How frequently should we reestimate DSGE models?*

Marcin Kolasa† Michał Rubaszek‡

September 9, 2013

Abstract

A common practice used by policy making institutions using DSGE models for forecasting is to re-estimate them only occasionally rather than every forecasting round. In this paper we ask how such a practice affects the accuracy of DSGE model-based forecasts. To this end we use a canonical medium-sized New Keynesian model and compare how its quarterly real-time forecasts for the US economy vary with the interval between consecutive re-estimations. We find that updating the model parameters only once a year does not lead to any significant deterioration in the accuracy of point forecasts. On the other hand, there are some gains from increasing the frequency of re-estimation if one is interested in the quality of density forecasts.

Keywords: forecasting; DSGE models; parameter updating
JEL Classification: C53; E37

*We are grateful to participants at the International Symposium on Forecasting in Seoul for useful comments. The views expressed in this paper are those of the authors and not necessarily of the institutions they represent.

†Corresponding author. National Bank of Poland and Warsaw School of Economics. Mail address: National Bank of Poland. ul. Świętokrzyska 11/21. 00-919 Warsaw. Poland. Tel.: +48 22 653 2465. Fax: +48 22 585 4374. E-mail: marcin.kolasa@nbp.pl

‡National Bank of Poland and Warsaw School of Economics. E-mail: michal.rubaszek@nbp.pl

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are currently the workhorse framework in macroeconomic analyses and forecasting. Their use is particularly widespread in policy making institutions, especially central banks. This might be due to the fact that the forecasting performance of DSGE models has been found to be relatively good in comparison to standard time series models as well as expert judgment (see e.g. Smets and Wouters, 2003; Adolfson, Lindé, and Villani, 2007; Rubaszek and Skrzypczynski, 2008; Edge, Kiley, and Laforte, 2010; Kolasa, Rubaszek, and Skrzypczynski, 2012; Wieland and Wolters, 2012; Del Negro and Schorfheide, 2012).

A common practice in the policy making institutions is to re-estimate the DSGE models only occasionally rather than every time a new forecast is produced. The main reason for such a practice is that the re-estimation of model complicates the communication between the modelers and policy makers as the difference between consecutive forecasts is affected both by new data release and changes in parameter estimates. Infrequent re-estimation eliminates the second source of the difference. The additional reason is that the estimation process of large DSGE models can be time consuming, especially during the periods of structural breaks. In contrast, the common practice in the DSGE-forecasting literature is to update model parameters quarterly, with the exception of few studies that were updating model parameters every four periods (Adolfson, Lindé, and Villani, 2007; Smets and Wouters, 2007; Christoffel, Coenen, and Warne, 2010) or even only once, i.e. at the beginning of the evaluation sample (Giannone, Monti, and Reichlin, 2010). However, none of the above studies analyzed how the frequency of model re-estimation affects the accuracy of forecasts it generates. While it is possible that frequent parameter updating does not necessarily improve forecasts obtained with econometric models (see e.g. Swamy and Schinasi, 1986), it is also possible that obsolete parameter estimates may have non-negligible costs in terms of the quality of predictions generated for some macroaggregates influencing the policy decisions.

The main question of this study is how often should we update model parameters? To answer the question we take a canonical medium-sized New Keynesian model and compare how its quarterly real-time forecasts for the US economy vary with the interval between consecutive re-estimations. Our main results, based on three key macroeconomic variables (output, inflation and the short-term interest rate), can be summarized as follows. We find that updating the model parameters only once a year does not lead to any significant deterioration in the accuracy of point forecasts. Even though there are some gains from increasing the frequency of re-estimation if one is interested in the quality of density forecasts, these gains are small. According to our results, much of the decrease in forecast accuracy that is observed when re-estimations become less frequent can be attributed to shorter sample rather than to data revisions.

The rest of this paper is organized as follows. Section two describes the model that we use in our investigation. Section three describes the forecasting contest. The results are discussed in section four. Section five concludes.

2 Model

Our investigation is based on the canonical medium-sized New Keynesian framework of Smets and Wouters (2007). It features utility maximizing households, profit maximizing firms, a fiscal authority financing exogenous spending with lump sum taxes, and a central bank setting short term interest rates according to a Taylor-like rule. The model incorporates a number of real and nominal rigidities, including habits in consumption, investment adjustment costs, time-varying capacity utilization, as well as wage and price stickiness with indexation.

The exact specification we use differs from Smets and Wouters (2007) only in that we additionally allow for trend investment-specific technological progress. This modification is aimed to account for the deviation between the average growth rate of real investment and that of other GDP components.[1] A full list of log-linearized model equations can be found in the Appendix.

