Online Appendix to Paper:
An Empirical Assessment of Optimal Monetary Policy in the Euro Area

Xiaoshan Chen [*]
University of Durham

Tatiana Kirsanova[†]
University of Glasgow

Campbell Leith[‡]
University of Glasgow

May 16, 2017

---

[*]Address: Durham University Business School, University of Durham, Durham, DH1 3LB; xiaoshan.chen@durham.ac.uk

[†]Address: Economics, Adam Smith Business School, Gilbert Scott Building, University of Glasgow, Glasgow G12 8QQ; e-mail tatiana.kirsanova@glasgow.ac.uk

[‡]Address: Economics, Adam Smith Business School, Gilbert Scott Building, University of Glasgow, Glasgow G12 8QQ; e-mail campbell.leith@glasgow.ac.uk

A The Complete Model

The complete system of non-linear equations describing the equilibrium are given by

$$N_t^\varphi \left(\frac{X_t}{A_t}\right)^\sigma = \frac{W_t}{A_t P_t}(1-\tau_t) \equiv w_t(1-\tau_t)$$

$$\left(\frac{X_t}{A_t}\right)^{-\sigma}\xi_t^{-\sigma} = \beta\mathbb{E}_t\left[\left(\frac{X_{t+1}}{A_{t+1}}\right)^{-\sigma}\frac{A_t}{A_{t+1}}\xi_{t+1}^{-\sigma}R_t\pi_{t+1}^{-1}\right]$$

$$N_t = \frac{Y_t}{A_t}\int_0^1\left(\frac{P_t(i)}{P_t}\right)^{-\eta}di$$

$$X_t = C_t - \theta C_{t-1}$$

$$Y_t = C_t$$

$$\tau_t W_t N_t = T_t$$

$$\frac{P_t^f}{P_t} = \frac{\eta}{\eta-1}\frac{\mathbb{E}_t\sum_{s=0}^\infty(\alpha\beta)^s\left(\frac{X_{t+s}\xi_{t+s}}{A_{t+s}}\right)^{-\sigma}mc_{t+s}\left(\frac{P_{t+s}\pi^{-s}}{P_t}\right)^\eta\frac{Y_{t+s}}{A_{t+s}}}{\mathbb{E}_t\sum_{s=0}^\infty(\alpha\beta)^s\left(\frac{X_{t+s}\xi_{t+s}}{A_{t+s}}\right)^{-\sigma}\left(\frac{P_{t+s}\pi^{-s}}{P_t}\right)^{\eta-1}\frac{Y_{t+s}}{A_{t+s}}}$$

$$mc_t = \frac{W_t}{A_t P_t}$$

$$P_t^b = P_{t-1}^*\pi_{t-1}$$

$$\ln P_{t-1}^* = (1-\zeta)\ln P_{t-1}^f + \zeta P_{t-1}^b$$

$$P_t^{1-\eta} = \alpha\left(\pi P_{t-1}\right)^{1-\eta} + (1-\alpha)\left(P_t^*\right)^{1-\eta}$$

$$\ln A_t = \ln\gamma + \ln A_{t-1} + \ln z_t$$

$$\ln z_t = \rho_z\ln z_{t-1} + \varepsilon_t^z$$

$$\ln\mu_t = \rho_\mu\ln\mu_{t-1} + \varepsilon_t^\mu$$

$$\ln\xi_t = \rho_\xi\ln\xi_{t-1} + \varepsilon_t^\xi$$

$$\ln(1-\tau_t) = \rho^\mu\ln(1-\tau_{t-1}) + (1-\rho^\mu)\ln(1-\tau) - \varepsilon_t^\mu$$

with an associated equation describing the evolution of price dispersion, $\Delta_t = \int_0^1\left(\frac{P_t(i)}{P_t}\right)^{-\eta}di$, which is not needed to tie down the equilibrium upon log-linearisation. The model is then closed with the addition of a description of monetary policy, which will either be rule based, or derived from various forms of optimal policy discussed in the main text.

