trigger bargaining if they wish. If they do their bargaining power is zero. The result is that workers with $b < r_0 = w_0$ will never trigger bargaining. Workers with $b \in (w_0, p]$ may, and if they do they will receive wages equal to their flow value of leisure: so, they won't care if they do or don't bargain.

The Shimer (2005) puzzle: amplification versus a shifting Beveridge Curve

In constructing a region of indeterminacy regarding the wage level in equilibrium our model bears resemblance to a set of models that utilize wage rigidity to solve the Shimer (2005) puzzle: e.g. Hall (2005), Gertler and Trigari (2009), and Kennan (2010). Rigidity in our model can be used toward this objective; however, our main goal is the generation of hysteresis. With that in mind, models that focus strictly on amplification categorically fall short.

The logic is simple, in the absence of a congestion effect, a negative shock large enough to cause a downward revision in wages will lead to an *increase in amplification*. This increase is concentrated fully on the up-side: at the new lower wage the vacancy rate will be higher for any given level of productivity than it was at the old higher wage. The result is a "job-full" recovery. In contrast, the data shows that severe contractions are associated with outward shifts in the Beveridge Curve or "job-less" recoveries. Our model builds a congestion effect that is triggered at the same time as a downward revision of wages. Thus we are able to generate both amplification and a "job-less" recovery following severe contractions. For us, amplification shows up in the later, "wage-less" phase of recovery.

References

Ahn, H. J. and J. D. Hamilton (2016, July). Heterogeneity and unemployment dynamics. Working Paper 22451, National Bureau of Economic Research.

Albrecht, J., F. Postel-Vinay, and S. Vroman (2013, 05). An Equilibrium Search Model Of Synchronized Sales. *International Economic Review 54*(2), 473–493.

Barnichon, R. (2010). Building a composite help-wanted index. *Economics Letters 109*(3), 175 – 178.

Barnichon, R. and A. Figura (2015). Labor market heterogeneity and the aggregate matching function. *American Economic Journal: Macroeconomics 7*(4), 222–49.

Barro, R. J. (1977, July). Long-term contracting, sticky prices, and monetary policy. *Journal of Monetary Economics 3*(3), 305–316.

Beaudry, P., D. Galizia, and F. Portier (2015, June). Reviving the limit cycle view of macroeconomic fluctuations. Working Paper 21241, National Bureau of Economic Research.

Blanchard, O. J. and L. H. Summers (1986a, March). Hysteresis and the European Unemployment Problem. In *NBER Macroeconomics Annual 1986, Volume 1*, NBER Chapters, pp. 15–90. National Bureau of Economic Research, Inc.

Blanchard, O. J. and L. H. Summers (1986b, June). Hysteresis and the european unemployment problem. Working Paper 1950, National Bureau of Economic Research.

Blanchard, O. J. and L. H. Summers (1987). Hysteresis in unemployment. *European Economic Review 31*(1-2), 288–295.

Cooper, R. (1994). Equilibrium Selection in Imperfectly Competitive Economies with Multiple Equilibria. *The Economic Journal 104*(426), 1106–1122.

Cooper, R. and A. John (1988). Coordinating Coordination Failures in Keynesian Models. *The Quarterly Journal of Economics 103*(3), 441–463.

Diamond, P. (1982). Aggregate Demand Management in Search Equilibrium. *Journal of Political Economy 90*(5), 881–94.

Diamond, P. and D. Fudenberg (1989, June). Rational Expectations Business Cycles in Search Equilibrium. *Journal of Political Economy 97*(3), 606–19.

Diamond, P. A. (1971, June). A Model of Price Adjustment. *Journal of Economic Theory 3*(2), 156–68.

Diamond, P. A. and A. Şahin (2015). Shifts in the Beveridge Curve. *Research in Economics 69*, 18–25.

Eeckhout, J. and I. Lindenlaub (2015). Unemployment cycles. IFS Working Papers W15/26, Institute for Fiscal Studies.

Elsby, M., B. Hobijn, and A. Şahin (2013). The Decline of the U.S. Labor Share. *Brookings Papers on Economic Activity 47*(2 (Fall)), 1–63.

Elsby, M. W. L. and M. D. Shapiro (2012, June). Why Does Trend Growth Affect Equilibrium Employment? A New Explanation of an Old Puzzle. *American Economic Review 102*(4), 1378–1413.

Gertler, M. and A. Trigari (2009, 02). Unemployment Fluctuations with Staggered Nash Wage Bargaining. *Journal of Political Economy 117*(1), 38–86.

Golosov, M. and G. Menzio (2015, November). Agency Business Cycles. NBER Working Papers 21743, National Bureau of Economic Research, Inc.

Hall, R. E. (2005, March). Employment Fluctuations with Equilibrium Wage Stickiness. *American Economic Review 95*(1), 50–65.

