Behavioral Heterogeneity in U.S. Inflation Dynamics – Appendix

A NKPC with heterogeneous expectations

The derivation outlined in this section follows the results presented in Kurz (2012) and Kurz et al. (2013). Under the Calvo (1983) pricing mechanism, the optimization problem of a firm with belief type i and producing good j reads as follows

$$\max_{P_{i,t}(j)} E_t^i \sum_{s=0}^{\infty} \omega^s Q_{t,t+s} \left(\frac{P_{i,t}(j)}{P_{t+s}} - mc_{t+s} \right) \left(\frac{P_{i,t}(j)}{P_{t+s}} \right)^{-\eta} Y_{t+s}.$$

When each firm producing a certain good $j \in [0,1]$ has a different subjective expectation of type $i \in [0,1]$, we can without loss of generality use a single index, say i, to denote expectation's type and produced good. Thus we can write firm's i optimization problem as

$$\max_{P_{i,t}} E_t^i \sum_{s=0}^{\infty} \omega^s Q_{t,t+s} \left(\frac{P_{i,t}}{P_{t+s}} - mc_{t+s} \right) \left(\frac{P_{i,t}}{P_{t+s}} \right)^{-\eta} Y_{t+s},$$

where $P_{i,t}$ denotes the price set by a firm with subjective expectations of type i producing good i. Defining $q^*_{i,t} \equiv P^*_{i,t}/P_t$, where $P^*_{i,t}$ is the profit-maximizing price, and log-linearizing the first order conditions of this problem around a zero inflation steady state leads to

$$\hat{q}^*_{i,t} + \hat{p}_t = (1-\omega\delta) E_t^i \sum_{s=0}^{\infty} (\omega\delta)^s (\widehat{mc}_{t+s} + \hat{p}_{t+s}), \tag{A.1}$$

where δ is the time discount factor and hatted variables denote log-deviations from steady state. Leading Eq. (A.1) and taking expectations on both sides we get

$$E_t^i(\hat{q}^*_{i,t+1} + \hat{p}_{t+1}) = (1-\omega\delta) E_t^i \sum_{s=0}^{\infty} (\omega\delta)^s (\widehat{mc}_{t+s+1} + \hat{p}_{t+s+1}), \tag{A.2}$$

where we assumed, as standard in the learning literature, that the law of iterated expectations holds at the individual level (see e.g., Evans and Honkapohja (2001), Branch and McGough (2009) and Kurz et al. (2013)). Rewriting Eq. (A.1) as

$$\hat{q}^*_{i,t} + \hat{p}_t = (1-\omega\delta)(\widehat{mc}_t + \hat{p}_t) + (1-\omega\delta) E_t^i \sum_{s=0}^{\infty} (\omega\delta)^{s+1} (\widehat{mc}_{t+s+1} + \hat{p}_{t+s+1})$$

and substituting Eq. (A.2) we get

$$\hat{q}^*_{i,t} = (1-\omega\delta)\widehat{mc}_t + \omega\delta E_t^i(\hat{q}^*_{i,t+1} + \hat{\pi}_{t+1}), \tag{A.3}$$

which corresponds to Eq. (1) is Section 2.1.

In order to derive an expression for the aggregate price

$$P_t = \left[\int_0^1 P_{i,t}^{(1-\eta)}\right]^{\frac{1}{1-\eta}},$$

we follow the literature and make the following assumption:

Assumption 1 (Kurz et al., 2013): Assume that the sample of firms allowed to adjust prices in each period is selected independently across agents, so that the distribution of firms in terms of output or belief is the same whether one looks at firms that adjust prices or at firms that do not adjust prices.

Given the Calvo pricing scheme, in each period only a set of firms $S_t \in [0,1]$ of measure $1-\omega$ adjust prices, while a set $S_t^c \in [0,1]$ of measure ω do not adjust. Using Assumption 1 we can write

$$P_t^{1-\eta} = \int_{S_t} P_{i,t}^{*(1-\eta)} di + \int_{S_t^c} P_{i,t-1}^{(1-\eta)} di = (1-\omega)\int_0^1 P_{i,t}^{*(1-\eta)} di + \omega P_{t-1}^{1-\eta},$$

or equivalently

$$1 = (1-\omega)\int_0^1 q_{i,t}^{*(1-\eta)} di + \omega\left(\frac{P_{t-1}}{P_t}\right)^{(1-\eta)}.$$

Log-linearizing the expression above we get

$$\pi_t = \frac{1-\omega}{\omega}\int_0^1 \hat{q}_{i,t}^{*} di. \tag{A.4}$$

Denote $\hat{q}_t \equiv \int_i \hat{q}_{i,t}^{*} di$ and integrate Eq. (A.3) on both sides to get[1]

$$\begin{aligned}
\hat{q}_t &= (1-\omega\delta)\widehat{mc}_t + \omega\delta\int_i E_t^i(\hat{q}_{i,t+1}^{*} + \pi_{t+1})di \\
&= (1-\omega\delta)\widehat{mc}_t + \omega\delta\int_i E_t^i(\hat{q}_{i,t+1}^{*} - \hat{q}_{t+1} + \hat{q}_{t+1} + \pi_{t+1})di \\
&= (1-\omega\delta)\widehat{mc}_t + \omega\delta\bar{E}_t(\hat{q}_{t+1} + \pi_{t+1}) + \omega\delta\int_i (E_t^i\hat{q}_{i,t+1}^{*} - E_t^i\hat{q}_{t+1})di, \qquad \text{(A.5)}
\end{aligned}$$

where $\bar{E}_t = \int_i E_t^i$ denotes the average expectations of firms. Recall from Eq. (A.4) that $\hat{q}_t = \omega/(1-\omega)\pi_t$ and substitute in Eq. (A.5) to get

$$\pi_t = \delta\bar{E}_t\pi_{t+1} + \gamma\widehat{mc}_t + \xi_t,$$

[1] For the aggregate price level, the Jensen's inequality is neglected since we consider, as standard in the literature, first-order approximation in log deviations around a steady state in which the inflation rate is zero. When the considered steady state is non-zero, even if the Jensen's inequality terms are neglected, the first-order approximation typically needs to be modified. This topic has been analyzed at length by e.g., Ascari (2004). However, for moderate inflation rates, the modification turns out not to be that important quantitatively (see, e.g., Coibion et al. (2012)).

where $\gamma \equiv (1-\omega)(1-\delta\omega)\omega^{-1}$ and $\xi_t = (1-\omega)\delta \int_i (E_t^i \hat{q}^*_{i,t+1} - E_t^i \hat{q}_{t+1}) di$, which corresponds to the NKPC in Eq. (4).

We now turn to the case considered in Section 2.3 in which there are two types of predictors, namely fundamental (f) and naive (n), indexed by $i \in [f, n]$. Since each firm hires labor from the same integrated economy-wide labor market, the pricing decision is homogeneous within type, so we now denote as $P^*_{i,t}$ the optimal price set by a firm of type i, and as $P^*_{i,t}(j)$ the optimal price set by a firm of type i, producing good j, in period t. Given Assumption 1, and denoting by $n_{f,t}$ the fraction of fundamentalist firms in period t, we can write the aggregate price as

$$P_t^{1-\eta} = (1-\omega)\left(\int_0^{n_{f,t}} P^*_{f,t}(j)^{(1-\eta)} dj + \int_{n_{f,t}}^1 P^*_{n,t}(j)^{(1-\eta)} dj\right) + \omega P_{t-1}^{1-\eta}$$
$$P_t^{1-\eta} = (1-\omega)(n_{f,t} P_{f,t}^{*(1-\eta)} + (1-n_{f,t}) P_{n,t}^{*(1-\eta)}) + \omega P_{t-1}^{1-\eta},$$

or equivalently

$$1 = (1-\omega)(n_{f,t} q_{f,t}^{*(1-\eta)} + (1-n_{f,t}) q_{n,t}^{*(1-\eta)}) + \omega \left(\frac{P_{t-1}}{P_t}\right)^{(1-\eta)}.$$

Log-linearizing the expression above yields

$$\pi_t = \frac{1-\omega}{\omega}(n_{f,t} \hat{q}^*_{f,t} + (1-n_{f,t}) \hat{q}^*_{n,t}). \tag{A.6}$$

Aggregating Eq. (A.3) and substituting in Eq. (A.6) leads to the following aggregate supply relationship

$$\pi_t = \delta(n_{f,t} E_t^f \pi_{t+1} + (1-n_{f,t}) E_t^n \pi_{t+1}) + \gamma \widehat{mc}_t + \xi_t \; . \tag{A.7}$$

Eq. (A.7), augmented with a composite error u_t, which includes the term ξ_t as well as potential errors due to measurement or other sources, corresponds to Eq. (12) in Section 2.3.

