The Effects of Monetary Policy on Stock Market Bubbles: Some Evidence

Jordi Gali*   Luca Gambetti†

ONLINE APPENDIX

The appendix describes the estimation of the time-varying coefficients VAR model. The model is estimated using the Gibbs sampling algorithm along the lines described in Del Negro and Primiceri (2013). Each iteration of the algorithm is composed of seven steps where a draw for a set of parameters is made conditional on the value of the remaining parameters. To clarify the notation, let $\mathbf{w}_t$ be a generic column vector. We denote $\mathbf{w}^T = [w_1', ..., w_T']'$. Below we report the conditional distributions used in the seven steps of the algorithm:

1. $p(\boldsymbol{\sigma}^T|\mathbf{x}^T, \boldsymbol{\theta}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \mathbf{s}^T)$
2. $p(\boldsymbol{\phi}^T|\mathbf{x}^T, \boldsymbol{\theta}^T, \boldsymbol{\sigma}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi})$
3. $p(\boldsymbol{\theta}^T|\mathbf{x}^T, \boldsymbol{\sigma}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi})$
4. $p(\boldsymbol{\Omega}|\mathbf{x}^T, \boldsymbol{\theta}^T, \boldsymbol{\sigma}^T, \boldsymbol{\phi}^T, \boldsymbol{\Xi}, \boldsymbol{\Psi})$
5. $p(\boldsymbol{\Xi}|\mathbf{x}^T, \boldsymbol{\theta}^T, \boldsymbol{\sigma}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Psi})$
6. $p(\boldsymbol{\Psi}_i|\mathbf{x}^T, \boldsymbol{\theta}^T, \boldsymbol{\sigma}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi})$, $i = 1, 2, 3, 4$
7. $p(\mathbf{s}^T|\mathbf{x}^T, \boldsymbol{\theta}^T, \boldsymbol{\sigma}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi})$ [1]

*CREI, Universitat Pompeu Fabra and Barcelona GSE. E-mail: jgali@crei.cat

†UAB and Barcelona GSE. E-mail: luca.gambetti@uab.es

[1] $\mathbf{s}_t$ be a $n \times 1$ vector whose elements can take integer values between 1 to 7 which is used for drawing σ^T.

Priors Specification

We assume that the covariance matrices $\mathbf{\Omega}$, $\mathbf{\Xi}$ and $\mathbf{\Psi}$ and the initial states, $\boldsymbol{\theta}_0$, $\boldsymbol{\phi}_0$ and $\log \boldsymbol{\sigma}_0$, are independent, the prior distributions for the initial states are Normal and the prior distributions for $\mathbf{\Omega}^{-1}$, $\mathbf{\Xi}^{-1}$ and $\mathbf{\Psi}_i^{-1}$ are Wishart. More precisely

$$
\begin{aligned}
\boldsymbol{\theta}_0 &\sim N(\hat{\boldsymbol{\theta}}, 4\hat{\mathbf{V}}_{\boldsymbol{\theta}}) \\
\log \boldsymbol{\sigma}_0 &\sim N(\log \hat{\boldsymbol{\sigma}}_0, \mathbf{I}_n) \\
\boldsymbol{\phi}_{i0} &\sim N(\hat{\boldsymbol{\phi}}_i, \hat{\mathbf{V}}_{\boldsymbol{\phi}_i}) \\
\mathbf{\Omega}^{-1} &\sim W(\underline{\mathbf{\Omega}}^{-1}, \underline{\rho}_1) \\
\mathbf{\Xi}^{-1} &\sim W(\underline{\mathbf{\Xi}}^{-1}, \underline{\rho}_2) \\
\mathbf{\Psi}_i^{-1} &\sim W(\underline{\mathbf{\Psi}}_i^{-1}, \underline{\rho}_{3i})
\end{aligned}
$$

where $W(\mathbf{S}, d)$ denotes a Wishart distribution with scale matrix $\mathbf{S}$ and degrees of freedom d and $\mathbf{I}_n$ is a $n \times n$ identity matrix where n is the number of variables in the VAR.

