ON THE STABILITY OF CALVO-STYLE PRICE-SETTING BEHAVIOR:

TECHNICAL APPENDIX

STÉPHANE LHUISSIER AND MARGARITA ZABELINA

This document contains i) the approximation, solution and estimation of our Markov-switching DSGE model; ii) data description and transformation; iii) a detailed description of each method for computing the marginal data density. This supplement is not self-contained. We strongly advise the readers to read the main paper.

Appendix A. Markov-switching DSGE model: Solution and Estimation

A.1. **Log-linearization.** The log-deviations of the stationary variable ζ_t from its steady state value is denoted $\hat{\zeta}_t$ and defined as $\hat{\zeta}_t = log\zeta_t - log\zeta$, except for $\hat{z} \equiv z_t - \gamma$.

A log-linear approximation of the solution to the firms' price-setting problem is expressed as follows

$$\hat{\pi}_t = \frac{\beta}{1+\gamma_p\beta}\mathrm{E}_t\hat{\pi}_{t+1} + \frac{\gamma_p}{1+\gamma_p\beta}\hat{\pi}_{t-1} + \frac{(1-\theta_p\beta)(1-\theta_p)}{\theta_p(1+\gamma_p\beta)}\hat{w}_t + \hat{\theta}_t \quad (1)$$

This above equation, known as the New-Keynesian Phillips Curve (NKPC), relates the current inflation to the lagged inflation $\tilde{\pi}_{t-1}$, the expected inflation rate $\mathrm{E}_t\tilde{\pi}_{t+1}$, and the real marginal cost $\tilde{s}_t$. The last block of parameters $\kappa = \frac{(1-\theta_p\beta)(1-\theta_p)}{\theta_p(1+\gamma_p\beta)}$ is widely interpreted as the slope of the Phillips curve; i.e., a measure of nominal rigidity. It is worth noting that this slope is inversely correlated with the parameter that determines the frequency of price changes, θ_p.

The other log-linearized equilibrium conditions are as follows

$$\begin{aligned}\hat{\lambda}_t =& \frac{h\beta e^{\gamma}}{(e^{\gamma}-h\beta)(e^{\gamma}-h)}E_t\hat{y}_{t+1} - \frac{e^{\gamma 2}+h^2\beta}{(e^{\gamma}-h\beta)(e^{\gamma}-h)}\hat{y}_t + \frac{he^{\gamma}}{(e^{\gamma}-h\beta)(e^{\gamma}-h)}\hat{y}_{t-1} \\ &+ \frac{h\beta e^{\gamma}\rho_z - he^{\gamma}}{(e^{\gamma}-h\beta)(e^{\gamma}-h)}\hat{z}_t + \frac{e^{\gamma}-h\beta\rho_b}{(e^{\gamma}-h\beta)}\hat{b}_t \end{aligned} \quad (2)$$

$$\hat{\lambda}_t = \hat{R}_t + E_t\left(\hat{\lambda}_{t+1} - \hat{\pi}_{t+1}\right) - \rho_z\hat{z}_t \quad (3)$$

Date: May 11, 2015.

$$\hat{w}_t = \eta \hat{y}_t + \hat{b}_t - \hat{\lambda}_t \tag{4}$$

where (2) is the marginal utility equation with $\hat{\lambda}_t$ denoting the marginal utility of consumption; (3) is the Euler equation; and (4) is the labor supply equation. The monetary policy rule is given by

$$\tilde{R}_t = \rho_R \tilde{R}_{t-1} + (1 - \rho_R) \left[\psi_\pi (\tilde{\pi}_{4,t} - \tilde{\pi}_t^*) + \psi_y (\tilde{y}_t - \tilde{y}_t^*) \right] + \varepsilon_{R,t} \tag{5}$$

where $\tilde{y}_t^*$ denotes the output of the economy with flexible prices. The equations (1), (2), (3), (4), and (5) describe the evolution of the economy conditional on the stochastic processes for the shocks $\hat{x}_t = \rho_x \hat{x}_t + \varepsilon_{x,t}$, with $x \in \{z, \theta, b, \pi^*\}$. The stochastic process for monetary policy has already been specified in (5).