B Data sources

Below we describe the data sources and the data definitions used in the paper.

Inflation is constructed using the quarterly Price Indexes for GDP from the September 2015 release of the NIPA Table 1.1.4, 1968:Q4 - 2015:Q2, which can be downloaded at www.bea.gov.

Output gap is constructed using the quarterly real GDP from the September 2015 release of the NIPA Table 1.1.3, 1968:Q4 - 2015:Q2, which can be downloaded at www.bea.gov. To construct our measure of the output gap we take logs and quadratically detrend.

Unit labor costs are constructed using the Bureau of Labor Statistics quarterly Unit Labor Costs series PRS85006113, 1968:Q4 - 2015:Q2, for the nonfarm business sector. The series can be downloaded at https://fred.stlouisfed.org. To construct our measure we take logs.

Labor share of income is constructed using the Bureau of Labor Statistics quarterly Labor Share Income series PRS85006173, 1968:Q4 - 2015:Q2, for the nonfarm business sector. The series can be downloaded at https://fred.stlouisfed.org.

Consumption-output ratio is constructed using the quarterly real GDP from the September 2015 release of the NIPA Table 1.1.3, 1968:Q4 - 2015:Q2, which can be downloaded at www.bea.gov. To construct our measure of the consumption-output ratio we take logs and linearly detrend.

C Diagnostic checks

Here we report diagnostic checks on the residuals of the NLS estimation of model (12). Fig. C.1 reports the time series of the residuals.

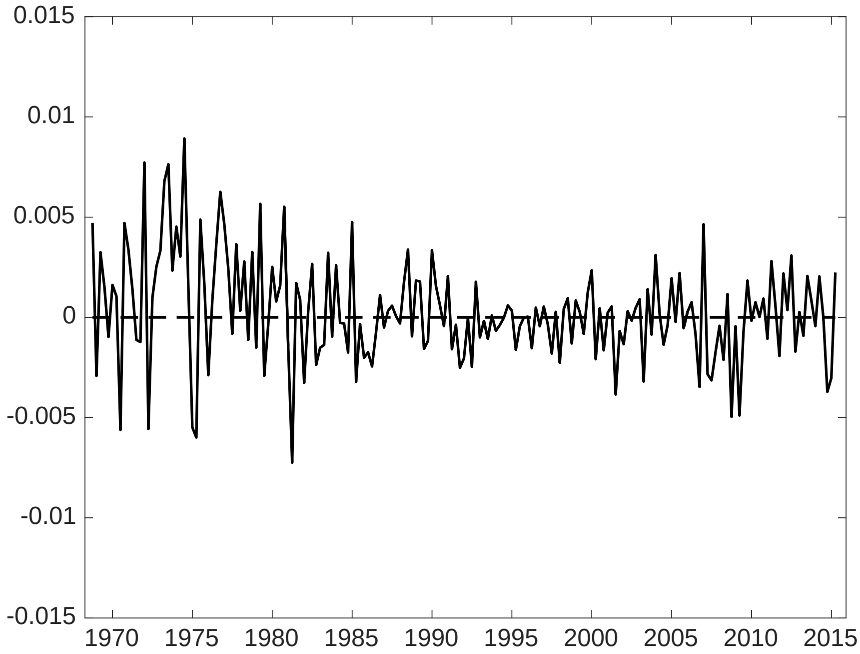

Figure C.1: Time series of residuals u_t

The result of Engle's heteroskedasticity test, reported in Table C.1, reveals the presence of heteroskedasticity up to lag 20 in the squared residuals. Standard errors in Table 1 are thus computed using White's heteroskedasticity-consistent covariance matrix estimator (HCCME).

Table C.1: Heteroskedasticity test

Null hypothesis: homoskedasticity			
F-statistic (20 lags):	48.03	p-value:	0.0004

Given the presence of heteroskedasticity, we perform the heteroskedasticity-robust F test for serial autocorrelation proposed by Davidson and MacKinnon (2009), p. 284. The results reported in Table C.2 show the absence of serial correlation in the residuals up to lag 20.

Table C.2: Serial correlation test

Null hypothesis: no serial correlation			
F-statistic (20 lags):	18.87	p-value:	0.530

D Adaptive learning with constant gain

In what follows we compare our benchmark model with an alternative model in which a representative agent updates her beliefs about future inflation using a constant gain algorithm. We consider three different perceived law of motions (PLM) used by the agent to forecast future inflation. The first PLM is given by

$$\pi_t = \phi_{0,t-1} + \phi_{1,t-1} mc_t + \epsilon_t, \tag{D.1}$$

and it corresponds to the MSV solution under rational expectations. The second PLM assumes a simple univariate AR(1) model for inflation dynamics and it takes the form

$$\pi_t = \phi_{0,t-1} + \phi_{1,t-1} \pi_{t-1} + \epsilon_t, \tag{D.2}$$

while the third PLM combines Eq. (D.1) and (D.2) and takes the form

$$\pi_t = \phi_{0,t-1} + \phi_{1,t-1} \pi_{t-1} + \phi_{2,t-1} mc_t + \epsilon_t. \tag{D.3}$$

As new data become available, agents update their estimates according to a constant gain algorithm

$$\begin{aligned} \hat{\phi}_t &= \hat{\phi}_{t-1} + g R_{t-1}^{-1} X_t (\pi_t - X_t' \hat{\phi}_{t-1}) \\ R_t &= R_{t-1} + g (X_{t-1} X_{t-1}' - R_{t-1}), \end{aligned} \tag{D.4}$$

where the first equation describes the recursive updating of the forecasting rule coefficients $\hat{\phi}_t = (\phi_{0,t}, \phi_{1,t})'$ in the case of Eqs. (D.1)–(D.2), and $\hat{\phi}_t = (\phi_{0,t}, \phi_{1,t}, \phi_{2,t})'$ in the case of Eq. (D.3), while the second equation describes the evolution of R_t, i.e., the matrix of second moments of the stacked regressors X_t. Parameter g denotes the constant gain. Given the PLMs in Eqs. (D.1)–(D.3), agents form expectations in the three different scenarios respectively as

$$E_t^g \pi_{t+1} = \phi_{0,t-1} + \phi_{1,t-1} mc_t \tag{D.5}$$

$$E_t^g \pi_{t+1} = \phi_{0,t-1}(1 + \phi_{1,t-1}) + \phi_{1,t-1}^2 \pi_{t-1} \tag{D.6}$$

$$E_t^g \pi_{t+1} = \phi_{0,t-1}(1 + \phi_{1,t-1}) + \phi_{1,t-1}^2 \pi_{t-1} + \phi_{2,t-1}(1 + \phi_{1,t-1}) mc_t, \tag{D.7}$$

where E^g denotes forecasts under constant gain learning. The actual law of motion (ALM) is then given by

$$\pi_t = \delta E_t^g \pi_{t+1} + \gamma mc_t + u_t. \tag{D.8}$$

In order to estimate Eq. (D.8) for a given PLM we follow the empirical strategy of Milani (2005). First, we estimate the agent's forecasts implied by the PLMs, i.e., Eqs. (D.5)

– (D.7), where the forecasting coefficients are updated over time through constant gain learning as described in Eq. (D.4). Initial estimates of coefficients ϕ_0 and regressors' second moments R_0 are obtained using pre-sample data as in Milani (2005) and Carceles-Poveda and Giannitsarou (2007). We fix the constant gain coefficient $g = 0.02$, which corresponds to the best-fitting constant gain coefficient estimated by Milani (2005) in a similar exercise, and which is also similar to the coefficient reported in Orphanides and Williams (2005) using survey data on expectations. We then substitute the resulting forecasts in Eq. (D.8) and estimate it using NLS.