Prior means and variances of the Normal distributions are calibrated using a time invariant VAR for $\mathbf{x}_t$ estimated using the first $\tau = 48$ observations. $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{V}}_{\boldsymbol{\theta}}$ are set equal to the OLS estimates. Let $\hat{\mathbf{\Sigma}}$ be the covariance matrix of the residuals $\hat{\mathbf{u}}_t$ of the initial time-invariant VAR. We apply the same decomposition discussed in the text $\hat{\mathbf{\Sigma}} = \hat{\mathbf{F}}\hat{\mathbf{D}}\hat{\mathbf{F}}'$ and set $\log \hat{\boldsymbol{\sigma}}_0$ equal to the log of the diagonal elements of $\hat{\mathbf{D}}^{1/2}$. $\hat{\boldsymbol{\phi}}_i$ is set equal to the OLS estimates of the coefficients of the regression of $\hat{\mathbf{u}}_{i+1,t}$, the $i+1$-th element of $\hat{\mathbf{u}}_t$, on $-\hat{\mathbf{u}}_{1,t}, ..., -\hat{\mathbf{u}}_{i,t}$ and $\hat{\mathbf{V}}_{\phi_i}$ equal to the estimated variances.

We parametrize the scale matrices as follows $\underline{\mathbf{\Omega}} = \underline{\rho}_1(\lambda_1 \hat{\mathbf{V}}^f_{\boldsymbol{\theta}})$, $\underline{\mathbf{\Xi}} = \underline{\rho}_2(\lambda_2 \mathbf{I}_n)$ and $\underline{\mathbf{\Psi}}_i = \underline{\rho}_{3i}(\lambda_3 \hat{\mathbf{V}}^f_{\boldsymbol{\phi}_i})$. The degrees of freedom for the priors on the covariance matrices $\underline{\rho}_1$ and $\underline{\rho}_2$ are set equal to the number of rows $\underline{\mathbf{\Omega}}^{-1}$ and $\mathbf{I}_n$ plus one respectively while $\underline{\rho}_{3i}$ is $i+1$ for $i = 1, ..., n-1$. We assume $\lambda_1 = 0.005$, $\lambda_2 = 0.01$ and $\lambda_3 = 0.01$. Finally $\hat{\mathbf{V}}^f_\theta$ and $\hat{\mathbf{V}}^f_{\phi_i}$ are obtained as $\hat{\mathbf{V}}_\theta$ and $\hat{\mathbf{V}}_{\boldsymbol{\phi}_i}$ but using the estimates of the whole sample.

Gibbs sampling algorithm

Let $\bar{T}$ be the total number of observations, in our case equal to 204. To draw realizations from the posterior distribution we use $T = \bar{T} - \tau/2$ observations starting from $\tau/2 + 1$.[2] The algorithm works as follows:

Step 1: sample $\boldsymbol{\sigma}^T$. The states $\boldsymbol{\sigma}^T$ are drawn using the algorithm of Kim, Shephard and Chib (1998, KSC hereafter). Let $\mathbf{x}_t^* \equiv \mathbf{F}_t^{-1}(\mathbf{x}_t - \mathbf{W}_t\boldsymbol{\theta}_t) = \mathbf{D}_t^{1/2}\mathbf{u}_t$, where $\mathbf{u}_t \sim N(0, \mathbf{I}_n)$, $\mathbf{W}_t = (\mathbf{I}_n \otimes \mathbf{w}_t)$, and $\mathbf{w}_t = [1\ \mathbf{x}'_{t-1}...\mathbf{x}'_{t-p}]$. Notice that conditional on $\mathbf{x}^T, \boldsymbol{\theta}^T$, and $\boldsymbol{\phi}^T$, $\mathbf{x}_t^*$ is observable. Therefore, by squaring and taking logs, we obtain the following state-space representation