In order to render this model stationary we need to scale certain variables by the non-stationary level of technology, A_t such that $k_t = K_t/A_t$ where $K_t = \{Y_t, C_t, W_t/P_t\}$. All other real variables are naturally stationary. Applying this scaling, the steady-state equilibrium conditions

reduce to:

$$
\begin{aligned}
N^{\varphi} X^{\sigma} &= w(1-\tau) \\
1 &= \beta R \pi^{-1}/\gamma = \beta r/\gamma \\
y &= N = c \\
X &= c(1-\theta) \\
\frac{\eta}{\eta-1} &= \frac{1}{w}.
\end{aligned}
$$

This system yields

$$
N^{\sigma+\varphi} (1-\theta)^{\sigma} = w(1-\tau). \tag{1}
$$

which can be solved for N. Note that this expression depends on the real wage w, which can be obtained from the steady-state pricing decision of our monopolistically competitive firms. In Appendix B we contrast this with the labour allocation that would be chosen by a social planner in order to fix the steady-state tax rate required to offset the net distortion implied by monopolistic competition and the consumption habits externality.

B   The Social Planner's Problem

The subsidy level that ensures an efficient long-run equilibrium is obtained by comparing the steady state solution of the social planner's problem with the steady state obtained in the decentralised equilibrium. The social planner ignores the nominal inertia and all other inefficiencies and chooses real allocations that maximise the representative consumer's utility subject to the aggregate resource constraint, the aggregate production function, and the law of motion for habit-adjusted consumption:

$$
\max_{\{X_t^*, C_t^*, N_t^*\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u \left(X_t^*, N_t^*, \xi_t, A_t \right)
$$
$$
\begin{aligned}
s.t. \ Y_t^* &= C_t^* \\
Y_t^* &= A_t N_t^* \\
X_t^* &= C_t^*/A_t - \theta C_{t-1}^*/A_{t-1}
\end{aligned}
$$

The optimal choice implies the following relationship between the marginal rate of substitution between labour and habit-adjusted consumption and the intertemporal marginal rate of substitution in habit-adjusted consumption

$$
(N_t^*)^{\varphi} (X_t^*)^{\sigma} = (1-\theta\beta) \mathbb{E}_t \left(\frac{X_{t+1}^* \xi_{t+1}}{X_t^* \xi_t} \right)^{-\sigma}.
$$

The steady state equivalent of this expression can be written as

$$(N^*)^{\varphi+\sigma} (1 - \theta)^{\sigma} = (1 - \theta\beta).$$

If we contrast this with the allocation achieved in the steady-state of our decentralised equilibrium given by equation (1), we can see that the two will be identical whenever the tax rate is set optimally to be

$$\tau^* \equiv 1 - \frac{\eta}{\eta - 1}(1 - \theta\beta).$$

Notice that in the absence of habits the optimal tax rate would be negative, such that it is effectively a subsidy which offsets the monopolistic competition distortion. However, for the estimated values of the habits parameter the optimal tax rate is positive as the policy maker wishes to prevent households from overconsuming.

C Derivation of Objective Function

Individual utility in period t is

$$\Gamma_0 = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{X_t^{1-\sigma} \xi_t^{-\sigma}}{1 - \sigma} - \frac{N_t^{1+\varphi} \xi_t^{-\sigma}}{1 + \varphi} \right)$$

where $X_t = c_t - \theta c_{t-1}$ is habit-adjusted aggregate consumption after adjusting consumption for the level of productivity, $c_t = C_t/A_t$.