A.2. **Solution method.** The solution proposed by Cho (2011) exploits the idea of the forward method for solving MS-DSGE models whereas the method by Farmer, Waggoner, and Zha (2009) and Farmer, Waggoner, and Zha (2011) exploit Newton's method to find all possible Minimum State Variable (MSV) solutions. When the model is determinate, both methods return the same solution. Using the algorithm solution of Farmer, Waggoner, and Zha (2011), we obtain the solution of the Markov-switching rational expectations model in the following way

$$f_t = V(s_t) F_1(s_t) f_{t-1} + V(s_t) G_1(s_t) \varepsilon_t \tag{6}$$

where

$$\begin{bmatrix} A(i)V(i) & \Pi \end{bmatrix} \begin{bmatrix} F_1(i) \\ F_2(i) \end{bmatrix} = B(i) \qquad \begin{bmatrix} A(i)V(i) & \Pi \end{bmatrix} \begin{bmatrix} G_1(i) \\ G_2(i) \end{bmatrix} = \Psi(i) \tag{7}$$

and

$$\left(\sum_{i=1}^{h} p_{i,j} F_2(i) \right) V(j) = 0 \tag{8}$$

We find an MSV equilibrium by finding the matrices V_i, then the matrices $F_{1,i}$, $F_{2,i}$, $G_{1,i}$, and $G_{2,i}$. If equation (8) is satisfied, we obtain a MSV equilibrium.

A.3. **Constructing the posterior distribution.** To form the posterior density, denoted $p(\theta|Y_T)$, we combine the overall likelihood function $p(Y_T|\theta)$ with the prior $p(\theta)$

$$p(\theta|Y_T) \propto p(Y_T|\theta) p(\theta) \tag{9}$$

where θ contains all the parameters. The evaluation of the overall likelihood function is obtained using the Kim and Nelson (1999) filter, which is a combination of the Kalman filter and the Hamilton (1989) filter. Let $p(y_t|s_t, s_{t-1}, \psi_{t-1}, \theta)$ the conditional likelihood function given s_t, s_{t-1} and the past information ψ_{t-1}. By integrating s_t and s_{t-1} out, the likelihood function at date] t is as follows

$$p(y_t|\psi_{t-1}, \theta) = \sum_{s_t} \sum_{s_{t-1}} p(y_t|s_t, s_{t-1}, \psi_{t-1}, \theta) \Pr[s_t, s_{t-1}|\psi_{t-1}] \tag{10}$$

with

$$\Pr[s_t, s_{t-1}|\psi_{t-1}] = \Pr[s_t|s_{t-1}] \Pr[s_{t-1}|\psi_{t-1}] \tag{11}$$

where $\Pr[s_t|s_{t-1}]$ is the transition probability described previously. We then update the joint probability term in the following way

$$\Pr[s_t, s_{t-1}|\psi_t] = \frac{f(y_t, s_t, s_{t-1}|\psi_{t-1})}{f(y_t|\psi_{t-1})} = \frac{f(y_t|s_t, s_{t-1}, \psi_{t-1}) \Pr(s_t, s_{t-1}|\psi_{t-1})}{f(y_t|\psi_{t-1})} \tag{12}$$

and finally obtain the probability term given the information at date t

$$\Pr[s_t|\psi_t] = \sum_{s_{t-1}} \Pr[s_t, s_{t-1}|\psi_{t-1}] \tag{13}$$

The conditional likelihood function, $p(y_t|s_t, s_{t-1}, \psi_{t-1})$, cannot be evaluated with the standard Kalman filter. If in the constant case, the updated forecasts of the unobserved state vector, β_t, and the updated mean squared error of forecast P_t depend only on the information set ψ_t, in a case with Markov switching elements, these forecasts are also conditioned on the unobserved state $s_t = j$ and $s_{t-1} = i$. It follows that, at each iteration, the number of β_t and P_t to consider increases, which makes the Kalman filter unfeasible. In each step, we then collapse these h^2 terms in order to make the evaluation feasible. This approximation allows to make inference on β_t based on information ψ_{t-1}, given only s_{t-1}. See Kim and Nelson (1999) for more details. The overall likelihood is

$$p(Y_T|\theta) = \prod_{t=1}^{T} p(y_t|\psi_{t-1}, \theta) \tag{14}$$

Once the parameters of the model are estimated, we follow Kim (1994) and Kim and Nelson (1999) and make inference on $s_T, (t = 1, ..., T)$, the smoothed probabilities, in the following way

$$\Pr[s_t = j|\psi_T] = \sum_{k=1}^{M} \Pr[s_t = j, s_{t+1} = i|\psi_T] \tag{15}$$

where

$$\Pr[s_t = j, s_{t+1} = i|\psi_T] = \frac{\Pr[s_{t+1} = i|\psi_t].\Pr[s_t = j|\psi_t].\Pr[s_{t+1} = i|s_t = j]}{\Pr[s_{t+1} = i|\psi_t]} \tag{16}$$

The advantage of such a method is that it allows us to infer the unobservable variable s_t using all the information in the sample.

