Correlated Disturbances and the Sources of U.S. Business Cycles

Vasco Curdia and Ricardo Reis*

FRB New York and Columbia University

February 2009

VERY PRELIMINARY AND INCOMPLETE

Abstract

The dynamic stochastic general equilibrium (DSGE) models used to study business cycles typically assume that exogenous disturbances are independent with a simple structure for serial correlation. This paper relaxes this tight restriction, by allowing for disturbances that have a rich contemporaneous and dynamic correlation structure. Our first contribution is a new Bayesian econometric method that uses conjugate conditionals to make the estimation of DSGE models with correlated disturbances feasible and quick. Our second contribution is a re-examination of the sources of U.S. business cycles, using two canonical models, one real and the other monetary. We find that allowing for correlated disturbances, the estimates of crucial parameters are more in line with other evidence, the impulse responses are closer to the results from vector autoregressions, and technology disturbances play a larger role in the business cycle. (*JEL* E30, E10)

*Contacts: vasco.curdia@ny.frb.org and rreis@columbia.edu.

A typical macroeconomic model takes as given some exogenous disturbances, specifies their impact on the behavior of economic agents, and makes predictions for some endogenous variables. Because the disturbances are exogenous to the theory, by definition they are unexplained, and must be taken as given, so it would be desirable to impose on them as few arbitrary restrictions as possible. However, the common practice is the opposite, with very strict assumptions imposed on the processes driving disturbances. This paper argues that these assumptions are unwarranted, develops new estimation techniques for models with a rich correlation structure for the disturbance vector, and uses two dynamic stochastic general-equilibrium (DSGE) models with correlated disturbances to understand U.S. business cycles.

Our first contribution is methodological. In the simultaneous-equation reduced-form macroeconomic model tradition (e.g. Fair, 2004), there has for long been a careful treatment of disturbances. In particular, at least since Cochrane and Orcutt (1949), time-series researchers have incorporated AR(1) disturbances into models, and since then it has become typical to consider cross-correlated disturbances (Zellner, 1962), rich dynamic cross and auto-correlations (Sims, 1980), and non-parametric estimates (Newey and West, 1986). This literature has found that arbitrary restrictions on disturbances can severely bias the estimates of key parameters, can distort predicted impulse responses, and can give a misleading impression of a model's fit to the data. Researchers that attempt to move their models forward by endogenizing incorrectly-identified disturbances will be led astray.

In contrast, in the more recent DSGE macroeconometric models, disturbances are almost always assumed to be independent first-order autoregressions, AR(1)s. The main justification for this assumption is that allowing for correlated disturbances increasesn the number of parameters to estimate, straining the capacity of current algorithms. In this paper, we develop new Bayesian likelihood econometric techniques to incorporate correlated disturbances in dynamics macroeconomic models. We show that the economic structure of the models implies that key conditional posterior distributions belong to the family of conjugate distributions with known analytical form. We propose a new *conjugate-conditionals algorithm* that exploits this knowledge to efficiently characterize the estimates. In particular, because the parameters associated with correlated disturbances are part of the conjugate conditional distributions, this algorithm overcomes the difficulty with allowing for correlated disturbances in these models.

Our second contribution is re-evaluate classic business-cycle questions, such as: What drives business cycles? What are the elasticities of intertemporal substitution and labor supply? How do hours respond to technology shocks? How does inflation respond to monetary shocks? What is the response of output to fiscal and monetary shocks? We use two canonical business cycle models, the real business cycle model of Christiano and Eichenbaum (1992) and the monetary business cycle model of Smets and Wouters (2007), extending them to allow for correlated disturbances, and estimating them on post-war U.S. data.

We focus on business-cycle DSGE models because the assumption that their disturbances are mutually uncorrelated is clearly incredible. A glaring example is disturbances to government spending, which in the data are certainly correlated with present and past disturbances to productivity or markups via automatic and discretionary fiscal stabilizers. More generally, whenever economists have measured the disturbances that go into business-cycle models, whether they are disturbances to technology (Solow, 1957, Kehoe and Prescott, 2002), disturbances to government spending (Barro, 1977, Rotemberg and Woodford, 1992), disturbances to labor supply (Parkin, 1988, Hall, 1997), or disturbances to investment productivity (Jorgenson, 1966, Greenwood, Hercowitz and Krusell, 1997), they have always found that these measures of the disturbances are cross and dynamically correlated in ways that are inconsistent with independent AR(1)s. Two striking examples of correlated disturbances were provided by Evans (1992) and Chari, Kehoe and McGrattan (2007). Evans estimated vector autoregressions using military spending to measure government-spending disturbance and using Solow residuals to measure technology disturbances, and found that government spending Granger-causes technology. Chari, Kehoe and McGrattan estimated a first-order autoregression, VAR(1), for the disturbances of a business-cycle model and found that most cross-correlations are large and statistically significant.

