Discussion of Cogley and Sargent's "Drifts and Volatilities: Monetary Policy and Outcomes in the Post WWII U.S."

Marco Del Negro*
Federal Reserve Bank of Atlanta

March 2003
preliminary!

Cogley and Sargent provide us with a very useful tool for empirical macroeconomics: a Gibbs sampler for the estimation of VARs with drifting coefficients and volatilities. The authors apply the tool to a VAR with three variables - inflation, unemployment, and the nominal interest rate - and two lags. This tool is a serious competitor to the identified-VAR-cum-Markov-switching technology recently developed by Sims (1999) and Sims and Zha (2002) for the study of economies that are subject to regime changes. However, the Gibbs sampler suffers from a curse of dimensionality: as more variables or more lags are added to the system, the computational burden of the estimation quickly grows out of proportion. My suggestions here are mainly aimed at making the tool more flexible, and hence more widely applicable.

Of the many variables that one may want to add to the three considered by Cogley and Sargent, one in particular stands out: commodity prices. This is for two

*I wish to thank Tim Cogley for providing the matlab programs and the data used in the paper, and Dan Waggoner and Tao Zha for helpful conversations. The views expressed here do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

reasons. First, in the identified VAR literature commodity prices play a key role in the identification of policy shocks (see Sims and Zha 1998). Second, it may be important to check the robustness of Cogley and Sargent's results to the inclusion of commodity prices. Sargent (1999) considers the evidence of parameter drift as a smoking gun in favor of his vindication of econometric policy evaluation story. Parameter drifts may be the outcome of learning-induced policy regime shifts, or of changes in the environment. In many people's mind the stagflation of the seventies is associated with the oil shocks. Are the results in Cogley and Sargent robust to the inclusion of commodity prices? Was the high inflation in the seventies due to bad luck (oil shocks - changes in the environment) or bad policy (regime shifts)? In the last section I present some results based on a VAR with commodity prices, and try to address at least the first of the two questions.

1 The Gibbs Sampler

1.1 Equation by equation

The goal of this section is to rewrite the Gibbs sampler in Cogley and Sargent so that it can be performed equation by equation, thereby reducing the computational burden of the enterprise and making it possible to include more variables and more lags. The notation is the same as in Cogley and Sargent, except where explicitly mentioned, so that the reader can refer to their definitions.

Much of the Gibbs sampler in Cogley and Sargent can already be performed equation by equation, the only exception being perhaps the most computationally intensive part, the draws from the posterior of θ^T conditional on all other parameters. The key to improve this part of the Gibbs sampler is to work with the structural form VAR parameters, as opposed to the reduced form parameters as Cogley and Sargent do. The measurement and the transition equations in Cogley and Sargent, reported here for convenience, are:

$$y_t = X_t'\theta_t + \epsilon_t \tag{1}$$

$$\theta_t = \theta_{t-1} + \nu_t. \tag{2}$$

where $X_t' = I_n \otimes x_t'$ and x_t includes all the regressors (i.e, the lags of y_t as well as the constant). The measurement equation can be equivalently rewritten as:

$$y_t = \Theta_t \; x_t + \epsilon_t \tag{3}$$

where the relationship between Θ_t and θ_t is given by $\theta_t = vec(\Theta_t')$. The innovations ν_t are normally distributed with covariance matrix Q. The innovations ϵ_t are also normally distributed, with variance that evolves over time:

$$\epsilon_t \sim N(0, R_t), \tag{4}$$

with

$$R_t = A^{0\ -1} H_t A^{0\ -1\ \prime}. \tag{5}$$

where H_t is diagonal with elements that vary over time according to a driftless, geometric random walk. As in Cogley and Sargent A^0 is lower triangular with ones on the diagonal (this matrix is called B in their paper), and H_t is diagonal with elements that vary over time according to a driftless, geometric random walk. The curse of dimensionality arises because θ_t is of dimension $k \times 1$, where $k =$ $(\#\text{ of variables})^2 \times (\#\text{ of lags})$.

