The Shimer puzzle and the Endogeneity of Productivity*

Regis Barnichon
CREI and Universitat Pompeu Fabra

This version: June 2014

Abstract

Shimer (2005) argues that the Mortensen-Pissarides (MP) model of unemployment lacks an amplification mechanism because it generates too little fluctuations in labor market variables given productivity shocks of plausible magnitude. While the literature has focused on ways to enhance the amplification mechanism of the MP model, this paper argues that part of the problem lies with the endogeneity of productivity. With variable capacity utilization in labor or capital, measured productivity can respond endogenously to non-technology shocks. Because such endogenous productivity movements are small relative to unemployment fluctuations, the cyclical component of measured labor productivity can fluctuate a lot less than unemployment. To illustrate quantitatively the possible importance of this mechanism, I use a New-Keynesian model with search unemployment and endogenous productivity movements caused by variable labor effort. Using a conservative calibration, the model generates an apparent elasticity between labor market variables and measured productivity that is three times larger than in the MP model. Using a calibration in the spirit of Hagedorn and Manovskii (2008) but with less extreme values, the model can match the data.

JEL classifications: E32, E37, J63, J64

Keywords: Unemployment Fluctuations, Labor productivity, Search and matching model, New-Keynesian model

*I would like to thank Shigeru Fujita, Jordi Gali, Wouter den Haan, Barbara Petrongolo, Chris Pissarides, John Roberts, Silvana Tenreyro, Thijs van Rens, and seminar participants for helpful comments and discussions. Any errors are my own. E-mail: rbarnichon@crei.cat.

1 Introduction

While the Mortensen-Pissarides (MP) search and matching model has become the standard theory of equilibrium unemployment, Shimer (2005) and Costain and Reiter (2008) have argued that a standard calibration of the model generates less than 10 percent of the observed business cycle fluctuations in unemployment, vacancy or labor market tightness (the vacancy-unemployment ratio) given productivity shocks of plausible magnitude. Put differently, the MP model generates a much too low labor market tightness-productivity elasticity.

This so-called "Shimer puzzle" has attracted a lot of interest in the literature, and a number of researchers have focused on ways to create more amplification, so that small exogenous productivity movements generate large fluctuations in unemployment and labor market tightness. The response of the literature has taken two main forms. One strand has explored ways to modify the model so that a given shock to productivity has a larger impact on labor market variables.[1] Another strand argues that the problem does not lie with the model itself, but with its calibration. Hagedorn and Manovskii (2008) propose an alternative calibration, particularly of the worker's value of non-market activity and the worker's bargaining power, in which the labor market tightness-productivity elasticity is much higher and the MP model can match the cyclical volatility of labor market variables.

This paper proposes an additional reason for the high labor market tightness-productivity elasticity apparent in the data: the endogeneity of labor productivity. There is substantial evidence that, because of labor hoarding and variable capacity utilization, some of the movements in productivity are in fact endogenous.[2] For example, when the firm is demand constrained in the short-run, firms can respond to changes in demand by adjusting their level of capacity utilization of inputs (capital or labor), and measured labor productivity fluctuates *endogenously* with aggregate demand and hence unemployment. But if the endogenous response of productivity is small compared to that of unemployment, it is natural to observe a high labor market tightness-productivity elasticity, and part of the Shimer puzzle is simply a by-product of the endogeneity of productivity.

Moreover, while the Shimer puzzle literature has focused on the magnitude of the labor market tightness-productivity elasticity, little attention has been devoted to the sign of the elasticity. A standard MP model implies that an increase in productivity raises labor market tightness, i.e., that the labor market tightness-productivity elasticity is always positive. However, a VAR with long-run restrictions shows that conditional on technology shocks, the labor

[1] See, among others, Shimer (2004), Hall (2005), Hall and Migrom (2006), Mortensen and Nagypal (2007, 2008), Costain and Reiter (2008), Eyigungor (2010).

[2] See, among others, Bils and Cho (1994), Burnside, Eichenbaum and Rebelo (1993), Burnside and Eichenbaum (1996) and Basu and Kimball (1997).

market tightness-productivity elasticity is negative, in contradiction with the MP model, but that conditional on non-technology shocks, the elasticity is positive.[3]

Thus, the MP model is confronted with not one but two challenges: it needs to match the magnitude as well as the sign of the labor market tightness-productivity elasticity. In other words, the Shimer puzzle, i.e. the inability of the MP model to match the elasticity's magnitude, needs to be discussed conditional on the nature of the shock and in the context of a model that is consistent with the signs of the conditional elasticities.

Embedding the search and matching model in a New Keynesian framework with nominal rigidities and variable labor effort allows the model to reproduce the signs of the conditional elasticities.

Conditional on aggregate demand (i.e., non-technology) shocks, the labor market tightness-productivity elasticity is positive. A positive aggregate demand shock raises labor market tightness as firms need more labor and post vacancies. It also raises productivity, because firms must increase hours per worker and effort to satisfy demand in the short-run as employment is subject to hiring frictions. This paper shows *analytically* that the magnitude of the labor market tightness-productivity elasticity depends on two factors: (i) the magnitude of the short run increasing returns to hours generated by variable labor effort, which is a function of the utility cost of longer hours and higher effort, and (ii) the magnitude of the trade-off between the intensive and the extensive labor margin, which depends on the worker's value of non-market activity and on the worker's bargaining power. The first factor (i) matters because the smaller the short run increasing returns to hours, the smaller the endogenous productivity movements and the larger the apparent labor market tightness-productivity elasticity. The second factor (ii) depends on the same two parameters that Hagedorn and Manovskii (2008) emphasized in the context of the standard MP model. This parallel between the two models is not surprising because these two parameters determine the match surplus and hence the incentive of firms to adjust employment in response to shocks. However, the sensitivity of the labor market tightness-productivity elasticity to these parameters is different in the two models, because the transmission mechanisms of the two models are different. In the standard MP model, firms react to *exogenous* technology shocks, but in this New-Keynesian model, firms react to changes in aggregate demand by adjusting employment as well as hours per worker and effort, which leads to *endogenous* movements in both labor market tightness and productivity.

Conditional on technology shocks, the labor market tightness-productivity elasticity is negative. Following a positive technology shock, aggregate demand does not increase as much as productivity when prices are sticky. Being more productive, firms use less labor, and labor mar-

[3]Moreover, the unconditional correlation between labor productivity and labor market tightness has been positive since the mid 80s, in contradiction with the MP model (Barnichon, 2010).

ket tightness goes down. I show that the magnitude of the labor market tightness-productivity elasticity conditional on technology shocks depends for a large part on the response of the central bank to the shock. The less the central bank accommodates a positive technology shock, the less aggregate demand will increase in response to the higher productivity level and the less labor firms will need. Thus, the Shimer puzzle conditional on technology shock is of a different kind than what the literature has focused on and, in the context of this New Keynesian model, comes down to the calibration of the central bank policy rule.

A conservative calibration exercise with a low income replacement ratio (40%) shows that in the New-Keynesian search model with endogenous productivity movements, the labor market tightness-productivity elasticity is three times larger than in the standard MP model. Importantly, this improvement comes from the endogeneity of productivity and not from a stronger amplification mechanism per se, and thus does not suffer from Costain and Reiter (2008) observation that raising the amplification of the MP model generates unrealistically strong effects of policies on the labor market.

Using a calibration with a higher income replacement ratio and a lower workers' bargaining weight in the spirit of Hagedorn and Manovskii (2008), the model can match the empirical labor market tightness-productivity elasticity. Importantly, because of a higher sensitivity of the labor market tightness-productivity elasticity to workers' bargaining weight in the New-Keynesian search model than in the MP model, the calibration uses less extreme parameter values than Hagedorn and Manovskii (2008). As a result, it does not suffer from Mortensen and Nagypal (2007) criticism that the extreme values advocated by Hagedorn and Manovskii (2008) imply that workers work for a minuscule (and, according to Mortensen and Nagypal, too small to be plausible) surplus.

This paper extends the recent work by Sveen and Weinke (2008) and Barnichon (2010) on the importance of aggregate demand shocks in explaining unemployment fluctuations.[4] Unlike the present paper, Sveen and Weinke (2008) consider the polar case of demand shocks without endogenous movements in productivity, an assumption which generates a de facto infinite elasticity conditional on demand shocks and hence artificially raises the model's ability to account for the Shimer puzzle. Moreover, while Sveen and Weinke (2008) study is unconditional on the nature of the shock, the present paper emphasizes the importance of conditioning on the nature of the shock.[5] The New Keynesian model used in this paper draws from Barnichon

[4]See also Furlanetto and Sveen (2009) for more evidence on the fact that technology shocks alone cannot explain labor market dynamics.

[5]In addition, while Sveen and Weinke's (2008) model introduces a separation between firms facing price stickiness and firms facing hiring frictions, the present model explicitly considers the interaction of nominal rigidities and hiring frictions in the spirit of Krause and Lubik (2007). Hence, it offers a response to Shimer's (2008) comment of Sveen and Weinke (2008) that the interaction of nominal and hiring frictions should not be ignored.

(2010) but the focus of this paper is different. While Barnichon (2010) shows quantitatively that changes in the volatility of aggregate demand shocks are behind movements in the sign of the unemployment-productivity correlation over time, this paper studies, both analytically and quantitatively, the effects of demand constraints and variable capacity utilization on the magnitude of the unemployment-productivity elasticity. Recently, Balleer (2009) has also stressed the importance of studying the Shimer puzzle conditionally on the nature of the shock and provided a thorough study of the effect of different types of technology shocks estimated from structural VARs.

The remainder of the paper is organized as follows: Section 2 discusses Shimer's (2005) puzzle; Section 3 presents a New-Keynesian model with search unemployment; Section 4 discusses analytically how the endogeneity of productivity can affect the labor market tightness-productivity elasticity; Section 5 quantifies the determinants of that elasticity, Section 6 presents simulation results; Section 7 offers some concluding remarks.

2 The Shimer puzzle

2.1 Shimer's (2005) evidence

Table 1 presents the standard deviations of unemployment, labor market tightness, hours per worker and labor productivity over 1951-2007.[6] As originally argued by Shimer (2005), the volatility of productivity is only a fraction of the volatility of labor market tightness. In fact, the ratio of standard-deviations of labor market tightness and productivity is $\sigma_\theta^{US}/\sigma_{lp}^{US} = 26$ where σ_x represents the standard-deviation of $\ln x$.

In the context of a standard MP model where productivity movements are the central driving force of unemployment fluctuations, Shimer (2005) shows that the standard deviations of unemployment, vacancies and productivity are of the same order of magnitude, and that $\sigma_\theta^{MP} \approx 2 \cdot \sigma_{lp}^{MP}$. Thus, the MP model generates less than 10 percent of the observed volatility in labor market tightness given productivity shocks of plausible magnitude.

2.2 Fixing the model to add more amplification

One way to reconcile the MP framework with the data is to modify the model so that it generates more amplification, i.e. that a given shock to productivity has a larger impact on unemployment. Mortensen and Nagypal (2007) provide a detailed review of the current effort in that direction, and I will only emphasize two influential examples. A first possibility,

[6] Labor productivity is measured as output per hour. I remove low-frequency movements using a standard HP-filter with $\lambda = 1600$. Alternatively, using $\lambda = 10^5$ as in Shimer (2005) does not change any of the results presented in this paper.

suggested by Hall (2005) and Shimer (2005), is to introduce real wage rigidity. In the standard MP model, the Nash bargaining real wage responds so much to movements in productivity that it effectively absorbs most of the changes in productivity. As a result, the surplus of the match responds only weakly to fluctuations in productivity. By introducing a degree of real wage rigidity, movements in productivity have a more substantial impact on the match surplus, on the incentives of firms to post vacancies and hence on equilibrium unemployment. However, this approach has been criticized, most notably by Pissarides (2007), on the account that wages in new matches –the wages that matter for job creation– are not rigid but display significant volatility, in line with the prediction of a standard MP model with flexible wages.