Linearisation up to second order yields

$$\begin{aligned}
\Gamma_0 &= \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\overline{X}^{1-\sigma} \left\{ \frac{1-\theta\beta}{1-\theta} \left(\widehat{c}_t + \frac{1}{2}\widehat{c}_t^2 \right) - \frac{1}{2}\sigma\widehat{X}_t^2 - \sigma\widehat{X}_t\widehat{\xi}_t \right\} \right. \\
&\quad \left. - \overline{N}^{1+\varphi} \left\{ \widehat{N}_t + \frac{1}{2}(1+\varphi)\widehat{N}_t^2 - \sigma\widehat{N}_t\widehat{\xi}_t \right\} \right) + tip(3).
\end{aligned}$$

where where $tip(3)$ includes terms independent of policy as well as terms of third order and higher. For every variable Z_t with steady state value Z we denote $\widehat{Z}_t = \log(Z_t/Z)$.

The second order approximation to the production function yields the exact relationship $\hat{N}_t = \hat{\Delta}_t + \hat{y}_t$, where $y_t = Y_t/A_t$ and $\Delta_t = \int_0^1 \left(\frac{P_t(i)}{P_t} \right)^{-\eta} di$. We substitute $\hat{N}_t$ out and follow Eser et al. (2009) in using

$$\sum_{t=0}^{\infty} \beta^t \hat{\Delta}_t = \frac{\alpha}{1 - \alpha\beta}\hat{\Delta}_{-1} + \frac{1}{2}\sum_{t=0}^{\infty} \beta^t \frac{\alpha\eta}{(1-\beta\alpha)(1-\alpha)} \left(\widehat{\pi}_t^2 + \frac{\zeta\alpha^{-1}}{(1-\zeta)} [\widehat{\pi}_t - \widehat{\pi}_{t-1}]^2 \right)$$

3

to yield

$$\Gamma_0 = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\overline{X}^{1-\sigma} \left\{ \frac{1-\theta\beta}{1-\theta} \left(\widehat{c}_t + \frac{1}{2}\widehat{c}_t^2 \right) - \frac{1}{2}\sigma \widehat{X}_t^2 - \sigma \widehat{X}_t \widehat{\xi}_t \right\} \right. $$

$$\left. - \overline{N}^{1+\varphi} \left(\begin{matrix} \widehat{y}_t + \frac{1}{2}\frac{\alpha\eta}{(1-\beta\alpha)(1-\alpha)} \left(\widehat{\pi}_t^2 + \frac{\zeta\alpha^{-1}}{(1-\zeta)} [\widehat{\pi}_t - \widehat{\pi}_{t-1}]^2 \right) \\ + \frac{1}{2}(1+\varphi)\widehat{y}_t^2 - \sigma \widehat{y}_t \widehat{\xi}_t \end{matrix} \right) \right) + tip(3).$$

The second order approximation to the national income identity yields

$$\widehat{c}_t + \frac{1}{2}\widehat{c}_t^2 = \widehat{y}_t + \frac{1}{2}\widehat{y}_t^2 + tip(3).$$

Finally, we use that in the efficient steady-state $\overline{X}^{1-\sigma}(1-\theta\beta) = (1-\theta)\overline{N}^{1+\varphi}$ and collect terms to arrive at

$$\Gamma_0 = -\frac{1}{2}\overline{N}^{1+\varphi}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left\{ \frac{\sigma(1-\theta)}{1-\theta\beta} \left(\widehat{X}_t + \widehat{\xi}_t \right)^2 + \varphi \left(\widehat{y}_t - \frac{\sigma}{\varphi}\widehat{\xi}_t \right)^2 \right.$$

$$\left. + \frac{\alpha\eta}{(1-\beta\alpha)(1-\alpha)} \left(\widehat{\pi}_t^2 + \frac{\zeta\alpha^{-1}}{(1-\zeta)} [\widehat{\pi}_t - \widehat{\pi}_{t-1}]^2 \right) \right\} + tip(3).$$