Finally, we test our benchmark model against the three alternative models outlined above. Denoting our benchmark switching model as HSM and the three alternative models corresponding to the different PLM specifications in Eqs. (D.1) – (D.3) respectively as $\tilde{\mathrm{M}}_1$, $\tilde{\mathrm{M}}_2$, $\tilde{\mathrm{M}}_3$ following the procedure described in Section 3.3, Table D.1 reports the results of the paired nonnested hypotheses tests, while Table D.2 shows the results of the joint nonnested hypothesis test.

Table D.1: Paired nonnested hypotheses tests and Bayesian Information Criteria

	HSM	$\tilde{\mathrm{M}}_1$	$\tilde{\mathrm{M}}_2$	$\tilde{\mathrm{M}}_3$
HSM vs. $\tilde{\mathrm{M}}_c$	–	0.974	0.219	0.089
$\tilde{\mathrm{M}}_c$ vs. HSM	–	0.000	0.000	0.000
BIC	**-1690**	−1436	−1654	−1653

Notes: The first row reports p-values from P tests of HSM model against each model $\tilde{\mathrm{M}}_c$ ($c \in \{1, 2, 3\}$). The second row reports p-values from P tests of each model $\tilde{\mathrm{M}}_c$ ($c \in \{1, 2, 3\}$) against HSM model. The third row reports the BIC for all models (best model shown in bold).

Table D.2: Joint nonnested hypotheses test

F-statistic:	7.387	p-value $\chi^2(3)$:	0.061

Notes: The null hypothesis is the joint insignificance of alternative models $\tilde{\mathrm{M}}_1$, $\tilde{\mathrm{M}}_2$, and $\tilde{\mathrm{M}}_3$ against HSM.

Consistently with the results in Section 3.3, the first row of Table D.1 shows that we never reject, with 95% confidence level, our baseline switching model (HSM) when tested against all alternative models ($\tilde{\mathrm{M}}_1$, $\tilde{\mathrm{M}}_2$, and $\tilde{\mathrm{M}}_3$). The second row of Table D.1 shows instead that we reject each of the alternative models when tested against the switching model. The results on the joint significance of the alternative models against the benchmark confirm the previous findings. Finally, the third row of Table D.1 reports the Bayesian Information Criterion for model selection. Consistently with the results of the nonnested hypotheses tests, the BIC selects the baseline switching model as the best model against the alternatives.

E Alternative forecasting rules

We first consider alternative BVARs in the estimation of the fundamentalist forecast. We focus the following priors:

- The Sims-Zha Normal-Wishart prior (P_1): a Normal prior for $A|\Sigma_\epsilon$, where Σ_ϵ is the variance of the term ϵ_t in Eq. (9), and a Wishart prior for Σ_ϵ^{-1};
- The Litterman/Minnesota prior (P_2): a Normal prior for $A|\Sigma_\epsilon$ with fixed Σ_ϵ;
- The Normal-Wishart prior (P_3): a Normal prior for $A|\Sigma_\epsilon$ and a Wishart prior for Σ_ϵ^{-1}.

The Sims-Zha Normal Wishart prior is widely used in the literature; see among others Sims and Zha (1998), Giannone et al. (2015), Carriero et al. (2015). The prior for $A \mid \Sigma_\epsilon$ is a normal distribution with expectation set to zero ($\mu_1 = 0$). The covariance prior specification of $A \mid \Sigma_\epsilon$ is determined by three parameters: λ_0, the overall tightness of beliefs on the errors ϵ_t; λ_1, the standard deviation around the matrix of the coefficients on the lagged variables; λ_3, the lag decay parameter. As the lag length increases, the coefficients are increasingly shrunk towards zero. As in Kadiyala and Karlsson (1997) and in the baseline specification of Carriero et al. (2015) we consider $\lambda_3 = 1$ (linear decay). The overall tightness parameter λ_0 in our specification is 0.2 as in Carriero et al. (2015) (as $\lambda_0 \to 0$ the prior is imposed exactly, as $\lambda_0 \to \infty$ the prior information does not influence the estimates). The parameter λ_1 implements additional shrinkage on the lags of the other variables than for the lags of the dependent variables. As in Carriero et al. (2015) we set $\lambda_1 = 1$ implying that no cross-variable shrinkage takes place. We follow common practice and use in the prior covariance of A the estimated covariance residuals based on univariate autoregressive models. Finally, the Sims-Zha Normal Wishart prior is also described by two more parameters μ_5 (the sum of coefficients dummy) and μ_6 (the initial observation dummy). The sum of coefficients dummy reflects the belief that when the average of lagged variables is at some level, the same level is likely to be a good forecast of future observations (this type of belief is implemented by introducing dummy observations in the specification of the model). The initial observation dummy parameter captures the fact that all values of all the variables are set equal to the corresponding average conditions up to the scaling factor μ_6. As $\mu_6 \to \infty$, the model allows for common trends. Following Carriero et al. (2015) and the references therein we consider $\mu_5 = 1$ and $\mu_6 = 1$. For a more detailed description of the Sims-Zha priors see Carriero et al. (2015) or Eviews Help (where our estimations were performed) regarding the Bayesian VAR estimation.

The Litterman/Minnesota prior, introduced by Litterman (1986) and Doan et al. (1984), is based on the assumption that Σ_ϵ is known, and Σ_ϵ is replaced by an estimate $\hat{\Sigma}_\epsilon$. We follow common practice and use univariate autoregressive models to obtain the estimate of Σ_ϵ. The prior for $A \mid \Sigma_\epsilon$ is a normal distribution with expectation set to zero ($\mu_1 = 0$). The covariance prior specification of $A \mid \Sigma_\epsilon$ is determined by three parameters: λ_1, the overall tightness; λ_2, the relative cross-variable weight; λ_3, the lag decay.

The third prior we consider is the Normal Wishart prior which sets as prior distribution of $A \mid \Sigma_\epsilon$ the normal distribution with mean zero ($\mu_1 = 0$) and variance that depends on Σ_ϵ (for which the prior is the inverted Wishart distribution) and the parameter λ_1.

Details on the hyper-parameters considered in the different prior specifications are summarized in Table E.1. The estimation results are reported in Table E.2.

Table E.1: Prior specifications

	Hyper-parameters						
Prior	λ_0	λ_1	λ_2	λ_3	μ_1	μ_5	μ_6
P_1: Sims-Zha Normal Wishart	0.2	1	-	1	0	1	1
P_2: Litterman/Minnesota	-	0.1	1	1	0	-	-
P_3: Normal Wishart	-	5	-	-	0	-	-

Notes: For the Sims-Zha Normal Wishart prior (labeled P_1): λ_0, the overall tightness of beliefs on the errors ϵ_t; λ_1, the standard deviation around the matrix of the coefficients on the lagged variables; λ_3, the lag decay parameter; μ_5, the sum of coefficients dummy; μ_6, the initial observation dummy. For the Litterman/Minnesota prior (labeled P_2): λ_1, the overall tightness; λ_2, the relative cross-variable weight; λ_3, the lag decay. For the conjugate Normal Wishart prior (labeled P_3): μ_1, the prior mean parameter; λ_1, the prior mean covariance parameter.

Table E.2: Estimation results using alternative BVARs for labor share of income

	$Y_t = [lsi_t, y_t]'$ $(1-3)$			$Y_t = [lsi_t, y_t, \pi_{t-1}, c_t/y_t]'$ $(4-6)$		
	(1)	(2)	(3)	(4)	(5)	(6)
β	9.095**	5.581***	4.004***	4.694***	4.821***	3.988***
	(4.571)	(1.760)	(1.097)	(1.409)	(1.466)	(1.095)
γ	0.001***	0.001***	0.017***	0.001***	0.004***	0.017***
	(0.0002)	(0.0001)	(0.005)	(0.001)	(0.001)	(0.005)
	R^2 from Inflation Equation					
	0.77	0.81	0.82	0.82	0.82	0.82

Notes: $lsi_t \equiv$ labor share, $y_t \equiv$ output gap, $\pi_{t-1} \equiv$ (past) inflation, $c_t/y_t \equiv$ detrended consumption-output ratio. Lag length (l_i) in BVAR specifications: $l_i = 4$ for $i = 1, \ldots, 6$. Standard errors are computed using White's heteroskedasticity-consistent covariance matrix estimator (HCCME). *, **, *** denote significance at the 10%, 5%, and 1% level. The BVARs in (1) and (4) use prior P_1, those in (2) and (5) use prior P_2, and those in (3) and (6) use prior P_3.