After a brief literature review, the paper is organized as follow. Section 1 introduces the real business-cycle model and uses it to present the conjugate-conditionals estimation method. The estimates of the model in the U.S. data show that disturbances are correlated in a particular way: government spending tends to strongly increase after a fall in productivity. This explains why hours tend to increase after a fall in productivity, why initial changes in productivity have a delayed and persistent effect on output, and why productivity has a larger role in the business cycle, three long-standing puzzles in the literature.

Section 2 present the estimation method more generally. We show that the conjugate conditionals arise in a broad class of equilibrium macroeconomic models. We discuss a few ways to exploit the knowledge of this known slice of the posterior distribution in making inferences.

Section 4 focuses on the monetary business-cycle model. [IN PROGRESS]

Section 5 concludes.

Literature review

The closest paper to this one is Ireland (2004). Following Sargent (1989), he adds measurement errors to the reduced-form equations of a DSGE model with many observables but a single technology disturbance. Ireland allows the measurement errors to follow a VAR(1) and, noting the equivalence of his model to a state-space model, estimates it by maximum likelihood and statistically tests for structural stability. We differ in several respects. First, our focus is on the exogenous disturbances of the model, not on measurement error (which we will even abstract from). A key distinction between disturbances and measurement errors is that the properties of the disturbance process affect the behavioral responses of the agents in the model, whereas the properties of the measurement error only affect the job of the econometrician. For instance, if technology disturbances are more persistent, agents in the model will engage in less intertemporal substitution in consumption and hours worked, altering the response of all endogenous variables, whereas more persistent measurement errors only mechanically drive a difference between endogenous variables and observations. Second, from an econometric perspective, while our model also has a state-space representation, the challenge in estimation comes in the state equation, whereas the innovation in Ireland is on dealing with the measurement equation. Third, we take a Bayesian approach, we allow for VARs of higher order than one, and we focus on implications for business cycles, rather than on structural stability.

Also close to this paper is Rabanal, Ramirez and Tuesta (2008), who model the disturbances in an open-economy business-cycle model as being cointegrated. They also emphasize that strong restrictions on disturbance processes are unwarranted and can severely distort inferences. The point in this paper applies more generally, and we apply it to a different set of models and questions.

With the exception of Chari, Kehoe, and McGrattan (2007), almost all closed-economy business-cycle models assume uncorrelated disturbances. Smets and Wouters (2007) move slightly beyond the independent AR(1)s by allowing some of their disturbances to follow an ARMA(1,1). We take a step further allowing for a much richer correlation between disturbances. In the open-economy literature, it is common to assume that technology and monetary disturbances are correlated across countries.[1] Our methods hold the promise of allowing that literature to move further in its treatment of cross-country relations.

Del Negro and Schorfheide (2008) merge the versatility of a VAR with the tight restrictions of a DSGE in an innovative method that uses the DSGE to provide priors for the VAR. They discuss the Smets and Wouters (2007) generalization of the disturbances to ARMA(1,1) processes, and they focus on the normative use of these models for policy analysis. We propose a new set of estimation methods that allow for much more general disturbance processes, and our focus is on the positive predictions of models. We see disturbances as measures of ignorance that we want to characterize flexibly, not just to sharpen the reliability of inferences in current models, but also to suggest the desirable properties of future models that endogenize these disturbances.

In terms of methods, our paper fits into a burgeoning literature extending our ability to estimate more general DSGE models. Justiniano and Primiceri (2008) allow for time-varying volatility, Ramirez and Villaverde (2007) consider non-normal shocks, and Binsbergen, Koijen, Ramirez and Villaverde (2008) deal with recursive non-expected utility preferences. Our methods are complementary to these. Chib and Ramamurthy (2008) propose a multiple-block Metropolis-Hasting approach to DSGE estimation, with some re-

[1] This literature is too large to do full justice to here, but see Lubik and Schorfheide (2003) for an early estimated open-economy DSGE model and Justiniano and Preston (2008) for a recent one.

semblances with ours. A key difference is that while our blocks are suggested by the structure of the model, in their work it is the statistical properties of the data that guides the blocking of parameters.