After normalising the coefficient on inflation to one, we can write the microfounded objective function as,

$$L_{micro} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left\{ \begin{matrix} \Phi_1 \left(\widehat{X}_t + \widehat{\xi}_t \right)^2 + \Phi_2 \left(\widehat{y}_t - \frac{\sigma}{\varphi}\widehat{\xi}_t \right)^2 \\ + \left(\widehat{\pi}_t^2 + \frac{\zeta\alpha^{-1}}{(1-\zeta)} [\widehat{\pi}_t - \widehat{\pi}_{t-1}]^2 \right) \end{matrix} \right\}, \tag{2}$$

where the weights on the two real terms are functions of model structural parameters, where $\Phi_1 = \frac{\sigma(1-\theta)}{1-\theta\beta} \frac{(1-\beta\alpha)(1-\alpha)}{\alpha\eta}$ and $\Phi_2 = \frac{\varphi(1-\beta\alpha)(1-\alpha)}{\alpha\eta}$.

D   A Bayesian Learning Rate Indicator

This section applies the Bayesian learning rate indicator proposed by Koop et al. (2013) to check the degree of parameter identification under discretion, commitment, and the simple rule with Markov switching rule parameters in Table 2. This indicator does not propose a 'Yes/No' answer to the question of whether a given parameter is identified. However, it indicates the degree of identification. This indicator is developed on the basis of Bayesian asymptotic theory. As sample size increases, the role of the prior vanishes and the posterior of the parameter asymptotically converges to its true value.

The advantage of this indicator is that it can be easily applied to models with Markov-switching parameters, since it requires only a few additional steps during an ordinary Bayesian

estimation. However, applying this indicator requires prior knowledge that a subset of model parameters is known to be identifiable. Therefore, we rely on results obtained using Iskrev (2010) that the fixed parameter versions of our model closed with either a simple rule or optimal policy are identifiable.

In developing this indicator Koop et al. (2013) assume Gaussian priors to obtain analytical solution of posterior precision when the sample period reaches infinity. However, for most DSGE models, the priors are non-Gaussian. Therefore, the exact expression of posterior precision is different from those illustrated in Koop et al. (2013). In applying this indicator to a DSGE model, Caglar et al. (2011) suggest treating the Hessian at the posterior mode as the measure of posterior precision. The technical details of this indicator can be found in Koop et al. (2013). Here, we focus on how this indicator is applied to our Markov switching models.

Let $\theta = [\theta_i, \theta_u]'$ be a vector of model parameters, with the assumption that θ_i is known to be identified, while the identification of θ_u is under question. Prior to applying the Bayesian learning rate indicator, we use Iskrev (2010) to determine how we split the model parameters into θ_i and θ_u. θ_u includes parameters that are associated with Markov switching in policy, shock variances and parameters in the transition matrix. These parameters cannot be incorporated in the Iskrev (2010) test. For both commitment and discretion $\theta_u = [p_{11}, p_{22}, q_{11}, q_{22}, \sigma_{\xi(s=1,2)},$ $\sigma_{\mu(s=1,2)}, \sigma_{z(s=1,2)}, \omega_{\pi(S=2)}]$, while for the simple rule with Markov-switching rule parameters θ_u $= [p_{11}, p_{22}, q_{11}, q_{22}, \sigma_{\xi(s=1,2)}, \sigma_{\mu(s=1,2)}, \sigma_{z(s=1,2)}, \sigma_{R(s=1,2)}, \psi_{1(S=1,2)}, \psi_{2(S=1,2)}, \rho^R_{(S=1,2)}].$[1]

To implement this indicator, we simulate samples of artificial data from each models. Models with Markov-switching parameters complicate the data generating processes (DGPs). To simulate data from a Markov-switching model, we need to set the model parameters and the probabilities of each regime. We set model parameters equal to posterior means in Table 2. Unlike when using a fixed parameter model to generate datasets as discussed in Koop et al. (2013) and Caglar et al. (2011), we cannot generate a single large dataset and then take subsets of it to produce smaller samples. This is because probabilities of different sample sizes have to correspond to the estimated transition probabilities ($p_{11}, p_{22}, q_{11}, q_{22}$).