Aside from the hyper-parameter values mentioned in Table E.1, we estimated the BVARs from Table E.2 using the priors P_1 and P_2 with larger λ_0 and λ_1 respectively. For the first BVAR with two variables (lsi_t, y_t) using prior P_1 with a larger λ_0 makes the estimate of β significant at 1%. Using the prior P_2, a larger λ_1 makes the estimate of γ insignificant, but the estimate of β remains significant at 1%. The estimation results for the BVAR with four variables are essentially the same as in Table E.2. Moreover, varying the variance of the Normal-Wishart prior (λ_1) for the BVARs in Table E.2 does not change our conclusions.

Table E.3 reports the estimation results when the alternative naive forecasting rule

$$E_t^n \pi_{t+1} = \frac{1}{4} \sum_{k=1}^{4} \pi_{t-k}$$

is implemented, for the different VAR specifications considered in Table 4.

Table E.3: Estimation results using alternative naive forecasting rule

	VAR specification					
	Recursive $\hat{A}_{(t)}$ (1–3)			Full sample $\hat{A}_{(T)}$ (4–6)		
	(1)	(2)	(3)	(4)	(5)	(6)
	$\begin{bmatrix} lsi_t \\ y_t \end{bmatrix}$	$\begin{bmatrix} lsi_t \\ y_t \\ \pi_{t-1} \end{bmatrix}$	$\begin{bmatrix} lsi_t \\ y_t \\ \pi_{t-1} \\ c_t/y_t \end{bmatrix}$	$\begin{bmatrix} lsi_t \\ y_t \end{bmatrix}$	$\begin{bmatrix} lsi_t \\ y_t \\ \pi_{t-1} \end{bmatrix}$	$\begin{bmatrix} lsi_t \\ y_t \\ \pi_{t-1} \\ c_t/y_t \end{bmatrix}$
β	4.312***	4.227***	4.333***	4.628***	4.552***	4.643***
	(1.520)	(1.416)	(1.474)	(1.715)	(1.541)	(1.649)
γ	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	R^2 from Inflation Equation					
	0.76	0.76	0.76	0.77	0.77	0.77

Notes: $lsi_t \equiv$ labor share, $y_t \equiv$ output gap, $\pi_{t-1} \equiv$ (past) inflation, $c_t/y_t \equiv$ detrended consumption-output ratio. Lag length (l_i) in VAR specifications ($i = 1, \ldots, 3$): $l_1 = 4$, $l_2 = l_3 = 2$. Standard errors are computed using White's heteroskedasticity-consistent covariance matrix estimator (HCCME). *, **, *** denote significance at the 10%, 5%, and 1% level. Columns 1–3 report estimates for coefficient matrix $\hat{A}_{(t)}$ updated in every period, while columns 4–6 report estimates for coefficient matrix $\hat{A}_{(T)}$, estimated over the full sample.

F Alternative measure of marginal costs

Our benchmark model considers the labor share of income as the driving variable in the NKPC. The motivation for this measure stems from the fact that the micro-foundations underpinning the NKPC imply that the correct driving variable for inflation is actually the real marginal cost. Using average unit labor costs (nominal compensation divided by real output) as a proxy for nominal marginal cost results in the labor share of income (nominal compensation divided by nominal output) as a proxy for real marginal cost.

In this section we consider a traditional output gap measure, defined as the log of real GDP from a quadratic trend, as the measure of inflationary pressure. Estimation results are reported in Tables F.1 – F.2, respectively for the benchmark naive forecasting rule $E_t^n \pi_{t+1} = \pi_{t-1}$ and the alternative naive forecasting rule $E_t^n \pi_{t+1} = \frac{1}{4}\sum_{k=1}^{4} \pi_{t-k}$.

Table F.1: Estimation results using alternative VARs for output gap

	VAR specification					
	Recursive $\hat{A}_{(t)}$ (1–3)			Full sample $\hat{A}_{(T)}$ (4–6)		
	(1)	(2)	(3)	(4)	(5)	(6)
	$\begin{bmatrix} y_t \\ c_t/y_t \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \\ \pi_{t-1} \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \\ \pi_{t-1} \\ lsi_t \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \\ \pi_{t-1} \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \\ \pi_{t-1} \\ lsi_t \end{bmatrix}$
β	5.883***	5.512***	5.187***	5.134***	5.163***	5.169***
	(1.827)	(1.480)	(1.635)	(1.511)	(1.550)	(1.559)
γ	0.008***	0.008***	−0.001	−0.0001	−0.0004	−0.0004
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
	R^2 from Inflation Equation					
	0.80	0.80	0.81	0.81	0.81	0.81

Notes: $y_t \equiv$ output gap, $lsi_t \equiv$ labor share, $\pi_{t-1} \equiv$ (past) inflation, $c_t/y_t \equiv$ detrended consumption-output ratio. Optimal lag length (l_i) in VAR specifications ($i = 1, \ldots, 3$): $l_1 = 4$, $l_2 = l_3 = 2$. Standard errors are computed using White's heteroskedasticity-consistent covariance matrix estimator (HCCME). *, **, *** denote significance at the 10%, 5%, and 1% level. Columns 1–3 report estimates for coefficient matrix $\hat{A}_{(t)}$ updated in every period, while columns 4–6 report estimates for coefficient matrix $\hat{A}_{(T)}$, estimated over the full sample.

Although the estimation results support the behavioral switching mechanism as all estimated β coefficients are positive and statistically significant, not all γ coefficients are statistically significant. This finding is in line with previous empirical work on the NKPC estimation. It has been quite difficult to obtain parameter estimates with the correct sign and of a plausible magnitude when the output gap is used as driving variable for the inflation process. Fuhrer and Moore (1995) and Galí and Gertler (1999), for example, find a negative and insignificant estimate of γ when the output gap is used as a proxy for real marginal costs. This can be due to the fact that, according to the micro-founded NKPC, real marginal cost is the relevant variable for inflationary pressure, while some theoretical restrictions are then required for real marginal costs to move with the output gap.

Table F.2: Estimation results using alternative naive forecasting rule

	VAR specification					
	Recursive $\hat{A}_{(t)}$ (1–3)			Full sample $\hat{A}_{(T)}$ (4–6)		
	(1)	(2)	(3)	(4)	(5)	(6)
	$\begin{bmatrix} y_t \\ c_t/y_t \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \\ \pi_{t-1} \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \\ \pi_{t-1} \\ lsi_t \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \\ \pi_{t-1} \end{bmatrix}$	$\begin{bmatrix} y_t \\ c_t/y_t \\ \pi_{t-1} \\ lsi_t \end{bmatrix}$
β	3.805***	3.499***	3.659***	4.199***	3.401***	3.294***
	(1.227)	(1.123)	(1.221)	(1.381)	(1.102)	(1.068)
γ	0.002***	0.0003	−0.001	0.002***	−0.0001	−0.0002
	(0.0005)	(0.0005)	(0.0003)	(0.0005)	(0.0003)	(0.0003)
	R^2 from Inflation Equation					
	0.72	0.71	0.73	0.73	0.71	0.71

Notes: $y_t \equiv$ output gap, $lsi_t \equiv$ labor share, $\pi_{t-1} \equiv$ (past) inflation, $c_t/y_t \equiv$ detrended consumption-output ratio. Optimal lag length (l_i) in VAR specifications ($i = 1 \ldots 3$): $l_1 = 4$, $l_2 = l_3 = 2$. Standard errors are computed using White's heteroskedasticity-consistent covariance matrix estimator (HCCME). *, **, *** denote significance at the 10%, 5%, and 1% level. Columns 1–3 report estimates for coefficient matrix $\hat{A}_{(t)}$ updated in every period, while columns 4–6 report estimates for coefficient matrix $\hat{A}_{(T)}$, estimated over the full sample.