Let us premultiply (3) by A^0 and obtain:

$$A^0 y_t = A_t \; x_t + u_t, \tag{6}$$

where $A_t = A^0\Theta_t$ are the so-called structural form coefficients and $u_t = A^0\epsilon_t$ is a vector of uncorrelated errors. Let us define $a_t \equiv vec(A_t') = (A^0 \otimes I_p)\theta_t$, where p is the number of regressors in each equation. Premultiplying (2) by $A^0 \otimes I_p$ delivers the transition equation for the structural parameters which is also a random walk:

$$a_t = a_{t-1} + \tilde{\nu}_t,\ \tilde{\nu}_t \sim N(0, \tilde{Q}), \tag{7}$$

with $\tilde{Q} = (A^0 \otimes I_p)\ Q\ (A^0 \otimes I_p)'$ (since Q is unrestricted in Cogley and Sargent this transformation does not alter any assumption). If one could start from scratch and assume that $\tilde{Q}$ is block diagonal (the innovations in 7 are correlated only within each equation), one could already draw the a_ts equation by equation. The u_t innovations are now orthogonal to each other (while the ϵ_t were not). Conditional on all other parameters (including A^0) the Kim-Nelson (1999)/Carter-Kohn (1994) Gibbs sampler described in section C.2.1 of Cogley and Sargent can be applied equation by equation (where each equation of (6) is the measurement equation) to obtain draws of a^T.[1]

Cogley and Sargent provide evidence that Q (and hence $\tilde{Q}$) is unlikely to be block diagonal - evidence that is consistent with the theory in Sargent (1999). If $\tilde{Q}$ is not block diagonal, one can proceed as follows. Let Ψ be a lower triangular matrix with ones on the diagonal such that $\tilde{Q} = \Psi^{-1} \tilde{H} \Psi^{-1\ \prime}$, where $\tilde{H}$ is diagonal. Let us premultiply (7) by Ψ and obtain:

$$\Psi\ (a_t - a_{t-1}) = \tilde{u}_t,\ \tilde{u}_t \sim N(0, \tilde{H}) \tag{8}$$

Call $a_{j,t}$ the jth element of a_t and a_j^T the whole history of $a_{j,t}$s. Likewise, call $a^{j-1,T}$ the whole histories of $a_{k,t}$ for $k = 1, .., j-1$. Since the $\tilde{u}_t$ are now uncorrelated across equations, we can draw a_j^T conditional on $a^{j-1,T}$ and on all other parameters (including Q and A^0, from which Ψ and $\tilde{H}$ can be obtained).[2] Given $a^{j,T}$, we can draw we can draw a_{j+1}^T, and so on. The transition equation for the parameter $a_{j,t}$ is:

$$a_{j,t} = a_{j,t-1} - \psi_{j,1}(a_{1,t} - a_{1,t-1})\ ..\ - \psi_{j,j-1}(a_{j-1,t} - a_{j-1,t-1}) + \tilde{u}_{j,t}. \tag{9}$$

The appendix shows how to change the procedure of section C.2.1 to take into account the time-varying constant in equation (9). Of course drawing each parameter at the time is likely to be inefficient. The approach can be easily modified to draw

[1] Given the draws for the a_t, the draws for θ_t can be computed using $\Theta_t = A^{0\ -1} A_t$. Hence one can just as easily impose the no-explosive-roots prior.

[2] Note that the random walk assumption is key for this approach to work: if $a_{j,t}$ depended on $a_{j+1,t-1}$, for instance, this would not work.

the $a_{j,t}$s block by block (where each block corresponds to the set of parameters belonging to a specific equation). This comes at the cost of more notation, so I will not pursue it here. The Gibbs sampler for the other parameters, namely H^T, σ, A^0, and Q, obtains as in Cogley and Sargent.