Another possibility, suggested by Hagedorn and Manovskii (2008), does not rely on real wage rigidity but uses a standard MP model with a different calibration than the one used in Shimer's. Hagedorn and Manovskii (2008) show that when the opportunity cost of employment is high, the job finding rate becomes very responsive to changes in productivity, and the MP model can quantitatively account for the magnitude of unemployment fluctuations. However, Hagedorn and Manovskii's (2008) calibration has been criticized by Costain and Reiter (2008) and Mortensen and Nagypal (2007). The former argue that the calibration generates unrealistically strong effects of policies on the labor market, while the latter note that the calibration implies that workers work for a minuscule (and, according to Mortensen and Nagypal, too small to be plausible) surplus.

While these two approaches are different, the underlying philosophy is the same: one needs to modify the MP model (either its equations or its calibration) so that the surplus of the match becomes more responsive to exogenous changes in productivity.

2.3 The conditional volatilities of productivity and labor market tightness

While the aforementioned literature considers productivity movements as exogenous, there is substantial evidence that, perhaps due to labor hoarding and variable capacity utilization, some of the movements in productivity are in fact endogenous.[7]

To identify the impact of exogenous changes in productivity on labor market variables, I impose long-run restrictions in structural VAR models to identify technological disturbances as in Gali (1999).[8] Technology shocks are the only shocks with a permanent impact on productivity, and I interpret transitory productivity movements as variations in capacity utilization.

[7] See, among others, Bils and Cho (1994), Burnside, Eichenbaum and Rebelo (1996) and Basu and Kimball (1997).

[8] See also Balleer (2009) and Canova, Lopez-Salido and Michelacci (2009) for more work on the effect of technology shocks in the context of VARs with long-run restrictions. In particular, Balleer (2009) tests and confirms the robustness of the long-run identification restriction used to identify the effect of technology shocks.

Specifically, I estimate the system

$$\begin{pmatrix} \Delta ln\ \frac{Y_t}{n_t h_t} \\ ln\ \theta_t \end{pmatrix} = C(L) \begin{pmatrix} \varepsilon_t^a \\ \varepsilon_t^m \end{pmatrix}$$

where $\frac{Y_t}{n_t h_t}$ is labor productivity defined as output per hour, θ_t the vacancy-unemployment ratio, $C(L)$ an invertible matrix polynomial and the vector of structural orthogonal innovations comprises technology shocks ε_t^a and non-technology shocks ε_t^m.[9]

Figure 1 presents the impulse response functions. The Shimer puzzle, i.e. the high value of $\left|\frac{\sigma_\theta^{US}}{\sigma_{lp}^{US}}\right|$, is clearly apparent for each shock, as the empirical standard deviation of labor market tightness is more than an order of magnitude larger than the standard deviation of output per hour. Moreover, technology shocks imply a negative labor market tightness elasticity while non-technology shocks imply a positive elasticity. In contrast, the MP model always implies a positive elasticity. Thus, the MP model is confronted with not one but two challenges: it needs to match the magnitude as well as the sign of the labor market tightness-productivity elasticity. In other words, the Shimer puzzle needs to be discussed *conditionally* on the nature of the shock and in the context of a model that can reproduce the conditional elasticities.

In the next section, I embed the search and matching model in a New-Keynesian framework with nominal rigidities and variable labor effort. Such a model can generate elasticities with the correct signs, allowing me to discuss the Shimer puzzle conditionally on the nature of the shock.

3 A New-Keynesian model with search unemployment

In a neoclassical setting, firms post vacancies depending on the return of the match. However, this needs not be the case when firms have to satisfy a given level of demand for their products. In a New-Keynesian setting with nominal rigidities, firms may have to hire more workers when demand is unexpectedly high even if productivity (and hence the match surplus) does not increase. Put differently, *the number of posted vacancies could increase without any change in productivity.*

In practice, firms also respond to higher demand by increasing capacity utilization of inputs (capital or labor), and measured labor productivity fluctuates *endogenously* with aggregate demand and hence unemployment. However, such endogenous productivity movements may

[9] Following Fernald (2007), I allow for two breaks in $\Delta \ln \left(\frac{y_t}{n_t h_t}\right)$, 1973:Q1 and 1997:Q1, and I filter the unemployment series with a quadratic trend. Fernald (2007) showed that the presence of a low-frequency correlation between labor productivity growth and unemployment, while unrelated to cyclical phenomena, could significantly distort the estimates of short run responses obtained with long run restrictions.

be small relative of unemployment fluctuations. If this is the case, the cyclical component of measured labor productivity fluctuates less than unemployment, and part of Shimer's puzzle can simply be a by-product of the endogeneity of productivity.

To illustrate this mechanism quantitatively, I use a New-Keynesian model with search unemployment and variable labor effort.[10] Importantly, while I only focus on variable labor effort to generate generating endogenous labor productivity movements for tractability reasons, the mechanism would also apply to variable capital utilization.

In this framework, unemployment fluctuations are the product of two disturbances: technology shocks and aggregate demand shocks (monetary policy or preference shocks). A positive technology shock permanently raises productivity but a positive aggregate demand shock also increases measured productivity (albeit temporarily), because firms increase labor effort to satisfy demand in the short run. As a result, measured labor productivity is the product of two components: permanent and temporary disturbances.

3.1 Households

There exists a continuum of households of measure one. To avoid distributional issues, I follow Merz (1995) and Andolfatto (1996) and assume that households form an extended family that pools its income and chooses per capita consumption and assets holding to maximize its expected lifetime utility. There are $1 - n_t$ unemployed workers who receive unemployment benefits b_t, and n_t employed workers who receive earnings $w_{it} = \omega_{it} h_{it} e_{it}$ from firm i for providing hours h_{it} and effort per hour e_{it} at a wage ω_{it} per unit of efficient hour. Denoting $g(h_{it}, e_{it})$ the individual disutility from working, the representative family maximizes

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\Gamma_t \ln(C_t) + \zeta_m \ln(\frac{M_t}{P_t}) - n_t \int_0^1 g(h_{it}, e_{it}) di \right] \tag{1}$$

subject to the budget constraint

$$\int_0^1 P_{jt} C_{jt} dj + M_t = \int_0^1 n_t w_{it} di + (1 - n_t) b_t + \Pi_t + M_{t-1} \tag{2}$$

with ζ_m a positive constant, M_t nominal money holdings and Π_t total transfers to the family. C_t is the composite consumption good index $C_t = \left(\int_0^1 C_{it}^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}$ where C_{it} is the quantity of good $i \in [0, 1]$ consumed in period t, P_{it} is the price of variety i, and $\varepsilon > 1$ is the elasticity of substitution among consumption goods. To introduce aggregate demand shocks, I use a stan-

[10] Walsh (2004), Krause and Lubik (2007) and Trigari (2009) are other important examples of New-Keynesian models with endogenous job destruction.

dard New-Keynesian short-cut and introduce preference shocks. Γ_t is an exogenous preference shifter evolving according to $\ln \Gamma_t = \rho_\Gamma \ln \Gamma_{t-1} + \varepsilon_t^\Gamma$. The aggregate price level is defined as $P_t = \left(\int_0^1 P_{it}^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}$. The disutility from supplying hours of work h_t and effort per hour e_t is the sum of the disutilities of the members who are employed. Following Bils and Cho (1994), the individual period disutility of labor takes the form $g(h_{it}, e_{it}) = \frac{\zeta_h}{1+\sigma_h} h_{it}^{1+\sigma_h} + h_{it} \frac{\zeta_e}{1+\sigma_e} e_{it}^{1+\sigma_e}$ where ζ_h, ζ_e, σ_h and σ_e are positive constants. The last term reflects disutility from exerting effort with the marginal disutility of effort per hour rising with the number of hours. An infinite value for σ_e generates the standard case with inelastic effort.

3.2 Firms and the labor market

Each differentiated good is produced by a monopolistically competitive firm using labor as the only input. There is a continuum of large firms distributed on the unit interval. At date t, each firm i hires n_{it} workers to produce a quantity $Y_{it} = A_t n_{it} L_{it}^\alpha$ where A_t is an aggregate technology index, L_{it} the effective labor input supplied by each worker and $0 < \alpha < 1$.[11] Effective labor input is a function of hours h_{it} and effort per hour e_{it} with $L_{it} = h_{it} e_{it}$.

Being a monopolistic producer, the firm faces a downward sloping demand curve $Y_{it}^d = (\frac{P_{it}}{P_t})^{-\varepsilon} Y_t$ and chooses its price P_{it} to maximize its value function given the aggregate price level P_t and aggregate output Y_t. Firms are subject to Calvo-type price setting, and each period a fraction ν of randomly selected firms cannot reset its price.

In a search and matching model of the labor market, firms post vacancies at a cost c_t, and unemployed workers search for jobs. The matching function takes the Cobb-Douglas form so that the flow m_t of successful matches within period t is given by $m_t = m_0 u_t^\eta v_t^{1-\eta}$where m_0 is a positive constant, $\eta \in (0, 1)$, u_t denotes the number of unemployed and v_t=$\int_0^1 v_{it} di$ the total number of vacancies posted by all firms. Accordingly, the probability of a vacancy being filled in the next period is $q(\theta_t) \equiv m(u_t, v_t)/v_t = m_0 \theta_t^{-\eta}$ where $\theta_t \equiv \frac{v_t}{u_t}$ is the labor market tightness. Hiring takes one period (so that employment n_{it} is predetermined at time t), and matches are destroyed at a constant rate λ, so that the law of motion for employment of firm i is given by $n_{it+1} = (1 - \lambda) n_{it} + q(\theta_t) v_{it}$.

When a firm and a worker meet, they must decide on the allocation of hours and effort to satisfy demand. It is assumed that both parties negotiate the hours/effort decision by choosing the optimal allocation, i.e., by choosing hours and effort per hour to satisfy demand at the

[11] The model does not explicitly consider capital for tractability reasons but the production technology can be rationalized by assuming a constant capital-worker ratio and a standard Cobb-Douglas production function $Y_{it} = A_t (n_{it} L_{it})^\alpha K_{it}^{1-\alpha}$.

lowest utility cost for the worker. More precisely, they solve

$$\min_{h_{it},e_{it}} \frac{\zeta_h}{1+\sigma_h} h_{it}^{1+\sigma_h} + h_{it}\frac{\zeta_e}{1+\sigma_e} e_{it}^{1+\sigma_e} \tag{3}$$

subject to satisfying demand $A_t n_{it} h_{it}^{\alpha} e_{it}^{\alpha} = Y_{it}^{d}$. The first-order conditions imply that effort per hour is a function of hours per worker with $e_{it} = e_0 h_{it}^{\frac{\sigma_h}{1+\sigma_e}}$ where e_0 a positive constant. Thus, changes in hours can proxy for changes in effort, and the firm production function can be rewritten

$$Y_{it} = Y_0 A_t n_{it} h_{it}^{\varphi} \tag{4}$$

with $y_0 = e_0^{\alpha}$ and

$$\varphi = \alpha\left(1 + \frac{\sigma_h}{1+\sigma_e}\right). \tag{5}$$

With $\varphi > 1$, the production function displays short run increasing returns to hours. In times of higher demand, firms respond by increasing hours and effort, which increases output per hour, i.e., measured labor productivity. This condition is critical to generate the procyclical response of measured productivity to aggregate demand shocks, and from now on, it is assumed that the model parameters ensure $\varphi > 1$.[12]