G Out-of-sample forecast

The UC-SV model of Stock and Watson (2007) is given by

$$\begin{aligned}
\pi_t &= \tau_t + v_t, \quad v_t = \sigma_{v,t}\epsilon_{v,t}, \quad \epsilon_{v,t} \sim N(0,1),\\
\tau_t &= \tau_{t-1} + \eta_t, \quad \eta_t = \sigma_{\eta,t}\epsilon_{\eta,t}, \quad \epsilon_{\eta,t} \sim N(0,1),\\
\log \sigma^2_{v,t} &= \log \sigma^2_{v,t-1} + \xi_{v,t}, \quad \xi_{v,t} \sim N(0,\phi_v),\\
\log \sigma^2_{\eta,t} &= \log \sigma^2_{\eta,t-1} + \xi_{\eta,t}, \quad \xi_{v,t} \sim N(0,\phi_\eta),
\end{aligned}$$

where $\sigma_{v,t}$ and $\sigma_{\eta,t}$ represent the stochastic volatility terms, and τ_t represents the trend inflation.

The iterated forecasts based on the UC-SV are obtained using two sample windows: recursive and rolling. Under the recursive scheme, the estimates are updated at each forecast origin using all available information. Under the rolling scheme, the estimates are updated at each forecast origin using the same number of observations in the window. Inflation forecasts $\hat{\pi}_{t+h}$ are constructed at each forecast origin and are then used to construct forecast errors as $\hat{u}_{t+h} = \pi_{t+h} - \hat{\pi}_{t+h}$.

The UC-SV model point forecasts at any forecast horizon h are given by the median of the posterior distribution of π_t (from each retained draw in the MCMC chain). More explicitly, we sample innovations to $\log \sigma^2_{v,t+h}$ and $\log \sigma^2_{\eta,t+h}$ from a normal distribution with variance ψ_v and ψ_η respectively, and use the random walk specifications to compute $\log \sigma^2_{v,t+h}$ and $\log \sigma^2_{\eta,t+h}$ from $\log \sigma^2_{v,t+h-1}$ and $\log \sigma^2_{\eta,t+h-1}$. Then, for each period of the forecast horizon, we sample shocks $\xi_{v,t+h}$ and $\xi_{\eta,t+h}$ from $N(0,1)$ which are used with $\sigma_{v,t+h}$ and $\sigma_{\eta,t+h}$ to compute the forecast draw of π_{t+h}. To compute the UC-SV point forecasts we used the Matlab code from http://cremfi.econ.qmul.ac.uk/efp/ and the prior specification mentioned therein. More precisely, we first applied the HP filter with smoothing parameter 1600 to filter π_t on the full sample from 1947:Q2 until 2015:Q2. Denote by τ_t^{HP} the trend obtained with the HP filter. Under the recursive scheme we used a pre-sample (of size $T_0 = 86$) from 1947:Q2 until 1968:Q3 to set up starting values and priors. Denote by $\hat{v}_t$ the residuals obtained by regressing τ_t^{HP} on π_t in the pre-sample, and let $\sigma_0^2 = T_0^{-1}\sum_{t=1}^{T_0} \hat{v}_t^2$. The value of the trend inflation in 1968:Q4 that is needed as an initial value in the state space representation is $\tau_0 = \tau_t^{HP}$ where t corresponds to 1968:Q4. The values of the log noise and log trend volatility in 1968:Q4 are: $\log \sigma^2_{v,0} \sim N((\bar{\mu}_v/\bar{\sigma} + \log h_v)s_v, s_v)$, where $s_v = \bar{\sigma}/(\bar{\sigma}+1)$, $\bar{\sigma} = 10$, $\bar{\mu}_v = \log \sigma_0^2$, and $h_v = (\pi_t - \pi_{t-1})^2 + 0.0001$ where t corresponds to 1969:Q1; $\log \sigma^2_{\eta,0} \sim N((\bar{\mu}_\eta/\bar{\sigma} + \log h_\eta)s_\eta, s_\eta)$, $\bar{\mu}_\eta = \log\left(T^{-1}\sum_{t=1}^{T}\left(\tau_t^{HP}\right)^2\right)$, where T is the length of the sample from 1968:Q4 to 2015:Q2, and $t = 1$ corresponds to 1968:Q4, $h_\eta = (\tau_t^{HP} - \tau_{t-1}^{HP})^2 + 0.0001$, where t corresponds to 1969:Q1. Moreover, $\phi_v \sim IG(1, 0.01)$, $\phi_\eta \sim IG(1, 0.0001)$. For the rolling scheme, the pre-sample starts in 1947:Q2 and increases from 1968:Q3 by one quarter at a time as the forecast origin shifts towards the end of the full sample. For both schemes the first forecast one-quarter ahead is for 1979:Q1.

For the HSM baseline model, we obtain direct and iterative forecasts based on both recursive and rolling windows. The model estimation under the rolling scheme is based on a window of 10 years (as in Stock and Watson (2009)), and under the recursive scheme, the estimation starts in 1968:Q4. To obtain the direct forecasts for a forecast horizon h, we first estimate:

$$\pi_t = \delta(n_{f,t-h}E^f_{t-h}\pi_{t-h+1} + (1-n_{f,t-h})E^n_{t-h}\pi_{t-h+1}) + \gamma^h mc_{t-h} + u_t,$$

where $mc_{t-h} = lsi_{t-h}$ in the HSM baseline model,

$$\begin{aligned}
\hat{A}_{(t-h)} &= (Z'_{t-h-1}Z_{t-h-1})^{-1}Z'_{t-h-1}Z_{t-h} \\
E^f_{t-h}\pi_{t-h+1} &= \gamma^h e'_1(I-\delta\hat{A}_{(t-h)})^{-1}\hat{A}_{(t-h)}Z_{t-h} \\
E^n_{t-h}\pi_{t-h+1} &= \pi_{t-h-1} \\
n_{f,t-h} &= \frac{1}{1+\exp\left(\beta^h\left(\frac{FE^f_{t-h-1}-FE^n_{t-h-1}}{FE^f_{t-h-1}+FE^n_{t-h-1}}\right)\right)} \\
FE^i_{t-h-1} &= \sum_{k=1}^{K}|E^i_{t-h-k-1}\pi_{t-h-k}-\pi_{t-h-k}|, \qquad \text{with } \ i=f,n,\ K=4,
\end{aligned}$$

and we denote the estimates by $\hat{\beta}^h$, $\hat{\gamma}^h$.

The h periods ahead forecast is:

$$\hat{\pi}_{t+h} = \delta(\hat{n}_{f,t}E^f_t\pi_{t+1} + (1-\hat{n}_{f,t})E^n_t\pi_{t+1}) + \hat{\gamma}^h mc_t,$$

where

$$\begin{aligned}
\hat{A}_{(t)} &= (Z'_{t-1}Z_{t-1})^{-1}Z'_{t-1}Z_t \\
E^f_t\pi_{t+1} &= \hat{\gamma}^h e'_1(I-\delta\hat{A}_{(t)})^{-1}\hat{A}_{(t)}Z_t \\
E^n_t\pi_{t+1} &= \pi_{t-1} \\
\hat{n}_{f,t} &= \frac{1}{1+\exp\left(\hat{\beta}^h\left(\frac{FE^f_{t-1}-FE^n_{t-1}}{FE^f_{t-1}+FE^n_{t-1}}\right)\right)} \\
FE^i_{t-1} &= \sum_{k=1}^{K}|E^i_{t-k-1}\pi_{t-k}-\pi_{t-k}|, \qquad \text{with } \ i=f,n,\ K=4.
\end{aligned}$$

The forecast errors at time $t-1$ for the fundamentalists and the random walk believers are:

$$FE^i_{t-1} = |E^i_{t-5}\pi_{t-4}-\pi_{t-4}| + |E^i_{t-4}\pi_{t-3}-\pi_{t-3}| + |E^i_{t-3}\pi_{t-2}-\pi_{t-2}| + |E^i_{t-2}\pi_{t-1}-\pi_{t-1}|,$$

where

$$FE^n_{t-1} = |\pi_{t-6}-\pi_{t-4}| + |\pi_{t-5}-\pi_{t-3}| + |\pi_{t-4}-\pi_{t-2}| + |\pi_{t-3}-\pi_{t-1}|,$$

$$\begin{aligned}
FE^f_{t-1} &= |\hat{\gamma}^h e'_1(I-\delta\hat{A}_{(t-5)})^{-1}\hat{A}_{(t-5)}Z_{t-5}-\pi_{t-4}| + |\hat{\gamma}^h e'_1(I-\delta\hat{A}_{(t-4)})^{-1}\hat{A}_{(t-4)}Z_{t-4}-\pi_{t-3}| \\
&+ |\hat{\gamma}^h e'_1(I-\delta\hat{A}_{(t-3)})^{-1}\hat{A}_{(t-3)}Z_{t-3}-\pi_{t-2}| + |\hat{\gamma}^h e'_1(I-\delta\hat{A}_{(t-2)})^{-1}\hat{A}_{(t-2)}Z_{t-2}-\pi_{t-1}|.
\end{aligned}$$

