

NBER WORKING PAPER SERIES

STICKY INFORMATION IN GENERAL EQUILIBRIUM

N. Gregory Mankiw
Ricardo Reis

Working Paper 12605
<http://www.nber.org/papers/w12605>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2006

This is an extended version of our paper with the same title published in the *Journal of the European Economic Association*, April-May 2007. It includes a lengthy appendix laying out the model, solving it, proving the propositions, and explaining the algorithms. All of the programs used are available at our websites. We are grateful to Tiago Berriel for excellent research assistance, and to Ruchir Agarwal and Mark Watson for useful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2006 by N. Gregory Mankiw and Ricardo Reis. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Sticky Information in General Equilibrium
N. Gregory Mankiw and Ricardo Reis
NBER Working Paper No. 12605
October 2006
JEL No. E10,E30

ABSTRACT

This paper develops and analyzes a general-equilibrium model with sticky information. The only rigidity in goods, labor, and financial markets is that agents are inattentive, sporadically updating their information sets, when setting prices, wages, and consumption. After presenting the ingredients of such a model, the paper develops an algorithm to solve this class of models and uses it to study the model's dynamic properties. It then estimates the parameters of the model using U.S. data on five key macroeconomic time series. It finds that information stickiness is present in all markets, and is especially pronounced for consumers and workers. Variance decompositions show that monetary policy and aggregate demand shocks account for most of the variance of inflation, output, and hours.

N. Gregory Mankiw
Department of Economics
Littauer 223
Harvard University
Cambridge, MA 02138
and NBER
ngmankiw@fas.harvard.edu

Ricardo Reis
Princeton University
Department of Economics
324 Bendheim Hall
Princeton, NJ 08544-1021
and NBER
rreis@princeton.edu

1. Introduction

Estimation and simulation of medium-sized macroeconomic models has increasingly attracted the attention of economists who study monetary policy and the business cycle.¹ This paper contributes to that effort by focusing on a model in which sticky information is the key imperfection that causes output to deviate from its long-run classical benchmark. In this otherwise standard dynamic stochastic general equilibrium model, information is updated sporadically by firms setting prices, workers setting wages, and consumers setting the level of spending.

Solution and estimation of a general equilibrium model with sticky information raises several thorny technical issues. We begin this paper outlining those issues and proposing solutions. Our first contribution is methodological. It consists of two propositions that provide an algorithm that efficiently solves medium-sized sticky-information models, derives their impulse response functions, and calculates their likelihood in a few seconds.

We then proceed to estimate the model using five key time-series: inflation, output, hours worked, wages, and an interest rate. The second contribution of this paper is to propose, implement, and compare two estimation strategies for the model: maximum likelihood and a Bayesian approach. The two strategies yield similar results.

We thus obtain estimates of how much information stickiness is needed to explain business cycle dynamics. We find that about a fifth of workers and consumers update their information sets every quarter, so the mean information lag for both household members is approximately five quarters. By contrast, firms are estimated to be much better informed when setting prices: about two-thirds update their information set every quarter.

The model also produces an estimated variance decomposition, which shows how much of the variation in each variable is attributable to each of the five shocks in the model. For inflation, over 80 percent of the variance is attributable to the monetary policy shock. For output growth and hours worked, the monetary policy shock is important, but so is the shock to aggregate demand. The other three shocks—to productivity, the goods markup, and the labor markup—are estimated to explain only a small fraction of the variance of inflation, output growth, and hours worked.

2. The model of the economy

¹See, for instance, Smets and Wouters (2003) and Levin, Onatksi, Williams and Williams (2006)

We study a general-equilibrium model with monopolistic competition and no capital accumulation, familiar in the literature on monetary policy. We assume a continuum of households with preferences that are additively separable and iso-elastic in consumption and leisure. Households live forever and wish to maximize expected discounted utility while being able to save and borrow by trading bonds between themselves. We think of households as having two members: a worker and a consumer. The workers sell labor to firms in a set of segmented markets for different labor varieties, where each worker is the sole provider of each variety. The consumers buy a continuum of varieties of goods from firms, which they value according to a Dixit-Stiglitz aggregator. There is a continuum of firms, each selling one variety of goods under monopolistic competition. Each firm operates a decreasing returns to scale technology in total labor input, which is a Dixit-Stiglitz aggregate of the different varieties of labor. Finally, monetary policy follows a Taylor rule.

Less common is our assumption on information. There are three agents making decisions in this economy: consumers, workers, and firms. We assume that each period, a fraction δ of consumers, a fraction ω of workers, and a fraction λ of firms, randomly drawn from their respective populations, obtain new information and calculate their optimal actions. This assumption of sticky information can be justified by costs of acquiring, absorbing and processing information (Reis, 2004, 2006) or by appealing to epidemiology (Carroll, 2002).²

We leave the detailed presentation of the model, the definition of an equilibrium and its log-linearization to the appendix. Here, we discuss the 5 key reduced-form relations. The first relation is the Phillips curve or aggregate supply curve:

$$p_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-j} \left[p_t + \frac{\beta(w_t - p_t) + (1-\beta)y_t - a_t}{\beta + \nu(1-\beta)} - \frac{\beta\nu_t}{(\nu-1)[\beta + \nu(1-\beta)]} \right]. \quad (1)$$

The price level (p_t) depends on past expectations of: its current value, real marginal costs, and desired markups.³ Marginal costs are higher: the higher are the real wages paid to workers ($w_t - p_t$), the more is produced (y_t) because of decreasing returns to scale ($\beta < 1$), and the lower is aggregate productivity (a_t). The desired markup falls with the elasticity of substitution across goods varieties (ν_t), which we allow to vary randomly over time.

²The optimal behavior of these inattentive agents and their interaction in markets raise some interesting challenges. We discuss these in Mankiw and Reis (2006).

³All variables with a t subscript refer to log-linearized values around their non-stochastic steady state. Without any subscript are fixed parameters and steady state values.

Unexpected shocks to any of these three variables only raise prices by λ since only this share of price-setters is aware of the news.

The second relation is the IS curve:

$$y_t = \delta \sum_{j=0}^{\infty} (1 - \delta)^j E_{t-j} (y_{\infty}^n - \theta R_t) + g_t, \quad (2)$$

where the long-run equilibrium output is $y_{\infty}^n = \lim_{i \rightarrow \infty} E_t (y_{t+i})$, and the long real interest rate is $R_t = E_t \left[\sum_{j=0}^{\infty} (i_{t+j} - \Delta p_{t+1+j}) \right]$. Higher expected future output raises wealth and increases spending, while higher expected interest rates encourage savings and lower spending. The impact of interest rates on spending depends on the intertemporal elasticity of substitution θ . We denote by g_t aggregate demand shocks, which in the model correspond to changes in government spending, but could also be modelled as changes in the desire for leisure. The higher is δ , the larger the share of informed consumers that respond to shocks immediately.

Next comes the wage curve:

$$w_t = \omega \sum_{j=0}^{\infty} (1 - \omega)^j E_{t-j} \left[p_t + \frac{\gamma(w_t - p_t)}{\gamma + \psi} + \frac{l_t}{\gamma + \psi} + \frac{\psi (y_{\infty}^n - \theta R_t)}{\theta(\gamma + \psi)} - \frac{\psi \gamma_t}{(\gamma + \psi)(\gamma - 1)} \right]. \quad (3)$$

The five determinants of nominal wages are split into the five terms on the right-hand side. First, nominal wages rise one-to-one with prices since workers care about real wages. Second, the higher are real wages elsewhere in the economy the higher is demand for a worker's variety of labor so the higher the wage she will demand. Third, the more labor is hired (l_t) the better it must be compensated since the marginal disutility of working rises. Fourth, higher wealth discourages work through an income effect, and higher interest rates promote it by giving a larger return on saved earnings today. The product of ψ , the Frisch elasticity of labor supply, and θ , the intertemporal elasticity of substitution, determine the strength of this intertemporal labor supply effect. Fifth and finally, if the elasticity of substitution across labor varieties (γ_t) rises, workers' desired markup falls so they lower their wage demands. If many workers are informed (ω is high), wages are instantly very responsive to changes in these determinants, whereas otherwise wages only respond gradually over time.

The fourth relation is a standard production function:

$$y_t = a_t + \beta l_t, \tag{4}$$

where β measures the extent of decreasing returns to scale from using more labor. The fifth and final relation is the Taylor rule:

$$i_t = \phi_y(y_t - y_t^n) + \phi_p \Delta p_t - \varepsilon_t, \tag{5}$$

where $y_t - y_t^n$ is the output gap, or the difference between actual output and its level if all agents were attentive, and ε_t are policy disturbances.