1 Real business cycles with correlated disturbances

We use the extension of the Prescott (1986) model of fluctuations due to Christiano and Eichenbaum (1992). Aside from being familiar, this model has three merits for our purposes. First, it is sufficiently simple that the effect of correlated disturbances can be grasped intuitively. Second, it has generated some puzzles that we can re-examine. And third, it only has a few parameters, which makes the estimation method transparent.

1.1 The model of fluctuations

A social planner chooses sequences of consumption and work, $\{C_t, N_t\}_{t=0}^{\infty}$, to maximize

$$E_0 \left[\sum_{t=0}^{\infty} \beta^t \left\{ \frac{\left[C_t \left(1 - N_t\right)^{\theta} \right]^{1-\gamma} - 1}{1-\gamma} + V(G_t) \right\} \right], \tag{1}$$

subject to

$$\begin{aligned} Y_t &= C_t + K_t - (1-\delta)K_{t-1} + G_t, & (2) \\ Y_t &= \left(A_t N_t\right)^{1-\alpha} K_{t-1}^{\alpha}. & (3) \end{aligned}$$

The notation is standard.[2] Utility increases with consumption and leisure and the benefits of government spending enter additively through the function $V(.)$, so they have no effect on the positive predictions of the model. Equation (2) states that output equals consumption plus investment plus government spending, and equation (3) is a neoclassical production function. We use this DSGE model to predict the business cycle in output and hours

[2] In particular: C_t is private consumption, G_t is government consumption, N_t is the fraction of hours in a quarter spent at work, K_t is capital, Y_t is output, A_t is total factor productivity, β is the discount factor, $1/\gamma$ is the intertemporal elasticity of substitution, θ determines the relative utility from leisure and consumption, δ is the geometric depreciation rate, and α is the labor share.

worked (Y_t, N_t) in response to disturbances to productivity and government spending (A_t, G_t).

Some of the parameters are easily pinned down by steady-state relations.[3] Two parameters, though, are not and they are crucial to the model's business-cycle predictions. The elasticity of intertemporal substitution, $1/\gamma$, determines the willingness of households to shift resources over time. It is a key determinant of how strongly savings and labor supply respond to persistent productivity changes, and thus of the model's ability to generate sizeable output fluctuations. The parameter θ pins down the steady-state elasticity of labor supply with respect to wages. It is the key determinant of the size of the fluctuations in hours worked. We collect these *economic parameters* in the vector $\varepsilon = (\gamma, \theta)$.

Collecting the disturbances in the vector $s_t = (\ln(A_t), \ln(G_t/\bar{G})) \equiv (\hat{A}_t, \hat{G}_t)$, they follow a vector autoregression of order k:

$$s_t = \Phi(L)s_{t-1} + e_t \;\; \text{with } e_t \sim N(0, \Omega), \tag{4}$$

where $\Phi(L) = \Phi_1 + .. + \Phi_k L^{k-1}$, the Φ_i are 2x2 matrices, and Ω is a positive-definite symmetric 2x2 matrix. This is a quite general representation; beyond assuming linearity and covariance stationarity, it merely assumes that the order k is lafrge enough to be able to approximate well any arbitrary Wold process.

In the literature, it is typically assumed that disturbances follow independent AR(1)s, which in our notation maps into k being one and Φ_1 and Ω both being diagonal. These assumptions are hard to accept in this context. Government spending is certainly not an independent process in the data, and via the payment of unemployment benefits or countercyclical fiscal policy, G_t typically responds to A_t at least with a lag. Moreover, as noted in the introduction, Evans (1992) found that in the U.S. data A_t responds with a lag to G_t, which could plausibly result from productive externalities of government spending on private-sector productivity. In this model, because G_t and A_t are exogenous, their

[3] In particular, the discount factor, β, is set at 0.995, to generate a steady-state annual real interest rate of 2%, the production parameter, α, is 0.33, to match the capital income share, the depreciation rate, δ, is 0.015 to roughly match econometric estimates and the average U.S. capital-ouput ratio, the average level of productivity, $\bar{A}$, is normalized to 1, and the average government spending $\bar{G}$ equals its historical average of 20% of GDP.

correlations cannot be explained but must be assumed. It is then desirable to assume as little as possible on these measures of our ignorance by using the more general specification in equation (4), and focus instead on the tight restrictions imposed by the model on the endogenous variables.