The iterative forecasts from the HSM baseline model are based on the one-step ahead forecasts from this model. Denote by $\hat{\beta}$ and $\hat{\gamma}$ the parameter estimates based on the sample up to time t. Notice that the model delivers also an estimate of π_t, i.e.,

$$\begin{aligned}\hat{\pi}_t &= \delta(\hat{n}_{f,t} E_t^f \pi_{t+1} + (1 - n_{f,t}) E_t^n \pi_{t+1}) + \hat{\gamma} mc_t \\ &= \delta(\hat{n}_{f,t} \hat{\gamma} e_1' (I - \delta \hat{A}_{(t)})^{-1} \hat{A}_{(t)} Z_t + (1 - \hat{n}_{f,t}) \pi_{t-1}) + \hat{\gamma} mc_t,\end{aligned}$$

where $\hat{n}_{f,t} = \left(1 + \exp\left(\hat{\beta}\left(\frac{FE_{t-1}^f - FE_{t-1}^n}{FE_{t-1}^f + FE_{t-1}^n}\right)\right)\right)^{-1}$.

The one period ahead ($h = 1$) forecast of inflation, $\hat{\pi}_{t+1}$, is obtained as:

$$\begin{aligned}\hat{\pi}_{t+1} &= \delta(\hat{n}_{f,t+1} E_t^f \pi_{t+2} + (1 - n_{f,t+1}) E_t^n \pi_{t+2}) + \hat{\gamma} \widehat{mc}_{t+1} \\ &= \delta(\hat{n}_{f,t+1} \hat{\gamma} e_1' (I - \delta \hat{A}_{(t)})^{-1} \hat{A}_{(t)}^2 Z_t + (1 - \hat{n}_{f,t+1}) \hat{\pi}_t) + \hat{\gamma} \widehat{mc}_{t+1},\end{aligned}$$

where $\widehat{mc}_{t+1} = \widehat{lsi}_{t+1}$ in baseline HSM) is the first variable from $E_t Z_{t+1} = \hat{A}_{(t)} Z_t$ and $\hat{n}_{f,t+1} = \left(1 + \exp\left(\hat{\beta}\left(\frac{FE_t^f - FE_t^n}{FE_t^f + FE_t^n}\right)\right)\right)^{-1}$.

The forecast errors at time t are:

$$\begin{aligned}FE_t^n &= |\pi_{t-5} - \pi_{t-3}| + |\pi_{t-4} - \pi_{t-2}| + |\pi_{t-3} - \pi_{t-1}| + |\pi_{t-2} - \hat{\pi}_t|, \\ FE_t^f &= |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-4)})^{-1} \hat{A}_{(t-4)} Z_{t-4} - \pi_{t-3}| + |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-3)})^{-1} \hat{A}_{(t-3)} Z_{t-3} - \pi_{t-2}| \\ &+ |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-2)})^{-1} \hat{A}_{(t-2)} Z_{t-2} - \pi_{t-1}| + |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-1)})^{-1} \hat{A}_{(t-1)} Z_{t-1} - \hat{\pi}_t|.\end{aligned}$$

The two-period ahead ($h = 2$) forecast of inflation, $\hat{\pi}_{t+2}$ is obtained as:

$$\begin{aligned}\hat{\pi}_{t+2} &= \delta(\hat{n}_{f,t+2} E_t^f \pi_{t+3} + (1 - n_{f,t+2}) E_t^n \pi_{t+3}) + \hat{\gamma} \widehat{mc}_{t+2} \\ &= \delta(\hat{n}_{f,t+2} \hat{\gamma} e_1' (I - \delta \hat{A}_{(t)})^{-1} \hat{A}_{(t)}^3 Z_t + (1 - \hat{n}_{f,t+2}) \hat{\pi}_{t+1}) + \hat{\gamma} \widehat{mc}_{t+2},\end{aligned}$$

where $\widehat{mc}_{t+2}$ is the first variable from $E_t Z_{t+2} = \hat{A}_{(t)}^2 Z_t$; $\hat{n}_{f,t+2} = \left(1 + \exp\left(\hat{\beta}\left(\frac{FE_{t+1}^f - FE_{t+1}^n}{FE_{t+1}^f + FE_{t+1}^n}\right)\right)\right)^{-1}$,

$$\begin{aligned}FE_{t+1}^n &= |\pi_{t-4} - \pi_{t-2}| + |\pi_{t-3} - \pi_{t-1}| + |\pi_{t-2} - \hat{\pi}_t| + |\pi_{t-1} - \hat{\pi}_{t+1}|, \\ FE_{t+1}^f &= |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-3)})^{-1} \hat{A}_{(t-3)} Z_{t-3} - \pi_{t-2}| + |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-2)})^{-1} \hat{A}_{(t-2)} Z_{t-2} - \pi_{t-1}| \\ &+ |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-1)})^{-1} \hat{A}_{(t-1)} Z_{t-1} - \hat{\pi}_t| + |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t)})^{-1} \hat{A}_{(t)} Z_t - \hat{\pi}_{t+1}|.\end{aligned}$$

The three-period ahead ($h = 3$) forecast of inflation is:

$$\begin{aligned}\hat{\pi}_{t+3} &= \delta(\hat{n}_{f,t+3} E_t^f \pi_{t+4} + (1 - \hat{n}_{f,t+3}) E_t^n \pi_{t+4}) + \hat{\gamma} \widehat{mc}_{t+3} \\ &= \delta(\hat{n}_{f,t+3} \hat{\gamma} e_1' (I - \delta \hat{A}_{(t)})^{-1} \hat{A}_{(t)}^4 Z_t + (1 - \hat{n}_{f,t+3}) \hat{\pi}_{t+3}) + \hat{\gamma} \widehat{mc}_{t+3},\end{aligned}$$

where $\widehat{mc}_{t+3}$ is the first variable from $E_t Z_{t+3} = \hat{A}_{(t)}^3 Z_t$; $\hat{n}_{f,t+3} = \left(1 + \exp\left(\hat{\beta}\left(\frac{FE_{t+2}^f - FE_{t+2}^n}{FE_{t+2}^f + FE_{t+2}^n}\right)\right)\right)^{-1}$,

$$\begin{aligned}FE_{t+2}^n &= |\pi_{t-3} - \pi_{t-1}| + |\pi_{t-2} - \hat{\pi}_t| + |\pi_{t-1} - \hat{\pi}_{t+1}| + |\hat{\pi}_t - \hat{\pi}_{t+2}|, \\ FE_{t+2}^f &= |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-2)})^{-1} \hat{A}_{(t-2)} Z_{t-2} - \pi_{t-1}| + |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t-1)})^{-1} \hat{A}_{(t-1)} Z_{t-1} - \hat{\pi}_t| \\ &+ |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t)})^{-1} \hat{A}_{(t)} Z_t - \hat{\pi}_{t+1}| + |\hat{\gamma} e_1' (I - \delta \hat{A}_{(t)})^{-1} \hat{A}_{(t)}^2 Z_t - \hat{\pi}_{t+2}|.\end{aligned}$$

Subsequent forecasts are obtained in a similar fashion.

Table 5 in the main text, and Tables G.1-G.3 below present the ratios of the RMSE of the baseline HSM, M_1, M_2, and M_3 models relative to the RMSE of the UC-SV model for forecast horizons of 1 – 4, 8, 12 and 16 quarters, as well as the results of the Diebold-Mariano test (Diebold and Mariano, 1995).

As it can be seen from Tables G.1 – G.3, the null of equal accuracy for the HSM baseline model and the M_1 model is not rejected for any forecast horizon. On the contrary, the null of equal accuracy in finite sample is rejected at 1% and 5% for one-quarter ahead forecast and two quarters ahead forecast for the M_2 model, while for the M_3 model the null of equal accuracy is rejected at 1% at 8, 12 and 16-quarters ahead forecasts. These conclusions are similar to those from Table 5.