1.2 Time-varying Covariances

Primiceri (2002) first introduces time-varying covariances, that is, a time-varying A^0 matrix, into the model of Cogley and Sargent. However, Primiceri works with the reduced form VAR parameters (the θ_ts) and his Gibbs sampler to the best of my knowledge still suffers from the same curse of dimensionality that affects the one of Cogley and Sargent. In order to make A^0 time-varying one has to take a stand on whether the primitive in terms of law of motion for the VAR parameters is equation (2) or equation (7). If the primitive is equation (2), then equation (7) becomes:

$$\begin{aligned} &a_t = F_t a_{t-1} + \tilde{\nu}_t,\ \tilde{\nu}_t \sim N(0, \tilde{Q}_t), \\ &F_t = (A_t^0 A_{t-1}^{0\ -1} \otimes I_p),\ \tilde{Q}_t = (A_t^0 \otimes I_p)\ Q\ (A_t^0 \otimes I_p)'. \end{aligned} \tag{10}$$

Since the matrix F_t is not lower diagonal, the equation by equation approach of the previous section cannot be applied.[3] If the primitive is equation (7) however, then F_t is the identity matrix. If A_t^0 is still lower triangular, one can incorporate the non-zero elements of A_t^0 into the a_t vector, and use (7) as the transition equation and:

$$A_t^0 y_t = A_t\ x_t + u_t$$

as the measurement equation. All can be done equation by equation again. An advantage of assuming equation (7) as the primitive is that one can allow for the shocks in A_t^0 and in A_t to be correlated, whereas in Primiceri they are orthogonal. One has to bear in mind that (7) coupled with a time-varying A^0 matrix implies that the reduced form parameters no longer follow a random walk.

[3] I wish to thank Tao Zha for pointing out a mistake in an earlier draft.

1.3 Stochastic Volatility in the Parameters' Law of Motion

The theory developed in Sargent (1999) postulates that the VAR coefficients vary slowly for long periods of time, and then suddenly drift away escaping the Nash equilibrium. During the escape phase innovations in the random walk process for the parameters appear to be much larger than in other periods. This observation suggests that the model in Cogley and Sargent may be missing an important feature: drifts in volatilities of the ν_t innovations. This feature can be incorporated in the Gibbs sampler described above.

Let us take equation (7) as the primitive law of motion for the structural parameters, with the difference that $\tilde{Q}$ is no longer constant over time:

$$a_t = a_{t-1} + \tilde{\nu}_t,\ \tilde{\nu}_t \sim N(0, \tilde{Q}_t), \tag{11}$$

where

$$\tilde{Q}_t = \Psi^{-1} \tilde{H}_t \Psi^{-1\,\prime}, \tag{12}$$

and $\tilde{H}_t$ is diagonal with elements that vary over time according to a driftless, geometric random walk. Conditional on the whole history of the stochastic volatilities, $\tilde{H}^T$, a^T can be drawn as described above. Conditional on a^T, $\tilde{H}^T$ can be drawn as described in the section C.2.5 of Cogley and Sargent, or in Kim, Shephard, and Chib (1998).

2 An Application: Adding Commodity Prices

In this section I apply the method described in section 1.1 and add commodity prices to the three variables VAR of Cogley and Sargent.[4] Aside from the inclusion of commodity prices, this application differs from Cogley and Sargent's in another dimension: the prior on the initial state θ_0 (in my case, a_0). The rationale for this

[4] Commodity prices are measured as the *FIBER Industrial Materials Index: All Items* and enter the VAR in log-differences. The data is obtained from Haver (mnemonic *PZRJOC*) for the same time period as in Cogley and Sargent. All other data was obtained from the authors.

change is as follows. Cogley and Sargent impose an implicit prior on their model (see Sims 2001): they rule out explosive roots. From a computational point of view, this implicit prior amounts to discarding those Gibbs sampler draws for which the roots are explosive. From simulations that I ran, it turns out that in their three variables VAR one throws out about one draw out of two, implying that the implicit prior is very costly in terms of computations. The problem grows worse in a four variable system, most likely because sampling variability increases with the additional regressors. I follow Cogley and Sargent in imposing the implicit prior, but I address the computational problem by choosing a different prior for the initial state. In Cogley and Sargent the prior for the initial state θ_0 is centered around the OLS estimates obtained from the period 1948.3-1958.4, with variance equal to its asymptotic variance. Since the pre-sample estimation period is relatively short, this variance is large (hence the prior is loose). To address directly the problem of sampling variability, I center around zero the prior for lags greater than one (and correspondingly make the prior variance matrix twice as tight, and diagonal), as in shrinkage estimators (James and Stein 1961).[5] I find that this shrinkage prior, in spite of being fairly loose, almost eliminates the need for the implicit prior in the sense that for very few draws the roots are explosive.