$$\mathbf{x}_t^{**} = 2\mathbf{r}_t + \boldsymbol{\upsilon}_t \tag{1}$$

$$\mathbf{r}_t = \mathbf{r}_{t-1} + \boldsymbol{\zeta}_t \tag{2}$$

where $\mathbf{x}_{i,t}^{**} = \log(\mathbf{x}_{i,t}^{*\,2})$, $\boldsymbol{\upsilon}_{i,t} = \log(\mathbf{u}_{i,t}^2)$ and $\mathbf{r}_t = \log \boldsymbol{\sigma}_{i,t}$.[3] The above system is non-normal since the innovation in (1) is distributed as $\log \chi^2(1)$. Following KSC, we use a mixture of 7 Normal densities with mean $m_j - 1.2704$, and variance v_j^2 (j=1,...,7) to approximate the system with a Gaussian one (see Table A1 for the values used). In practice the algorithm of Carter and Kohn (1994, CK henceforth) is used to draw $\mathbf{r}_t$ using, as density of $\boldsymbol{\upsilon}_{i,t}$, the one indicated by the value of $\mathbf{s}_{i,t}$: $(\boldsymbol{\upsilon}_{i,t}|\mathbf{s}_{i,t} = j) \sim N(m_j - 1.2704, v_j^2)$. More precisely $\mathbf{r}_t$ is drawn from $N(\mathbf{r}_{t|t+1}, \mathbf{R}_{t|t+1})$, where $\mathbf{r}_{t|t+1} = E(\mathbf{r}_t|\mathbf{r}_{t+1}, \mathbf{x}^t, \boldsymbol{\theta}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \mathbf{s}^T,)$ and $\mathbf{R}_{t|t+1} = Var(\mathbf{r}_t|\mathbf{r}_{t+1}, \mathbf{x}^t, \boldsymbol{\theta}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \mathbf{s}^T)$ are the conditional mean and variance obtained from the backward recursion equations.

[2] We start the sample from $\tau/2 + 1$ instead of τ in order to not to lose too many data points.

[3] We do not use any offsetting constant since given that the variables are in logs multiplied by 100, we do not have numerical problems.

Table A1

j	q_j	m_j	v_j^2
1.0000	0.0073	-10.1300	5.7960
2.0000	0.1056	-3.9728	2.6137
3.0000	0.0000	-8.5669	5.1795
4.0000	0.0440	2.7779	0.1674
5.0000	0.3400	0.6194	0.6401
6.0000	0.2457	1.7952	0.3402
7.0000	0.2575	-1.0882	1.2626

Step 2: sample $\boldsymbol{\phi}^T$. Let $\hat{\mathbf{x}}_t = \mathbf{x}_t - \mathbf{W}_t\boldsymbol{\theta}_t$. The $i+1$-th $(i = 1, ..., n-1)$ equation of the system $\mathbf{F}_t^{-1}\hat{\mathbf{x}}_t = \mathbf{D}_t^{1/2}\mathbf{u}_t$ can be written as

$$\hat{\mathbf{x}}_{i+1,t} \quad = \quad -\hat{\mathbf{x}}_{[1,i],t}\boldsymbol{\phi}_{i,t} + \sigma_{i,t}\mathbf{u}_{i+1,t} \quad i = 2, ..., n \tag{3}$$

where $\boldsymbol{\sigma}_{i,t}$ and $\mathbf{u}_{i,t}$ are the ith elements of $\boldsymbol{\sigma}_t$ and u_t respectively and $\hat{\mathbf{x}}_{[1,i],t} = [\hat{\mathbf{x}}_{1,t}, ..., \hat{\mathbf{x}}_{i,t}]$. Conditional on $\boldsymbol{\theta}^T$ and $\boldsymbol{\sigma}^T$, equation (3) is the observable equation of a state-space model where the states are $\boldsymbol{\phi}_{i,t}$. Moreover, since $\boldsymbol{\phi}_{i,t}$ and $\boldsymbol{\phi}_{j,t}$ are independent for $i \neq j$, the algorithm of CK can be applied equation by equation to draw $\boldsymbol{\phi}_{i,t}$ from a $N(\boldsymbol{\phi}_{i,t|t+1}, \boldsymbol{\Phi}_{i,t|t+1})$, where $\boldsymbol{\phi}_{i,t|t+1} = E(\boldsymbol{\phi}_{i,t}|\boldsymbol{\phi}_{i,t+1}, \mathbf{x}^t, \boldsymbol{\theta}^T, \boldsymbol{\sigma}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi})$ and $\boldsymbol{\Phi}_{i,t|t+1} = Var(\boldsymbol{\phi}_{i,t}|\boldsymbol{\phi}_{i,t+1}, \mathbf{x}^t, \boldsymbol{\theta}^T, \boldsymbol{\sigma}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi})$.