Table G.1: Ratios of RMSE (iterated forecasts, rolling scheme) to UC-SV model

Model	$h=1$	$h=2$	$h=3$	$h=4$	$h=8$	$h=12$	$h=16$
HSM	0.913	0.887	0.947	0.965	0.983	0.973	0.928
M_1	0.963	0.927	0.948	0.983	0.943	0.981	0.965
M_2	1.758***	1.536*	1.395	1.242	0.880	0.667	0.608
M_3	1.045	1.000	1.031	1.096	1.098	1.154**	1.208***

Notes: See the notes of Table 5.

Table G.2: Ratios of RMSE (direct forecasts, recursive scheme) to UC-SV model

Model	$h=1$	$h=2$	$h=3$	$h=4$	$h=8$	$h=12$	$h=16$
HSM	1.099	1.049	1.024	1.059	1.176	1.197	1.082
M_1	1.048	1.036	1.039	1.075	0.974	0.890	0.922
M_2	2.090***	1.916***	1.769***	1.619**	1.171	0.889	0.869
M_3	1.051	1.016	1.001	1.052	1.093	1.096	1.127*

Notes: See the notes of 5.

Table G.3: Ratios of RMSE (direct forecasts, rolling scheme) to UC-SV model

Model	$h=1$	$h=2$	$h=3$	$h=4$	$h=8$	$h=12$	$h=16$
HSM	0.941	0.918	0.945	1.007	1.051	1.197	0.958
M_1	0.963	0.927	0.948	0.983	0.943	0.981	0.965
M_2	1.811***	1.597**	1.482*	1.348	0.951	0.719	0.685
M_3	1.045	1.000	1.031	1.096	1.098	1.154**	1.208***

Notes: See the notes of Table 5.

Fig. G.1 below presents the rolling RMSE for the baseline HSM, M_1, M_2, M_3 and the UC-SV models using iterated forecasts and the recursive scheme for a forecast horizon of four quarters ($h = 4$). The rolling RMSE is computed using a weighted centered 15-quarter window (as in Stock and Watson (2009)):

$$\text{rolling RMSE}(t) = \left(\sum_{s=t-7}^{t+7} \frac{K\left(\frac{|s-t|}{8}\right)\left(\pi_{s+h} - \hat{\pi}_{s+h|s}\right)^2}{\sum_{s=t-7}^{t+7} K\left(\frac{|s-t|}{8}\right)} \right)^{1/2}, \tag{G.1}$$

where K is the biweight kernel, $K(x) = (15/16)(1-x^2)^2 1_{|x|\geq 1}$, and $\hat{\pi}_{s+h|s}$ is the pseudo out-of-sample forecast of π_{s+h} made using data until s.

Fig. G.1 panel (a) indicates that the HSM forecasts improved over the UC-SV model forecasts over most of the 1980s. Over the 1990s and 2000s the HSM baseline and UC-SV models forecasts are similar, while during the financial crisis originated in 2007/8 the UC-SV model had smaller RMSE than the HSM. The conclusions are similar for the M_1 model (Fig. G.1 panel (b)), M_2 model (Fig. G.1 panel (c)), and M_3 model (Fig. G.1 panel (d)), except that the UC-SV model forecasts outperforms these models over the early/mid-2000s. These conclusions are in general the same for other forecast horizons $h = 1, \ldots, 16$, and for the direct forecasts with rolling scheme. One exception are the iterated/direct forecasts with rolling scheme for which the HSM baseline, M_1 and M_2 models have in general a similar RMSE to the UC-SV for the years after 2010 (although not for all $h = 1, \ldots, 16$). These further results are available upon request from the authors.

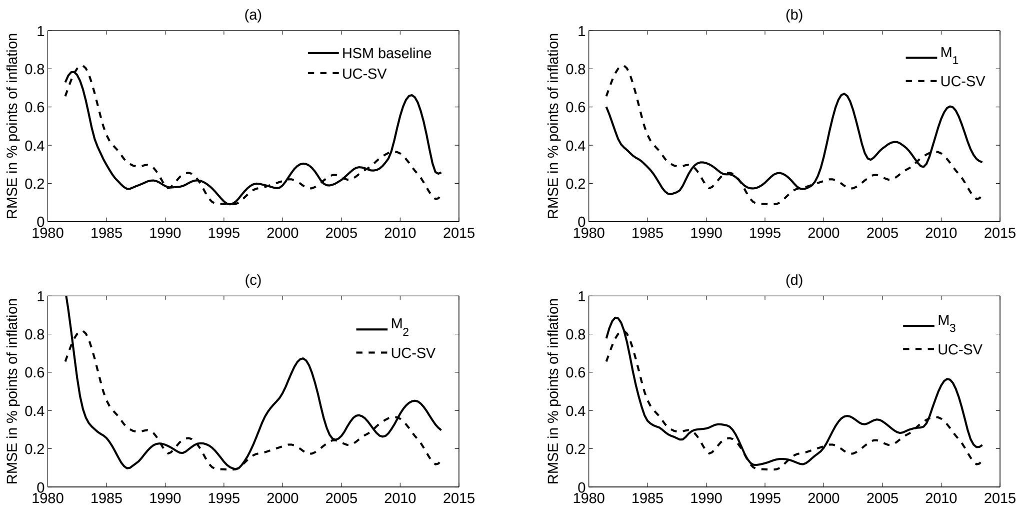

Figure G.1: Rolling Root Mean Square Errors for Inflation Forecasts, HSM baseline model (panel(a)), M_1 model (panel(b)), M_2 model (panel(c)), M_3 model (panel(d)), compared to the UC-SV model; forecast horizon $h = 4$; iterated forecasts, recursive scheme.

Fig. G.2 below presents the results of the (Giacomini and Rossi, 2010) fluctuations test for equal accuracy at each point in time of the HSM baseline model, M_1, M_2 and M_3

models relative to the benchmark UC-SV model, using iterated forecast and the recursive window scheme. The test is computed with the squared error loss metric and a forecast window size b, where b is the integer part of μT^* with T^* the out-of-sample size (for the four quarters ahead forecast ($h = 4$) case which we report below, we have $T^* = 143$). To compute the critical values of the test we take $\mu = 0.1$ (see Table 1 of Giacomini and Rossi (2010)), hence $b = 14$ quarters. For the recursive window, the critical values are computed as described on page 601 of Giacomini and Rossi (2010). As in Clark and Doh (2014) we compute the variance for the test statistic using the full sample of forecasts and the pre-whitened quadratic spectral kernel of Andrews and Monahan (1992).

Fig. G.2 panel (a) shows the fluctuations test statistic of the HSM baseline model relative to the benchmark UC-SV model. The figure reveals some instabilities in the relative forecast accuracy of the HSM model. A test statistic above the upper critical values of 5% and 10% indicates that the HSM baseline model has more accurate forecasts than the UC-SV model (this is the case for most of the 1980s). A test statistic below the lower critical values indicates that the benchmark UC-SV model has more accurate forecasts than the HSM baseline model (this is the case for the period following the 2007/8 financial crisis). A test statistic within the upper and lower critical values indicates equal accuracy (this is the case for the 1990s, 2000s and the recent years). A similar conclusion can be drawn about model M_1 relative to the benchmark UC-SV model, except that model M_1 is less accurate in the early 2000s (see Fig. G.2 panel (b)). Fig. G.2 panel (c) indicates that model M_2 is less accurate than the UC-SV model in the early 1980s and 2000s, more accurate than the UC-SV model in the late 1980, and equally accurate the rest of the time (the 1990s, late 2000s, the 2007/8 financial crisis period and recent years). Finally, Fig. G.2 panel (d) indicates that model M_3 is as accurate as the UC-SV model most of the time except mid-1980s (when it is more accurate than the UC-SV model) and the period following the 2007/8 financial crisis (when it is less accurate than the UC-SV model).

We have also considered other values of μ (hence higher b) and other values of h, and the results (available upon request) indicate the fact that there are instabilities in the relative forecast accuracies of the models along the lines mentioned above although with some minor variations. Moreover, the conclusions are the same when iterated forecasts with rolling window scheme are used (in which case the critical values from Table 1 of Giacomini and Rossi (2010) are used), or when direct forecast of the HSM baseline, M_1 – M_3 models with both recursive and rolling window scheme are used. These further results are available upon request.