Firms and workers bargain individually about earnings and split the surplus in shares determined by an exogenous worker's bargaining weight γ. As shown in the Appendix, when firms hire many workers and when there are decreasing returns to hours, the wage negotiation problem resembles the intra-firm bargaining problem of Stole and Zwiebel (1996) and earnings w_{it} satisfies the differential equation

$$w_{it} = \gamma\left(\frac{h_{it}}{\varphi}\frac{\partial w_{it}}{\partial h_{it}} + c_t\theta_t\right) + (1-\gamma)\left(b_t + \frac{g(h_{it}, e_{it})}{\upsilon_t}\right) \tag{6}$$

of which a solution is given by

$$w_{it} = \gamma c_t \theta_t + (1-\gamma) b_t + (1-\gamma)\varkappa \frac{h_{it}^{1+\sigma_h}}{\upsilon_t} \tag{7}$$

with $\upsilon_t = \frac{1}{C_t}$ the marginal utility of consumption and $\varkappa = \frac{\zeta_h \frac{1+\sigma_h+\sigma_e}{(1+\sigma_h)\sigma_e}}{1-\frac{\gamma}{\varphi}(1+\sigma_h)} > 0$, so that earnings increase with hours per worker at the rate $1+\sigma_h$. While the earnings equation (25) is a weighted average of both parties surpluses and is similar to other bargained wages derived in search models (Pissarides, 2001), the firm's surplus is *not* given by the marginal product of

[12] This condition holds with sufficiently high marginal product of efficient hour (high α) or high effort elasticity with respect to hours (high $\frac{\sigma_h}{1+\sigma_e}$).

labor, because once the firm has chosen its price, it is demand constrained and a marginal worker will not increase the firm's revenue. Instead, the gross surplus of an additional match is given by $-n_{it}\frac{\partial w_{it}}{\partial n_{it}} = \frac{h_{it}}{\varphi}\frac{\partial w_{it}}{\partial h_{it}} > 0$ (the first term of (25)). As discussed in the next section, it captures the reduction in the wage bill obtained by substituting the flexible but expensive (because of convex disutility costs in hours and effort) intensive labor margin with the extensive margin.[13]

Given the aggregate price level, firm i will choose a sequence of price $\{P_{it}\}$ and vacancies $\{v_{it}\}$ to maximize the expected present discounted value of future profits

$$E_t \sum_j \beta^j \frac{u'(C_{t+j})}{u'(C_t)} \left[\frac{P_{i,t+j}}{P_{t+j}} Y^d_{i,t+j} - n_{i,t+j} w_{i,t+j} - c_{t+j} v_{i,t+j} \right] \tag{8}$$

subject to the Calvo price setting rule, the demand constraint $Y^d_{it} = (\frac{P_{i,t}}{P_t})^{-\varepsilon} Y_t$, the hours-effort choice (4), the law of motion for employment $n_{it+1} = (1-\lambda)n_{it} + q(\theta_t)v_{it}$ and the bargained wage (26).

3.3 Technological progress and the central bank

To be consistent with the long run identifying assumption made in Section 2, the technology index series is non-stationary with a unit root originating in technological innovations. Technology evolves according to $A_t = \bar{A}_t e^{a_t}$ with $\bar{A}_t = (1+g_a)\bar{A}_{t-1}$ the deterministic component with growth rate g_A and a_t the stochastic component with $a_t = a_{t-1} + \varepsilon^a_t$. $\varepsilon^a_t \sim N(0, \sigma^a)$ is a technology shock with a permanent impact on productivity. Consistent with a growing economy and zero inflation in "steady-state", money supply evolves according to $M_t = \bar{A}_t e^{m_t}$ with $\Delta m_t = \rho_m \Delta m_{t-1} + \varepsilon^m_t + \tau^{cb}\varepsilon^a_t$, $\rho_m \in [0,1]$ and $\varepsilon^m_t \sim N(0,\sigma^m)$. I interpret ε^m_t as monetary policy shock. As in Gali (1999), when $\tau^{cb} \neq 0$, the monetary authority responds in a systematic fashion to technology shocks.[14]

3.4 Closing the model

Averaging firms' employment, total employment evolves according to $n_{t+1} = (1-\lambda)n_t + v_t q(\theta_t)$. The labor force being normalized to one, the number of unemployed workers is $u_t = 1 - n_t$. In this non-stationary economy, unemployment benefits and vacancy posting costs grow in line with technology so $c_t = cA_t$ and $b_t = bA_t$. Finally, as in Krause and Lubik (2007), vacancy posting costs are distributed to the aggregate households so that $C_t = Y_t$ in equilibrium.

[13] The model is well behaved only if $\varkappa > 0$. This imposes that $1 - \frac{\gamma}{\varphi}(1+\sigma_h) > 0$, which will be verified by the calibrated parameters.

[14] Using a Taylor rule where the interest rate responds to deviations of inflation from its steady state and to the level and the growth rate of the output gap (as in, e.g., Smets and Wouters, 2007) gives similar results.

4 The Shimer puzzle in a New-Keynesian setting

Using the New-Keynesian model with search unemployment and endogenous productivity movements (henceforth, NKMP), I now revisit the Shimer puzzle in a world with demand constraints and endogenous productivity movements. Specifically, I discuss how demand constraints and endogenous productivity affect the measured labor market tightness-productivity elasticity.

This section shows analytically that, even when technology is held constant, productivity can move endogenously in response to changes in aggregate demand and lead to a Shimer-type puzzle in which measured labor productivity is a lot less volatile than labor market tightness. The relative volatilities of labor market tightness and *measured* productivity is then shown to depend on two key factors: (i) the value of the short run increasing returns to hours parameter $\varphi=\alpha(1+\frac{\sigma_h}{1+\sigma_e})$, and (ii) the magnitude of the trade-off between the intensive and the extensive labor margin, which depends on the worker's value of non-market activity and on workers' bargaining power.

4.1 The endogeneity of productivity

The production function takes the form $Y_t = Y_0 A_t n_t h_t^{\varphi}$ so that labor productivity $lp_t \equiv \frac{Y_t}{n_t h_t}$ can be written:[15]

$$\ln lp_t = (\varphi - 1)\ln h_t + \ln A_t + \ln Y_0. \tag{9}$$

Thus, the behavior of measured productivity depends on shocks to technology A_t *as well as* on (i) the behavior of the intensive margin h_t and (ii) the value of φ. The closer is φ to 1, the smaller are the endogenous movements in productivity for given movements in hours per worker, i.e., the larger is the hours per worker-productivity elasticity.

To relate labor productivity and labor market variables through (9), I need to link hours per worker to vacancy posting and labor market tightness. Such a link is provided by the vacancy posting condition, which I discuss next.

[15] Importantly, in the context of our model where employment is a state variable and where hours can be adjusted instantaneously, the natural definition of labor productivity is output per hour. Analyzing an alternative definition –output per worker– also sometimes used in the literature, would require an extension (and complication) of the model, because in the present model output and output per worker are perfectly correlated (since employment is a state variable). Specifically, one would need to allow firms to adjust employment instantaneously (as in e.g., Michaillat, 2012) and have decreasing returns in employment, as in Michaillat (2012) or Elsby and Michaels (2013). With decreasing returns in employment, the general point of the paper would go through: measured output per worker would fluctuate endogenously with aggregate demand and employment, but such endogenous productivity movements would be smaller than employment fluctuations, thereby explaining part of the Shimer puzzle.

4.2 The vacancy posting condition and the intensive margin-extensive margin trade-off

The vacancy posting condition captures the trade-off between the intensive and the extensive labor margins. Indeed, because hiring is subject to time consuming frictions, a trade-off emerges between the less flexible extensive margin (n_{it}) and the intensive margin (h_{it} and e_{it}), which is flexible but more costly because of convex utility costs in hours and effort. The vacancy posting condition is given by

$$\frac{c_t}{q(\theta_t)} = E_t\beta_{t+1}\left[\chi_{it+1} + \frac{c_{t+1}}{q(\theta_{t+1})}(1-\lambda)\right] \tag{10}$$

with $c_t = \frac{c}{\upsilon_t}$ and χ_{it}, the shadow value of a marginal worker, given by

$$\begin{aligned}\chi_{it} &= -\frac{\partial n_{it}w_{it}}{\partial n_{it}} = -n_{it}\frac{\partial w_{it}}{\partial n_{it}} - w_{it} \\ &= (1-\gamma)\varkappa\frac{1+\sigma_h}{\varphi}\frac{h_{it}^{1+\sigma_h}}{\upsilon_t} - w_{it}. \end{aligned} \tag{11}$$

Each firm posts vacancies until the expected cost of hiring a worker $\frac{c_t}{q(\theta_t)}$ equals the expected discounted future benefits $\left\{\chi_{it+j}\right\}_{j=1}^{\infty}$ from an extra worker. Once the firm has chosen its price, it is demand constrained, and the flow value of a marginal worker is not his contribution to revenue but his reduction of the firm's wage bill. The first term of χ_{it} ($-n_{it}\frac{\partial w_{it}}{\partial n_{it}}$) captures the gross surplus of an additional match and represents the savings due to the decrease in hours and effort achieved with an additional worker. Indeed, hiring an extra worker allows the firm to reduce hours per worker for all its workers, and through (26) to lower the wage of all its workers. The second term of χ_{it} is the wage payment going to an extra worker.

Using the wage equation, the marginal worker's value takes the form

$$\chi_{it} = -\gamma c_t\theta_t - (1-\gamma)b_t + (1-\gamma)\left(\frac{1+\sigma_h}{\varphi} - 1\right)\varkappa\frac{h_{it}^{1+\sigma_h}}{\upsilon_t}. \tag{12}$$

Provided that $\frac{1+\sigma_h}{\varphi} > 1$, the worker's marginal value increases with hours per worker. Since $h_t = \left(\frac{Y_t^d}{A_t n_t y_0}\right)^{\frac{1}{\varphi}}$ and n_t is a state variable, the firm relies on the intensive margin to satisfy demand in the short-run, and the level of hours per worker captures "demand pressures" and the firm's incentives to post vacancies. With $\frac{1+\sigma_h}{\varphi} > 1$, the longer hours are, the larger is the wage bill reduction obtained with an extra worker.[16] As hours increase because of higher

[16] $1+\sigma_h > \varphi$ captures the fact that, absent hiring frictions, the firm would rather hire an extra-worker than use the intensive margin because the cost of longer hours increases faster than output. This property of the model captures the fact that although it is easier to increase the workload of an employee than to hire a new one,

demand for the firm's products, the worker's marginal value increases, and the firm posts more vacancies to increase employment.

Thus, (10) and (12) capture the trade-off between the intensive and the extensive margin, and determine the relative volatilities of labor market tightness and hours per worker.

4.3 The labor market tightness-productivity elasticity

I can now combine the two previous results (9) and (10) to study the parameters determining the labor market tightness-productivity elasticity. To get closed-form expressions and put the discussion in the perspective of the Shimer puzzle literature, I proceed as in Shimer (2005), Hagedorn and Manovskii (2008) and Mortensen and Nagypal (2007), and consider the model without aggregate shocks. After a little bit of algebra left for the Appendix, one can show that the elasticity between labor market tightness and hours per worker in the NKMP model is given by

$$\epsilon_{\theta,h}^{NKMP} = \varphi \frac{p}{p-b} \kappa \tag{13}$$

with $\kappa \equiv \frac{\frac{1}{q}(1-\beta(1-\lambda))+\gamma\beta\theta}{\frac{\sigma}{q}(1-\beta(1-\lambda))+\gamma\beta\theta}$ and $p \equiv (\frac{1+\sigma_h}{\varphi} - 1)\varkappa h^{1+\sigma_h} y$ the gross surplus of a match net of the disutility cost of hours.