All estimations are done using Bayesian methods and seven standard quarterly macroeconomic variables for the US: output, consumption, investment, wages, hours worked, inflation and the interest rate. Full definitions and sources of the real-time data used are given in the Appendix. The prior assumptions are identical to those in Smets and Wouters (2007) and also listed in the Appendix. The posterior distributions are approximated using the Metropolis-Hastings algorithm with 250,000 replications, out of which we drop the first 50,000.

3 Forecasting contest

We compare the accuracy of forecasts generated by the DSGE model, the parameters of which are re-estimated in five variants:

update 1Q: the baseline scheme in which the model is re-estimated each quarter,

update 2Q: estimation is repeated when the data for the second or forth quarter are available,

update 1Y: the model is re-estimated when the full-year data are available,

update 2Y: the model is re-estimated only in even years, when the full-year data are available,

fixed: the parameters are estimated once and kept constant throughout the forecast evaluation sample.

. *Check this*. It should be emphasized, however, that for the periods in which the parameters of the model in a given forecasting scheme are not updated, the Kalman filtering

[1]Smets and Wouters (2007) deal with this discrepancey in long-run trends by defining real investment as nominal investment deflated with the GDP deflator. We cannot follow this path since there is no nominal investment series in our real-time database.

and smoothing is applied to compute the realization of shocks in the sample.

Our investigation proceeds in three steps. First, we collect the real time data (RTD) describing the functioning of the US economy in the period between 1966:1 and 2011:4. The data are taken from the Philadelphia Fed "Real-Time Data Set for Macroeconomists", which is described in more detail by Croushore and Stark (2001). The use of the RTD enables us to control for both reasons that justify frequent re-estimation, which we call *sample* and *vintage* effects. To be more precise, let θ_T^v be the vector of parameters estimate based on the sample $1:T$ from the vintage date v. The difference between the actual estimates and those from q quarters ago, both with the use of the latest available vintage, can be decomposed into:

$$\theta_T^v - \theta_{T-q}^{v-q} = \underbrace{(\theta_T^v - \theta_{T-q}^v)}_{\text{sample effect}} + \underbrace{(\theta_{T-q}^v - \theta_{T-q}^{v-q})}_{\text{vintage effect}}. \tag{1}$$

The use of latest available data (LAD) in the estimation process would allow to analyze only the impact of the *sample effect* on the accuracy of forecasts.

In the second stage, for each quarter from the period 1989:4 - 2011:3, we draw from the predictive density. In particular, given 200,000 MH draws of θ from the posterior, we select 5,000 different values θ_i and for each i we draw ten times from the predictive density $p(Y^*|\theta_i)$. The point forecast is calculated as a mean of these draws, whereas the density forecast statistics are calculated on the basis of the 50,000 draws from the predictive density. The forecasting scheme is recursive, the evaluation sample spans from 1990:1 to 2011:4 and the maximum forecast horizon is twelve quarters. More specifically, the first set of forecasts is generated for the period 1990:1-1992:4 with the model estimated on the sample spanning 1966:1-1989:4. The second set of forecasts is for the period 1990:2-1993:1 with the model estimated on the sample 1966:1-1990:1 (baseline model) or 1966:1-1989:4 (remaining models). The third set of forecasts is for the period 1990:3-1993:2 with the model estimated on the sample 1966:1-1990:2 (baseline and update 2Q models) or

1966:1-1989:4 (remaining models), etc. Since our dataset ends in 2011:4 the forecasts are evaluated on the basis of 77 (for 12-quarter ahead forecasts) to 88 (for 1-quarter ahead forecasts) observations.

In the third stage, we assess the quality of forecasts for the key three US macroeconomic time series: output, inflation and the interest rate. Given that the maximum forecast horizon is relatively long, the comparisons are for output and price levels rather than in growth rates. The realizations, which we call *actuals*, are taken from the latest available vintage, i.e. the data released in 2012:1.

4 Results

In this section we report the relative accuracy of point and density forecasts, which is evaluated with the root mean squared forecast error (RMSFE) and log predictive scores (LPS).

4.1 Point forecasts

We begin the comparison by analyzing point forecasts. In Table 1 we report the values for RMSFEs, both using the real-time and latest available data. In the former case we control for the *sample* and *vintage* effects, while in the latter for the *sample* effect only. The numbers for the baseline updating scheme represent the values of the RMSFE and the remaining numbers are expressed as ratios so that values above unity indicate that a given scheme underperforms the baseline. Moreover, to provide a rough gauge of whether the RMSFE ratios are significantly different from unity we report the results of the Diebold-Mariano (1995) test.