These 5 equations give the equilibrium values for output, wages, prices, labor, and nominal interest rates as a function of shocks to aggregate productivity growth, aggregate demand, goods markups, labor markups, and monetary policy. We assume that each of these shocks follows an autoregressive process of order 1 with coefficients $\rho_{\Delta a}$, ρ_g , ρ_ν , ρ_γ , and ρ_ε , and is subject to innovations $e_t^{\Delta a}$, e_t^g , e_t^ν , e_t^γ , and e_t^ε , that are independent and normally distributed with standard deviations $\sigma_{\Delta a}$, σ_g , σ_ν , σ_γ , and σ_ε .

3. Solving for the economy's dynamics

Our model fits into the general class of linear rational expectations models for which there are several ready-to-use solution algorithms. However, none of them is particularly useful to solve the sticky-information model. The model involves both an infinite number of past expectations of the present through sticky information, as well as present expectations of variables at an infinite number of future dates through intertemporal smoothing. This double infinity implies that the state-space of the model has an infinite dimension, which current algorithms cannot handle.⁴

We have developed a general algorithm that can solve this and much larger general-equilibrium models with sticky information in a few seconds. It is based on the following result, which comes from using a method of undetermined coefficients and exploiting the recursiveness of the model's dynamics:

Proposition 1. *Letting $s \in S = \{\Delta a, g, \nu, \gamma, \varepsilon\}$ denote the different shocks, then $p_t =$*

⁴Recently, Wang and Wei (2006) proposed an ingenious method to adapt existing algorithms to solve sticky-information models. We leave a systematic comparison of their method with the one in this paper for future research.

$\sum_{s \in S} \sum_{n=0}^{\infty} \hat{p}_n(s) e_{t-n}^s$ where $\hat{p}_n(s)$ is a scalar measuring the impact of shock s at lag n on the price level. The undetermined coefficients solve the second-order difference equation:

$$A_{n+1} \hat{p}_{n+1}(s) - B_n \hat{p}_n(s) + \phi_p \hat{p}_{n-1}(s) = C_n(s) \text{ for } n = 0, 1, 2, \dots \quad (6)$$

$$\text{with boundary conditions} \quad : \quad \hat{p}_{-1} = 0 \text{ and } \lim_{n \rightarrow \infty} (\hat{p}_n - \hat{p}_{n-1}) = 0.$$

The coefficients A_n and B_n do not depend on the shock, while $C_n(s)$ does; all depend on the parameters and are given in the appendix.

The appendix describes our algorithm to solve this difference equation and finds, in corollary 1, the solution for the other variables in the model as a function of the price dynamics.

Figure 1 shows the responses of inflation, the output gap, and labor to one-standard-deviation shocks to monetary policy, aggregate productivity growth, and aggregate demand.⁵ In response to a monetary expansion, output and labor increase as the economy enters a boom. Inflation rises gradually, following the hump-shaped pattern that has been found in empirical work. Noticeably, inflation is more persistent than output, another robust feature of the data that many monetary models have trouble reproducing. The model fits well the facts on how the economy responds to monetary policy shocks.

In response to a positive technological shock, inflation falls but converges rapidly to its previous level. Interestingly positive productivity shocks in this economy lead to recessions, just as in sticky price models (Gali, 1999). However, this is not a robust feature of the sticky-information model: for different parameter values, we can get a boom following a technological improvement. Finally, a positive innovation to aggregate demand raises inflation, output, and labor.

4. Estimating the model

We use U.S. quarterly data from 1954:3 to 2006:1 for the non-farm business sector. We measure wages using the total compensation per hour and labor input using total hours. We divide output and hours by the total civilian non-institutional population and deflate nominal variables using the implicit price deflator for the nonfarm business sector. Changes in the log of this deflator are our measure of inflation, and the effective federal funds rate measures the nominal interest rate.

⁵The parameters are set at the maximum-likelihood estimates in Table 1, described in the next section.

Using these data, we build series for de-measured inflation, output growth, nominal interest rates, real wage growth, and hours. These are our observables, collected in the vector $x_t = (\Delta p_t, \Delta y_t, l_t, i_t, \Delta(w_t - p_t))'$. The sticky-information general-equilibrium model implies that $x_t = \sum_{i=0}^{\infty} \Phi_i e_{t-i}$ where e_t is the vector of shocks $(e_t^\varepsilon, e_t^{\Delta a}, e_t^g, e_t^l, e_t^\gamma)'$ and the Φ_i are 5x5 matrices of coefficients, found in proposition 1 and corollary 1. The question we ask in this section is how to estimate the vector of parameters of the model using these data.

We estimate our model using both maximum likelihood and Bayesian methods.⁶ The key input into these methods is the likelihood function, which in standard dynamic models with a state-space solution can be evaluated using the Kalman filter. The solution of the model using proposition 1 does not have a convenient state-space representation, so we use instead the following result:

Proposition 2. *Given a sample of data of length T , let X be the $5T \times 1$ vector that vertically stacks the x_t , and let Ω be the $5T \times 5N$ matrix that vertically stacks $[\Phi_{j-1} \ \Phi_{j-2} \ \dots \ \Phi_0 \ \Phi_1 \ \dots \ \Phi_N]$ from $j = 1$ to $j = T$. Finally, let Σ be a diagonal matrix with $(\sigma_\varepsilon^2, \sigma_{\Delta a}^2, \sigma_g^2, \sigma_l^2, \sigma_\gamma^2)$ in the diagonal and I_N be an identity matrix of size N . The log-likelihood function is:*

$$\mathcal{L} = -2.5T \ln(2\pi) - 0.5 \ln |\Omega(I_N \otimes \Sigma)\Omega'| - 0.5 X' (\Omega(I_N \otimes \Sigma)\Omega')^{-1} X \quad (7)$$

The main difficulty with evaluating this expression is that inverting the large $5T \times 5T$ matrix $\Omega(I_N \otimes \Sigma)\Omega'$ is both slow and subject to potentially large numerical errors. The appendix shows how to evaluate (7) without inverting this matrix by instead solving a recursive linear system of equations. This provides an algorithm to evaluate the log-likelihood function quickly and reliably.

Turning to estimation, we set the value of 9 out of the 20 parameters. Namely, we set the intertemporal elasticity of substitution to 1 (the King-Plosser-Rebelo, 1988, utility function) to guarantee that hours are stationary, the Frisch elasticity of labor supply to 4, and the labor share to 2/3. Using the production function, we can then measure the aggregate productivity shocks exactly, and estimate that $\rho_{\Delta a} = .350$ and $\sigma_{\Delta a} = .010$. We set $\phi_y = 0.33$ and $\phi_p = 1.24$ to match Rudebusch's (2002) estimates of the Taylor rule, and using these we estimate that $\rho_\varepsilon = .918$ and $\sigma_\varepsilon = .012$.

We start our estimation of the model by finding the set of parameters that maximize

⁶See An and Schorfheide (forthcoming) and Canova (forthcoming) for recent surveys on the estimation of dynamic stochastic general equilibrium models.

the likelihood function. Table 1 presents the estimates. Curiously, we estimate a value for the elasticity of substitution between goods that is higher and a value for the elasticity of substitution between labor that is lower than what is typically assumed. The implied price markup is only 3% and the implied wage markup is 31%, whereas usually these are calibrated to values between 5% and 20%. A second feature to note is that most estimates are quite precise, with tight confidence intervals.

Our main focus of interest are the measures of information stickiness. We estimate that firms are relatively attentive, updating their information about every 4 months, whereas consumers and workers are quite inattentive, only updating their plans about every 16 months. We test the null hypothesis that both members of a household, the consumer and the worker, update their information at the same time. The likelihood ratio statistic is .089, which has a p-value of .23 in the χ_1^2 distribution. The data do not reject this plausible hypothesis.⁷

Table 2 presents the variance decompositions associated with these estimates. Noticeably, the variance of inflation, output, and hours is almost entirely accounted for solely by monetary and aggregate demand shocks. Shocks to productivity and price markups are relatively unimportant for these three variables, but play a role on the fluctuations of interest rates and real wages. Wage markups are on average large, but their fluctuations explain little of the variance of any of the variables.

Next we estimate our model using Bayesian methods instead. We see the main virtue of these methods as allowing us, through the priors, to focus on an area of the parameter space that we are particularly interested in. In our case, this area corresponds to the typical calibrations of these models. For instance, we pick priors for the average substitutability of goods and labor that imply average markups that are with 95% confidence between 6% and 21%, the values commonly assumed in the literature. For the parameters of inattentiveness, we instead opt for a flat prior in order to impose as little as possible on the data. The priors for the correlation and the variance of shocks are similar to those on the literature, although they are more diffuse than usual.