Using the definition of labor productivity and *holding technology constant* gives $\ln lp_t = (\varphi - 1)\ln h_t$, so that the labor market tightness-productivity elasticity is non-zero and given by

$$\epsilon_{\theta,lp}^{NKMP} = \frac{\varphi}{\varphi - 1} \frac{p}{p-b} \kappa. \tag{14}$$

To interpret (13) and (14), it is instructive to compare them to the labor market tightness-productivity elasticity in the context of a standard MP model. Using the present paper's notations, Hagedorn and Manovskii (2008) show that in a standard MP model without aggregate uncertainty, the elasticity of labor market tightness with respect to productivity is

$$\epsilon_{\theta,lp}^{MP} = \frac{A}{A-b} \kappa \tag{15}$$

with A the gross surplus of a match in the MP model.

While similar, (14) and (15) differ in two points. The first difference is the multiplicative term $\frac{\varphi}{\varphi-1}$. In the NKMP model, and unlike the standard MP model, productivity does not play any causal role in (14) but instead responds endogenously (alongside labor market tightness) to changes in aggregate demand. The smaller φ, the smaller the endogenous movements in

overtime hours are more expensive than regular ones because of convex disutility costs of hours. The model's parameters will verify $\frac{1+\sigma_h}{\varphi} - 1 > 0$.

measured productivity, the larger $\frac{\varphi}{\varphi-1}$ and the larger is $\epsilon_{\theta,lp}^{NKMP}$.

The second difference is that A appears to have been substituted for p in (14). Indeed, p and A play similar roles in each model as both capture the gross surplus of a match:[17] p captures the gross surplus of an additional match net of the disutility cost of hours in the NKMP model, and A captures the gross surplus of a match in the MP model.[18] However, the surplus is different in each model. In the NKMP model, the firm's surplus is not given by the marginal product of labor A, because once the firm has chosen its price, it is demand constrained and a marginal worker will not increase the firm's revenue. Instead, the value of a marginal worker is to allow the firm to lower the wage bill of satisfying a given level of demand by substituting the flexible but expensive intensive margin (hours and effort) with the cheaper extensive margin (employment). This trade-off between the intensive and the extensive margin is apparent in the expression of p as $p \propto \frac{1+\sigma_h}{\varphi} - 1$, which measures the difference between the hours per worker margin and the employment margin in terms of the cost of providing the required amount of output. The intensive margin displays increasing returns with $\varphi > 1$ but its cost increases at the rate $1 + \sigma_h$ so that the cost of producing a given quantity Y^d increases at the rate $\frac{1+\sigma_h}{\varphi} > 1$. For the extensive margin, on the other hand, both output and costs increase linearly, so that the rate is one. The larger the difference between the two rates, i.e., the larger $\frac{1+\sigma_h}{\varphi} - 1$, the stronger is the firm's incentive to avoid increases in hours per worker, and the larger the value of a marginal worker.

5 The effect of the endogeneity of productivity

By comparing (14) and (15), one can easily compare the performances of the standard MP model and the NKMP model in terms of the relative volatilities of labor market tightness and labor productivity. To highlight the importance of the endogeneity of productivity in the context of the Shimer puzzle literature, I first consider a conservative (in terms of $\epsilon_{\theta,lp}^{MP}$) calibration and compare estimates of $\epsilon_{\theta,lp}$ across the MP and the NKMP models. Then, I study analytically the determinants of $\epsilon_{\theta,lp}^{NKMP}$.

[17] One can also notice the symmetry between the two models by considering the gross and net surpluses of an additional match. In the MP model, the gross surplus of a match is A and the net surplus is $\chi = A - w = A(1-\gamma) - \gamma c\theta - (1-\gamma)b$. In the NKMP model, the contribution of a match is to reduce the cost of satisfying a given level of demand, and the gross surplus of a marginal worker is $(1-\gamma)\varkappa\frac{1+\sigma_h}{\varphi}\frac{h_{it}^{1+\sigma_h}}{\lambda_t}$ with the net surplus given by $\chi = (1-\gamma)\varkappa\frac{1+\sigma_h}{\varphi}h^{1+\sigma_h}y - w = p(1-\gamma) - \gamma c\theta - (1-\gamma)b$ with $p = (\frac{1+\sigma_h}{\varphi} - 1)\varkappa h^{1+\sigma_h}y$.

[18] Another subtle difference is that in the standard MP model (15), it is the gross surplus that matters, not the gross surplus net of the disutility of work as in the NKMP model (14). This is simply because no disutility of work is assumed in the standard MP model (in addition to the value of unemployment benefits/home production b), whereas the NKMP model also features a convex disutility cost of working.

5.1 A first calibration

The quarterly discount factor β is set to 0.99 and the returns to labor α to 0.64, as typically used in the literature (e.g., Bils and Cho, 1994). The markup of prices over marginal costs is assumed to average 10 percent, which amounts to setting ε equal to 11. Consistent with Bils and Klenow (2004), firms reset their price every 2 quarters. Turning to the labor market, the matching function elasticity is set to η=0.5. The quarterly job separation rate λ is set to 0.1, consistent with US evidence that jobs last for two years and a half (Shimer, 2005), and the job finding rate is set to 0.6, implying a steady-state unemployment rate of 10 percent.[19] I set ζ_h and ζ_e so that in steady-state, h=e=1. As a baseline calibration, I set $\sigma_h = 2$ (i.e., an hours per worker elasticity of 0.5) and $\sigma_e = 0.5$ to fix a value for $\varphi = 1.5$ roughly in the middle of the 1.3-1.6 range reported by Basu and Kimball (1997), and I use Shimer's conservative income replacement ratio $\rho_b = \frac{b}{w} = 0.4$. Finally, the bargaining weight γ is set to to a middle value 0.5.

The values for $\epsilon_{\theta,lp}$ for the MP and the NKMP models become

$$\begin{cases} \epsilon_{\theta,lp}^{NKMP} = 5.7 \\ \epsilon_{\theta,lp}^{MP} = 1.9 \end{cases} \tag{16}$$

so that the elasticity of labor market tightness with respect to measured productivity is three times bigger in the NKMP model.

Importantly, the difference between (14) and (15) comes from the endogeneity of productivity, not from different amplification mechanisms as the calibration delivers similar values for $\frac{1}{1-b/A}$ and $\frac{1}{1-b/p}$ (about 1.6) (and $1/\kappa = 1.2$ being identical for the two models). As a result, the difference between (14) and (15) owes mostly to $\frac{\varphi}{\varphi-1} = 3$ and the fact that measured productivity is an endogenous variable that fluctuates with aggregate demand. As the endogenous response of productivity to shocks is relatively small, measured labor productivity fluctuates less than labor market tightness, and part of Shimer's puzzle is a by-product of the endogeneity of productivity. The smaller the short-run increasing returns to scale parameter, the less productivity fluctuates and the larger the ratio of the variance of labor market tightness to the variance of measured productivity. Put differently, the smaller φ, the less the NKMP model suffers from a Shimer-type puzzle.

Importantly, this significant improvement in the performance of the search and matching model with respect to $\epsilon_{\theta,lp}$ does not come at the cost of unrealistic responses of unemployment to changes in labor market policies. Costain and Reiter (2008) showed that increasing the amplification mechanism of the standard MP model through $\frac{A}{A-b}$ in order to match $\epsilon_{\theta,lp}^{US}$ implies

[19] As in Merz (1995), Andolfatto (1996), den Haan, Ramey, and Watson (2000) and others, model unemployment includes those individuals registered as inactive that are actively searching.

a stronger (and unrealistic) effect of policy changes on labor market variables. In contrast, in the present model, the improvement in performances comes from the additional term $\frac{\varphi}{\varphi-1}$, so that the response of labor market variables to policy changes is unchanged.

5.2 The role of the income replace ratio (b), workers' bargaining weight (γ) and short-run increasing to hours (φ)

The standard calibration presented above generated similar low values for $\frac{A}{A-b}$ and $\frac{p}{p-b}$, which allowed me to highlight the key role played by the endogeneity of productivity and its impact on the labor market tightness-productivity elasticity. However, as Hagedorn and Manovskii (2008) argued, other calibrations of the MP model are possible.

I now show that, just as in the MP model, calibrating the values of b and γ is of crucial importance. However, we will also see that the sensitivity of the labor market tightness-productivity elasticity to b and γ is different, because the transmission mechanisms of the two models are different. In the standard MP model, firms react to *exogenous* technology shocks, but in this New-Keynesian model, firms react to changes in aggregate demand by adjusting employment as well as hours per worker and effort, which leads to *endogenous* movements in both labor market tightness and productivity. In other words, the implications of demand constraints and variable capacity utilization for the Shimer puzzle may be more or less large depending on the values of b and γ. Finally, I will emphasize the importance of a new parameter: φ. The short-run increasing returns to hours parameter depends on the values of both σ_h and σ_e.

In this section, I proceed in a parallel fashion to Hagedorn and Manovskii (2008) and discuss the calibration of b, γ, σ_h and σ_e.

5.2.1 The income replacement ratio

As Hagedorn and Manovskii (2008) pointed out for the standard MP model, unemployment benefits b play a key role in (15) to determine the elasticity of labor market variables with respect to productivity. Indeed, for the range of plausible parameter values, κ is between 1 and 2 for values of γ between 0 and 1, so that the important parameter driving the elasticity of θ_t with respect to productivity is $A-b$, the difference between the gross surplus of a match and the opportunity cost of employment. The closer is A to b, the smaller the surplus and the more effect a given change in A has on the surplus in percentage terms and the larger is the firm's incentive to adjust vacancies. In the NKMP model, a similar reasoning is at play, and the closer is b to the surplus p, the bigger is $\epsilon_{\theta,lp}$.

Since the calibration of b is of first importance, it is helpful to restate the discussion in terms of income replacement ratio ρ_b with b=$\rho_b w$. After a little bit of algebra and the approximation $\frac{1}{q}(1-\beta(1-\lambda)) \ll \gamma\theta\beta$ for γ not too small, I get

$$\epsilon_{\theta,lp}^{NKMP} = \frac{\varphi}{\varphi-1}\frac{p}{p-\frac{\rho_b}{n\mu}} \tag{17}$$

which has a similar form to (14).[20]

It is clear from (17) that $\frac{\partial \epsilon_{\theta,lp}^{NKMP}}{\partial \rho_b} > 0$.[21] Using a higher income replacement ratio with ρ_b=0.7, a value used for instance in Costain and Reiter (2008), Mortensen and Nagypal (2007), and Sveen and Weinke (2008), and γ=0.5, I get

$$\begin{cases} \epsilon_{\theta,lp}^{NKMP} = 10.5 \\ \epsilon_{\theta,lp}^{MP} = 3.8 \end{cases}$$

As in the standard MP model, raising the value of unemployment benefits closer to the surplus of a match increases $\epsilon_{\theta,lp}$. However, raising ρ_b does little to $\epsilon_{\theta,lp}^{NKMP}/\epsilon_{\theta,lp}^{MP}$ because it raises the performance of both models through the same mechanism: the smaller the surplus, the larger the percentage changes in profits and the more volatile is labor market tightness.

5.2.2 The bargaining weight

Unlike the standard MP model in which γ has no direct effect on $\epsilon_{\theta,lp}^{MP}$, γ plays a crucial role in this NKMP model because it affects not only the wage w as in the standard MP model, but also the gross surplus of a marginal worker.[22]

In the NKMP model, the gross surplus of a marginal worker is the reduction in the wage bill obtained by substituting the intensive margin (hours and effort) with the extensive one (employment). This reduction depends on the share of the surplus going to the worker. With a lower bargaining power of the worker, changes in hours per worker have a stronger effect on the firm's vacancy posting condition through (26) (and hence on labor market tightness) because the firm can obtain a larger reduction in the wage with a marginal worker when the wage is more responsive to changes in h.