One argument for assuming uncorrelated AR(1)s is that it reduces the number of parameters. Letting σ denote vector of *statistical parameters* in $\Phi(L)$ and Ω that describe the dynamics of the disturbances, with independent AR(1)s, σ has four elements. With unrestricted correlated disturbances, there are $3 + 4k$ statistical parameters. There is a curse of dimensionality as k increases, since the computational complexity of most estimation algorithms explodes even for modest values of k. However, as we show next, this is not a limitation of the theory, but rather of the particular algorithms being used.

1.2 Estimating the model

Log-linearizing the solution of the model around a non-stochastic steady state:

$$x_t = \lambda_1 \hat{K}_{t-1} + \lambda_2 s_t, \tag{5}$$

$$\hat{K}_t = \lambda_3 \hat{K}_{t-1} + \lambda_4 s_t, \tag{6}$$

where $x_t = (\hat{Y}_t, \hat{L}_t)$ are the observables, and a hat over a variable denotes its log-deviation. The state vector of the problem includes the exogenous s_t and the endogenous capital stock $\hat{K}_t$, and the λ_i are conformable matrices of coefficients that are functions of both ε and σ. These functions can be complicated, but are nowadays easily computed by many algorithms. Substituting out the unobserved capital stock, the reduced-form of the DSGE is:

$$x_t = \lambda_3 x_{t-1} + \lambda_2 s_t + (\lambda_1\lambda_4 - \lambda_3\lambda_2) s_{t-1}, \tag{7}$$

$$s_t = \Phi(L) s_{t-1} + e_t \text{ with } e_t \sim N(0, \Omega), \tag{8}$$

together with initial conditions s_0, x_0, and a transversality condition.

Recently, it has become popular to estimate models like this one by taking a Bayesian perspective.[4] Starting with a prior distribution for the parameters, $q(\varepsilon, \sigma)$, we use the reduced-form (7)-(8) to compute the likelihood function $\mathcal{L}(\{x\}|\,\varepsilon, \sigma)$, and then use Bayes rule to obtain the posterior distribution for the parameters:

$$p(\varepsilon, \sigma|\,\{x\}) = \mathcal{L}(\{x\}|\,\varepsilon, \sigma)q(\varepsilon, \sigma)/p(\{x\}).$$

The marginal posterior density of the data $p(\{x\})$ is unknown, and there is no convenient analytical form for the distribution $p(\varepsilon, \sigma|\,\{x\})$, so it must be characterized numerically. This is usually done with Markov Chain Monte Carlo (MCMC) algorithms, that draw a new (ε, σ) pair from an approximate distribution conditional on the last draw, in a way that ensures convergence of the draws to the posterior distribution.

The typical algorithm used is a random-walk Metropolis. At step j, this draws a proposal $(\varepsilon, \sigma)^{(j)}$ from a normal density with mean $(\varepsilon, \sigma)^{(j-1)}$ and some pre-defined covariance matrix, accepting this draw with a probability that depends on the ratio $p(\varepsilon, \sigma)^{(j)}/p(\varepsilon, \sigma)^{(j-1)}$, keeping $(\varepsilon, \sigma)^{(j-1)}$ in case of rejection. This algorithm is robust in the sense that it usually explores well the posterior distribution with minimal input from the researcher. The other side to this robustness is that, because it uses almost no knowledge of what the posterior should look like, the algorithm can take many draws to converge. Experience with DSGE models has found that it takes millions of draws to converge if there are more than ten parameters to estimate. With correlated disturbances, this algorithm quickly hits the curse of dimensionality.

We propose an alternative algorithm that avoids the curse of dimensionality by exploiting the economic structure of the model. Because its central observation is to use knowledge that some conditional posterior distributions are conjugate, we label it the *conjugate-conditionals* algorithm. It is based on three observations.