Table 1 contains the results, which turn out to be similar to the maximum-likelihood

⁷With $\delta = \omega$, the wage curve can be re-written instead as:

$$w_t = \delta \sum_{j=0}^{\infty} (1 - \delta)^j E_{t-j} \left[p_t + \frac{\gamma(w_t - p_t) + l_t - \psi\gamma_t/(\gamma - 1)}{\gamma + \psi} \right] + \frac{\psi(y_t - g_t)}{\theta(\gamma + \psi)}. \quad (8)$$

results. As expected, the difference between the markups on goods and labor is not as extreme as before, as our prior heavily penalizes those extreme results. Also as expected, our diffuse priors lead to wider credible sets. However, the estimates of inattentiveness are relatively similar: consumers and workers update their information every 5 to 6 quarters, whereas firms update every 1.5 quarters. Table 2 shows the variance decompositions using these Bayesian estimates. These are similar to the maximum-likelihood conclusions, with the exception of shocks to goods markups, which now account for a larger share of the variance of all variables.

5. Conclusion

In Mankiw and Reis (2002) we proposed a new way to model sluggish macroeconomic adjustment. In this paper we have explored how this approach can be used in an empirical dynamic stochastic general equilibrium model.

One lesson from our estimation (and also emphasized in Mankiw and Reis, 2006) is that information stickiness is pervasive: it applies to firms, workers, and consumers. Some recent research has estimated empirical dynamic stochastic general equilibrium models with sticky information on the part of firms, assuming fully informed workers and consumers.⁸ Our results suggest that these models were misspecified; this misspecification can potentially explain reported poor fits of the model. Although more work is needed before reaching a final verdict, we believe the assumption of sticky information remains a promising tool for applied macroeconomists.

⁸See Trabandt (2003) and Keen (2003) for early attempts to build DSGE models with sticky information on the part of firms, and Andres et al (2005), Korenok and Swanson (2005), Kiley (2005) and Laforte (2005) for estimations. Coibion (2006) finds that sticky information on the part of consumers is important to explain inflation dynamics.

Appendix

This appendix sets out the model in the paper formally, solves it, proves the propositions, and describes the algorithms that we used.

A.1. The economic environment

The model is similar to the one in Mankiw and Reis (2006), but allows for shocks to the elasticities of substitution between varieties of goods and labor. The reader is referred to that article for a more detailed exposition and a description of the intuition behind our assumptions and optimal behavior. Here, we are brief.

There are three types of agents: consumers, workers and firms, of which there is a continuum evenly distributed in the unit interval. Consumers and workers share a household and strive to maximize the same utility function subject to the same budget constraint. Within consumers, there are two members: a shopper, that decides the allocation of spending across the different varieties and is always attentive, and a planner that decides total expenditure and is often inattentive. Firms have two departments: a purchasing department that is always attentive and chooses how much of each variety of labor to hire, and a sales department that is only sporadically attentive and sets the price of the firm's output.

These agents meet in three sets of markets. In the labor market, workers sell their labor to firms; in the goods market, firms sell their goods to consumers; and in the savings market consumers trade bonds between themselves. Monetary and aggregate demand policy follow exogenous rules and close the model.

To lay down the model formally, we start by describing the market clearing conditions and policy processes, then set out the attentive agents' problem, and finally write down the inattentive agents' problem.

Policy and market clearing. We assume that the government consumes a common fraction of each good in the economy. This is financed by lump-sum taxes that keep the budget balanced at all dates. Therefore, the market clearing condition in the market for goods' variety i is:

$$G_t \int_0^1 C_{t,j}(i) dj = Y_{t,i}, \quad (9)$$

where $1 - 1/G_t$ is the fraction of output consumed by the government, $C_{t,j}(i)$ is the consumption of variety i by agent j at time t , and $Y_{t,i}$ is the total production of good i at time t . The fraction G_t is stochastic and shocks to it can be interpreted broadly as aggregate

demand shocks.⁹

The market clearing condition in each labor variety i is:

$$L_{t,i} = \int_0^1 N_{t,j}(i) dj, \quad (10)$$

where $L_{t,i}$ is the total labor supply of variety i at time t , and $N_{t,j}(i)$ is the labor demand by firm j of variety i at time t .

Monetary policy sets interest rates according to:

$$i_t \equiv \log [E_t (\Pi_{t+1} P_t / P_{t+1})] = \phi_y \log \left(\frac{Y_t}{Y_t^n} \right) + \phi_p \log \left(\frac{P_t}{P_{t-1}} \right) - \varepsilon_t. \quad (11)$$

The definition of the nominal interest rate follows the Fisher relation, whereas policy is set according to a Taylor rule. The new notation is P_t for the price level, Π_{t+1} for the gross real interest rate between t and $t+1$, Y_t^n for the equilibrium level of output if all are attentive, and ε_t to discretionary policy shocks. Note that a positive ε_t corresponds to an expansionary shock. The coefficient ϕ_y is positive reflecting a desire for stabilization, and $\phi_p > 1$ to respect the Taylor principle and lead to a determinate solution for inflation. This rule by itself leaves the price level indeterminate, but we peg the initial price level at $P_{-1} = 1$ ensuring determinacy.

Finally, we define total output and total labor as the aggregators across all varieties:¹⁰

$$Y_t = \left(\int_0^1 Y_{t,i}^{\frac{\hat{\nu}_t-1}{\hat{\nu}_t}} di \right)^{\frac{\hat{\nu}_t}{\hat{\nu}_t-1}}, \quad (12)$$

$$L_t = \left(\int_0^1 L_{t,i}^{\frac{\hat{\gamma}_t-1}{\hat{\gamma}_t}} di \right)^{\frac{\hat{\gamma}_t}{\hat{\gamma}_t-1}}. \quad (13)$$

Attentive agents. Consumer's shopper j at date t solves:

$$\min_{\{C_{t,j}(i)\}_{i \in [0,1]}} \int_0^1 P_{t,i} C_{t,j}(i) di \text{ s.t. } C_{t,j} = \left(\int_0^1 C_{t,j}(i)^{\frac{\hat{\nu}_t-1}{\hat{\nu}_t}} di \right)^{\frac{\hat{\nu}_t}{\hat{\nu}_t-1}}. \quad (14)$$

The price of each variety of goods is $P_{t,i}$, and the consumer values them according to a Dixit-Stilitz utility function, with a stochastic elasticity of substitution $\hat{\nu}_t$. The standard

⁹Shocks to the utility of leisure relative to consumption enter the model in a similar way to G_t .

¹⁰Note that using instead the definitions $Y_t = \int_0^1 Y_{t,i} di$ and $L_t = \int_0^1 L_{t,i} di$ leads to the same results up to a first-order approximation.

solution to this problem is:

$$C_{t,j}(i) = C_{t,j}(P_t(i)/P_t)^{-\hat{\nu}_t} \text{ with } P_t = \left(\int_0^1 P_t(i)^{1-\hat{\nu}_t} di \right)^{\frac{1}{1-\hat{\nu}_t}}, \quad (15)$$

so that P_t is the static price index. Summing over all consumers and using the market clearing condition gives the total demand for variety i :

$$Y_{t,i} = (P_t(i)/P_t)^{-\hat{\nu}_t} \bar{C}_t G_t, \text{ where } \bar{C}_t \equiv \int_0^1 C_{t,j} dj. \quad (16)$$

The purchasing department of firm j at date t minimizes expenditures given a Dixit-Stiglitz production function that aggregates labor of different varieties into a labor aggregate:

$$\min_{\{N_{t,j}(i)\}_{i \in [0,1]}} \int_0^1 W_{t,i} N_{t,j}(i) di \text{ s.t. } N_{t,j} = \left(\int_0^1 N_{t,j}(i)^{\frac{\hat{\gamma}_t-1}{\hat{\gamma}_t}} di \right)^{\frac{\hat{\gamma}_t}{\hat{\gamma}_t-1}}. \quad (17)$$

$W_{t,i}$ is the wage paid to labor variety i and $\hat{\gamma}_t$ is the stochastic elasticity of substitution across labor varieties. The solution is:

$$N_{t,j}(i) = N_{t,j}(W_t(i)/W_t)^{-\hat{\gamma}_t} \text{ with } W_t = \left(\int_0^1 W_t(i)^{1-\hat{\gamma}_t} di \right)^{\frac{1}{1-\hat{\gamma}_t}}, \quad (18)$$

where W_t is the static wage index. Summing over all firms and using the market clearing condition, we obtain the demand for labor variety i :

$$L_{t,i} = (W_t(i)/W_t)^{-\hat{\gamma}_t} \bar{N}_t, \text{ where } \bar{N}_t \equiv \int_0^1 N_{t,j} dj. \quad (19)$$

Inattentive agents. We start by considering the problem facing the pricing department of a firm that last updated its information j periods ago. We assume that each period, a randomly drawn fraction of firms λ updates their information, so there are $\lambda(1-\lambda)^j$ firms in this situation. They choose a nominal price to maximize expected real profits:

$$\max_{P_{t,j}} E_{t-j} \left[\frac{P_{t,j} Y_{t,j}}{P_t} - \frac{W_t N_{t,j}}{P_t} \right] \quad (20)$$

$$\text{s.t.: } Y_{t,j} = A_t N_{t,j}^\beta \text{ and (16)} \quad (21)$$

The first constraint is the production function, where β measures the degree of returns to

scale. We interpret this model with no capital accumulation as one in which there is a fixed stock of capital, so β corresponds to the labor share. Aggregate productivity A_t is stochastic. The second constraint is the demand for the firm's product in (16). After some rearranging, the first-order condition of this optimization problem is:

$$P_{t,j} = \frac{E_{t-j} [\hat{\nu}_t W_t N_{t,j} / P_t]}{E_{t-j} [\beta (\hat{\nu}_t - 1) Y_{t,j} / P_t]}. \quad (22)$$