We generate data samples with $T = 100, 1000, 10000$ and 20000. In order to ensure our implementation of this indicator is as comparable as possible across models, we use the same seed for the random number generator for DGPs in each case.

Tables D1, D2 and D3 present the normalised posterior precision of parameters included in θ_u under discretion, commitment and a simple rule. As discussed in Koop et al. (2013), we observe

---

[1] We set $\omega_{\pi(S=1)} = 1$, therefore $\omega_{\pi(S=1)}$ is not included in θ_u under optimal policy.

that posterior precision need not rise monotonically with T. The posterior precision may, in fact, fall before rising depending on prior type. However, Koop et al. (2013) show that the normalised posterior precision of an unidentified parameter will shrink to zero very quickly as T increases. To make our results robust, we double our sample size from $T = 10000$, the largest sample size used in Koop et al. (2013) to $T = 20000$. It can be seen that none of the normalised posterior precision in θ_u collapse to zero when $T = 20000$. This indicates that our model parameters are reasonably well identified.

Table D1: Posterior precision divided by sample size (Discretion)

Parameters	$n = 100$	$n = 1000$	$n = 10000$	$n = 20000$
Parameters associated with the MS mechanism				
$\omega_{\pi(S=2)}$	5.246	3.280	1.355	0.733
$\sigma_{\xi(s=1)}$	2.022	3.584	2.938	2.778
$\sigma_{\xi(s=2)}$	2.969	0.959	1.628	1.505
$\sigma_{\mu(s=1)}$	7.324	8.812	5.512	7.017
$\sigma_{\mu(s=2)}$	4.447	1.151	1.768	1.815
$\sigma_{z(s=1)}$	7.628	8.525	4.567	8.475
$\sigma_{z(s=2)}$	4.480	1.210	1.704	1.645
p_{11}	8.244	1.692	1.735	1.274
p_{22}	35.245	11.901	3.209	1.804
q_{11}	19.865	2.834	4.836	5.573
q_{22}	12.903	15.956	10.427	11.448

Table D2: Posterior precision divided by sample size (Commitment)

Parameters	$n = 100$	$n = 1000$	$n = 10000$	$n = 20000$
Parameters associated with the MS mechanism				
$\omega_{\pi(S=2)}$	8.262	4.963	4.195	2.766
$\sigma_{\xi(s=1)}$	1.171	1.770	4.881	2.657
$\sigma_{\xi(s=2)}$	3.405	1.184	0.857	0.817
$\sigma_{\mu(s=1)}$	0.154	0.207	0.383	0.318
$\sigma_{\mu(s=2)}$	0.506	0.226	0.251	0.152
$\sigma_{z(s=1)}$	2.969	12.158	11.467	9.317
$\sigma_{z(s=2)}$	4.175	2.602	3.618	2.113
p_{11}	6.023	20.369	20.935	16.786
p_{22}	15.055	14.846	8.503	5.381
q_{11}	10.221	13.311	10.774	8.300
q_{22}	2.451	12.749	14.382	12.000

Table D3: Posterior precision divided by sample size (simple rule)