Using (14), it is easy to see that $\frac{\partial \epsilon_{\theta,lp}^{NKMP}}{\partial \gamma} < 0$.[23]

[20] The approximation behind (17) is only used to clarity of exposition. The Appendix shows that the result $\frac{\partial \epsilon_{\theta,lp}^{NKMP}}{\partial \rho_b} > 0$ is always true, and I always use the exact formulation to calculate the values of $\epsilon_{\theta,lp}^{NKMP}$.

[21] In all my comparative statics exercises, I focus on the effect of parameter changes on $\epsilon_{\theta,lp}^{NKMP}$ while abstracting from any equilibrium effect, in a similar fashion to Costain and Reiter (2008). See the Appendix for more details.

[22] However, γ does play a crucial role in the calibration strategy followed by Hagedorn and Manovskii (2008) through its effect the wage-productivity elasticity that Hagedorn and Manovskii target.

[23] Again, I use (17) for clarity of exposition, and for the range of plausible parameter values, $\frac{\partial \epsilon_{\theta,lp}^{NKMP}}{\partial \gamma} < 0$ is

Using the baseline calibration but the smaller bargaining weight $\gamma = 0.05$ as used by Hagedorn and Manovskii (2008), one gets

$$\begin{cases} \epsilon_{\theta,lp}^{NKMP} = 14.4 \\ \epsilon_{\theta,lp}^{MP} = 2.2 \end{cases}$$

so that the NKMP model now generates a volatility that 6.5 times larger than a standard MP model. As γ was lowered from 0.5 to 0.05, the labor market tightness-productivity elasticity increased faster in the NKMP model than in the MP model. Since Hagedorn and Manovskii (2008) match the empirical elasticity using $\gamma = 0.05$, we can already see that the NKMP model will need less extreme values of γ to match the data.

Intuitively, the sensitivity of $\epsilon_{\theta,lp}^{NKMP}$ to γ comes from the fact that γ tilts the trade-off between the intensive and extensive margin. The lower is workers' bargaining power, the closer is the wage to workers' reservation wage, which depends on the disutility of hours and effort. Thus, a decrease in workers' bargaining power raises the sensitivity of the wage to changes in hours and effort. Firms become more reluctant to increase hours and effort given the steeper wage schedule they face, and the extensive margin becomes more volatile (and the intensive margin less volatile). With a less volatile intensive margin, endogenous productivity movements are smaller, and the labor market tightness-productivity elasticity is larger.

One can now combine a higher income-replacement ratio with a lower bargaining weight. The previous calibration kept the income-replacement ratio $\rho_b = 0.4$, a value that Hagedorn and Manovskii (2008) found too low because it does not allow for the "value of leisure" or "home production" forgone when employed, in addition to the unemployment benefits. For instance, using a higher income-replacement ratio value $\rho_b = 0.7$ and γ=0.35 generates

$$\epsilon_{\theta,lp}^{NKMP} = 24.8$$

so that the model is close to matching the empirical labor market tightness-productivity elasticity conditional on non-technology shocks.

5.2.3 The short-run increasing returns to hours parameter

Compared to the standard MP model, a critical new parameter in this NKMP model with endogenous productivity is $\varphi = \alpha(1 + \frac{\sigma_h}{1+\sigma_e})$, the short-run increasing returns to hours parameter. The choice of φ used in the baseline calibration was in the middle of the range reported by Basu and Kimball (1997). I now discuss how changing φ through σ_h or σ_e affects $\epsilon_{\theta,lp}^{NKMP}$.

true with the exact expression of $\epsilon_{\theta,lp}^{NKMP}$.

Since $\epsilon_{\theta,lp} = \frac{\varphi}{\varphi-1}\epsilon_{\theta,h}$, the total effect of σ_h or σ_e on $\epsilon_{\theta,lp}$ is a combination of two effects. First, a lower σ_h or a higher σ_e lowers φ and raises $\frac{\varphi}{\varphi-1}$, which implies that, for a given volatility of the labor market variables (i.e. a given $\epsilon_{\theta,h}$), measured productivity is less volatile and $\epsilon_{\theta,lp}$ is higher. Second, changing σ_h or σ_e affects $\epsilon_{\theta,h}$ because it modifies the trade-off between the hours per worker and the employment margin. Because of these two effects, the impact of changes in σ_h or σ_e on $\epsilon_{\theta,lp}$ may be indeterminate. In fact, after some algebra left for the Appendix, one can show that $\frac{\partial\epsilon_{\theta,lp}^{NKMP}}{\partial\sigma_h}$ is always negative but that $\frac{\partial\epsilon_{\theta,lp}^{NKMP}}{\partial\sigma_e}$ may or may not be positive.

The effect of σ_h: $\frac{\partial\epsilon_{\theta,lp}^{NKMP}}{\partial\sigma_h} < 0$. A higher σ_h unequivocally lowers $\epsilon_{\theta,lp}$ because both effects go in the same direction. Differentiating (14) with respect to σ_h gives[24]

$$\frac{\partial\epsilon_{\theta,lp}^{NKMP}}{\partial\sigma_h} = \frac{-1}{(\varphi-1)^2}\frac{p}{p-b}\underbrace{\frac{\partial\varphi}{\partial\sigma_h}}_{>0} - \frac{\varphi}{\varphi-1}\frac{b}{(p-b)^2}\underbrace{\frac{\partial p}{\partial\sigma_h}}_{>0} < 0.$$

The first term on the right-hand side corresponds to the fact that increasing σ_h raises φ, leading to larger movements in productivity, which, holding $\epsilon_{\theta,h}$ constant, decreases $\epsilon_{\theta,lp}$. The second term on the right-hand side captures the effect of σ_h on p, the gross surplus of an additional match net of the disutility cost of hours. Because $p \propto \left(\frac{1+\sigma_h}{\varphi} - 1\right)$, a higher σ_h has two effects on p. First, a higher σ_h increases the utility *cost* of hours per worker $(1+\sigma_h)$, which tilts the trade-off between the hours per worker margin and the employment margin towards the employment margin and raises p. Second, a higher σ_h raises the *benefit* of using the hours per worker margin because it raises φ, which generates higher returns to hours, and ceteris paribus, lowers the value of a marginal worker. However, because $\frac{\partial}{\partial\sigma_h}\left(1 - \frac{\varphi}{1+\sigma_h}\right)>0$, the cost increases faster than the benefit and $\frac{\partial p}{\partial\sigma_h}>0$. With p higher and further away from b, profits are larger and changes are smaller in percentage terms. As a result, the employment margin becomes less volatile compared to the hours per worker margin. In other words, a higher σ_h lowers $\epsilon_{\theta,h}$, and hence $\epsilon_{\theta,lp}^{NKMP}$.

For instance, using the conservative baseline calibration but reducing σ_h from 2 to 1.5 to get φ=1.3 –the lower bound of the plausible range identified by Basu and Kimball (1997)– $\epsilon_{\theta,lp}^{NKMP}$ increases from 5.7 to 8.2, which is 4.3 times larger than $\epsilon_{\theta,lp}^{MP}$.

[24] See the Appendix for the final expression.

The effect of σ_e: $\frac{\partial \epsilon_{\theta,lp}^{NKMP}}{\partial \sigma_e} < 0$ **or** > 0. Raising σ_e has an indeterminate effect as two effects go in opposite directions. Differentiating (14) with respect to σ_e gives

$$\frac{\partial \epsilon_{\theta,lp}^{NKMP}}{\partial \sigma_e} = \frac{-1}{(\varphi-1)^2}\frac{p}{p-b}\underbrace{\frac{\partial \varphi}{\partial \sigma_e}}_{<0} - \frac{\varphi}{\varphi-1}\frac{b}{(p-b)^2}\underbrace{\frac{\partial p}{\partial \sigma_e}}_{>0} < 0 \text{ or } > 0.$$

The first-term on the right hand side captures the fact that raising σ_e lowers φ, which lowers the volatility of measured productivity and raises $\epsilon_{\theta,lp}$. The second term on the right hand side is positive ($\frac{\partial p}{\partial \sigma_e}$>0) because increasing σ_e raises the benefit of using the hours per worker margin (increasing φ) without affecting its cost (1+σ_h). Hence, raising σ_e increases the value of a marginal worker. With p higher and further away from b, profits are larger and changes are smaller in percentage terms. As a result, the employment margin becomes less volatile compared to the hours per worker margin, and $\epsilon_{\theta,h}$ and $\epsilon_{\theta,lp}$ are lower.

While the effect of σ_e on $\epsilon_{\theta,lp}^{NKMP}$ is a priori indeterminate, in practice, raising σ_e increases $\epsilon_{\theta,lp}^{NKMP}$. For instance, using the baseline calibration and increasing σ_e from 0.5 to 1 (so that φ decreases from 1.5 to 1.3) raises $\epsilon_{\theta,lp}^{NKMP}$ from 5.7 to 8.5, which is 4.5 times larger than $\epsilon_{\theta,lp}^{MP}$.

5.3 Taking stock

The previous analysis shows that, across a wide range of plausible parameter values, the NKMP model can generate large values of $\epsilon_{\theta,lp}$ when technology is held constant. Our most conservative calibration implies a three-fold increase in $\epsilon_{\theta,lp}$ compared to a standard MP model.

However, our analytical results for $\epsilon_{\theta,lp}^{NKMP}$ held technology constant and left aside technology shocks.[25] To evaluate the properties of the model conditional on both aggregate demand and technology shocks, we now resort to numerical simulations.

6 Simulation

I first consider the unconditional performance of the NKMP model under the baseline (conservative) calibration with ρ_b=0.4, γ=0.5, and $\varphi = 1.5$. Then, using the previous discussion as a guide, I present one possible calibration that uses plausible parameter values that can match

[25] Technology was held constant in order to highlight the role played by the endogeneity of productivity from simple closed-form expressions. In order to include changes in technology in the discussion, one would need to map technology shocks ε_t^a (as modeled in search and matching models and as identified in a long-run VARs) to the cyclical component of measured labor productivity lp_t recovered from a low-frequency trend in output per hour (Shimer, 2005, Hagedorn and Manovskii, 2008). However, this complicates the analysis significantly because the cyclical component of measured labor productivity lp_t need not only identify exogenous technology shocks but also some of the endogenous movements in productivity.

the impulse response functions presented in Figure 1 as well as the unconditional elasticity $\epsilon^{US}_{\theta,lp}$.

Before simulating the model, I need to specify the data generating process for technology and non-technology shocks. As in Sveen and Weinke (2008), I set the quarterly standard-deviation of monetary shocks σ^m to 0.002. I set the standard deviation of technology shocks σ^a to 0.007 in line with the estimate from the structural VAR, and I set the standard deviation of preference shocks σ^Γ to match the volatility of output. The growth rate of technology (and money supply) is set to a=0.5% a quarter so that the economy is growing by 2% on average each year. A money growth autocorrelation parameter ρ_m of 0.5 is in line with the first autocorrelations of M1 and M2 growth in the US. There is little microevidence for τ^{cb}, the degree of monetary policy accommodation to technology shocks, and I preliminary use $\tau^{cb} = -0.5$ as in Barnichon (2010) but will later consider values of τ^{cb} ranging from 1 to -1. Finally, I set the autocorrelation for the preference shock process to 0.93 (Gali and Rabanal 2004, Sveen and Weinke, 2008).