Next, consider the problem of an inattentive consumer's planner. If she updates her plan at date t , she chooses a plan for current and future consumption to solve:

$$V(A_t) = \max_{\{C_{t+i,i}\}} \left\{ \sum_{i=0}^{\infty} \xi^i (1-\delta)^i \frac{C_{t+i,i}^{1-1/\theta} - 1}{1-1/\theta} + \xi \delta \sum_{i=0}^{\infty} \xi^i (1-\delta)^i E_t [V(A_{t+1+i})] \right\}, \quad (23)$$

$$\text{s.t.} \quad : \quad A_{t+1+i} = \Pi_{t+1+i} \left(A_{t+i} - C_{t+i,i} + \frac{W_{t+1+i} L_{t+1+i} + T_{t+1+i}}{P_{t+1}} \right) \text{ for } i=0,1, \quad (24)$$

and a no-Ponzi scheme condition. $V(\cdot)$ is the value function of the agent that depends on her wealth A_t . The parameter ξ is the discount factor, while δ is the probability at each date that the consumer updates her plan. The coefficient θ is the intertemporal elasticity of substitution so preferences are iso-elastic. Preferences are also additively separable in consumption and leisure, but since the consumer does not control labor supply, the term in leisure drops out of her problem. The budget constraint assumes that wages are received at the beginning of the period so they earn interest before the next period. Finally $T_{t,\cdot}$ denote both lump-sum taxes as well as the payments from an insurance contract that all agents sign at the beginning of time that ensures that they all have the same wealth at the start of each period. This is a standard assumption in these models to avoid tracking the wealth distribution over time. The optimality conditions are:

$$\xi^i (1-\delta)^i C_{t+i,i}^{-1/\theta} = \xi \delta \sum_{k=i}^{\infty} \xi^k (1-\delta)^k E_t [V'(A_{t+1+k}) \bar{\Pi}_{t+i,t+1+k}] \text{ for } i=0,1,2,\dots \quad (25)$$

$$V'(A_t) = \xi \delta \sum_{k=0}^{\infty} \xi^k (1-\delta)^k E_t [V'(A_{t+1+k}) \bar{\Pi}_{t,t+1+k}]. \quad (26)$$

We denote by $\bar{\Pi}_{t+i,t+1+k} = \prod_{z=t+i}^{t+k} \Pi_{z+1}$ the compound return between $t+i$ and $t+1+k$. Combining (25) for $i=0$ with (26) one learns that $C_{t,0}^{-1/\theta} = V'(A_t)$. Writing (26) recursively and using these results one gets the first condition below. Condition (25) for $i=j$ and (26)

for date $t + j$ imply the second result:

$$C_{t,0}^{-1/\theta} = \xi E_t \left[R_{t+1} C_{t+1,0}^{-1/\theta} \right], \quad (27)$$

$$C_{t+j,j}^{-1/\theta} = E_{t-j} \left[C_{t+j,0}^{-1/\theta} \right], \quad (28)$$

Finally, we turn to workers. They solve a similar problem to consumers:

$$\hat{V}(A_t) = \max_{\{W_{t+i,i}\}} \left\{ - \sum_{i=0}^{\infty} \xi^i (1-\omega)^i E_t \left(\frac{L_{t+i,i}^{1+1/\psi} + 1}{1 + 1/\psi} \right) + \xi \omega \sum_{i=0}^{\infty} \xi^i (1-\omega)^i E_t \left[\hat{V}(A_{t+1+i}) \right] \right\} \quad (29)$$

s.t. (24) and (19) (30)

where $\hat{V}(\cdot)$ is the value function perceived by the worker, ω is the probability each period that she will update her information, and ψ is the Frisch elasticity of labor supply in the iso-elastic utility function. The worker faces as constraints the same budget as the consumer, as well as the total demand for her services. The optimality conditions are:

$$\xi^i (1-\omega)^i E_t \left(\hat{\gamma}_{t+i} L_{t+i,i}^{1+1/\psi} \right) / W_{t+i,i} = \xi \omega \sum_{k=i}^{\infty} \xi^k (1-\omega)^k E_t \left[V'(A_{t+1+k}) \bar{\Pi}_{t+i,t+1+k} (\hat{\gamma}_{t+i} - 1) L_{t+i,i} / P_{t+i} \right] \quad \text{for } i=0,1,2,\dots \quad (31)$$

$$\hat{V}'(A_t) = \xi \omega \sum_{k=0}^{\infty} \xi^k (1-\omega)^k E_t \left[\hat{V}'(A_{t+1+k}) \bar{\Pi}_{t,t+1+k} \right]. \quad (32)$$

Combining (31) for $i = 0$ with (32) one learns that

$$W_{t,0} = \frac{\hat{\gamma}_t}{\hat{\gamma}_t - 1} \times \frac{P_t L_{t,0}^{1/\psi}}{\hat{V}'_t(A_t)}. \quad (33)$$

Writing (32) recursively and using these results one gets the first condition below. Condition (31) for $i = j$ and (32) for date $t + j$ imply the second result:

$$\frac{\hat{\gamma}_t}{\hat{\gamma}_t - 1} \times \frac{L_{t,0}^{1/\psi} P_t}{W_{t,0}} = \xi E_t \left(R_{t+1} \times \frac{\hat{\gamma}_{t+1}}{\hat{\gamma}_{t+1} - 1} \times \frac{L_{t+1,0}^{1/\psi} P_{t+1}}{W_{t+1,0}} \right), \quad (34)$$

$$W_{t+i,i} = \frac{E_t \left(\hat{\gamma}_{t+i} L_{t+i,i}^{1/\psi} \right)}{E_t \left(\hat{\gamma}_{t+i} L_{t+i,i} L_{t+i,i}^{1/\psi-1} / W_{t+i,0} \right)}. \quad (35)$$

Monopolistically competitive equilibrium. A competitive equilibrium of this economy is

an allocation of total expenditures, consumption of varieties, labor supplied of the different varieties, and output produced of each variety such that consumers, workers and firms all behave optimally, monetary policy follows the Taylor rule, and all markets clear.

A.2. The log-linearized economy and shocks

We log-linearize the equilibrium conditions around the non-stochastic steady state. Small caps denote the log-deviations of the respective large-cap variable from this steady state, with the exceptions of: ν_t and γ_t which are the log-deviations of $\hat{\nu}_t$ and $\hat{\gamma}_t$, r_t which is the log-deviation of the short rate $E_t[\Pi_{t+1}]$, and R_t which is the log-deviation of the long rate $\lim_{k \rightarrow \infty} E_t[\bar{\Pi}_{t,t+1+k}]$.

Log-linearizing the market clearing conditions and policy rules, we get:

$$y_t = g_t + c_t, \quad (36)$$

$$i_t = \phi_y(y_t - y_t^n) + \phi_p \Delta p_t - \varepsilon_t, \quad (37)$$

$$i_t = r_t + E_t(\Delta p_{t+1}), \quad (38)$$

$$c_t = \delta \sum_{j=0}^{\infty} (1 - \delta)^j c_{t,j}, \quad (39)$$

From the attentive agents' section:

$$y_{t,j} = y_t - \nu(p_{t,j} - p_t), \quad (40)$$

$$p_t = \lambda \sum_{j=0}^{\infty} (1 - \lambda)^j p_{t,j}, \quad (41)$$

$$l_{t,j} = l_t - \gamma(w_{t,j} - w_t), \quad (42)$$

$$w_t = \omega \sum_{j=0}^{\infty} (1 - \omega)^j w_{t,j} \quad (43)$$

From the inattentive firm's problem:

$$y_{t,j} = a_t + \beta l_{t,j}, \quad (44)$$

$$\begin{aligned} p_{t,j} &= E_{t-j} [w_t - (y_{t,j} - n_{t,j}) - \nu_t / (\bar{\nu} - 1)] \\ &= E_{t-j} \left[p_t + \frac{\beta(w_t - p_t) + (1 - \beta)y_t - a_t - \nu_t \beta / (\bar{\nu} - 1)}{\beta + \bar{\nu}(1 - \beta)} \right]. \end{aligned} \quad (45)$$

The second expression uses the two constraints in (21) to eliminate firm-specific variables.