Parameters	$n = 100$	$n = 1000$	$n = 10000$	$n = 20000$
Parameters associated with the MS mechanism				
$\rho^R_{(S=1)}$	27.191	30.109	29.181	29.731
$\rho^R_{(S=2)}$	2.089	1.469	1.852	3.036
$\psi_{1(S=1)}$	0.538	0.573	0.409	0.504
$\psi_{1(S=2)}$	0.717	1.558	1.764	1.777
$\psi_{2(S=1)}$	2.042	0.605	0.237	0.258
$\psi_{2(S=2)}$	2.636	0.697	0.304	0.327
$\sigma_{R(s=1)}$	48.769	63.814	70.160	67.808
$\sigma_{R(s=2)}$	6.584	6.019	4.348	5.173
$\sigma_{\xi(s=1)}$	1.518	1.967	2.483	2.380
$\sigma_{\xi(s=2)}$	1.007	0.907	0.776	0.703
$\sigma_{\mu(s=1)}$	1.025	0.238	0.015	0.016
$\sigma_{\mu(s=2)}$	0.286	0.041	0.006	0.008
$\sigma_{z(s=1)}$	9.299	8.299	9.649	10.309
$\sigma_{z(s=2)}$	1.035	1.019	0.761	1.047
p_{11}	36.516	21.990	11.710	15.341
p_{22}	5.293	3.395	3.785	4.816
q_{11}	8.116	9.110	11.263	11.229
q_{22}	3.379	2.246	1.762	2.015

E Implicit Interest Rate Rule

This section outlines how, in principle, we can construct an interest rate rule underpinning discretion, and estimates that rule without imposing the cross-equation restrictions implied by discretion.

There are numerous ways of representing the policy rules implied by discretion, which will rarely be unique, even although the equilibrium implied by discretionary policy will be. To consider potential functional forms of an instrument rule we employ the following Lagrangian representation of the policy problem under discretion:

$$
\begin{aligned}
L \;=\; & \frac{1}{2}\left\{ \begin{array}{c} \omega_1\left((1-\theta)^{-1}(\widehat{y}_t - \theta\widehat{y}_{t-1}) + \widehat{\xi}_t\right)^2 + \omega_2\left(\widehat{y}_t - \frac{\sigma}{\varphi}\widehat{\xi}_t\right)^2 \\ \omega_{\pi,s_t}\left(\widehat{\pi}_t^2 + \frac{\zeta\alpha^{-1}}{(1-\zeta)}[\widehat{\pi}_t - \widehat{\pi}_{t-1}]^2\right) + \omega_3\left(\Delta\widehat{R}_t\right)^2 \end{array}\right\} + \beta\mathbb{E}_t V_{t+1} \\
& + \lambda_t^1\left[\begin{array}{c} (1-\theta)^{-1}(\widehat{y}_t - \theta\widehat{y}_{t-1}) - (1-\theta)^{-1}(\mathbb{E}_t\widehat{y}_{t+1} - \theta\widehat{y}_t) \\ + \frac{1}{\sigma}\left(\widehat{R}_t - \mathbb{E}_t\widehat{\pi}_{t+1} - \mathbb{E}_t\widehat{z}_{t+1}\right) + (1-\rho)\widehat{\xi}_t \end{array}\right] \\
& + \lambda_t^2[\widehat{\pi}_t - \chi_f\beta\mathbb{E}_t\widehat{\pi}_{t+1} - \chi_b\widehat{\pi}_{t-1} - \kappa_c(\sigma(1-\theta)^{-1}(\widehat{y}_t - \theta\widehat{y}_{t-1}) + \varphi\widehat{y}_t + \widehat{\mu}_t)],
\end{aligned}
$$

where, due to the linear-quadratic nature of our policy problem, the expectations variables are a linear function of the states which include $\widehat{\pi}_t$ and $\widehat{y}_t$, while the value function, V_t, will be quadratic in the states. The first order condition for $\widehat{R}_t$ is

$$\omega_3 \Delta \widehat{R}_t - \lambda_t^1 \left[\frac{1}{1-\theta} \frac{\partial \mathbb{E}_t \widehat{y}_{t+1}}{\partial \widehat{R}_t} + \frac{1}{\sigma} \frac{\partial \mathbb{E}_t \widehat{\pi}_{t+1}}{\partial \widehat{R}_t} - \frac{1}{\sigma} \right] - \lambda_t^2 \chi_f \beta \frac{\partial \mathbb{E}_t \widehat{\pi}_{t+1}}{\partial \widehat{R}_t} + \beta \frac{\partial \mathbb{E}_t V_{t+1}}{\partial \widehat{R}_t} = 0.$$