6.1 A conservative calibration

In a first (conservative) calibration exercise, I set $\rho_b = 0.4$, as in Shimer (2005), $\gamma = \eta = 0.5$ to satisfy the Hosios (1990) condition, and $\varphi = 1.5$ from $\sigma_h = 2$ $\sigma_e = 0.5$. Following Shimer (2005), I detrend the model generated productivity series with an HP-filter ($\lambda = 1600$). Table 1 reports the summary statistics for the simulated labor market variables over 50 years of data, simulated 1000 times. Confirming the results from our steady-state analysis, simulated labor market tightness is 6.3 times more volatile than the cyclical component of labor productivity.[26]

6.2 A calibration in the spirit of Hagedorn and Manovskii (2008)

I now consider a calibration in the spirit of Hagedorn and Manovskii (2008) with a higher income replacement ratio and lower worker's bargaining weight. The ability of the NKMP model to account for the conditional and unconditional Shimer puzzle improves dramatically. Using $\rho_b = 0.7$ and $\gamma = 0.35$ and holding $\varphi = 1.5$ constant, Table 1 shows that we get $\frac{sd(\theta)}{sd(lp)} = 25.5$, in line with its empirical value.

Importantly, by using much less extreme parameter values than Hagedorn and Manovskii's (HM, 2008) ($\gamma = 0.35$, $\rho_b = 0.7$ versus $\gamma = 0.05$, $b = 0.95$ in HM), this calibration alleviates some of the criticisms addressed to HM's approach. In particular, it relieves the tensions highlighted by Mortensen and Nagypal (2007) and Costain and Reiter (2008). As Mortensen

[26] A value of 6.3 is slightly higher than 5.7, the elasticity reported in the previous section. The difference is due to the steady state assumption and the constant technology level assumed previously.

and Nagypal (2007) point out, HM's (2008) calibration with $b = 0.95$ implies that the surplus from working is minuscule (and according to Mortensen and Nagypal, too small to be plausible) with workers working for a $\frac{w-b}{b} = 1.7\%$ surplus. In contrast, in this calibration, workers working for a much more significant surplus $\frac{w-b}{b} = \frac{1-\rho_b}{\rho_b} = 42\%$. Costain and Reiter (2008) also argue that the high ρ_b used by HM (2008) implies that changes in unemployment insurance have too strong effects on unemployment. The present model alleviates this issue, because as we saw in Section 5, a significant fraction of the "amplification" of the NKMP model is not caused by a stronger amplification mechanism but is instead a by-product of the endogeneity of productivity. That latter effect is captured by the extra term $\frac{\varphi}{\varphi-1}$ in(14), which unlike $\epsilon_{\theta,h}^{NKMP}$ or $\epsilon_{\theta,lp}^{MP}$, does not influence the effect of policy changes on the labor market. The volatility of hours per worker is too high in the model. This is due to the simplifying assumption that employment is a state variable in the model so that, in response to a shock, all labor adjustment initially takes place along the intensive margin. Moreover, all the short-run adjustment occurs through the labor intensive margin. However, in practice, firms may also adjust their utilization level of capital. Since the present NKMP model does not model capital and variable capital utilization, it must generate extra volatility in hours per worker and effort in order to match the endogenous movements in productivity. Introducing capital and variable capacity utilization would help the model match the volatility of hours per worker by reducing the needs for firms to vary the hours margin, while still generating a large tightness-productivity elasticity.

I now turn to the conditional elasticities, and Figure 1 shows the impulse response functions generated by the NKMP model after technology shocks and monetary policy shocks. Unlike the standard MP model which always generates a positive value for $\epsilon_{\theta,lp}$, the NKMP model generates conditional elasticities with the correct signs: positive following non-technology shocks but negative following technology shocks.

Impulse responses to technology shocks: Conditional on technology shocks, $\epsilon_{\theta,lp}^{NKMP}$ is negative. Following a positive technology shock, aggregate demand does not increase as much as productivity because prices are sticky and because the central bank does not accommodate the shock. As a result, aggregate demand is sticky in the short run. Being more productive, firms need less labor, post fewer vacancies and labor market tightness declines.

The less accommodating the central bank, the less aggregate demand will adjust in response to the higher productivity level and the less labor firms will need. Figure 1 shows the effect of varying τ^{cb}, the degree of monetary policy accommodation to technology shocks, on the impulse responses following technology shocks. We can see that the magnitude of the labor market-tightness-productivity elasticity depends for a large part on the reaction of the central bank, rather than on the amplification properties of the model as in the standard MP model. A central bank that fully accommodates technology shocks (τ^{cb}=1) would see labor market tightness increasing following technology shocks. In contrast, a central bank pursuing a contractionary monetary policy following technology shocks (τ^{cb}=−1) would see labor market tightness decreasing following technology shocks. Thus, the Shimer puzzle conditional on technology shock is of a different kind than what the literature has focused on, as whether the model can or cannot match the data depends for a large part on the central bank reaction function. The NKMP model can reasonably match $\epsilon_{\theta,lp}^{US}$ with τ^{cb}=−0.5.[27]

Impulse responses to aggregate demand shocks: Conditional on aggregate demand shocks, $\epsilon_{\theta,lp}^{NKMP}$ is positive. A positive aggregate demand shock raises labor market tightness as firms need more labor and post vacancies. It also raises productivity, because firms must increase hours per worker and effort to satisfy demand in the short-run as employment is subject to hiring frictions.

With this calibration, the NKMP model can match the empirical impulse responses of labor productivity and labor market tightness. Nonetheless, labor market tightness displays too little persistence, a standard problem with search models of unemployment already pointed out by Fujita and Ramey (2004). This is due to the excessively rapid response of vacancies; and incorporating sunk costs for vacancy creation as in Fujita and Ramey (2004) would presumably correct this shortcoming.

7 Conclusion

While the standard MP model implies that labor market tightness and productivity should display volatilities of similar magnitude, in the data, the former is about 25 times more volatile than the latter. The search literature has interpreted this discrepancy as a sign that the standard MP model generates a too weak amplification mechanism.

This paper proposes a new reason for this so-called "Shimer puzzle": the endogeneity of measured labor productivity. Measured productivity can move endogenously when firms respond to changes in aggregate demand and adjust their level of capacity utilization of inputs (capital or labor). If the endogenous response of productivity is small, the cyclical component of measured productivity may be less volatile than labor market variables such as labor market tightness.

[27] A negative value for τ^{cb} may be surprising but, as Gali and Rabanal (2004) argue, potential output is difficult to observe for the policy maker, and some positive technology shocks may have been misinterpreted, leading the central bank to pursue a contractionary policy. Indeed, Orphanides (2002) claims that the Great Inflation of the 1970's "could be attributed to [...] an adverse shift in the natural rate of unemployment that could not have been expected to be correctly assessed for some time."

To capture this idea theoretically, I present a New-Keynesian model with search unemployment and variable labor effort. I *analytically* study the key parameters behind the labor market tightness-productivity elasticity and show that the relative volatilities of labor market tightness and *measured* productivity depend on two sets of key factors: (i) the magnitude of the short run increasing returns to hours generated by variable labor effort, which is a function of the utility cost of longer hours and higher effort, and (ii) the magnitude of the trade-off between the intensive and the extensive labor margin, which depends on the worker's value of non-market activity and on the worker's bargaining power.

To illustrate quantitatively the possible contribution of the endogeneity of productivity to the Shimer puzzle, I calibrate and simulate the model. With a conservative calibration (in particular, a low income-replacement ratio), the labor market tightness-productivity elasticity is three times larger than that implied by a standard MP model. Using a calibration with a higher income replacement ratio and a lower workers' bargaining weight in the spirit of Hagedorn and Manovskii (2008), the model can match the empirical elasticity. However, the calibration uses less extreme values than Hagedorn and Manovskii (2008), and thus relieves some of the tensions associated with Hagedorn and Manovskii's (2008) calibration.

Appendix

(Non-stationary) Equilibrium

In this non-stationary model economy, I rescale the non-stationary variables with the technology index A_t. Denoting rescaled variables with lower-case letters, the frictionless economy is described by the following system with 5 equations and 5 unknowns θ, y, h, e and n:

$$\begin{aligned}
y &= \left(\frac{Y}{A}\right) = y_0 n h^{\varphi} \\
e &= e_0 h^{\frac{\sigma_h}{1+\sigma_e}} \\
\beta\chi &= \frac{c}{q(\theta)}\left(1-\beta(1-\lambda)\right) \\
\chi &= -\gamma c\theta - (1-\gamma)b + (1-\gamma)\left(\frac{1+\sigma_h}{\varphi}-1\right)\varkappa h^{1+\sigma_h} \\
1 &= \mu n \frac{1+\sigma_h}{\varphi}(1-\gamma)\varkappa h^{1+\sigma_h-\varphi} \\
n &= \frac{\theta q(\theta)}{\lambda+\theta q(\theta)}
\end{aligned}$$

where y_0, e_0 and $\varkappa$ are positive constants defined previously.

Wage bargaining

Firms and workers bargain individually about income w_{it} and split the surplus in shares determined by an exogenous bargaining weight γ (as in e.g. Krause and Lubik, 2007 and Trigari, 2009).

On the firm's side, the surplus $J_i(w_{it})$ obtained from a marginal worker equals his marginal contribution to profits so

$$\begin{aligned}
J_i(w_{it}) &= \frac{\partial\left(\frac{P_{it}}{P_t}y_{it} - w_{it}n_{it}\right)}{\partial n_{it}} + E_t\beta_{t+1}(1-\lambda)J_i(w_{it+1}) \\
&= \frac{h_{it}}{\varphi}\frac{\partial w_{it}}{\partial h_{it}} - w_{it} + E_t\beta_{t+1}(1-\lambda)J_i(w_{it+1}) \qquad (18)
\end{aligned}$$

with w_{it} the wage bill per worker, υ_t the marginal utility of consumption and $\beta_{t+1} = \beta\frac{\upsilon_t}{\upsilon_t}$ the stochastic discount factor. In a context of monopolistic competition and infrequent price adjustment, once the firm has set a price, its revenue is independent of n_{it}. Therefore, the contribution of the marginal worker to flow profits is given, not by the marginal revenue product of the worker ($\frac{\partial\left(\frac{P_{it}}{P_t}y_{it}\right)}{\partial n_{it}} = \frac{\partial\left(\left(\frac{P_{it}}{P_t}\right)^{1-\varepsilon}Y_t\right)}{\partial n_{it}} = 0$), but by the marginal reduction in the wage bill

$(-\frac{\partial(w_{it}n_{it})}{\partial n_{it}} = -n_{it}\frac{\partial(w_{it})}{\partial h_{it}}\frac{\partial h_{it}}{\partial n_{it}} - w_{it} = \frac{h_{it}}{\varphi}\frac{\partial w_{it}}{\partial h_{it}} - w_{it})$. If the worker walked away from the job, given the impossibility of hiring a replacement immediately, the firm would need to increase the number of hours of (and therefore the wage payments to) all other workers in order to meet its demand.