The RTD results show that the accuracy of *update 2Q* and *update 1Y* schemes is not significantly different from the baseline. However, for the *update 2Y* and *fixed* schemes the ratios for output, inflation and the interest rate tend to be above unity, where in many cases the difference is significant. This brings us to the first conclusion: the parameters of the DSGE model can be re-estimated once a year without a significant loss of point forecasts accuracy. However, less frequent re-estimation leads to a significant deterioration of forecasts quality.

The comparison of the RTD and LAD results for the baseline scheme indicates the use if the RTD inflates the RMSFEs for output and inflation by about 10% in comparison to the LAD case. The comparison of the RMSFE ratios for the remaining schemes shows that they are broadly the same for the RTD and LAD cases. This brings us to the second conclusion: the RTD result (re-estimate the model at least once a year) is driven to a large extent by the *sample* effect. The *vintage* effect is of lower importance.

4.2 Density forecasts

We complement the discussion on point forecasts accuracy with an evaluation of density forecasts. The aim is to check to what extent the analyzed forecasts provide a realistic description of actual uncertainty.

Let $p(Y_{t+h}|t,j)$ and $p(y_{t+h}|t,j)$ be the predictive density and predictive score of a h-step ahead forecast formulated at time t by the model updated with scheme j. We follow Adolfson, Lindé, and Villani (2007) and assume that $p(Y_{t+h}|t,j)$ is Gaussian, the moments of which can be approximated using the sample of draws from the predictive density. This enables us to compute the average log predictive score (LPS) of h-step ahead forecasts from the j-th scheme as:

$$S_{j,h} = \frac{1}{T-P-h+1}\sum_{t=P}^{T-h} \ln p(y_{t+h}|t,j). \qquad (2)$$

where T and P stand for full sample and in-sample length and $T-P-h+1$ is the number of h-step ahead predictions.

In Tables 2 and 3 we report the values of $S_{j,h}$ for each of the three key macroeconomic variables separately and for three variables together, respectively. As in the case of the RMSFE, we present the results for models estimated with the RTD as well as with the LAD. The numbers for the baseline model represent the values of the LPSs, whereas the remaining numbers are expressed as differences so that values below zero indicate that a given model underperforms the baseline. To provide a rough gauge of whether these differences are significantly different from zero, we report the results of the Amisano and Giacomini (2007) test.

The RTD results show that decreasing the frequency of model re-estimation leads to a deterioration in the quality of density forecasts for interest rates. However, the decrease in the fit, although significant, is not huge. For example, the probability of realizations of forecasts from the model updated with the *update 1Y* scheme is less than 1% lower in comparison to the baseline. In the case of output the deterioration is not significant for the *U2Q* scheme, whereas for inflation the decline in the accuracy of forecasts is significant only for *update 2Y* and *fixed* schemes. As regards the multivariate density forecasts, which takes into account the covariances between forecasts for single variables, the Table 3 suggests that the model can be re-estimated once a year without a loss of forecast precision. The above findings for the density forecasts, although mixed, bring us to the third conclusion: the DSGE model should be re-estimated once a quarter, but less frequent updating (one a year) does not lead to a serious deterioration in the accuracy of density forecasts.

As in the case of point forecasts, the comparison of the RTD and LAD results for the baseline scheme shows that the use if the RTD declines the probability of realization of forecasts for output and inflation in the range from 5 to 15% in comparison to the LAD case. The decline is most sizable for the short-term inflation forecasts. The comparison of the LPS differences for the remaining schemes indicates that they are broadly the same for the RTD and LAD cases. This brings us to the forth conclusion (similar to the second one): the RTD result (re-estimate the model at least once a year) is driven to a large extent by the *sample* effect. The *vintage* effect is of lower importance.

To summarize, the general conclusions that can be drawn from the above exercise is that the DSGE model can be updated once a year without a serious loss in forecast accuracy. The additional conclusion is that the main source of loss in forecast accuracy when updating becomes less frequent is related to the *sample* rather than the *vintage* effect.

5 Conclusions

The results of this study show that the common practice in the policy making institutions to re-estimate the DSGE models only occasionally is justified by the following arguments. Updating the model parameters only once a year does not lead to any significant deterioration in the accuracy of forecasts. Additionally, infrequent re-estimation facilitates the communication between the modelers and policy makers. The above conclusion, however, has to be interpreted with caution. Following Hendry (1980), forecasting has a lot in common with alchemy, the art of transmuting base metals into noble ones. In certain circumstances, in times of a financial crisis for instance, it might be justified to update model parameters frequently or even change the suite of models used for forecasting.

References

Adolfson, Malin, Jesper Lindé, and Mattias Villani. 2007. "Forecasting Performance of an Open Economy DSGE Model." *Econometric Reviews* 26 (2-4):289–328.