Appendix B. Data

The data used for estimation includes quarterly data from the third quarter of 1954 to the second quarter of 2009. Inflation π_t is the first log-difference of the GDP deflator; the nominal interest rate R_t is the Federal Funds rate; and the output growth Δy_t is the first log-difference of real per capita GDP. This latter is obtained by dividing real GDP (GDPC96) by population (LF and LH). All data comes from the St. Louis Federal Reserve Bank database (FRED). The series are reported in Figure 1.

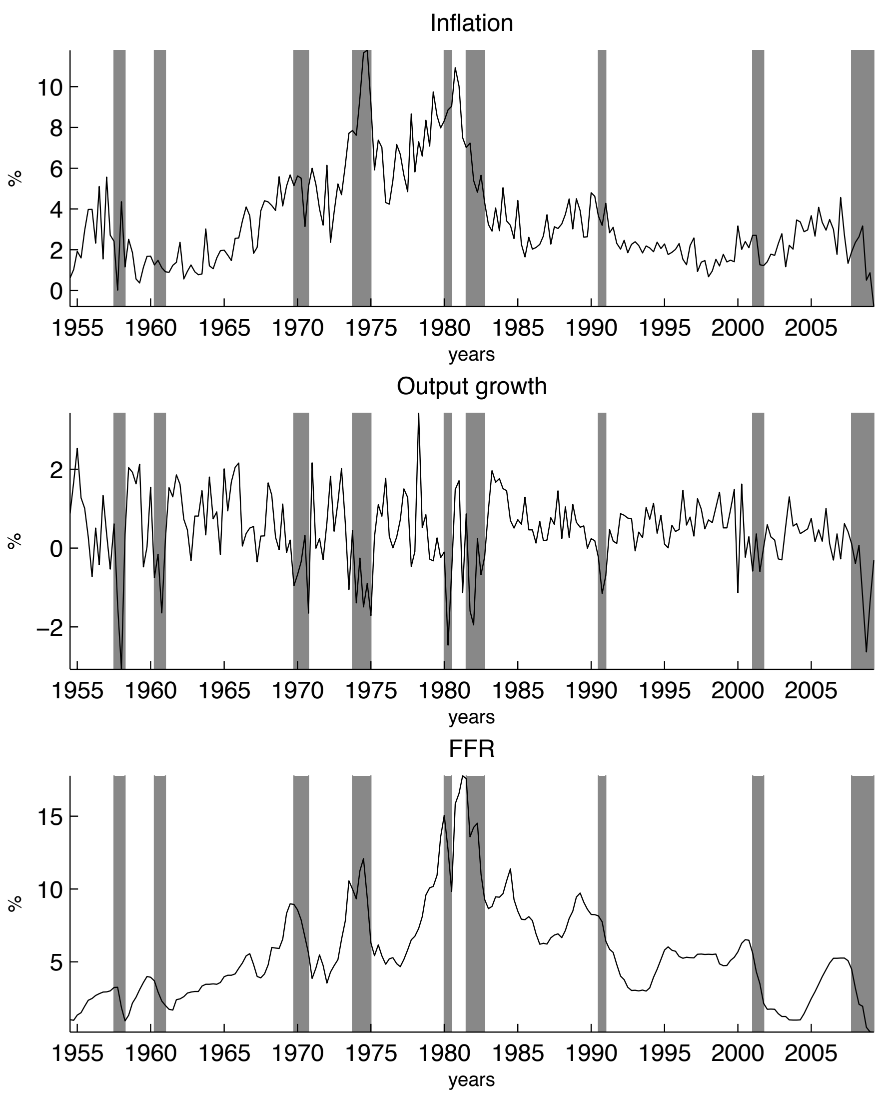

Figure 1. Sample period: 1954.Q3-2009.Q2. The shaded grey columns denote the NBER recessions.

Appendix C. Marginal Data Densities

In Bayesian analysis, Marginal Data Density (MDD) is a tool commonly used for comparison between models. The general idea behind this is as follows: We know that posterior density can be written as

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)} \tag{17}$$

We know that true posterior is

$$\int P(\theta|Y)d\theta = 1 \tag{18}$$

which makes

$$\int P(Y|\theta)P(\theta)d\theta = P(Y) \tag{19}$$

and if we use some proposal density $P_{prop}(\theta)$ which integrates to 1 we can deduce that

$$\int \frac{P_{prop}(\theta)}{P(Y|\theta)P(\theta)} \frac{P(Y|\theta)P(\theta)}{P(Y)} d\theta = \frac{1}{P(Y)} \tag{20}$$

we can define $m(\theta) = \frac{P_{prop}(\theta)}{P(Y|\theta)P(\theta)}$ and since we know that $\frac{P(Y|\theta)P(\theta)}{P(Y)} d\theta = P(\theta|Y)$ and is true density integrating to 1 we can conclude that

$$\frac{1}{P(Y)} = \sum_{i=0}^{\infty} m(\theta_i) \tag{21}$$

The closer proposed density is to posterior kernel, more accurate results are obtained. In addition when one looks at the formulas, it is clear that the higher marginal data density, the closer estimated posterior is to the "true" posterior. Thus it allows us to compare models in a most efficient way.