A vacancy is filled with probability $q(\theta_t)$ and remains open otherwise. With c_t the cost of keeping a vacancy open at date t, the value $V_i(w_{it})$ of posting a vacancy in terms of current consumption is given by

$$V_i(w_{it}) = -c_t + E_t\beta_{t+1}[q(\theta_t)J_i(w_{it+1}) + (1 - q(\theta_t))V_i(w_{it+1})] \tag{19}$$

Note that the firm will post vacancies as long as the value of a vacancy is greater than zero. In equilibrium, $V_i(w_{it}) = 0$ so that

$$\frac{c_t}{q(\theta_t)} = E_t\beta_{t+1}[J_i(w_{it+1})]. \tag{20}$$

Turning to the worker's problem, denote $W_i(w_{it})$ and U_t the value of being respectively employed and unemployed in units of consumption goods. The worker's asset value of being matched to firm i is

$$W_i(w_{it}) = w_{it} - \frac{1}{\upsilon_t}\left(\frac{\zeta_h}{1+\sigma_h}h_{it}^{1+\sigma_h} + h_t\frac{\zeta_e}{1+\sigma_e}e_{it}^{1+\sigma_e}\right) + E_t\beta_{t+1}[(1-\lambda)W_i(w_{it+1}) + \lambda U_t] \tag{21}$$

and the value of being unemployed U_t is

$$U_t = b_t + E_t\beta_{t+1}\left[\int_0^1 \theta_t q(\theta_t)\frac{v_{jt}}{v_t}W_j(w_{it+1})dj + (1 - \theta_t q(\theta_t))U_{t+1}\right] \tag{22}$$

with b_t the value of home production or unemployment benefits. A worker receives earnings w_{it} minus the disutility of labor, and has a probability λ of becoming unemployed next period. When unemployed, a worker receives b_t, has a probability $\theta_t q(\theta_t)\frac{v_{jt}}{v_t}$ to find a job next period with firm j and a probability $1 - \theta_t q(\theta_t)$ to remain unemployed.

The negotiated income w_{it} satisfies $w_{it} = \underset{w_{it}}{\text{argmax}}\ (W_i(w_{it}) - U_t)^{\gamma}(J_i(w_{it}))^{1-\gamma}$ so that the surplus-sharing rule implies

$$W_i(w_{it}) - U_t = \frac{\gamma}{1-\gamma}J_i(w_{it}). \tag{23}$$

Denoting the worker's surplus $S_{it} = W_i(w_{it}) - U_t$, I can write

$$\begin{aligned}
S_{it} &= w_{it} - \frac{1}{\upsilon_t}\left(\frac{\zeta_h}{1+\sigma_h}h_{it}^{1+\sigma_h} + h_t\frac{\zeta_e}{1+\sigma_e}e_{it}^{1+\sigma_e}\right) - b_t \\
&\quad + E_t\beta_{t+1}[-\int_0^1 \theta_t q(\theta_t)\frac{v_{jt}}{v_t}S_{jt+1}dj + (1-\lambda)E_t\beta_{t+1}S_{it+1} \\
&= w_{it} - \frac{g(h_{it}, e_{it})}{\upsilon_t} - b_t \\
&\quad + E_t\beta_{t+1}\frac{\gamma}{1-\gamma}[\int_0^1 \theta_t q(\theta_t)\frac{v_{jt}}{v_t}J_j(w_{it+1})dj + (1-\lambda)J_i(w_{it+1})] \text{ using (23)} \\
&= w_{it} - \frac{g(h_{it}, e_{it})}{\upsilon_t} - b_t \\
&\quad + \frac{\gamma}{1-\gamma}\frac{c_t}{q(\theta_t)}(1-\lambda-\theta_t q(\theta_t)) \text{ with (20)}
\end{aligned} \tag{24}$$

Combining (24) with (23), (18) and (20), earnings per worker satisfies

$$w_{it} - b_t - \frac{g(h_{it}, e_{it})}{\upsilon_t} + \frac{\gamma}{1-\gamma}\frac{c_t}{q(\theta_t)}(1-\lambda-\theta_t q(\theta_t)) = \frac{\gamma}{1-\gamma}\left(-w_{it} + \frac{h_{it}}{\varphi}\frac{\partial w_{it}}{\partial h_{it}} + (1-\lambda)\frac{c_t}{q(\theta_t)}\right)$$

or after rearranging,

$$w_{it} = \gamma\left(\frac{h_{it}}{\varphi}\frac{\partial w_{it}}{\partial h_{it}} + c_t\theta_t\right) + (1-\gamma)\left(b_t + \frac{g(h_{it}, e_{it})}{\upsilon_t}\right). \tag{25}$$

While the income equation (25) is a weighted average of both parties surpluses and is similar to other bargained wages derived in e.g. Krause and Lubik (2007) or Trigari (2009), the firm's surplus is *not* given by the marginal product of labor. Indeed, once the firm has chosen its price, it is demand constrained and a marginal worker will not increase the firm's revenue. Instead, the first term of (25) is given by $-\frac{\partial w_{it}}{\partial n_{it}} = \frac{h_{it}}{\varphi}\frac{\partial w_{it}}{\partial h_{it}}$, the change in the wage bill caused by substituting the intensive margin (hours and effort) with the extensive one (employment).

A solution to (25) is given by

$$w_{it} = \gamma c_t\theta_t + (1-\gamma)b_t + (1-\gamma)\varkappa\frac{h_{it}^{1+\sigma_h}}{\upsilon_t} \tag{26}$$

with $\varkappa = \frac{\zeta_h\frac{1+\sigma_h+\sigma_e}{(1+\sigma_h)\sigma_e}}{1-\frac{\gamma}{\varphi}(1+\sigma_h)}$.

Closed-form expressions for the labor market tightness-productivity elasticity

In the standard MP model without aggregate uncertainty, the value for the vacancy posting cost comes out of the steady-state conditions once a value has been chosen for b, and we have

$$c = \frac{\beta(1-\gamma)\,(A-b)}{\frac{1}{q}\,(1-\beta(1-\lambda))+\gamma\beta\theta} \tag{27}$$

so that the elasticity of labor market tightness with respect to productivity in the standard MP model is given by

$$d\ln\theta_t = \frac{A}{A-b}\underbrace{\frac{\frac{\sigma}{q}\,(1-\beta(1-\lambda))+\gamma\beta\theta}{\frac{1}{q}\,(1-\beta(1-\lambda))+\gamma\beta\theta}}_{\kappa=}d\ln A_t \tag{28}$$

which is the central point of Hagedorn and Manovskii's (2008) discussion.

In the NKMP model without aggregate uncertainty, rewriting the vacancy posting condition and the wage equation in the New-Keynesian model in steady state delivers

$$c = \frac{\beta(1-\gamma)\left(\left(\frac{1+\sigma_h}{\varphi}-1\right)\varkappa\frac{h^{1+\sigma_h}}{\lambda}-b\right)}{\frac{1}{q}\,(1-\beta(1-\lambda))+\gamma\beta\theta}$$

which, using the price setting condition, can be written as

$$c = \frac{\beta(1-\frac{\varphi}{1+\sigma_h})\frac{1}{n\mu}-\beta(1-\gamma)b}{\frac{1}{q}\,(1-\beta(1-\lambda))+\gamma\beta\theta}. \tag{29}$$

Using (29), combining the wage equation and the vacancy posting condition in the model without aggregate shock and implicit differentiation gives

$$\begin{aligned} d\ln\theta_t &= \varphi\frac{\left(\frac{1+\sigma_h}{\varphi}-1\right)\varkappa h^{1+\sigma_h}y}{\left(\frac{1+\sigma_h}{\varphi}-1\right)\varkappa h^{1+\sigma_h}y-b}\kappa d\ln h_t & (30)\\ &= \varphi\frac{p}{p-b}\kappa d\ln h_t & (31)\end{aligned}$$

where $p \equiv \left(\frac{1+\sigma_h}{\varphi}-1\right)\varkappa h^{1+\sigma_h}y$. With $y_t = y_0 A_t n_t h_t^{\varphi}$, labor productivity is given by $lp_t = \frac{y_t}{n_t h_t}$, so that, if technology A_t is held constant, $d\ln lp_t = (\varphi-1)\,d\ln h_t$ and the elasticity of labor market tightness with respect to *measured* productivity is given by

$$\epsilon_{\theta,lp}^{NKMP} = \frac{\varphi}{\varphi - 1}\frac{p}{p-b}\kappa. \tag{32}$$

Using the price setting condition, it is also possible to rewrite (32) as

$$\epsilon_{\theta,lp}^{NKMP} = \frac{\varphi}{\varphi - 1}\frac{(1-\frac{\varphi}{1+\sigma_h})\frac{1}{1-\gamma}\frac{1}{\mu n}}{(1-\frac{\varphi}{1+\sigma_h})\frac{1}{1-\gamma}\frac{1}{\mu n} - b}\kappa \tag{33}$$

so that $\epsilon_{\theta,lp}^{NKMP}$ is a function of n and exogenous parameters.

The effect of the income replacement ratio Using that $\rho_b = b/w$ and combining (33) with (26), I can write

$$\epsilon_{\theta,lp}^{NKMP} = \frac{\varphi}{\varphi - 1}\frac{p}{\frac{1}{n\mu}\left(-\rho_b + \frac{1}{1-\gamma}\left(1 - \frac{\varphi}{1+\sigma_h}\right)\right)\left(1 - \rho_b(1-\gamma)(1 - \frac{\gamma\theta\beta}{\frac{1}{q}(1-\beta(1-\lambda))+\beta\gamma\theta})\right)^{-1}}\kappa. \tag{34}$$

Using the approximation $\frac{1}{q}(1-\beta(1-\lambda)) \ll \gamma\theta\beta$ for γ not too small, I get $\frac{\gamma\theta\beta}{\frac{1}{q}(1-\beta(1-\lambda))+\beta\gamma\theta}) \simeq 1$ and

$$\begin{aligned}
\epsilon_{\theta,lp}^{NKMP} &\simeq \frac{\varphi}{\varphi - 1}\frac{p}{\frac{1}{n\mu}\left(-\rho_b + \frac{1}{1-\gamma}\left(1 - \frac{\varphi}{1+\sigma_h}\right)\right)}\kappa \\
&\simeq \frac{\varphi}{\varphi - 1}\frac{p}{p - \frac{\rho_b}{n\mu}}.
\end{aligned}$$

Differentiating (34) with respect to ρ_b gives

$$\begin{aligned}
\frac{\partial\epsilon_{\theta,lp}^{NKMP}}{\partial\rho_b} &= \frac{\varphi}{\varphi - 1}\frac{1}{n\mu(1-\gamma)^2}\frac{1}{\left[\frac{1}{1-\gamma} + \rho_b(-1 + \frac{\gamma\beta\theta}{\frac{1}{q}(1-\beta(1-\lambda))+\beta\gamma\theta})\right]^2} \\
&\quad \cdot(1 - \underbrace{\left(1 - \frac{\varphi}{1+\sigma_h}\right)}_{<1}\underbrace{\left(1 - \frac{\gamma\beta\theta}{\frac{1}{q}(1-\beta(1-\lambda)) + \beta\gamma\theta}\right)}_{<1} \\
&> 0
\end{aligned}$$

so that $\frac{\partial\epsilon_{\theta,lp}^{NKMP}}{\partial\rho_b} > 0$ and the closer is ρ_b to $(1-\frac{\varphi}{1+\sigma_h})\frac{1}{1-\gamma}$, the larger $\epsilon_{\theta,lp}^{NKMP}$. As in Costain and Reiter (2008), I isolate the effect of parameter changes on the elasticity without including their effect on the steady-state level of employment (the only endogenous variable entering

(33)). This amounts to assuming that the matching efficiency constant m_0 is adjusted across calibrations to hold the job finding rate $\theta q(\theta)$ constant so that $n = \frac{\theta q(\theta)}{\lambda + \theta q(\theta)}$ is constant. As a result, $\frac{\partial n}{\partial \rho_b} = 0$.

The effect of σ_h Differentiating (33) with respect to σ_h gives

$$\begin{aligned}
\frac{\partial \epsilon_{\theta,lp}^{NKMP}}{\partial \sigma_h} &= \frac{-1}{(\varphi-1)^2}\frac{p}{p-b}\frac{\partial \varphi}{\partial \sigma_h} + \frac{1}{\varphi-1}\frac{\partial p/(p-b)}{\partial \sigma_h} \\
&= \frac{-1}{(\varphi-1)^2}\frac{\alpha}{1+\sigma_e}\frac{p}{p-b} - \frac{1}{\varphi-1}\frac{b}{(p-b)^2}\frac{1}{1-\gamma}\frac{\alpha\sigma_e}{(1+\sigma_h)^2(1+\sigma_e)} < 0
\end{aligned}$$

where the steady-state level of employment n is held constant as in the previous subsection.