Amisano, Gianni and Raffaella Giacomini. 2007. "Comparing Density Forecasts via Weighted Likelihood Ratio Tests." *Journal of Business & Economic Statistics* 25:177–190.

Christoffel, Kai, Günter Coenen, and Anders Warne. 2010. "Forecasting with DSGE models." Working Paper Series 1185, European Central Bank.

Croushore, Dean and Tom Stark. 2001. "A real-time data set for macroeconomists." *Journal of Econometrics* 105 (1):111–130.

Del Negro, Marco and Frank Schorfheide. 2012. "DSGE Model-Based Forecasting." In *Handbook of Economic Forecasting*, *Handbook of Economic Forecasting*, vol. 2, edited by G. Elliott and A. Timmermann. Elsevier.

Diebold, Francis X. and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13 (3):253–63.

Edge, Rochelle M., Michael T. Kiley, and Jean-Philippe Laforte. 2010. "A comparison of forecast performance between Federal Reserve staff forecasts, simple reduced-form models, and a DSGE model." *Journal of Applied Econometrics* 25 (4):720–754.

Giannone, Domenico, Francesca Monti, and Lucrezia Reichlin. 2010. "Incorporating conjunctural analysis in structural models." In *The Science and Practice of Monetary Policy Today*, edited by Volker Wieland. Springer Berlin Heidelberg, 41–57.

Hendry, David F. 1980. "Econometrics-Alchemy or Science?" *Economica* 47 (188):387–406.

Kolasa, Marcin, Michal Rubaszek, and Pawel Skrzypczynski. 2012. "Putting the New Keynesian DSGE model to the real-time forecasting test." *Journal of Money, Credit and Banking* 44 (7):1301–1324.

Rubaszek, Michal and Pawel Skrzypczynski. 2008. "On the forecasting performance of a small-scale DSGE model." *International Journal of Forecasting* 24 (3):498–512.

Smets, Frank and Raf Wouters. 2003. “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area.” *Journal of the European Economic Association* 1 (5):1123–1175.

Smets, Frank and Rafael Wouters. 2007. “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *American Economic Review* 97 (3):586–606.

Swamy, P.A.V.B. and Garry J. Schinasi. 1986. “Should fixed coefficients be reestimated every period for extrapolation?” International Finance Discussion Papers 287, Board of Governors of the Federal Reserve System (U.S.).

Wieland, Volker and Maik H. Wolters. 2012. “Foreacasting and Policy Making.” In *Handbook of Economic Forecasting*, *Handbook of Economic Forecasting*, vol. 2, edited by G. Elliott and A. Timmermann. Elsevier.

Tables and figures

Table 1: Root Mean Squared Forecast Errors (RMSFE)

	$H=1$	$H=2$	$H=4$	$H=6$	$H=8$	$H=12$
	Real time data results					
	Output					
update 1Q	0.66	1.17	2.03	2.61	3.08	4.14
update 2Q	1.00	1.00	1.00	1.00	1.00	1.01
update 1Y	1.01	1.01	1.01	1.01	1.01	1.01
update 2Y	1.03	1.02	1.03*	1.04**	1.04***	1.04**
Fixed	1.01	0.99	1.14*	1.35**	1.51***	1.55***
	Prices					
update 1Q	0.26	0.47	0.87	1.30	1.77	2.76
update 2Q	1.00	0.99	0.99	1.00	1.00	1.01
update 1Y	1.00	0.99	1.00	1.00	1.00	1.02
update 2Y	1.00	1.00	1.01	1.03	1.04	1.08**
Fixed	1.00	1.10	1.39**	1.62***	1.73***	1.81***
	Interest rate					
update 1Q	0.12	0.22	0.38	0.46	0.52	0.57
update 2Q	1.01	1.01	1.00	1.00	1.00	1.00
update 1Y	1.01	1.01	1.01	1.00	1.00	0.99
update 2Y	1.03**	1.04**	1.02*	1.02	1.01	0.98
Fixed	1.28***	1.30***	1.30***	1.27***	1.22***	1.16***
	Latest available data results					
	Output					
update 1Q	0.61	1.07	1.92	2.50	2.88	3.65
update 2Q	1.01	1.01	1.01	1.00	1.01	1.01
update 1Y	1.01	1.02	1.01	1.01	1.02	1.01
update 2Y	1.03	1.03	1.03	1.03	1.04	1.03
Fixed	1.13	1.11	1.02	0.97	0.94	0.92
	Prices					
update 1Q	0.21	0.37	0.77	1.20	1.72	2.84
update 2Q	1.01	1.01	1.01	1.00	0.99	0.99
update 1Y	1.02	1.03	1.02	1.02	1.02	1.03
update 2Y	1.03	1.05	1.05	1.04	1.04	1.05
Fixed	1.06	1.09	1.20**	1.31***	1.39***	1.48***
	Interest rate					
update 1Q	0.12	0.23	0.39	0.48	0.54	0.59
update 2Q	1.02	1.01	1.00	1.00	1.00	1.00
update 1Y	1.04**	1.03	1.01	1.00	0.99	1.00
update 2Y	1.02	1.03	1.02	1.01	1.01	0.99
Fixed	1.24**	1.24**	1.24***	1.26***	1.29***	1.34***