From the inattentive consumer's problem:

$$c_{t,0} = E_t(c_{t+1,0} - \theta r_t), \quad (46)$$

$$c_{t,j} = E_{t-j}(c_{t,0}), \quad (47)$$

and from the inattentive worker's problem:

$$w_{t,0} - p_t - l_{t,0}/\psi + \gamma_t/(\bar{\gamma} - 1) = E_t[-r_t + w_{t+1,0} - p_{t+1} - l_{t+1,0}/\psi + \gamma_{t+1}/(\bar{\gamma} - 1)] \quad (48)$$

$$w_{t,j} = E_{t-j}(w_{t,0}). \quad (49)$$

There are five source of shocks in the model: monetary policy, aggregate productivity growth, aggregate demand, goods markups, and labor markups. We assume that each follows an independent AR(1):

$$\varepsilon_t = \rho_\varepsilon \varepsilon_{t-1} + e_t^\varepsilon, \quad \Delta a_t = \rho_{\Delta a} \Delta a_{t-1} + e_t^{\Delta a}, \quad (50)$$

$$g_t = \rho_g g_{t-1} + e_t^g, \quad \nu_t = \rho_\nu \nu_{t-1} + e_t^\nu, \quad \gamma_t = \rho_\gamma \gamma_{t-1} + e_t^\gamma, \quad (51)$$

where the shocks $e_t^s \sim N(0, \sigma_s^2)$ are i.i.d. over time, $E[e_t^s e_{t+k}^s] = 0$ for $k \neq 0$, and independent of each other, $E[e_t^s e_t^{s'}] = 0$ for $s \neq s'$.

A.3. The attentive equilibrium

The attentive equilibrium is the one that obtains when $\delta = \omega = \lambda = 1$, so all are attentive. Following convention, we refer to variables in this equilibrium as being at their “natural” levels and superscript them with n . Note however that this is not a Pareto optimal equilibrium since there is monopoly power. Moreover, note that because the elasticities of substitution change, markups change as well, so the “natural” levels of output or labor are not a constant fraction of their Pareto optimal levels.

If all are attentive, all are identical, so: $y_{t,j} = y_t$, $l_{t,j} = l_t$, $p_{t,j} = p_t$, $w_{t,j} = w_t$, and

$c_{t,j} = c_t$. The model then collapses into the system of 6 equations:

$$\begin{aligned}
y_t^n &= g_t + c_t^n, \\
E_t(\Delta p_{t+1}^n) &= \phi_p \Delta p_t^n - r_t^n - \varepsilon_t, \\
y_t^n &= a_t + \beta l_t^n, \\
0 &= \beta(w_t^n - p_t^n) + (1 - \beta)y_t^n - a_t - \nu_t \beta / (\bar{\nu} - 1), \\
c_t^n &= -\theta r_t + E_t(c_{t+1}^n), \\
z_t &= -r_t^n + E_t(z_{t+1}), \text{ with } z_t \equiv w_t^n - p_t^n - l_t^n / \psi + \gamma_t / (\bar{\gamma} - 1),
\end{aligned}$$

in 6 variables $(y_t^n, c_t^n, l_t^n, p_t^n, r_t^n, w_t^n)$. The solution for output is:

$$\begin{aligned}
y_t^n &= \Xi_a a_t + \Xi_g g_t + \Xi_\gamma \gamma_t + \Xi_\nu \nu_t, \text{ where:} \\
\Xi_a &\equiv (1 + 1/\psi) / (1 + 1/\psi + \beta/\theta - \beta), \quad \Xi_g \equiv (\beta/\theta) / (1 + 1/\psi + \beta/\theta - \beta), \\
\Xi_\gamma &\equiv (\beta/(\gamma - 1)) / (1 + 1/\psi + \beta/\theta - \beta), \quad \Xi_\nu \equiv (\beta/(\nu - 1)) / (1 + 1/\psi + \beta/\theta - \beta).
\end{aligned}$$

Using the solution for output, the solution for the other real variables follows: $l_t^n = (y_t^n - a_t) / \beta$, $w_t^n - p_t^n = y_t^n - l_t^n - \nu_t \beta / (\bar{\nu} - 1)$, $c_t^n = y_t^n - g_t$, and $\theta r_t^n = E_t(y_{t+1}^n - y_t^n - g_{t+1} + g_t)$. Note that for hours to remain bounded, $E(l_t^n) = 0$, requires the King-Plosser-Rebelo (1988) restriction, $\theta = 1$. In this case, output and real wages increase proportionally with aggregate productivity and labor is independent of productivity shocks. This economy respects the classical dichotomy as real variables are determined independently of monetary shocks.

Finally, the solution for inflation in the $\theta = 1$ case is:

$$\begin{aligned}
\Delta p_t^n &= \frac{\rho_a \Delta a_t}{\phi_p - \rho_a} + \frac{(1 - \beta / (1 + 1/\psi))(1 - \rho_g) g_t}{\phi_p - \rho_g} - \frac{(1 - \rho_\gamma) \gamma_t}{(\bar{\gamma} - 1)(1 + 1/\psi)(\phi_p - \rho_\gamma)} \\
&\quad - \frac{(1 - \rho_\nu) \nu_t}{(\bar{\nu} - 1)(1 + 1/\psi)(\phi_p - \rho_\nu)} + \frac{\varepsilon_t}{\phi_p - \rho_\varepsilon}. \tag{52}
\end{aligned}$$

Expansionary monetary policy, higher productivity growth, higher government spending, and higher markups all raise inflation. Nominal interest rates are $i_t^n = r_t^n + E_t(\Delta p_{t+1}^n)$, and the equilibrium is fully characterized.

A.4. The sticky-information equilibrium

Starting with (41), replace $y_{t,j}$ using (40) and $p_{t,j}$ using (45) and rearrange to get the

AS curve in (1). Denoting by mc_t (real marginal costs) the fraction on the right-hand side, it can be re-arranged to obtain a sticky-information Phillips curve:

$$\Delta p_t = \frac{\lambda mc_t}{1 - \lambda} + \lambda \sum_{j=0}^{\infty} (1 - \lambda)^j E_{t-1-j} (\Delta p_t + \Delta mc_t). \quad (53)$$

Next, starting with (46), iterate forward and take the limit as time goes to infinity. Then, use the definition of the long rate R_t and the fact that complete insurance plus the fact that eventually all become aware of shocks implies that $\lim_{i \rightarrow \infty} E_t (c_{t+i,0}) = \lim_{i \rightarrow \infty} E_t [y_{t+i}^n] \equiv y_t^\infty$. Then, using this solution to replace for $c_{t,0}$ in (47) and (39) gives an expression for aggregate consumption. Replacing it in (36) and using the fact that g_t is stationary gives the IS curve in (2).

Very similar steps, using the expressions for $w_{t,0}$ in (48), for $w_{t,j}$ in (49), the aggregator for w_t in (43) and replacing out $l_{t,j}$ using (42), gives the wage curve in (3). Aggregating (44) over j gives the aggregate production function in (4). Finally, the expressions for the nominal interest rate in (37) and (38) give the Taylor rule in (5).

These 5 equations together with the initial condition $p_{-1} = 0$ define an equilibrium in the 5 variables $(y_t, p_t, w_t, l_t, i_t)$ as function of the five stochastic variables $(\varepsilon_t, \Delta a_t, g_t, \nu_t, \gamma_t)$

A.5. Proof of Proposition 1

Using a method of undetermined coefficients, the solution for the generic variable $z_t \in Z_t = \{y_t, p_t, w_t, l_t, i_t\}$ as a function of the innovations e_{t-n}^s for all non-negative n and $s \in S = \{\Delta a, g, \nu, \gamma, \varepsilon\}$ is:

$$z_t = \sum_{s \in S} \sum_{n=0}^{\infty} \hat{z}_n(s) e_{t-n}^s, \quad (54)$$

where $\hat{z}_n(s)$ is the undetermined coefficient measuring the impact of shock s at lag n on variable z . Define the following auxiliary parameters that will be useful:

$$\Lambda_n = \lambda \sum_{i=0}^n (1 - \lambda)^i, \quad \Delta_n = \delta \sum_{i=0}^n (1 - \delta)^i, \quad \text{and} \quad \Omega_n = \omega \sum_{i=0}^n (1 - \omega)^i; \quad (55)$$

$$\Psi_n = \frac{\theta \Delta_n \left\{ [\psi + \gamma(1 - \Omega_n)] \left[\frac{\beta + \nu(1 - \beta)}{\Lambda_n} - \nu(1 - \beta) \right] - \beta \psi \Omega_n \right\}}{(1 - \beta)(\gamma + \psi)\theta \Delta_n + \Omega_n \{ \theta \Delta_n [1 - \gamma(1 - \beta)] + \psi \beta \}} \equiv \frac{\Psi_n^{num}}{\Psi_n^{den}}. \quad (56)$$