C.1. **Geweke (1999) method.** We follow Geweke method for our first calculation and choose $P_{prop}(\theta)$ to be truncated Normal. First, we run a random-walk Metropolis-Hastings algorithm and generate a significant number of posterior draws θ_t. Using twenty percent of these draws, we compute certain statistics, such as mode $\hat{\theta}$ for each estimated parameter and the analogue of the variance-covariance matrix, we center it around $\hat{\theta}$ instead of the mean.

$$V = \frac{1}{T} \sum_{t=1}^{T} (\theta_t - \hat{\theta})(\theta_t - \hat{\theta})' \tag{22}$$

The reason for this choice of centering is the fact that in the Markov-switching world mean is often located in a low probability region. It stands to reason that if everything is centered around the mean truncation used for this method, it would cut "too much" of the distribution when creating proposal distribution. Thus most of the posterior draws would fall within the zero-probability region of posterior distribution.

After we obtain these statistics, we create proposal density using truncated Normal centered around $\hat{\theta}$ and scaled by $V^{(-1)}$ Using the rest of the posterior draws, we evaluate posterior and proposal densities at these draws, obtaining $P_{post}(\theta)$ and $P_{prop}(\theta)$. Using these values, we proceed by computing

$$MDD = \frac{1}{T}\sum_{t=1}^{T}(\frac{P_{prop}(\theta_t)}{P_{post}(\theta_t)}) \tag{23}$$

C.2. **Waggoner and Zha (2008) method.** Since in practice posteriors estimated for the parameters in a Markov-switching framework are often highly non-Gaussian, we are using a new modified harmonic mean method for proposed by Sims, Waggoner, and Zha (2008)calculations of MDD. We first proceed by generating posterior draws from the posterior distribution using the Random-Walk Metropolis-Hastings algorithm. We then make proposal draws from normal distribution as our model is fairly small and Gaussian approximation may give accurate results, even though in the original Sims, Waggoner, and Zha (2008) method elliptical distribution is used (it includes Gaussian density as a special case). The following procedure could be used for elliptical distribution:

$$g(\theta) = \frac{\Gamma(k/2)}{2\pi^{k/2}|det(\hat{S})|}\frac{f(r)}{r^{k-1}} \tag{24}$$

where Γ is a standard gamma function and $f(r)$ is a one-dimensional density defined on the positive reals. Calculations can be done in the following way:

(1) Calculate the statistics of posterior draws from a Metropolis-Hastings algorithm using 20percent of all draws (all other calculations are done using the remaining 80 percent of draws). For centering, we used posterior mode $\hat{\theta}$. Calculate scale $\hat{S} = \sqrt{\hat{\Omega}}$, where $\hat{\Omega}$ is a variance-covariance matrix and radius is

$$r^{(i)} = \sqrt{(\theta^{(i)} - \hat{\theta})'\hat{\Omega}^{-1}(\theta^{(i)} - \hat{\theta})} \tag{25}$$

Using this radius, calculate other statistics:

- c_1 such that 1 percent of $r^i \leq c_1$
- c_{10} such that 10 percent of $r^i \leq c_{10}$ and

- c_{90} such that 90percent of $r^i \leq c_{10}$

From these statistics calculate parameters a, b and v

$$a = c_1 \qquad b = \frac{c_{90}}{0.9^{\frac{1}{v}}} \qquad v = \frac{\ln(1/9)}{\ln(c_{10}/c_{90})} \tag{26}$$

(2) Using these values, evaluate function $g(\theta)$ at posterior draws. It is calculated as:

$$f(r) = \begin{cases} \frac{vr(i)^{(v-1)}}{b^v - a^v} & \text{if } r^{(i)} \in (a,b) \\ 0 & \text{elsewhere} \end{cases} \tag{27}$$

(3) Calculate proposal draws from elliptical density and evaluate them at posterior density. First, simulate draws x from the standard normal distribution. Second, generate draw u identically and independently from the uniform distribution between $[0,1]$. Then we form

$$r = (u(b^v - a^v) + a^v)^{\frac{1}{v}} \tag{28}$$

Using these x and r we calculate proposal draws

$$\theta_{proposal} = \frac{r}{||x||}\hat{S}x + \hat{\theta} \tag{29}$$

where x is the random normal and evaluate posterior at these draws.