Notes: For the baseline the RMSFEs are reported in levels, whereas for the remaining models they appear as the ratios so that the values above unity indicate that a given model has a higher RMSFE than the baseline. To provide a rough guidance of whether the ratios are different from unity. we use the Diebold-Mariano test. where the long-run variance is calculated with the Newey-West method. Asterisks ***, **and *denote the 1%, 5% and 10% significance levels, respectively.

Table 2: Average univariate log predictive scores

	$H = 1$	$H = 2$	$H = 4$	$H = 6$	$H = 8$	$H = 12$
	Real time data results					
	Output					
update 1Q	-1.123	-1.665	-2.205	-2.468	-2.634	-2.900
update 2Q	-0.003	0.000	-0.003	-0.002	-0.002	-0.007
update 1Y	-0.009**	-0.007	-0.012*	-0.013*	-0.011	-0.009
update 2Y	-0.021**	-0.021**	-0.031***	-0.031**	-0.028**	-0.024*
Fixed	-0.123*	-0.091*	-0.116**	-0.211**	-0.298**	-0.389**
	Prices					
update 1Q	-0.146	-0.800	-1.512	-1.939	-2.239	-2.642
update 2Q	0.000	0.002	0.002	0.001	0.000	-0.003*
update 1Y	-0.002	0.000	-0.001	-0.001	-0.002	-0.008*
update 2Y	-0.007*	-0.005	-0.008	-0.011*	-0.015**	-0.027***
Fixed	-0.137***	-0.209***	-0.300***	-0.356***	-0.387***	-0.427***
	Interest rate					
update 1Q	0.320	-0.130	-0.549	-0.736	-0.839	-0.938
update 2Q	-0.003**	-0.004**	-0.003	-0.003	-0.002	0.001
update 1Y	-0.008***	-0.009***	-0.009***	-0.006	-0.003	0.004
update 2Y	-0.019***	-0.023***	-0.024***	-0.021**	-0.018	0.006
Fixed	-0.176***	-0.204***	-0.212***	-0.193***	-0.159***	-0.103***
	Latest available data results					
	Output					
update 1Q	-1.084	-1.604	-2.156	-2.427	-2.580	-2.806
update 2Q	-0.005	-0.008	-0.007	-0.004	-0.005	-0.005
update 1Y	-0.011	-0.015	-0.014	-0.013	-0.016	-0.012
update 2Y	-0.021*	-0.026	-0.029*	-0.031**	-0.035**	-0.030**
Fixed	-0.140***	-0.123***	-0.074	-0.053	-0.055	-0.039
	Prices					
update 1Q	0.011	-0.677	-1.438	-1.890	-2.209	-2.638
update 2Q	-0.005	-0.004	-0.002	-0.001	-0.000	-0.001
update 1Y	-0.010	-0.011	-0.010	-0.010	-0.011	-0.015
update 2Y	-0.017	-0.021**	-0.023**	-0.023**	-0.023**	-0.029**
Fixed	-0.093***	-0.142***	-0.195***	-0.230***	-0.256***	-0.301***
	Interest rate					
update 1Q	0.331	-0.124	-0.551	-0.740	-0.839	-0.927
update 2Q	-0.005**	-0.005*	-0.003	-0.001	0.000	0.002
update 1Y	-0.012***	-0.014**	-0.009	-0.003	0.000	0.000
update 2Y	-0.016***	-0.019***	-0.021**	-0.019	-0.019	0.004
Fixed	-0.162***	-0.171***	-0.182***	-0.201***	-0.222***	-0.254***

Notes: For the baseline model LPSs are reported in levels, whereas for the remaining models they appear as the differences so that the values below zero indicate that a given model has a lower LPS than the baseline. To provide a rough guidance of whether the differences are different from zero, we use the Amisano and Giacomini (2007) test, where the long-run variance is calculated with the Newey-West method. Asterisks ***, **and *denote the 1%, 5% and 10% significance levels, respectively.