The effect of σ_e Differentiating (33) with respect to σ_e gives

$$\begin{aligned}
\frac{\partial \epsilon_{\theta,lp}^{NKMP}}{\partial \sigma_e} &= \frac{1}{(\varphi-1)^2}\frac{p}{p-b}\frac{\partial \varphi}{\partial \sigma_e} + \frac{1}{\varphi-1}\frac{\partial p/(p-b)}{\partial \sigma_h} \\
&= \frac{1}{(\varphi-1)^2}\frac{\alpha\sigma_h}{(1+\sigma_e)^2}\frac{p}{p-b} - \frac{1}{\varphi-1}\frac{1}{(p-b)^2}b\frac{1}{1-\gamma}\frac{\alpha\sigma_h}{(1+\sigma_h)(1+\sigma_e)^2} > 0 \text{ or } < 0.
\end{aligned}$$

where the steady-state level of employment n is held constant as in the previous subsection.

References

[1] Andolfatto, D. “Business Cycles and Labor-Market Search,” *American Economic Review*, 86(1), 1996.

[2] Balleer, A. “New Evidence, Old Puzzles: Technology Shocks and Labor Market Fluctuations,” Working Paper, 2009.

[3] Barnichon, R. “Building a composite Help-Wanted index,” *Economics Letters*, 2010.

[4] Barnichon, R. “Productivity and Unemployment over the Business Cycle,” *Journal of Monetary Economics*, 2010.

[5] Basu, S. and M. Kimball. “Cyclical Productivity with Unobserved Input Variation,” *NBER Working Papers* 5915, 1997.

[6] Bils, M. and J. Cho. “Cyclical Factor Utilization,” *Journal of Monetary Economics*, 33, 319-354, 1994.

[7] Bils, M. and P. Klenow. “Some Evidence on the Importance of Sticky Prices,” *Journal of Political Economy*, 112(5), pp. 947–985, 2004.

[8] Burnside, C., Eichenbaum, M. “Factor-Hoarding and the Propagation of Business-Cycle Shocks,” *American Economic Review,* 86(5), pp. 1154-1174, 1996.

[9] Burnside, C., Eichenbaum, M. and S. Rebelo. “Labor Hoarding and the Business Cycle,” *Journal of Political Economy*, 01(2), pp. 245-73, 1993.

[10] Canova, F., D. Lopez-Salido and C. Michelacci, “The Ins and Outs of Unemployment: A Conditional Analysis,” Working Paper, 2009.

[11] Costain, J. and M. Reiter. “Business cycles, unemployment insurance, and the calibration of matching models,” *Journal of Economic Dynamics and Control*, vol. 32(4), pp. 1120-1155, 2008.

[12] den Haan, W., G. Ramey and J. Watson. “Job Destruction and Propagation of Shocks,” *American Economic Review*, 90(03), 2000.

[13] Elsby, M. and R. Micahels "Marginal Jobs, Heterogeneous Firms and Unemployment Flows," *American Economic Journal: Macroeconomics*, January 2013.

[14] Eyigungor, B. “Specific Capital and Vintage Effects on the Dynamics of Unemployment and Vacancies,” *American Economic Review*, 100(3): 1214–37, 2010.

[15] Fujita, S. and G. Ramey. "Job Matching and Propagation," *Journal of Economic Dynamics and Control*, Vol. 31, pp. 3671-3698, 2007.

[16] Furlanetto F. and T. Sveen. "The Unemployment Correlation Puzzle," Working Paper, 2009.

[17] Galí, J. "Technology, Employment and The Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review*, 89(1), 1999.

[18] Galí, J and P. Rabanal. "Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar U.S. Data?," *NBER Macroeconomics Annual*, 2004.

[19] Hagedorn, M. and I. Manovskii. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited," *American Economic Review*, 98(4), 2008.

[20] Hall, R. "Employment Fluctuations with Equilibrium Wage Stickiness," *American Economic Review*, 95(1), pp. 50-65, 2005.

[21] Hall, R. "Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years," *NBER Macroeconomics Annual*, pp. 101-137, 2005.

[22] Hall, R. and P. Milgrom. "The Limited Influence of Unemployment on the Wage Bargain," *American Economic Review*, 98(4), pp. 1653-1674, 2008.

[23] Krause, M and T. Lubik. "The (Ir)relevance of Real Wage Rigidity in the New Keynesian Model with Search Frictions", *Journal of Monetary Economics*, 54(3), 2007.

[24] Mertz, M. "Search in the Labor Market and the Real Business Cycle," *Journal of Monetary Economics*, 49, 2002.

[25] Michaillat, P. "Do matching frictions explain unemployment? Not in bad times," *American Economic review*, 102 (4), 1721–1750, 2012

[26] Mortensen, D. and E. Nagypal. "More on Unemployment and Vacancy Fluctuations," *Review of Economic Dynamics*, 10(3), 2007.

[27] Mortensen, D. and E. Nagypal. "Labor-market Volatility in Matching Models with Endogenous Separations," *Scandinavian Journal of Economics*,109(4), 2008.

[28] Mortensen, D and C. Pissarides. "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61, 397-415, 1994.

[29] Orphanides, A. "Monetary-Policy Rules and the Great Inflation," *American Economic Review*, vol. 92(2), pages 115-120, 2001.

[30] Pissarides, C. *Equilibrium Unemployment Theory*, 2nd Edition, MIT Press, 2001

[31] Pissarides, C. "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?," The Walras-Bowley lecture, North American Summer Meetings of the Econometric Society, Duke University, June 21-24, 2007

[32] Schor J., "Does Work Intensity Respond to Macroeconomic Variables? Evidence from British Manufacturing," Working Paper, 1987.

[33] Shimer, R. "The Consequences of Rigid Wages in Search Models," *Journal of the European Economic Association (Papers and Proceedings)*, 2: 469-479, 2004.

[34] Shimer, R. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95(1), pp. 25-49, 2005.

[35] Shimer, R. "Comment on: New Keynesian perspectives on labor market dynamics," *Journal of Monetary Economics*, vol. 55(5), pages 931-935, 2008.

[36] Stole, L and J. Zwiebel "Intra-firm Bargaining under Non-binding Contracts," Review of Economic Studies, Wiley Blackwell, vol. 63(3), pages 375-410, 1996.

[37] Sveen, T. and L. Weinke. "New Keynesian perspectives on labor market dynamics," *Journal of Monetary Economics*, vol. 55(5), pages 921-930, 2008.

[38] Trigari, A. "Equilibrium Unemployment, Job Flows and Inflation Dynamics," *Journal of Money, Credit and Banking*, 2009.

[39] Walsh, C. "Labor Market Search and Monetary Shocks," in *Elements of Dynamic Macroeconomic Analysis*, S. Altug, J. Chadha, and C. Nolan, Cambridge University Press, 2004, 451-486.

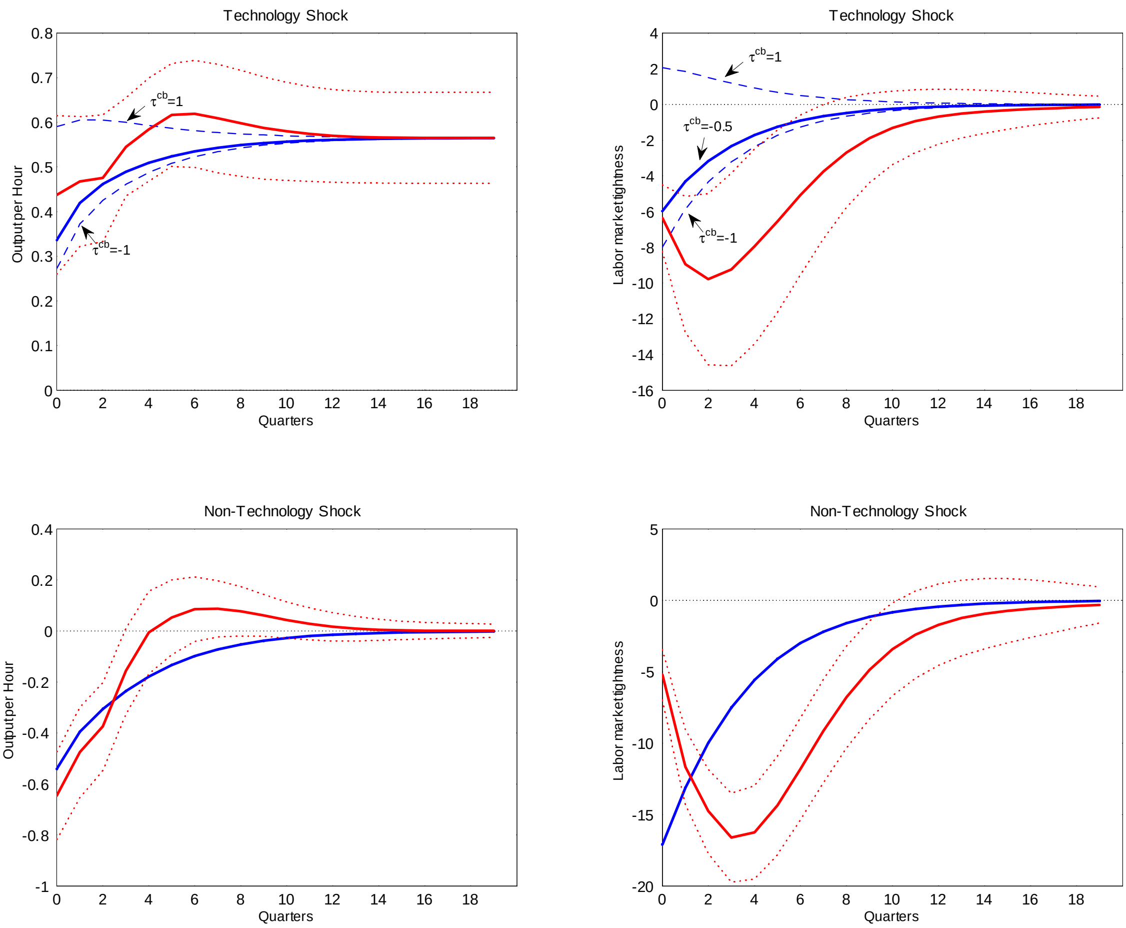

Figure 1: Empirical (red plain line) and model (blue dotted line) impulse response functions to a technology and a non-technology shock. Red dotted lines around the empirical responses represent the 95% confidence interval. Blue dashed lines around the baseline ($\tau^{cb} = -0.5$) model impulse responses represent the impulse responses when the monetary policy reaction function ranges from $\tau^{cb} = 1$ (fully accommodative) to $\tau^{cb} = -1$ (contractionary).

Table 1: Standard-deviations of US and model data, 1951-2007

	u	θ	h	lp
US data	0.071	0.260	0.011	0.010
Calibration 1 ρ_b=0.4, γ=0.5, φ=1.5	0.018	0.070	0.025	0.011
Calibration 2 ρ_b=0.7, γ=0.35, φ=1.5	0.067	0.255	0.023	0.010

Notes: Unemployment *u* is constructed by the BLS from the Current Population Survey (CPS). Labor market tightness is the vacancy-unemployment ratio with vacancy posting taken from the composite Help-Wanted index presented in Barnichon (2010). Hours per worker is derived from subtracting (log) employment from (log) total hours in the non-farm business sector from the CES. Average labor productivity *lp* is seasonally adjusted real average output per hour in the non-farm business sector. All variables except unemployment are reported in logs as deviations from an HP trend with smoothing parameter λ=1600.