Howitt, P. (1986). The Keynesian Recovery. *The Canadian Journal of Economics / Revue canadienne d'Economique 19*(4), 626–641.

Juhn, C., K. M. Murphy, and R. H. Topel (2002). Current Unemployment, Historically Contemplated. *Brookings Papers on Economic Activity 2002*(1), 79–116.

Juhn, C., K. M. Murphy, R. H. Topel, J. L. Yellen, and M. N. Baily (1991). Why has the Natural Rate of Unemployment Increased over Time? *Brookings Papers on Economic Activity 1991*(2), 75–142.

Kaplan, G. and G. Menzio (2013, February). Shopping Externalities and Self-Fulfilling Unemployment Fluctuations. NBER Working Papers 18777, National Bureau of Economic Research, Inc.

Karabarbounis, L. and B. Neiman (2014). The global decline of the labor share. *The Quarterly Journal of Economics 129*(1), 61–103.

Kennan, J. (2010). Private information, wage bargaining and employment fluctuations. *The Review of Economic Studies 77*(2), 633–664.

Krugman, P. (1991). History Versus Expectations. *The Quarterly Journal of Economics 106*(2), 651–667.

Morris, S. and H. S. Shin (2000, September). Global Games: Theory and Applications. Cowles Foundation Discussion Papers 1275R, Cowles Foundation for Research in Economics, Yale University.

Mukoyama, T., C. Patterson, and A. Sahin (2016, August). Job search behavior over the business cycle. Staff Reports 689, Federal Reserve Bank of New York.

Murphy, K. M. and R. Topel (1997). Unemployment and Nonemployment. *The American Economic Review 87*(2), 295–300.

Pissarides, C. A. and B. Petrongolo (2001, June). Looking into the black box: A survey of the matching function. *Journal of Economic Literature 39*(2), 390–431.

Schaal, E. and M. Taschereau-Dumouchel (2015). Coordinating Business Cycles. 2015 Meeting Papers 178, Society for Economic Dynamics.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4*(1), 25–55.

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

Shimer, R. (2012, April). Reassessing the Ins and Outs of Unemployment. *Review of Economic Dynamics 15*(2), 127–148.

Summers, L. H. (1992, October). *Understanding Unemployment*. The MIT Press.

Vives, X. (1990). Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics 19*(3), 305 – 321.

Vives, X. (2005). Complementarities and Games: New Developments. *Journal of Economic Literature 43*(2), 437–479.

A Appendix

A.1 Proof of Proposition 2

The setting is simple enough to complete the proof by construction. Without loss of generality, I consider the case that $r_0 = r^*(w_0)$. To see this note that if $w_0 = w^*(r_0)$ then by definition of the *simultaneous best-response dynamics* we have $w_1 = w_0$ and $r_1 = r^*(w_1)$. Thus we can construct the proof of the $w_0 = w^*(r_0)$ case from the proof of the $r_0 = r^*(w_0)$ case simply by resetting the indices.

Let $W^+ = \{w \in [\underline{b}, \bar{b}] : w \geq w^*(r)\}$ be set of wages that weakly exceed the firm's best response. Similarly, let $R^+ = \{r \in [\underline{b}, \bar{b}] : r \geq r^*(w)\}$ be the set of participation thresholds that weakly exceed the worker's best response. Also let $A^+ = W^+ \cap R^+$ be the set of wage and participation threshold pairs such that both wage and participation threshold weakly exceed the corresponding player's best response.

Lemma 1. *A* simultaneous best-response dynamics *starting at any a_0 in A^+ converges monotonically downwards to $\{w^C, r^C\}$.*

Proof. This follows from the monotonicity if the best response functions in this region. □

Now, We have already shown that $w^*(r_k) > r_k$ if $r_k < r^L$; $w^*(r_k) < r_k$ if $r_k > r^C$; $w^*(r_k) = r_k$ if $r_k \in [r^L, r^C]$ and $r^*(w_k) = w_k$ for all w_k.

Now we can construct an iteration of the *simultaneous best-response dynamics*–$w_{k+1} = w^*(r_k)$–and consider the three possible cases:

1. $w_{k+1} > w^C$: in this case we have $r_{k+2} = w_{k+1}$ and $w_{k+2} = w_{k+1}$. So $a_{k+2} \in A^+$ and Lemma 1 shows that from this point on the *simultaneous best-response dynamics* converges monotonically downwards to $\{w^C, r^C\}$.

2. $w_{k+1} \in [w^L, w^C]$: in this case we have $r_{k+2} = w_{k+1}$, $w_{k+2} = w_{k+1}$, so $w_{k+2} = r_{k+2} \in [w^L, w^C]$. Thus w_{k+2} and r_{k+2} are mutual best response and the *simultaneous best-response dynamics* has converged at $k + 2$ to at a wage in $[w^L, w^C]$.