Table 3: Average log predictive scores for 3 variables

	$H=1$	$H=2$	$H=4$	$H=6$	$H=8$	$H=12$
	Real time data results					
update 1Q	-3.882	-5.682	-7.763	-8.901	-9.71	-11.26
update 2Q	0.001	0.004	-0.005	-0.007	-0.015	-0.030
update 1Y	0.001	0.006	0.000	-0.008	-0.009	-0.004
update 2Y	-0.023	-0.034*	-0.049*	-0.046	-0.038	-0.002
Fixed	-0.732***	-1.131***	-1.990***	-2.877***	-3.637***	-4.865***
	Latest available data results					
update 1Q	-3.796	-5.569	-7.593	-8.742	-9.596	-11.29
update 2Q	0.007	0.007	0.002	0.001	0.002	-0.013
update 1Y	0.015	0.021	0.015	0.014	0.011	0.016
update 2Y	0.007	-0.001	-0.015	-0.012	-0.015	0.028
Fixed	-0.385***	-0.522***	-0.758***	-1.035***	-1.386***	-2.116***

Notes: For the baseline model LPSs are reported in levels, whereas for the remaining models they appear as the differences so that the values below zero indicate that a given model has a lower LPS than the baseline. To provide a rough guidance of whether the differences are different from zero, we use the Amisano and Giacomini (2007) test, where the long-run variance is calculated with the Newey-West method. Asterisks ***, **and *denote the 1%, 5% and 10% significance levels, respectively.

Appendix

A List of model equations

This section lays out the full system of log-linearized equations that make up the DSGE model analyzed in the main text. The specification is the same as used by Smets and Wouters (2007). The only difference concerns trend investment specific technological progress, which we include to account for a secular trend in the real investment to output ratio observed in the data.

This modification means that for the model to have a well defined steady state, the trending real variables need to be normalized both with neutral (γ) and investment specific (γ_i) technological progress (see e.g. Greenwood, Herzowitz and Krussel, 1997). In particular, if we define $\gamma_y = \gamma\gamma_i^{\frac{\alpha}{1-\alpha}}$, the stationarization is as follows: $y_t = Y_t/\gamma_y^t$, $y_t^p = Y_t^p/\gamma_y^t$, $c_t = C_t/\gamma_y^t$, $i_t = I_t/(\gamma_y\gamma_i)^t$, $k_t = K_t/(\gamma_y\gamma_i)^t$, $k_t^s = K_t^s/(\gamma_y\gamma_i)^t$, $w_t = W_t/(P_t\gamma_y^t)$, $r_t^k = R_t^k\gamma_i^t/P_t$ and $q_t = Q_t\gamma_i^t$, where Y_t is output, Y_t^p is potential output, C_t is consumption, I_t is investment, K_t is capital, K_t^s is capital services, W_t is nominal wage, R_t^k is the nominal rental rate on capital, Q_t is the real price of capital and P_t is the price level. As regards the remaining endogenous variables showing up in the equations below, l_t stands for labor, r_t is the nominal interest rate, π_t is inflation, μ_t^p is price markup, μ_t^w is wage markup and z_t is capital utilization.

The model is driven by seven stochastic disturbances: total factor productivity ε_t^a, investment specific technology ε_t^i, risk premium ε_t^b, exogenous spending ε_t^g, price markup ε_t^p, wage markup ε_t^w, and monetary policy ε_t^r. The two markup shocks are assumed to follow ARMA(1,1) processes. All remaining disturbances are modeled as first-order autoregressions, except that the government spending shock additionally depends on the current innovation to total factor productivity. All variables presented in the equations below are expressed as log deviations from the non-stochastic steady state. The parameters are defined in section D. Stars in subscripts indicate the steady state values, which are functions of deep model parameters.

Aggregate resource constraint

$$y_t = (1 - (\gamma_y\gamma_i - 1 + \delta)k_* y_*^{-1} - g_y)c_t + (\gamma_y\gamma_i - 1 + \delta)k_y i_t + r_*^k k_* y_*^{-1} z_t + g_y\varepsilon_t^g \tag{A.1}$$

Consumption Euler equation

$$c_t = \frac{\lambda}{\lambda+\gamma_y}c_{t-1} + \frac{\gamma_y}{\lambda+\gamma_y}E_t c_{t+1} + \frac{\gamma_y(\sigma_c-1)\frac{w_* l_*}{c_*}}{\sigma_c(\gamma_y+\lambda)(1+\lambda_w)}(l_t - E_t l_{t+1})$$
$$-\frac{\gamma_y-\lambda}{\sigma_c(\gamma_y+\lambda)}(r_t - E_t\pi_{t+1}) + \varepsilon_t^b \quad \text{(A.2)}$$