Figure 7 in the Cogley and Sargent paper displays the normalized spectrum for inflation. The figure makes two important points: i) The evidence in favor of parameter drifts found in Cogley and Sargent (2001) is robust to the inclusion of stochastic volatility in the VAR innovations; ii) Inflation persistence increased dramatically in the late-seventies, and then dropped sharply following the Volcker disinflation. Figure 7-A plots the normalized spectrum for inflation for the model with commodity prices. The big picture is unchanged. The evidence in favor of parameter drift is just as strong as in Cogley and Sargent and the pattern of inflation

[5] My prior is then similar to the Minnesota prior of Doan, Litterman, and Sims (1984), except that I do not center around 1 the prior for the first lag. The prior for the first lag is centered around the same value of Cogley and Sargent. I use the same number of Gibbs sampler draws as the authors.

persistence is very similar.[6] Most other results in their paper are also qualitatively robust to the inclusion of commodity prices. A different picture than the ones they present however presents a potential challenge to their story. Figure 7-B plots the normalized spectra for the four series in the VAR. The figure shows that the spectrum has changed over time for all four series, but most notably for inflation and commodity prices. Strikingly, the increase in persistence in inflation coincides with the increase in persistence in commodity prices. Is this evidence that the increase in the inflation persistence in the late seventies was due to commodity prices (changes in the environment) rather than bad policy (regime shifts)? This question deserves more extensive research to be addressed - research that can be done using the tools provided by Cogley and Sargent.

References

Carter, Christopher K. and Robert Kohn. 1994. "On Gibbs sampling for state space models." Biometrica 81(3), 541-553.

Doan, Thomas, Robert Litterman, and Christopher A. Sims, 1984, "Forecasting and Conditional Projections Using Realistic Prior Distributions". Econometric Reviews, 3, 1-100.

James, W. and C. Stein, "Estimation with quadratic loss." In "Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability," 1. Berkeley: University of California Press.

Kim, Chang-Jin, and Charles R. Nelson. 1999. "State-space models with regime switching." The MIT Press. Cambridge, Massachussets.

Kim, Sangjoon, Neil Shephard, Siddhartha Chib. 1998. "Stochastic volatility:

[6]There are minor differences however. The peak of the normalized spectrum for inflation in the late seventies does not occur at zero frequency, but at frequencies corresponding to five years cycles. This difference is entirely due to the shrinkage priors, as can be appreciated by running the model without commodity prices.

Likelihood inference and comparison with ARCH models." Review of Economic Studies 65, 361-393.

Primiceri, Giorgio E.. 2002. "Time varying structural vector autoregressions and monetary policy." Mimeo. Princeton University.

Sargent, Thomas J. 1999. "The conquest of American inflation." Princeton University Press: Princeton, New Jersey.

Cogley, Timothy and Thomas J. Sargent. 2001. "Evolving post World War II U.S. inflation dynamics." NBER macroeconomics Annual 16, 331-337.

Sims, Christopher A.. 1999. "Drifts and breaks in monetary policy." Discussion paper. Princeton University. `http://www.princeton.edu/~sims`.

Sims, Christopher A.. 2001. "Comment on Cogley and Sargent's 'Evolving post World War II U.S. inflation dynamics.' " NBER macroeconomics Annual 16, 373-379.

Sims, Christopher A., and Tao Zha. 1998. "Does monetary policy generate recessions?" Federal Reserve Bank of Atlanta Working Paper 1998-12.

Sims, Christopher A., and Tao Zha. 2002. "Macroeconomic switching." Discussion paper. Princeton University. `http://www.princeton.edu/~sims`.