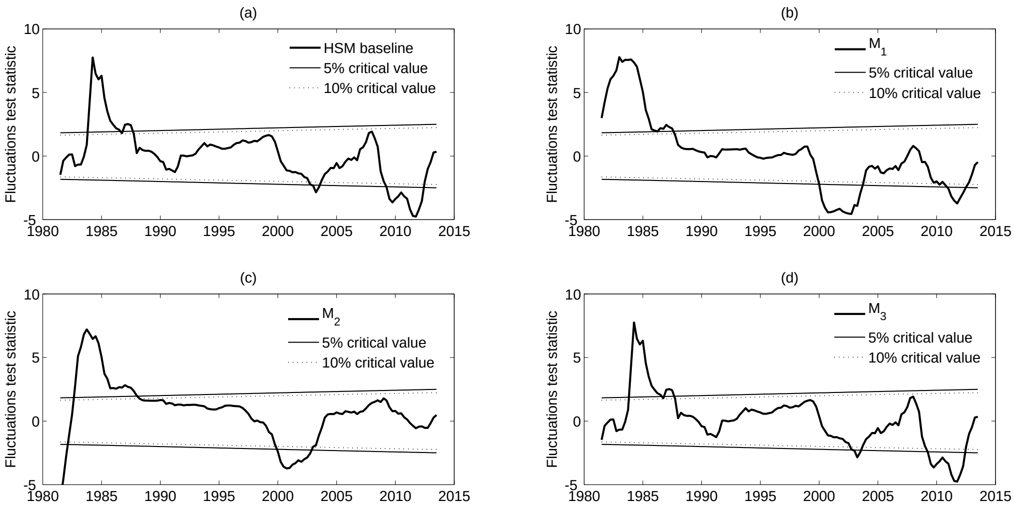

Figure G.2: Fluctuations test statistic for HSM baseline (panel (a)), M_1 (panel (b)), M_2 (panel (c)), M_3 (panel (d)) relative to the UC-SV model; forecast horizon $h = 4$; iterated forecasts, recursive scheme.

H Additional analysis and extensions

Costly fundamental predictor

In this section we extend the baseline switching model by allowing for adoption costs for the more sophisticated fundamentalist predictor. In particular, the fitness of the fundamentalist rule is given by

$$U_{f,t} = -\frac{FE_t^f}{FE_t^f + FE_t^n} - C,$$

where parameter C denotes the cost of the fundamentalists' prediction strategy. This way of introducing costs is in line with the theoretical work of Brock and Hommes (1997) and the empirical implementation of Branch (2004). We then estimate the following model

$$\begin{aligned}
\pi_t &= \delta(n_{f,t}E_t^f\pi_{t+1} + (1-n_{f,t})E_t^n\pi_{t+1}) + \gamma mc_t + u_t \\
E_t^f\pi_{t+1} &= \gamma e_1'(I - \delta\hat{A}_{(t)})^{-1}\hat{A}_{(t)}Z_t \\
E_t^n\pi_{t+1} &= \pi_{t-1} \\
n_{f,t} &= \frac{\exp\left(\beta\left(-\frac{FE_{t-1}^f}{FE_{t-1}^f+FE_{t-1}^n} - C\right)\right)}{\exp\left(\beta\left(-\frac{FE_{t-1}^f}{FE_{t-1}^f+FE_{t-1}^n} - C\right)\right) + \exp\left(\beta\left(-\frac{FE_{t-1}^n}{FE_{t-1}^f+FE_{t-1}^n}\right)\right)} \\
FE_{t-1}^i &= \sum_{k=1}^{K}|E_{t-k-1}^i\pi_{t-k} - \pi_{t-k}|, \qquad \text{with } i = f, n,
\end{aligned}$$

where the simple naive predictor is assumed to be freely available. The estimation results are reported in Table H.1. Our estimates report a positive adoption cost for the more

Table H.1: NLS estimates of baseline model with costs

Parameter	β	γ	C
Estimate	3.753^{**}	0.001^{***}	0.185
Std. error	1.875	0.0003	0.192
R^2 from Inflation Equation	0.82		

Notes: Standard errors are computed using White's heteroskedasticity-consistent covariance matrix estimator (HCCME). *, **, *** denote significance at the 10%, 5%, and 1% level.

complex fundamentalist rule, although not statistically significant. The estimated values and significance of parameters β and γ are in line with the baseline results reported in Table 1.

"Real-time" detrending

In this section we perform an additional robustness check. In Section 3 the baseline model is estimated using data which have been detrended over the full sample. We now re-estimate the baseline model assuming "real-time" detrending, i.e., variables in the fundamentalists' VAR are detrended in each period t. The results are reported in Table H.2. The results

Table H.2: NLS estimates of model (12) with real-time detrending

Parameter	β	γ
Estimate	5.756^{***}	0.005^{***}
Std. error	1.690	0.0008
R^2 from Inflation Equation	0.82	

Notes: Standard errors are computed using White's heteroskedasticity-consistent covariance matrix estimator (HCCME). *, **, *** denote significance at the 10%, 5%, and 1% level.

are broadly in line with the baseline estimation reported in Table 1.

References

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*, 953 – 966.

Ascari, G. (2004). Staggered prices and trend inflation: some nuisances. *Review of Economic Dynamics 7*(3), 642 – 667.

Branch, W. (2004). The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations. *The Economic Journal 114*(497), 592–621.

Branch, W. and B. McGough (2009). Monetary policy in a new keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control 33*, 1036–1051.

Brock, W. A. and C. H. Hommes (1997). A rational route to randomness. *Econometrica 65*(5), 1059–1095.

Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics 12*, 983–998.

Carceles-Poveda, E. and C. Giannitsarou (2007). Adaptive learning in practice. *Journal of Economic Dynamics and Control 31*(8), 2659–2697.

Carriero, A., T. E. Clark, and M. M. (2015). Bayesian vars: specification choices and forecast accuracy. *Journal of Applied Econometrics 30*, 46–73.

Clark, T. E. and T. Doh (2014). Evaluating alternative models of trend inflation. *International Journal of Forecasting 30*, 426–448.

Coibion, O., Y. Gorodnichenko, and J. Wieland (2012). The optimal inflation rate in new keynesian models: Should central banks raise their inflation targets in light of the zero lower bound? *The Review of Economic Studies 79*(4), 1371–1406.

Davidson, R. and J. G. MacKinnon (2009). *Econometric Theory and Methods.* Oxford University Press, Oxford.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253–263.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projection using realistic prior distribution. *Econometric Reviews 3*, 1–100.

Evans, G. W. and S. Honkapohja (2001). *Learning and Expectations in Macroeconomics.* Princeton University Press.

Fuhrer, J. and G. Moore (1995). Inflation persistence. *Quarterly Journal of Economics 110*, 127–59.

Galí, J. and M. Gertler (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics 44*, 195–222.

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25*, 595 – 620.

Giannone, D., M. Lenza, and P. G. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*, 436–451.

Kadiyala, K. and S. Karlsson (1997). Numerical methods for estimation and inference in bayesian var-models. *Journal of Applied Econometrics 12*, 99–132.

Kurz, M. (2012). A new keynesian model with diverse beliefs. Technical report, Stanford University.

Kurz, M., G. Piccillo, and H. Wu (2013). Modeling diverse expectations in an aggregated new keynesian model. *Journal of Economic Dynamics and Control 37*, 1403–1433.

Litterman, R. (1986). Forecasting with bayesian vector autoregressions: five years of experience. *Journal of Business and Economic Statistics 4*, 25–38.

Milani, F. (2005). Adaptive Learning and Inflation Persistence. Working papers, University of California-Irvine, Department of Economics.

Orphanides, A. and J. C. Williams (2005). The decline of activist stabilization policy: Natural rate misperceptions, learning, and expectations. *Journal of Economic Dynamics and Control 29*(11), 1927–1950.

Sims, C. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review 39*, 949–968.

Stock, J. H. and M. W. Watson (2007). Why has u.s. inflation become harder to forecast? *Journal of Money, Credit and Banking 39*, 3–33.

Stock, J. H. and M. W. Watson (2009). Phillips curve inflation forecasts. In J. Fuhrer, K. Y., L. J., and O. G. (Eds.), *Understanding Inflation and the Implication for the Monetary Policy*. Cambridge MIT press.