Investment Euler equation

$$i_t = \frac{1}{1+\beta\gamma_y^{1-\sigma_c}}i_{t-1} + \frac{\beta\gamma_y^{1-\sigma_c}}{1+\beta\gamma_y^{1-\sigma_c}}E_t i_{t+1} + \frac{1}{(1+\beta\gamma_y^{1-\sigma_c})\gamma_y^2\gamma_i^2\varphi}q_t + \varepsilon_t^i \quad \text{(A.3)}$$

Value of capital

$$q_t = \frac{\beta(1-\delta)}{\gamma_y^{\sigma_c}\gamma_i}E_t q_{t+1} + \frac{\gamma_y^{\sigma_c}\gamma_i - \beta(1-\delta)}{\gamma_y^{\sigma_c}\gamma_i}E_t r_{t+1}^k - (r_t - E_t\pi_{t+1}) + \frac{\sigma_c(\gamma_y+\lambda)}{\gamma_y-\lambda}\varepsilon_t^b \quad \text{(A.4)}$$

Aggregate production function

$$y_t = \phi_p(\alpha k_t^s + (1-\alpha)l_t + \varepsilon_t^a) \quad \text{(A.5)}$$

Capital services

$$k_t^s = k_{t-1} + z_t \quad \text{(A.6)}$$

Optimal capacity utilization

$$z_t = \frac{1-\psi}{\psi}r_t^k \quad \text{(A.7)}$$

Capital accumulation

$$k_t = \frac{1-\delta}{\gamma_y\gamma_i}k_{t-1} + \frac{\gamma_y\gamma_i - 1 + \delta}{\gamma_y\gamma_i}i_t + (\gamma_y\gamma_i - 1 + \delta)(1+\beta\gamma_y^{1-\sigma_c})\gamma_y\gamma_i\varphi\varepsilon_t^i \quad \text{(A.8)}$$

Price markup

$$\mu_t^p = \alpha(k_t^s - l_t) + \varepsilon_t^a - w_t \quad \text{(A.9)}$$

Phillips curve

$$\pi_t = \frac{\iota_p}{1+\beta\gamma_y^{1-\sigma_c}\iota_p}\pi_{t-1} + \frac{\beta\gamma_y^{1-\sigma_c}}{1+\beta\gamma_y^{1-\sigma_c}\iota_p}E_t\pi_{t+1}$$
$$-\frac{(1-\beta\gamma_y^{1-\sigma_c}\xi_p)(1-\xi_p)}{(1+\beta\gamma_y^{1-\sigma_c}\iota_p)\xi_p((\phi_p-1)\varepsilon_p+1)}\mu_t^p + \varepsilon_t^p \quad \text{(A.10)}$$

Input cost minimization

$$r_t^k = -(k_t - l_t) + w_t \tag{A.11}$$

Wage markup

$$\mu_t^w = w_t - (\sigma_l l_t + \frac{1}{\gamma_y - \lambda}(\gamma_y c_t - \lambda c_{t-1})) \tag{A.12}$$

Real wage dynamics

$$\begin{aligned} w_t = & \frac{1}{1+\beta\gamma_y^{1-\sigma_c}} w_{t-1} + \frac{\beta\gamma_y^{1-\sigma_c}}{1+\beta\gamma_y^{1-\sigma_c}}(E_t w_{t+1} - E_t \pi_{t+1}) - \frac{1+\beta\gamma_y^{1-\sigma_c}\iota_w}{1+\beta\gamma_y^{1-\sigma_c}}\pi_t \\ & + \frac{\iota_w}{1+\beta\gamma_y^{1-\sigma_c}}\pi_{t-1} - \frac{(1-\beta\gamma_y^{1-\sigma_c}\xi_w)(1-\xi_w)}{(1+\beta\gamma_y^{1-\sigma_c})\xi_w((\phi_w - 1)\varepsilon_w + 1)}\mu_t^w + \varepsilon_t^w \end{aligned} \tag{A.13}$$

Taylor rule

$$r_t = \rho r_{t-1} + (1-\rho)[r_\pi \pi_t + r_y(y_t - y_t^p)] + r_{\Delta y}(\Delta y_t - \Delta y_t^p) + \varepsilon_t^r \tag{A.14}$$

B Data

The source of all data used to estimate the model is the "Real-Time Data Set for Macroeconomists" (RTDSM) database maintained by the Federal Reserve Bank of Philadelphia. The only exception is the short-term interes rate, which is not subject to revisions and taken from the Federal Reserve Board statistics. The exact definitions follow below (RTDSM codes in parentheses).

Ouptut: Real gross domestic product (ROUTPUT) divided by civilian noninstitutional population (POP).

Consumption: Real personal consumption expenditures (RCON) divided by civilian noninstitutional population (POP).

Investment: Real gross private domestic nonresidential investment (RINVBF) divided by civilian noninstitutional population (POP).