Step 3: sample $\boldsymbol{\theta}^T$. Consider the state-space representation

$$\mathbf{x}_t \quad = \quad \mathbf{W}_t\boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t \tag{4}$$

$$\boldsymbol{\theta}_t \quad = \quad \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t. \tag{5}$$

$\boldsymbol{\theta}_t$ is drawn from a $N(\boldsymbol{\theta}_{t|t+1}, \mathbf{P}_{t|t+1})$, where $\boldsymbol{\theta}_{t|t+1} = E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathbf{x}^t, \boldsymbol{\sigma}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi})$ and $\mathbf{P}_{t|t+1} = Var(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathbf{x}^t, \boldsymbol{\sigma}^T, \boldsymbol{\phi}^T, \boldsymbol{\Omega}, \boldsymbol{\Xi}, \boldsymbol{\Psi})$ are obtained using the CK algorithm.

Step 4: sample $\mathbf{\Omega}$. A draw is obtained as follows: $\mathbf{\Omega} = (\mathbf{M}\mathbf{M}')^{-1}$ where $\mathbf{M}$ is an $(n^2p + n) \times \overline{\rho}_1$ matrix whose columns are independent draws from a $N(0, \overline{\mathbf{\Omega}}^{-1})$ where $\overline{\mathbf{\Omega}} = \underline{\mathbf{\Omega}} + \sum_{t=1}^{T} \Delta\boldsymbol{\theta}_t(\Delta\boldsymbol{\theta}_t)'$ (see Gelman et. al., 1995).

Step 5: sample $\mathbf{\Xi}$. As above, $\mathbf{\Xi} = (\mathbf{M}\mathbf{M}')^{-1}$ where $\mathbf{M}$ is an $n \times \overline{\rho}_2$ matrix whose columns are independent draws from a $N(0, \overline{\mathbf{\Xi}}^{-1})$ where $\overline{\mathbf{\Xi}} = \underline{\mathbf{\Xi}} + \sum_{t=1}^{T} \Delta \log \boldsymbol{\sigma}_t(\Delta \log \boldsymbol{\sigma}_t)'$.

Step 6: sample $\mathbf{\Psi}_i$ $i = 1, ..., 5$. As above, $\mathbf{\Psi}_i = (\mathbf{M}\mathbf{M}')^{-1}$ where $\mathbf{M}$ is an $i \times \overline{\rho}_{3i}$ matrix whose columns are independent draws from a $N(0, \overline{\mathbf{\Psi}}_i^{-1})$ where $\overline{\mathbf{\Psi}}_i = \underline{\mathbf{\Psi}}_i + \sum_{t=1}^{T} \Delta\boldsymbol{\phi}_{i,t}(\Delta\boldsymbol{\phi}_{i,t})'$.

Step 7: sample $\mathbf{s}^T$. Each $\mathbf{s}_{i,t}$ is independently sampled from the discrete density $Pr(\mathbf{s}_{i,t} = j | \mathbf{x}_{i,t}^{**}, \mathbf{r}_{i,t}) \propto q_j f_N(\mathbf{x}_{i,t}^{**} | 2\mathbf{r}_{i,t} + m_j - 1.2704, v_j^2)$, where $f_N(x|\mu, \sigma^2)$ denotes the Normal pdf with mean μ and variance σ^2, and q_j is the probability reported in Table A1 associated to the j-th density.

We make 22000 draws discarding the first 20000 and collecting one out of two of the remaining 2000 draws. The results presented in the paper are therefore based on 1000 draws from the posterior distribution. Parameters convergence is assessed using trace plots.

References

Carter, Chris K., and Robert Kohn (1994): ”On Gibbs Sampling for State Space Models,” *Biometrika*, 81:541-553.

Del Negro, Marco and Giorgio Primiceri (2013): ”Time-Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum” Federal Reserve Bank of New York Staff Report No. 619.

Gelman Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, London.

Kim Sangjoon, Neil Shepard and Siddhartha Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models", *Review of Economic Studies*, 65, 361-393.