First, while we are interested in the parameters, there is also uncertainty on the realization of the innovations $\{e_t\}$ and the disturbances $\{s_t\}$. We can therefore focus on characterizing the joint posterior $p(\varepsilon, \sigma, \{s\}|\,\{x\})$, from which the marginal distribution of interest follows immediately.

[4]See Fernandez-Villaverde (2009) for a survey and a summary of the virtues of the Bayesian approach.

Second, note that conditional on the parameters, then the reduced-form (7)-(8) is a state-space system. The conditional distribution $p(\{s\}|\{x\}, \varepsilon, \sigma)$ is normal with mean and variance given by the Kalman filter. Moreover, conditional on the disturbances $\{s_t\}$, equation (8) is a standard vector autoregression, for which there are conjugate priors. In particular, if the prior distribution for $(\Phi(L), \Omega)$ is normal-inverse-Wishart, then the posterior distribution $p(\sigma|\{s\})$ is also normal-inverse-Wishart with analytical expressions for its moments.

Third, note that sampling from the joint posterior can be broken into three steps. Starting from a draw $(\varepsilon^{(j-1)}, \sigma^{(j-1)}, \{s\}^{(j-1)})$, we can draw $s^{(j)}$ from the conditional, $p(\{s\}^{(j)} \big| \{x\}, \varepsilon^{(j-1)}, \sigma^{(j-1)})$, then draw $\sigma^{(j)}$ from the conditional $p(\sigma^{(j)} \big| \{x\}, \varepsilon^{(j-1)}, \{s\}^{(j)})$, and finally draw $\varepsilon^{(j)}$ from the conditional $p(\varepsilon^{(j)} \big| \{x\}, \sigma^{(j)}, \{s\}^{(j)})$. This is the Gibbs-sampling method of breaking parameters in blocks, and in our case, there is a natural and convenient separation between disturbances, statistical parameters, and economic parameters.

Combining these three observations provides our algorithm. It draws from the expanded parameter vector $(\varepsilon, \sigma, \{s\})$ in turn, exploiting the knowledge that two of the three needed conditional distributions are known analytically. Only the conditional for ε is unknown, but this involves only two parameters, regardless of the assumptions on the disturbances. Allowing for correlated disturbances may dramatically increase the number of parameters in σ, but because the conditional posterior distribution for σ is known analytically, the curse of dimensionality is broken. Estimating a DSGE with correlated disturbances is not significantly harder than one with independent AR(1) disturbances, because it is not harder to draw from normals and inverse-Wishart distributions of higher dimension. Because it uses our knowledge of particular slices of the posterior distribution that we are trying to characterize, this algorithm should be more efficient than the standard Metropolis algorithm.[5]

[5]The statement has to be qualified, because it is possible that the co-dependence between ε and σ is so strong that the Metropolis algorithm ends up dominating the Gibbs-sampler. In our experience, this is not the typical case.

1.3 Data, priors, and the efficiency of the algorithm

We estimate the model on U.S. data for non-farm business sector hours and output per capita. The data is quarterly, HP-filtered, and goes from 1948:1 to 2008:2, although we use the data before 1960:1 only to calibrate the priors.

The priors are summarized in table 1. Informed by the estimation results of Christiano and Eichenbaum (1992), the prior modes for the economic parameters are $\gamma = 1.5$ and $\theta = 2.5$ and they have a gamma distribution. For the statistical parameters, the modes of the four AR(1) parameters (the diagonal terms of Φ_0 and Ω_0) are set to match four moments in the the data from 1948:1 to 1959:4, the two variances and serial correlations of output and hours. For the remainder statistical parameters, we consider two cases. In the first case, we follow the literature and assume independent AR(1)s. The priors for all of the correlated-disturbance terms is zero with zero variance. We include this case both because it provides the comparison point for the correlated-disturbances case, and because it provides an illustration of the relative efficiency of our new algorithm. Our focus is on the correlated case, and we present results for a unrestricted VAR(1). The three non-diagonal elements in Φ_0 and Ω still have a prior mode of zero, but have a non-zero variance set according to the extension of the Minnesota prior by Kadiyala and Karlsson (1997), tighter around zero the further we move from the diagonal.[6] We focus on the VAR(1) case because posterior odds tests selected this model over higher-order VARs, but the inferences seem to be robust.[7]

Our first set of results address the efficiency of the conjugate-conditionals algorithm versus the Metropolis random-walk. We simulated data of the same length as the sample using the priors for the independent AR(1), estimated the model on the simulated data using the two algorithms with four parallel chains, and then compared their relative efficiency at converging to the posterior distribution.[8]

[6] Section 3 discusses the prior in more details as well as alternatives within the conjugate-conditionals family.