Focus first on the impact of monetary policy shocks. The AS in (1) implies that

$$\left[\frac{\beta + \nu(1 - \beta)}{\Lambda_n} - \nu(1 - \beta) \right] \hat{p}_n(\varepsilon) = (1 - \beta)\hat{y}_n + \beta\hat{w}_n \quad (57)$$

for all n . The IS in equation (2) implies that:

$$\hat{y}_n = -\Delta_n \theta \sum_{i=0}^{\infty} \hat{r}_{n+i}, \quad (58)$$

for all n . The wage curve in (3) in turn implies that:

$$(\gamma + \psi - \Omega_n \gamma) \hat{w}_n = \Omega_n \psi \hat{p}_n + \Omega_n \left(\hat{y}_n / \beta - \psi \sum_{i=0}^{\infty} \hat{r}_{n+i} \right). \quad (59)$$

for all n . Using these three equations to substitute out $\sum_{i=0}^{\infty} \hat{r}_{n+i}$ and \hat{w}_n and first-differencing (58) gives the two equations:

$$\hat{y}_n = \Psi_n \hat{p}_n, \quad (60)$$

$$\theta \hat{r}_n = \frac{\hat{y}_{n+1}}{\Delta_{n+1}} - \frac{\hat{y}_n}{\Delta_n}. \quad (61)$$

where we used the definition of the parameter Ψ_n . Finally, using the Taylor rule: (5):

$$\phi_p \hat{p}_n = \phi_y \hat{y}_{n+1} + (1 + \phi_p) \hat{p}_{n+1} - \hat{p}_{n+2} - \hat{r}_{n+1} - \rho_e^{n+1}, \quad (62)$$

and replacing for \hat{y}_n and \hat{r}_n gives the solution in the proposition with:

$$A_n = 1 + \frac{\Psi_n}{\theta \Delta_n}, \quad B_n = A_n + \phi_y \Psi_n + \phi_p, \quad C_n(\varepsilon) = \rho_e \quad (63)$$

You can go through the exact same steps for the other four shocks. It should be evident though that the only change is that there is a new term on the right-hand side of (60) which we denote by $\Upsilon_n(s)$, and that the term in ρ_e no longer appears on the right-hand side of

(62). The solution will therefore have the same form, but now:

$$C_n(s) = \begin{cases} \frac{1}{1-\rho_a} \left[(1-\rho_a^{n+1}) \left(\phi_y \Upsilon_n(a) - \phi_y \Xi_a + \frac{\Upsilon_n(a)}{\theta \Delta_n} \right) - \frac{(1-\rho_a^{n+2}) \Upsilon_{n+1}(a)}{\theta \Delta_{n+1}} \right] & \text{for } s = a \\ \left[\frac{(1-\Upsilon_{n+1}(g)) \rho_g}{\theta \Delta_{n+1}} - \frac{1-\Upsilon_n(g)}{\theta \Delta_n} + \phi_y (\Upsilon_n(g) - \Xi_g) \right] \rho_g^n & \text{for } s = g \\ \left[\left(\phi_y + \frac{1}{\theta \Delta_n} \right) \Upsilon_n(\gamma) - \phi_y \Xi_\gamma - \frac{\Upsilon_{n+1}(\gamma) \rho_\gamma}{\theta \Delta_{n+1}} \right] \rho_\gamma^n & \text{for } s = \gamma \\ \left[\left(\phi_y + \frac{1}{\theta \Delta_n} \right) \Upsilon_n(\nu) - \phi_y \Xi_\nu - \frac{\Upsilon_{n+1}(\nu) \rho_\nu}{\theta \Delta_{n+1}} \right] \rho_\nu^n & \text{for } s = \nu \end{cases}, \quad (64)$$

and

$$\Upsilon_n(s) = \begin{cases} \theta \Delta_n [\gamma + \psi + \Omega_n(1 - \gamma)] / \Psi_n^{den} & \text{for } s = a \\ \beta \psi \Omega_n / \Psi_n^{den} & \text{for } s = g \\ \beta \theta \psi \Omega_n \Delta_n / \Psi_n^{den} (\gamma - 1) & \text{for } s = \gamma \\ \beta \theta \Delta_n [\psi + \gamma(1 - \Omega_n)] / \Psi_n^{den} (\nu - 1) & \text{for } s = \nu \end{cases}. \quad (65)$$

Finally, turning to the boundary conditions, the first follows from the initial condition that ensures the determinacy of the price level. The second condition follows from the fact that as the time after a shocks goes to infinity, all become aware of the shock, inflation approaches its natural level, and this in turn tends to zero since Δp_t^n is stationary. This concludes the proof.

A.6. Algorithm to solve for the sticky-information equilibrium

In principle, solving the second-order difference equation should be easy. In practice, we found that shooting algorithms (including multiple shooting alternatives) or the extended path method were often unreliable. Small numerical imprecisions are compounded by both of these algorithms leading them to quickly diverge away from the solution. As an alternative we found that solving the system of linear equations implied by the difference equation and the boundary conditions:

$$\begin{pmatrix} -B_0 & A_1 & \dots & 0 & 0 & 0 \\ \phi_p & -B_1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -B_{N-2} & A_{N-1} & 0 \\ 0 & 0 & \dots & \phi_p & -B_{N-1} & A_N \\ 0 & 0 & \dots & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \hat{p}_0(s) \\ \hat{p}_1(s) \\ \dots \\ \hat{p}_{N-2}(s) \\ \hat{p}_{N-1}(s) \\ \hat{p}_N(s) \end{pmatrix} = \begin{pmatrix} C_0(s) \\ C_1(s) \\ \dots \\ C_{N-2}(s) \\ C_{N-1}(s) \\ 0 \end{pmatrix} \quad (66)$$

was fast and reliable even for a very large N . Because the matrix of coefficients is sparse

and recursive, this system poses no difficulties to most equation-solving programs. Note that A_n and B_n are bounded and non-zero: tedious algebra shows that $\Psi_n > 0$ for all finite n and $\lim_{n \rightarrow \infty} \Psi_n = 0$ and that as long as all parameters are finite, so is Ψ_n . Therefore, $A_n \neq 0$, $B_n \neq 0$, $\lim_{n \rightarrow \infty} A_n = 1$, and $\lim_{n \rightarrow \infty} B_n = 1 + \phi_p$. The system of equations is therefore well-behaved. We have found that setting $N = 1000$, Matlab can find the solution in less than 5 seconds, and that the ignored terms $\hat{z}_n(s)$, for $n > 1000$, are typically lower than 10^{-15} .

With a solution for prices, after tedious algebra, we can find:

Corollary 1. *The coefficients for output, real interest rates, nominal interest rates, real wages and labor are:*

$$\hat{y}_n(s) = \Psi_n \hat{p}_n(s) + \begin{cases} \frac{\Upsilon_n^a (1 - \rho_a^{n+1})}{1 - \rho_a} & \text{for } s = a \\ \Upsilon_n^s \rho_j^n & \text{for } s = g, \gamma, \nu \\ 0 & \text{for } s = \varepsilon \end{cases} \quad (67)$$

$$\hat{r}_n(s) = \frac{\hat{y}_{n+1}(s)}{\theta \Delta_{n+1}} - \frac{\hat{y}_n(s)}{\theta \Delta_n} + \begin{cases} \frac{g_n}{\theta \Delta_n} - \frac{g_{n+1}}{\theta \Delta_{n+1}} & \text{for } s = g \\ 0 & \text{for } s = a, \gamma, \nu, \varepsilon \end{cases} \quad (68)$$

$$\hat{i}_n(s) = \hat{r}_n(s) + \hat{p}_{n+1}(s) - \hat{p}_n(s) \quad (69)$$

$$(\hat{w}_n - \hat{p}_n)(s) = [1 + \nu(1/\beta - 1)] (1/\Lambda_n - 1) \hat{p}_n(s) + (1 - 1/\beta) \hat{y}_n(s) + \begin{cases} \frac{1 - \rho_a^{n+1}}{\beta(1 - \rho_a)} & \text{for } s = a \\ \frac{\rho_\nu^n}{\nu - 1} & \text{for } s = \nu \\ 0 & \text{for } s = g, \gamma, \varepsilon \end{cases} \quad (70)$$

$$\hat{l}_n(s) = \frac{\hat{y}_n(s)}{\beta} - \begin{cases} \frac{1 - \rho_a^{n+1}}{\beta(1 - \rho_a)} & \text{for } s = a \\ 0 & \text{for } s = g, \gamma, \nu, \varepsilon \end{cases} \quad (71)$$

A.7. Proof of Proposition 2

From the properties of the exogenous shocks: $e_t \sim N(0_5, \Sigma)$ and $E(e_t e'_{t-j}) = 0$ for any $j \neq 0$. The properties of the normal distribution then imply that $X \sim N(0_{5T}, \Omega(I_N \otimes \Sigma)\Omega')$.

The likelihood function follows from the density of the multivariate normal.

A.8. Two algorithms to evaluate the log-likelihood function.

The formula for the log-likelihood function in proposition 2 involves inverting a $5T \times 5T$ matrix $V = \Omega(I_5 \otimes \Sigma)\Omega'$, which is both slow and subject to potentially large numerical errors. There are two approaches to getting around this problem.