Hours: Aggregate weekly hours (H) divided by civilian noninstitutional population (POP), normalized to average one over the estimation sample.

Wages: Nominal wage and salary disbursements (WSD) divided by civilian noninstitutional population (POP) and deflated by the price index for gross domestic product (P).

Price level: Price index for gross domestic product (P).

Interest rate: Federal funds rate.

C Measurement equations

The following equations relate the model variables to their empirical conterparts defined in section B:

$$\Delta \log Output_t = \gamma_y - 1 + y_t - y_{t-1} \tag{C.1}$$

$$\Delta \log Consumption_t = \gamma_y - 1 + c_t - c_{t-1} \tag{C.2}$$

$$\Delta \log Investment_t = \gamma_y \gamma_i - 1 + i_t - i_{t-1} \tag{C.3}$$

$$\log Hours_t = \bar{l} + l_t \tag{C.4}$$

$$\Delta \log Wages_t = \gamma_y - 1 + w_t - w_{t-1} \tag{C.5}$$

$$\Delta \log PriceLevel_t = \bar{\pi} + \pi_t \tag{C.6}$$

$$InterestRrate_t = \beta^{-1} \gamma_y^{\sigma_c} (1 + \bar{\pi}) - 1 + r_t \tag{C.7}$$

D Calibration and prior assumptions

The calibrated parameters are reported in Table D.1, while Tables D.2 and D.3 describe the prior assumptions used in Bayesian estimation.

Table D.1: Calibrated parameters

Parameter	Value	Description
δ	0.025	Depreciation rate
g_y	0.18	Exogenous spending share in output
λ_w	1.5	Steady-state wage markup
ε_p	10	Kimball aggregator curvature in the goods market
ε_w	10	Kimball aggregator curvature in the labor market

Table D.2: Prior assumptions - structural parameters

Parameter	Type	Mean	Std.	Description
φ	normal	4	1.5	Investment adjustment cost curvature
σ_c	normal	1.5	0.37	Inv. elasticity of intertemporal substitution
h	beta	0.7	0.1	Habit persistance
ξ_w	beta	0.5	0.1	Calvo probability for wages
σ_l	normal	2	0.75	Inv. Frisch elasticity of labor supply
ξ_p	beta	0.5	0.1	Calvo probability for prices
ι_w	beta	0.5	0.15	Wage indexation
ι_p	beta	0.5	0.15	Price indexation
ψ	beta	0.5	0.15	Capacity utilization cost curvature
ϕ	normal	1.25	0.12	Steady-state price markup
r_π	normal	1.5	0.25	Weight on inflation in Taylor rule
ρ	beta	0.75	0.1	Interest rate smoothing
r_y	normal	0.12	0.05	Weight on output gap in Taylor rule
$r_{\Delta y}$	normal	0.12	0.05	Weight on output gap change in Taylor rule
$\bar{\pi}$	gamma	0.62	0.1	Steady-state inflation rate
$100(\beta^{-1}-1)$	gamma	0.25	0.1	Rate of time preference
$\bar{l}$	normal	0	2	Steady-state hours worked
$100(\gamma_y-1)$	normal	0.4	0.1	Trend growth of output
$100(\gamma_i-1)$	normal	0.3	0.1	Trend growth of relative price of investment
α	normal	0.3	0.05	Capital share

Table D.3: Prior assumptions - shocks

Parameter	Type	Mean	Std.	Description
σ_a	inv. gamma	0.1	2	Volatility of productivity shock
σ_b	inv. gamma	0.1	2	Volatility of risk premium shock
σ_g	inv. gamma	0.1	2	Volatility of exogenous spending shock
σ_i	inv. gamma	0.1	2	Volatility of investment specific shock
σ_r	inv. gamma	0.1	2	Volatility of monetary policy shock
σ_p	inv. gamma	0.1	2	Volatility of price markup shock
σ_w	inv. gamma	0.1	2	Volatility of wage markup shock
ρ_a	beta	0.5	0.2	Persistence of productivity shock
ρ_b	beta	0.5	0.2	Persistence of risk premium shock
ρ_g	beta	0.5	0.2	Persistence of exogenous spending shock
ρ_i	beta	0.5	0.2	Persistence of investment specific shock
ρ_r	beta	0.5	0.2	Persistence of monetary policy shock
ρ_p	beta	0.5	0.2	Persistence of price markup shock
ρ_w	beta	0.5	0.2	Persistence of wage markup shock
μ_p	beta	0.5	0.2	Moving average term in price markup
μ_w	beta	0.5	0.2	Moving average term in wage markup
ρ_{ga}	beta	0.5	0.2	Impact of productivity on exogenous spending