(4) One can calvulate the weighting function in the same way for any proposal dencity used.

$$h(\theta) = \frac{\chi_{\Theta_L}(\theta)}{q_L} g(\theta) \tag{30}$$

is a truncated proposal distribution. Truncation is done using $\hat{q}_L$, which is estimated as probability that the posterior evaluated at proposal draws falls within the region:

$$\Theta_L = \{\theta : p(Y_t|\theta)p(\theta) \geq L\} \tag{31}$$

$\chi_{\Theta_L}(\theta)$ is an indicator function, which is equal to 1 when posterior density evaluated at posterior draw falls within Θ_L and 0 otherwise. From here we can assess the overlap between posterior density and weighting function and calculate the marginal likelihood:

$$p(Y_T)^{-1} = \int_{\Theta} \frac{h(\theta)}{p(Y_t|\theta)p(\theta)} p(\theta|Y_T) d(\theta) \tag{32}$$

Defining $m(\theta)=\frac{h(\theta)}{p(Y_t|\theta)p(\theta)}$ and using the Monte Carlo integration, we get

$$\hat{p}(Y_T)^{-1} = \frac{1}{N}\sum_{i=1}^{N} m(\theta^{(1)}) \tag{33}$$

In order to check the robustness of our conclusions based on this methodology, we also use the truncated normal method proposed by Geweke (1999) and find that even though magnitudes of MDD are different, the rankings of the models stay the same.

We are using the procedure described above for truncation; however instead of an eliptical distribution, we are using a normal distribution.

C.3. **Bridge method.** Meng and Wong (1996) propose a generalization of the importance sampling method; the so-called "bridge sampling". This technique combines the Markov Chain Monte Carlo (MCMC) draws from the posterior probability density function (pdf) with the draws from the weighting function (or importance density) through a bridge function $\alpha(.)$ that reweighs both functions. Their method is based on the following result:

$$p(Y_T) = \frac{\mathrm{E}_q(\alpha(\theta)p^*(\theta))}{\mathrm{E}_p(\alpha(\theta)h(\theta))} \tag{34}$$

where $\alpha(\theta)$ is an arbitrary function and $p^*(\theta)$ the posterior kernel such that $p^*(\theta|Y_t) = p(Y_T|\theta)p(\theta)$.

It follows that the estimator $[\hat{p}(Y_T)]$ is called the general bridge sampling estimator

$$\hat{p}(Y_T) = \frac{\frac{1}{N_p}\sum_{j=1}^{N_p}\alpha(\theta^j)p^*(\theta^{(j)})}{\frac{1}{N_h}\sum_{i=1}^{N_p}\alpha(\theta^i)h(\theta^{(i)})} \tag{35}$$

where N_h is the number of draws from the weighting density and N_p is the number of draws from the posterior distribution.

Once all draws from the importance density $h(\theta)$ and MCMC draws from the posterior density $p(\theta|Y_T)$ have been made, one can easily calculate $\hat{p}(Y_T)$. Meng and Wong (1996) proposes the following bridge function:

$$\alpha(\theta) \propto \frac{1}{N_h h(\theta) + N_p p(\theta|Y_T)} \tag{36}$$

References

Cho, S. (2011): "Characterizing Markov-switching Rational Expectations Models," *Working paper*.

Farmer, R. E. A., D. F. Waggoner, and T. Zha (2009): "Understanding Regime-switching Rational Expectations Models," *Journal of Economic Theory*, 144, 1849–1867.

——— (2011): "Minimal State Variable Solutions to Markov-switching Rational Expectation Models," *Journal of Economic and Dynamic Control*, 35(12), 2150–2166.

Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Review*, 18(1), 1–73.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Kim, C.-J. (1994): "Dynamic Linear Models with Markov-Switching," *Journal of Econometrics*, 60, 1–22.

Kim, C.-J., and C. R. Nelson (1999): *State-space Models with Regime Switching*.

Meng, X., and W. H. Wong (1996): "Simulating Ratios of Normalizing Constants Via a Simple Identity: A Theoretical Exploration," *Statistica Sinica*, (6), 831–860.

Sims, C. A., D. F. Waggoner, and T. Zha (2008): "Methods for Inference in Large Multiple-equation Markov-switching Models," *Journal of Econometrics*, 146, 255–274.