The first approach is common in the estimation of ARMA models. A Choleski decomposition gives $V = LL'$, where L is a lower triangular matrix. Defining $L\tilde{X} = X$, we can construct the \tilde{X} vector easily since this is a recursive system of equations. Moreover, $\ln|V| = |LL'| = |L|^2 = 2 \sum_{j=1}^{5T} \ln(l_j)$, where l_j is the j^{th} element in the diagonal of L . The first equality uses the Choleski decomposition, the second the properties that the determinant of the product of two square matrices is equal to the product of the determinants and that the determinant of a matrix is equal to the determinant of its transpose, and the third equality uses the fact that for triangular matrices the determinant equals the product of the elements in the diagonal and the properties of logarithms. The log-likelihood function then becomes:

$$\mathcal{L} = -2.5T \ln(2\pi) - \sum_{j=1}^{5T} \ln(l_j) - 0.5\tilde{X}'\tilde{X}. \quad (72)$$

We have found that this algorithm takes less than 5 seconds to execute.

There is an alternative algorithm that is sometimes faster but only applicable if we have a long data series. If T is large, then setting $N = T$ leads to a negligible error in the solution of the model. In this case, defining $\Omega\hat{X} = X$, then $X'(\Omega(I_5 \otimes \Sigma)\Omega')^{-1}X = \hat{X}'(I_5 \otimes \Sigma^{-1})\hat{X}$. Because Σ is diagonal with σ_k^2 as its k^{th} element, this expression equals $\sum_{k=1}^5 \sum_{j=1}^T \hat{x}_{5(j-1)+k}^2 / \sigma_k^2$. Moreover, $\ln|V| = 2 \ln|\Omega| + \ln(I_5 \otimes \Sigma) = 2 \ln|\Omega| + T \sum_{k=1}^5 \ln(\sigma_k^2)$. Therefore, the log-likelihood function becomes:

$$\mathcal{L} = -2.5T \ln(2\pi) - \ln|\Omega| - 0.5T \sum_{k=1}^5 \ln(\sigma_k^2) - 0.5 \sum_{k=1}^5 \sum_{j=1}^T \hat{x}_{5(j-1)+k}^2 / \sigma_k^2. \quad (73)$$

One difficulty with applying this algorithm is that it requires solving the linear system of $5T$ equations $\Omega\hat{X} = X$. While in principle this could be difficult and numerically imprecise, we have typically found that Matlab is able to do it well.

A.9. Estimation

The goal is to estimate the vector of parameters $\theta = (\nu, \gamma, \rho_g, \sigma_g, \rho_\nu, \sigma_\nu, \rho_\gamma, \sigma_\gamma, \delta, \omega, \lambda)$. The parameter space is $(1, +\infty)$ for ν and γ , $(-1, 1)$ for ρ_g, ρ_ν , and ρ_γ , $(0, +\infty)$ for σ_g, σ_ν , and σ_γ and $(0, 1]$ for δ, ω , and λ .

Maximum likelihood estimates come from maximizing \mathcal{L} with respect to these parameters. We tried several algorithms (Matlab's simplex-search algorithm, fminsearch, Matlab's modified Newton-Raphson algorithm, fmincon, and Chris Sims's alternative modified

Newton-Raphson algorithm, `csmmwel`), and we started them from several points. While there were several local maxima, especially close to the boundaries of the parameter space, the clear global maximum found by all the algorithms is reported in table 1. For an estimate of the variance-covariance matrix of the estimates, we use the inverse of the Hessian: $-(\partial^2 \mathcal{L} / \partial \theta \partial \theta')^{-1}$.

Bayesian estimates come from postulating a prior density for the parameters $f(\theta)$, and computing their posterior density using Bayes law: $f(\theta | X) \propto \exp(\mathcal{L}) f(\theta)$. The priors we use are described in the notes of table 1. There is no closed-form solution for the posterior, which must be simulated numerically using Markov Chain Monte Carlo methods. We developed two ways to do so.

The first uses the Gibbs sampler and exploits the fact that conditional on the remaining parameters, the posterior for $(\sigma_g^2, \sigma_\nu^2, \sigma_\gamma^2)$ is known. Since the prior for each of these parameters is an independent inverse-gamma distribution with parameter $(\tau_{k,1}, \tau_{k,2})$, the density of the prior is proportional to:

$$\prod_{k=3}^5 (\sigma_k^2)^{-0.5\tau_{k,1}+1} \exp(-0.5\tau_{k,2}/\sigma_k^2). \quad (74)$$

Multiplying by the likelihood in (73) shows that the posterior for $(\sigma_g^2, \sigma_\nu^2, \sigma_\gamma^2)$ conditional on the other parameters is proportional to:

$$\prod_{k=3}^5 (\sigma_k^2)^{-0.5(\tau_{k,1}+T)+1} \exp \left[-0.5 \left(\frac{\tau_{k,2}\tau_{k,1} + \sum_{j=1}^T \hat{y}_{5(j-1)+k}^2}{\tau_{k,1} + T} \right) / \sigma_s^2 \right]. \quad (75)$$

Therefore, the posterior is also proportional to the product of three independent inverse-gamma distributions, with parameters $(\tau_{k,1} + T, (\tau_{k,2}\tau_{k,1} + \sum_{j=1}^T \hat{x}_{5(j-1)+k}^2) / (\tau_{k,1} + T))$. The density of the other parameters conditional on the variance does not have a known density, so it must be simulated using a Metropolis algorithm. The Gibbs sampler therefore alternatively draws variances from the product of inverse gamma densities conditional on the other parameters, and then uses a Metropolis step to draw these other parameters conditional on the variances.

The second approach is to use a Metropolis random-walk algorithm for all 11 parameters. The proposal function is a multivariate normal with mean equal to the last draw and variance covariance matrix proportional to its maximum-likelihood estimate. A natural

starting point is the vector of maximum-likelihood estimates.

One would expect that the first approach dominates the second. The pure Metropolis algorithm must learn the shape of all of the distributions, whereas the Gibbs algorithm exploits the knowledge that the density of the variances conditional on the other parameters is a product of inverse gamma distributions. However, for our particular application, we found that the pure Metropolis algorithm converged faster than the Gibbs algorithm. The results in tables 1 and 2 were therefore generated using it. We started 5 Metropolis chains, one at the maximum-likelihood value and the other 4 at overly dispersed draws from the multivariate normal. Multiplying the maximum-likelihood variance-covariance matrix by 0.75 led to an acceptance rate of 25% for the Metropolis algorithm. Each chain was ran for 50,000 draws and the first 30,000 were discarded. It took only about 3 days using in two parallel Pentium 4, 3.2 Ghz computers to obtain these 250,000 draws—this confirms the speed of the algorithms behind propositions 1 and 2. We monitored the scale reduction factors proposed by Brooks and Gelman (1998). The largest of these factors was 1.010 supporting convergence, and plots of the between and within variances confirmed it. We therefore proceeded to mix the draws from the 5 samples to obtain 100,000 independent draws from the posterior density.

References

- An, Sungbae, and Frank Schorfheide (forthcoming). “Bayesian Analysis of DSGE models.” *Econometric Reviews*, forthcoming.
- Brooks, Stephen P., and Andrew Gelman (1998) “General Methods for Monitoring Convergence of Iterative Simulations.” *Journal of Computational and Graphical Statistics*, 7 (4), 434-455.
- Canova, Fabio (forthcoming). *Applied Macroeconomic Research*. Princeton University Press: Princeton.
- Coibion, Olivier (2006). “Inflation Inertia in Sticky Information Models.” *Contributions to Macroeconomics*, 6 (1), article 1.
- Gali, Jordi (1999). “Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?” *American Economic Review*, 89 (1), 249-271.
- Kiley, Michael (2005). “A Quantitative Comparison of Sticky-Price and Sticky-Information Models of Price Setting.” Working paper, Federal Reserve Board.
- King, Robert G., Charles Plosser, and Sergio T. Rebelo (1988). “Production, Growth and Business Cycles I. The Basic Neoclassical Model.” *Journal of Monetary Economics*, 21, 195-232.
- Korenok, Oleg, and Norman R. Swanson (2005). “The Incremental Predictive Information

- Associated with Using Theoretical New Keynesian DSGE Models vs. Simple Linear Econometric Models.” *Oxford Bulletin of Economics and Statistics*, 67 (1), 905-930.
- Laforte, Jean-Philippe (2005). “Pricing Models: A Bayesian DSGE approach for the US Economy.” Working paper, Federal Reserve Board.
- Levin, Andrew T., Alexei Onatski, John C. Williams and Noah Williams (2006). “Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models.” In *NBER Macroeconomics Annual 2005*, edited by Mark Gertler and Kenneth Rogoff, MIT Press: Cambridge.
- Mankiw, N. Gregory and Ricardo Reis (2002). “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve.” *Quarterly Journal of Economics*, 117 (4), 1295-1328.
- Mankiw, N. Gregory and Ricardo Reis (2006) “Pervasive Stickiness.” *American Economic Review*, 96 (2), 164-169.
- Nelson, Edward, Javier Andrés and David López-Salido (2005). “Sticky-Price Models and the Natural Rate Hypothesis.” *Journal of Monetary Economics*, 52 (5), 1025-1053.
- Reis, Ricardo (2004). “Inattentive Consumers.” NBER Working Paper No. 10883.
- Reis, Ricardo (2006). “Inattentive Producers,” *Review of Economic Studies*, 73 (3), 793-821.
- Rudebusch, Glenn D. (2002). “Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia.” *Journal of Monetary Economics*, 49 (6), 1161-1187.
- Smets, Frank and Raf Wouters (2003). “An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area.” *Journal of the European Economic Association*, 1 (5), 1123-1175.
- Wang, Pengfei and Yie Wen (2006). “Solving Linear Difference Systems with Lagged Expectations by a Method of Undetermined Coefficients.” FRB Saint Louis Working Paper No. 2006-003c.