A Appendix

The Carter-Kohn procedure of section C.2.1 in Cogley and Sargent can be modified as follows to take into account the time-varying constant in equation (9), $\mu_{j,t} = -\psi_{j,1}(a_{1,t} - a_{1,t-1}) \, .. \, - \psi_{j,j-1}(a_{j-1,t} - a_{j-1,t-1})$.

$$
\begin{aligned}
P_{t|t-1} &= P_{t-1|t-1} + \tilde{H}_{jj}, \\
K_t &= (C + P_{t|t-1}X_t)(X_t' P_{t|t-1} X_t + R_t)^{-1}, \\
\theta_{t|t} &= \mu_{j,t} + \theta_{t-1|t-1} + K_t(\tilde{y}_t - X_t'(\mu_{j,t} + \theta_{t-1|t-1})), \\
P_{t|t} &= P_{t|t-1} - K_t X_t' P_{t|t-1}, \\
\mathbb{E}[\theta_t | y^t, \theta_{t+1}, V] &= \theta_{t|t} + P_{t|t}(P_{t|t} + \tilde{H}_{jj})^{-1}(\theta_{t+1} - \theta_{t|t} - \mu_{j,t+1}), \\
\mathrm{var}[\theta_t | y^t, \theta_{t+1}, V] &= P_{t|t} - P_{t|t}(P_{t|t} + \tilde{H}_{jj})^{-1} P_{t|t}.
\end{aligned}
$$

where $\tilde{y}_t$ is the ith row of $A^0 y_t$.

Figure 7-A: Normalized Spectrum for Inflation

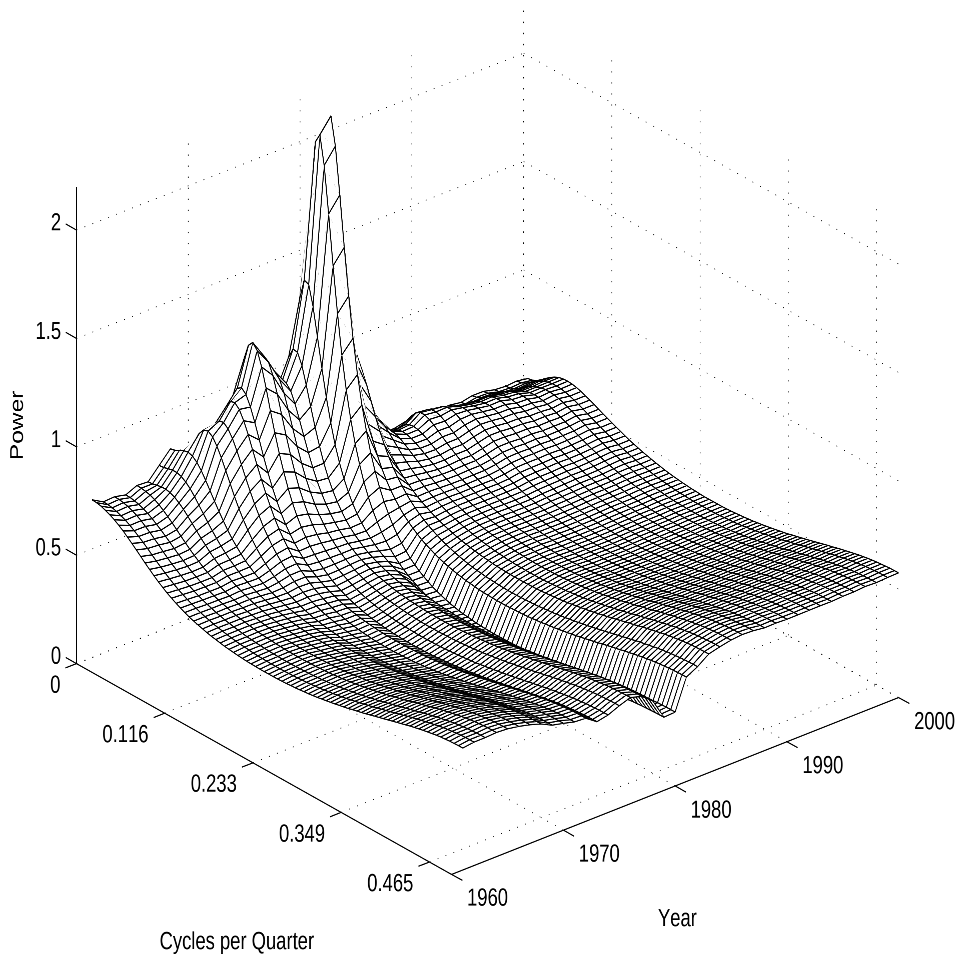

Figure 7-B: NORMALIZED SPECTRA