3. $w_{k+1} < w^L$: in this case we have $r_{k+2} = w_{k+1}$ and $r_{k+2} > r_k$ since r^* is monotone. In this case we do not have convergence at $k+2$; however, w_{k+3} must fall into one of these three cases. Thus, by induction, w converges monotonically upward toward w^L if case 1 or 2 does not occur for any k. □

A.2 Value Functions

We can write out the asset value equations faced by a generic worker:

$$\begin{aligned}
\rho\mathcal{I}(b, w_0) &= b \\
\rho\mathcal{U}(b, w_0) &= b + f(\theta)max\{0, [\mathcal{W}(b, w_0) - \mathcal{U}(b, w_0)]\} \\
\rho\mathcal{W}(b, w_0) &= w_0 + \delta[\mathcal{U}(b, w_0) - \mathcal{W}(b, w_0)]
\end{aligned}$$

The first, $\mathcal{I}(b, w_0)$, captures the asset equation for an inactive worker with flow value of leisure b. This worker simply consumes b.[45] The second, $\mathcal{U}(b, w_0)$, captures the asset equation for a searching workers. While searching the worker consumes flow value b and at hazard $f(\theta)$ receive job offers which yield option value $max\{0, [\mathcal{W}(b, w_0) - \mathcal{U}(b, w_0)]\}$. The third, $\mathcal{W}(b, w_0)$ captures the asset equation for employed workers, these consume flow value w_0 and at hazard δ are separated to unemployment.

We begin by solving for the generic worker's reservation wage conditional on participating. This sets the value of unemployment equal to the value of employment: $\mathcal{W}(b, w_0) = \mathcal{U}(b, w_0)$. Thus we have a reservation wage equal to this generic workers realization of the flow value of leisure: b. In the event that w_0 exceeds b this worker anticipates higher value from the employed state than the unemployed state: given the opportunity she accepts all wage offers.[46] If w_0 falls short of b then the worker anticipates lower value from the employed state than the unemployed state: given the opportunity she rejects all wage offers.

[45] Note that in steady state this is independent of the wage level. When adding shocks we note that as long as there is no barrier to reentering the unemployment pool following a shock the independence is preserved.

[46] As we have seen in the two player game and will see again here, information frictions prevent the firm from observing b and thus from setting $w = b$ for each worker and extracting the full rent.

Now we turn to considering the threshold participation decision. We can observe that if $b < w_0$ then the worker strictly prefers unemployment to inactivity and employment to unemployment: $I(b, w_0) < U(b, w_0) < W(b, w_0)$. Conversely if $b > w_0$ then the worker is indifferent between inactivity and unemployment and strictly prefers either of these to employment: $I(b, w_0) = U(b, w_0) > W(b, w_0)$. Thus we have that whenever $b < w_0$ the worker always prefers to participate and accepts employment whenever matched. Meanwhile if $b > w_0$ the worker is indifferent between participation and non-participation and strictly prefers not to accept employment whenever matched. As in the two player game we break the worker's indifference with Assumption 1.

We can also write out the asset value equations faced by a generic firm:

$$\begin{aligned}
\rho\mathcal{V}(w, r_0) &= -c + q(\theta)\Big[\mathbb{I}_{\{w \le r_0\}} \tfrac{uH(w)}{uH(r_0)+i(1-H(r_0))} \\
&\quad + (1 - \mathbb{I}_{\{w \le r_0\}}) \tfrac{uH(r_0)+i[H(w)-H(r_0)]}{uH(r_0)+i(1-H(r_0))}\Big][\mathcal{J}(w, r_0) - \mathcal{V}(w, r_0)] \\
\rho\mathcal{J}(w, r_0) &= p - w + \delta[\mathcal{V}(w, r_0) - \mathcal{J}(w, r_0)]
\end{aligned}$$