Figure 1: Impulse responses to one-standard-deviation shocks

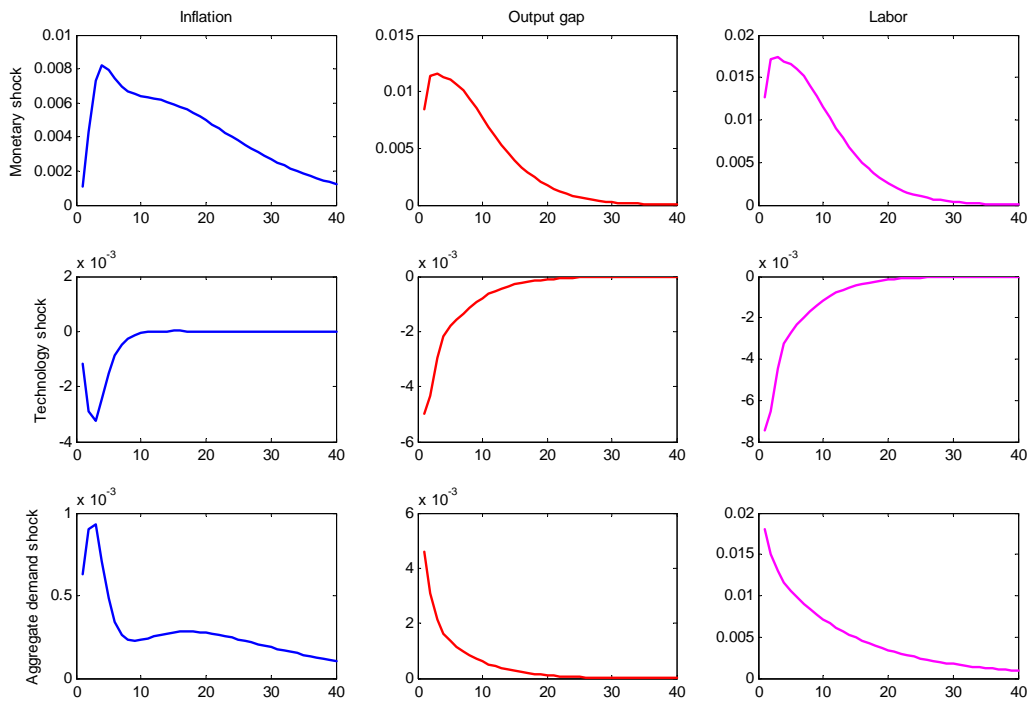


Table 1. Parameter estimates

	<u>Maximum likelihood</u>			<u>Prior</u>				<u>Bayesian posterior</u>		
	Estimate	Standard error	95% confidence interval	Density	Mean	Standard error	95% coverage set	Median	Standard error	95% coverage set
ν	34.068	1.000	[32.109 ; 36.027]	1+G	11	3.162	[5.795 ; 18.085]	20.547	2.781	[15.554 ; 26.408]
γ	4.196	.626	[2.970 ; 5.422]	1+G	11	3.162	[5.795 ; 18.085]	6.884	1.438	[4.542 ; 10.245]
ρ_g	.938	.021	[.897 ; .979]	B	.7	.224	[.198 ; .991]	.950	.022	[.904 ; .988]
σ_g	.014	.002	[.010 ; .018]	IG ^{1/2}	.222	.114	[.107 ; .507]	.015	.002	[.012 ; .019]
ρ_v	.630	.019	[.593 ; .666]	B	.7	.224	[.198 ; .991]	.676	.023	[.628 ; .719]
σ_v	1.819	.252	[1.325 ; 2.313]	IG ^{1/2}	.222	.14	[.107 ; .507]	1.289	.242	[.887 ; 1.838]
ρ_γ	.667	.035	[.599 ; .735]	B	.7	.224	[.198 ; .991]	.638	.043	[.534 ; .701]
σ_γ	.187	.047	[.094 ; .279]	IG ^{1/2}	.222	.114	[.107 ; .507]	.347	.122	[.184 ; .674]
δ	.184	.026	[.133 ; .234]	U	.5	.289	[.025 ; .975]	.176	.027	[.134 ; .242]
ω	.195	.011	[.173 ; .217]	U	.5	.289	[.025 ; .975]	.210	.016	[.182 ; .244]
λ	.702	.015	[.673 ; .731]	U	.5	.289	[.025 ; .975]	.657	.023	[.612 ; .703]

Notes: Sample size is 206. The calibrated coefficients are $\beta=2/3$, $\psi=4$, $\theta=1$, $\phi_y=.33$, $\phi_\pi=1.24$, $\rho_e=.918$, $\sigma_e=.012$, $\rho_{\Delta a}=.350$, $\sigma_{\Delta a}=.010$. Maximum likelihood estimates come from using a modified Newton-Raphson search algorithm to maximize the log-likelihood, and the standard errors from inverting the Hessian at the optimum. The confidence intervals come from the cumulative density function of the multivariate normal. For the prior densities we used the gamma (G), the beta (B), the inverse gamma (IG) and the uniform (U) distributions, with parameters (10,1), (2.24,.96), (2.02,.06) and (0,1) respectively. The posterior moments are based on 100,000 draws from the posterior, which come from mixing 5 independent simulations of 50,000 draws, each with the first 30,000 draws discarded to ensure convergence.

Table 2. Variance decompositions

Maximum likelihood estimates and 95% confidence intervals					
<u>Variable</u>	<u>Shock</u>				
	Monetary	Aggregate productivity	Aggregate demand	Goods markup	Labor markup
Inflation	.896 [.826 ; .937]	.028 [.019 ; .041]	.004 [.002 ; .009]	.070 [.033 ; .130]	.003 [.001 ; .006]
Output growth	.247 [.171 ; .320]	.153 [.115 ; .198]	.436 [.257 ; .610]	.101 [.040 ; .185]	.064 [.024 ; .132]
Hours	.551 [.319 ; .663]	.032 [.014 ; .067]	.336 [.179 ; .612]	.041 [.012 ; .084]	.041 [.010 ; .116]
Interest rate	.506 [.352 ; .646]	.066 [.048 ; .089]	.017 [.008 ; .031]	.295 [.163 ; .438]	.117 [.053 ; .200]
Wage growth	.183 [.136 ; .244]	.262 [.195 ; .360]	.016 [.007 ; .033]	.479 [.300 ; .605]	.061 [.027 ; .099]

Bayesian median estimates and 95% credible sets					
<u>Variable</u>	<u>Shock</u>				
	Monetary	Aggregate productivity	Aggregate demand	Goods markup	Labor markup
Inflation	.835 [.717 ; .906]	.028 [.019 ; .040]	.005 [.003 ; .010]	.129 [.066 ; .242]	.002 [.001 ; .004]
Output growth	.213 [.155 ; .274]	.133 [.101 ; .164]	.446 [.309 ; .590]	.150 [.080 ; .263]	.048 [.021 ; .107]
Hours	.469 [.237 ; .603]	.028 [.010 ; .057]	.398 [.238 ; .703]	.063 [.023 ; .144]	.025 [.006 ; .076]
Interest rate	.408 [.251 ; .569]	.057 [.039 ; .077]	.016 [.010 ; .027]	.416 [.266 ; .585]	.093 [.043 ; .188]
Wage growth	.093 [.053 ; .146]	.192 [.125 ; .271]	.035 [.008 ; .027]	.663 [.514 ; .786]	.035 [.017 ; .070]

Notes: Maximum likelihood estimates come from using the MLE parameter estimates. The confidence intervals are the 2.5% and 97.5% percentiles from 1,000 draws taken from a multivariate normal distribution with mean and variance-covariance equal to their MLE estimates. Bayesian estimates are the median, 2.5% and 97.5% percentiles cell-by-cell using 100,000 draws from the posterior density.