the first order condition for output $\widehat{y}_t$ is given by,

$$\frac{\omega_1}{1-\theta} \left(\frac{1}{1-\theta} \widehat{y}_t - \frac{\theta}{1-\theta} \widehat{y}_{t-1} + \widehat{\xi}_t \right) + \omega_2 \left(\widehat{y}_t - \frac{\sigma}{\varphi} \widehat{\xi}_t \right)$$
$$+ \lambda_t^1 \left[\frac{1+\theta}{1-\theta} - \frac{1}{1-\theta} \frac{\partial \mathbb{E}_t \widehat{y}_{t+1}}{\partial \widehat{y}_t} - \frac{1}{\sigma} \frac{\partial \mathbb{E}_t \widehat{\pi}_{t+1}}{\partial \widehat{y}_t} \right]$$
$$+ \lambda_t^2 \left[\widehat{\pi}_t - \chi_f \beta \mathbb{E}_t \widehat{\pi}_{t+1} - \chi_b \widehat{\pi}_{t-1} - \frac{\kappa_c \sigma}{1-\theta} - \kappa_c \varphi \right] + \beta \frac{\partial \mathbb{E}_t V_{t+1}}{\partial \widehat{y}_t} = 0.$$

and the first order condition for inflation, $\widehat{\pi}_t$, is

$$\omega_{\pi,s_t} \widehat{\pi}_t + \omega_{\pi,s_t} \frac{\zeta \alpha^{-1}}{(1-\zeta)} \left(\widehat{\pi}_t - \widehat{\pi}_{t-1} \right) - \lambda_t^1 \left[\frac{1}{1-\theta} \frac{\partial \mathbb{E}_t \widehat{y}_{t+1}}{\partial \widehat{\pi}_t} + \frac{1}{\sigma} \frac{\partial \mathbb{E}_t \widehat{\pi}_{t+1}}{\partial \widehat{\pi}_t} \right]$$
$$+ \lambda_t^2 \left[1 - \chi_f \beta \frac{\partial \mathbb{E}_t \widehat{\pi}_{t+1}}{\partial \widehat{\pi}_t} \right] + \beta \frac{\partial \mathbb{E}_t V_{t+1}}{\partial \widehat{\pi}_t} = 0.$$

In principle, we could use the first order conditions for $\widehat{y}_t$ and $\widehat{\pi}_t$ to eliminate the LMs, λ_t^1 and λ_t^2, from the first order condition for $\widehat{R}_t$ to get an implied instrument rule under discretion. However, to write such an instrument rule is complicated and difficult to compare informatively with the estimated simple rules.

Nevertheless, we can see that the implied instrument rule under discretion is a linear function of the following arguments:

$$R_t = f(\widehat{R}_{t-1}, \widehat{\pi}_t, \widehat{\pi}_{t-1}, \widehat{y}_t, \widehat{y}_{t-1}, \widehat{\xi}_t, \widehat{z}_t, \widehat{\mu}_t, \widehat{\xi}_{t-1}, \widehat{z}_{t-1}, \widehat{\mu}_{t-1}).$$

where the rule is a function of the contemporaneous values of all endogenous variables and all states. However, one can manipulate this further, as in Clarida et al. (1999) by substitution of either the IS curve or the NKPC, to show that the interest rate is a function of expected inflation and output, current inflation and output and all states,

$$R_t = f(\widehat{R}_{t-1}, E_t \widehat{\pi}_{t+1}, \widehat{\pi}_t, \widehat{\pi}_{t-1}, E_t \widehat{y}_{t+1}, \widehat{y}_t, \widehat{y}_{t-1}, \widehat{\xi}_t, \widehat{z}_t, \widehat{\mu}_t, \widehat{\xi}_{t-1}, \widehat{z}_{t-1}, \widehat{\mu}_{t-1}).$$