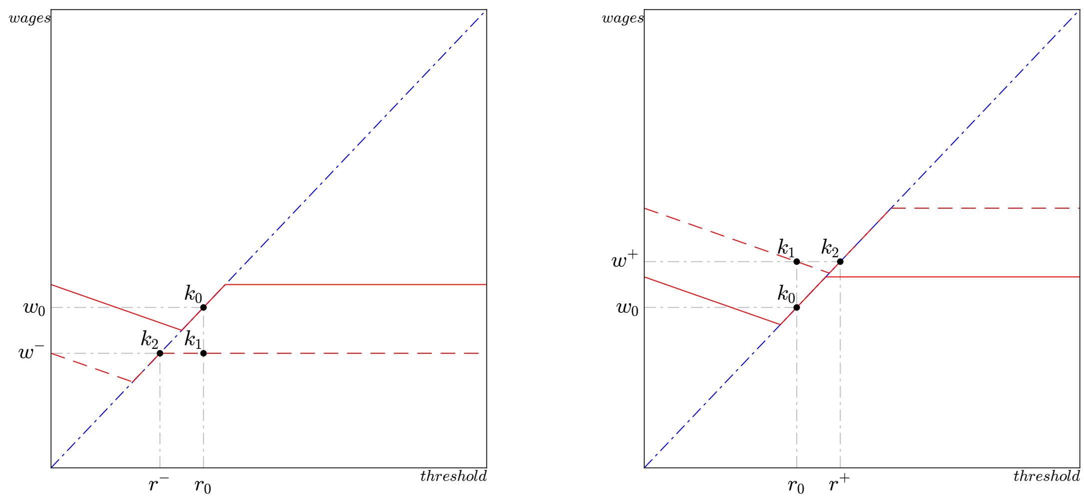
The first asset equation, $\mathcal{V}(w, r_0)$, captures the value of a open vacancy. The vacancy costs the firm a flow of c and with hazard $q(\theta)$ the firm meets and makes the posted wage offer w_0 to a worker. The bracketed term follows the logic of the firm's problem in the two player game: higher wage offers are accepted by a larger fraction of the workers that the firm might meet with the return to increasing the wage offer being discontinuous at the expected participation threshold. The final term is the option value of forming a match. The second asset equation, $\mathcal{J}(w, r_0)$ captures the value of a filled job. This is the flow rent $(p - w)$ and the hazard of separation times the option value of separation.

We posit free entry into vacancy creation. This drives the value of a vacancy to zero. Thus the firm's objective function in the two-sided game is isomorphic to the objective function of the firm in the two-player game.[47] Note that this implies that $v = \frac{\delta}{\delta + q(\theta)\Lambda(w^*)}$.

[47]It may seem odd to maximize a value that, in equilibrium, is zero. This is, however, a straightforward application of the envelope theorem.

Figure 12: Large Productivity Shocks.

Negative Shock | Positive Shock



Note:

A.3 Revisions to Wages and Participation

The left panel of figure 12 illustrates the shift in each firm's best response function in response to a reduction in productivity as well as the sequence of simultaneous best responses defined by the *simultaneous best-response dynamics*. The pre-shock mutual best responses are plotted as continuous lines. The reduction in productivity (left panel) induces a downward revision in each firm's best response (dashed lines) and leaves unchanged each worker's best response. At k_0 both worker and firm hold beliefs consistent with the pre-shock equilibrium. In the first iteration the firm best responds selecting $w_{k_1} = w^C(p^-)$: expected participation at k_0 exceeds the optimal wage choice given the negative revision to productivity. Also in the first iteration the worker best responds selecting $r_{k_1} = r_0$ since the initial beliefs at k_0 lie on the worker's best response function. In iteration two the worker best-responds to the iteration one wage choice setting $r_{k_2} = r^*(w_{k_1}) = w_{k_1}$. Also in iteration two the firm best responds to the iteration one threshold choice setting $w_{k_2} = w_{k_1}$ and the *simultaneous best-response* *dynamics* reaches convergence at k_2.[48]

The right panel of figure 12 illustrates the shift in the firm's best response function in this case as well as the sequence defined by the *simultaneous best-response dynamics*. The pre-shock mutual best responses are plotted as continuous lines. The increase in productivity induces an upward revision in each firm's best responses (dashed lines) and leaves unchanged each worker's best response. At k_0 both worker and firm hold believes consistent with the pre-shock equilibrium. In the first iteration the firm best responds selecting $w_{k_1} = w^*(r_0|p^+)$. We know that $w_0 < w_{k_1}$ since the positive revision in productivity shifted the best response function of firms toward higher wages. In the case depicted, where the distribution of flow values of leisure is uniform, $w_{k_1} \in (w^L, w^C)$.[49] Also in the first iteration the worker best responds selecting $r_{k_1} = r_0$ since the initial beliefs at k_0 lie on the worker's best response function. In iteration two the worker best-responds to the iteration one wage choice setting $r_{k_2} = r^*(w_{k_1}) = w_{k_1}$. Also in iteration two the firm best responds to the iteration one threshold choice setting $w_{k_1} = w_{k_1}$. In the example shown $w_1 \in (w^L, w^C)$ and so the *simultaneous best-response dynamics* reaches convergence at k_2.

[48]We see that the *simultaneous best-response dynamics* converges in two steps. Rapid convergence is guaranteed by the fact that the firm has a local satiation point at w^C: for expected participation thresholds above this point best response is independent of the firm's expectation. Further, that the sequence converges to w^C is guaranteed by the (weak) positive monotonicity of both best response functions in the region above w^C.

[49]Whenever the sufficient condition holds we also know that $w_{k_1} \leq w^C(p^+)$ with equality only for $r_0 = \underline{b}$. Also, whenever $H(b)$ is weakly concave we can show that that $w_{k_1} \geq w^L(p^+)$.