Therefore, we proceed by estimating a very general interest rate rule containing all these terms:

$$
\begin{aligned}
\widehat{R}_t &= \rho^R_{s_t}\widehat{R}_{t-1} + (1 - \rho^R_{s_t})\left(\begin{array}{c} \psi_{1,s_t}\widehat{\pi}_t + \psi_{2,s_t}\widehat{y}_t + \psi_{3,s_t}\widehat{\pi}_{t-1} + \psi_{4,s_t}\widehat{y}_{t-1} \\ +\psi_{5,s_t}E_t\widehat{\pi}_{t+1} + \psi_{6,s_t}E_t\widehat{y}_{t+1} \end{array} \right) \\
&\quad +\psi_7\widehat{z}_t + \psi_8\widehat{\mu}_t + \psi_9\widehat{\xi}_t + \psi_{10}\widehat{z}_{t-1} + \psi_{11}\widehat{\mu}_{t-1} + \psi_{12}\widehat{\xi}_{t-1} + \varepsilon^R_t,
\end{aligned}
$$

where we allow Markov-switching in parameters of lagged interest rates, expected, current and lagged output and inflation, and we also allow the interest rate to directly respond shocks. Specifically, the priors of $\rho^R_{s_t}$, ψ_{1,s_t} and ψ_{2,s_t} are the same as we reported in Table 1 in the paper, while for the priors of other parameters they are set to follow the normal distribution with wide standard deviations.

We find that the likelihood at the mode of this general rule is superior to discretion, but it is over-parameterised that discretion remains dominant in terms of marginal data density, which is the correct criterion to compare different models within the Bayesian estimation framework. Therefore, generalising the interest rate rule tends to improve the likelihood, but at the cost of increasing model complexity.

Table E1 decomposes the marginal data density (which underpins the Bayes factor comparisons of model fit) into the likelihood at the mode and the penalty associated with over-parameterisation. The results suggest that, in terms of likelihood, discretion marginally improves upon a simple rule with switches in parameters, but that the latter is penalised due to the larger number of parameters such that discretion is 'decisively' preferred to the simple rule in terms of Bayes Factors.[2] The rule with switches in the inflation target has fewer parameters and so faces a milder penalty, but the underlying likelihood is less favourable which again accounts for the relative success of discretion.

References

Caglar, E., J. S. Chadha, and K. Shibayama (2011). Bayesian Estimation of DSGE models: Is the Workhorse Model Identified? Working Paper No 1125, University of Kent.

Clarida, R. H., J. Galí, and M. Gertler (1999). The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature 37*, 1661–1707.

---

[2]Discretion requires estimation of the 4 objective function parameters (ω_1, ω_2, ω_3 and ω_π) while the simple rule contains 3 parameters (ρ^R, ψ_1 and ψ_2) which vary across regimes making 6 policy parameters in total.

Table E1: Model Comparison

Model	likelihood at mode	penalty	MDD
Discretion	-320.023	-50.283	-367.407
Rule - Parameters	-320.761	-56.887	-375.859
Rule - Target	-331.301	-48.834	-384.169
Very General Rule	-303.618	-81.727	-376.615

Note: penalty is calculated as $\frac{d}{2} \ln(2\pi) + \frac{1}{2} \ln(|V|)$, where d is the dimension of the parameter vector and $|V|$ is the determinant of covariance matrix of the posterior.

Eser, F., C. Leith, and S. Wren-Lewis (2009). When is monetary policy all we need? University of Oxford Working Paper No.430.

Iskrev, N. (2010). Local identification in DSGE models. *Journal of Monetary Economics 52*(7), 189–202.

Koop, G., M. H. Pesaran, and R. P. Smith (2013). On Identification of Bayesian DSGE Models. *Journal of Business and Economic Statistics 31*(3), 300–314.