[7] STILL IN PROGRESS.

[8] We have also compared the efficiency of the two algorithms using actual, rather than simulated data, and in the cases with correlated shocks. Because in the independent AR(1) case, the number of statistical parameters is smaller, the advantage of the conjugate-conditionals method should be smaller, so we focus on this case to be conservative.

We used four metrics to assess convergence. First, the R statistic of Gelman and Rubin (1992), which compares the variance of each parameter estimate between and within chains, to estimate the factor by which these could be reduced by continuing to take draws. This statistic is always larger or equal than one, and a cut-off of 1.01 is often used. We report the maximum of these statistics across all the parameters. Second, the number of effective draws, *neff*, in each chain for each parameter, which corrects for the serial correlation across draws following Geweke (1992). The larger this is, the more efficient the algorithm, and we again report the minimum of these statistics. Third, the number of effective draws in total, *mneff*, which combines the previous two corrections applied to the mixed simulations from the four chains (Gelman et al, 1998: 298), where again we report the minimum across parameters. Finally, the number of rejections at the 5% level of the z-test that the mean of the parameter draws in two separated parts of the chain is the same. This is the separated means test, *SPM*, of Geweke (1992) and fewer rejections implies being closer to convergence.

Figure 1 shows the results.[9] In the horizontal axes are the number of draws, and in the vertical axes the convergence metrics. The conjugate-conditionals algorithm clearly dominates the Metropolis random-walk. The number of effective draws is almost always higher, and commonly used thresholds like 1.01 for R, 300 for $neff$, or 2000 for $mneff$, are reached earlier. Since in this case, disturbances are uncorrelated, these figures provide a conservative estimate on the improvement to be had in switching to the conjugate-conditionals algorithm. When the disturbances are as rich as a VAR(4), the benefits from the conjugate-conditional approach over a random-walk Metropolis are larger.

1.4 Estimates of correlated disturbances

Starting with the independent AR(1)s case, the first panel of table 2 reports moments of the posterior distributions, and the top panel of figure 2 plots the prior and posterior distributions for each parameter. The median intertemporal elasticity of substitution is 1/1.6, which despite being lower than the prior is still substantially higher than the usual

[9] The proposal density for ε in the conjugate-conditionals algorithm is a random-walk Metropolis. The covariance matrix for the Metropolis algorithm is the Hessian at the mode of the posterior (found by numerical maximization), multiplied by a scale factor to obtain approximately a 20% acceptance rate. This is updated after 20,000 draws to the covariance matrix of these draws, and the algorithm is then re-started. We report the draws in this second run, after discarding the initial 12,500 for burn-in.

value of 0.2 that comes from Euler-equation estimates (Hall, 1988, Yogo, 2004). The posterior distribution for θ is close to the prior, and its median of 2.1 implies a Frisch elasticity of labor supply of 0.97.[10]

Figure 3 shows the impulse responses to one standard-deviation innovations to the two disturbances, as well as the unconditional variance decomposition. The impulse responses to productivity shocks display two known puzzles that have been associated with this model. In response to an improvement in productivity, output increases both because of the higher productivity, and also because the representative household chooses to work longer today when the returns to working are higher. As Cogley and Nason (1995) noted, the persistence of the output response closely mirrors the persistence of the technology disturbance, whereas most other estimates of these responses are more gradual. In turn, Gali (1998), Francis and Ramey (2005), and Basu, Fernald and Kimball (2006) estimated that hours fall after improvements in productivity, while Uhlig (2004) and Dedola and Neri (2007) find a response of hours close to zero, unlike the strong positive response in figure 3.

The variance decompositions present a third puzzle, since government spending disturbances account for three quarters of the variance of output and over 90% of the variance of hours. As noted by Christiano and Eichenbaum (1992), it is the "aggregate-demand" shock that drives the business cycle in this model, against most VAR studies (e.g., Shapiro and Watson, 1986), which attribute a larger role to productivity.

A fourth puzzling result comes from the impulse response of output to government spending. More public spending on the one hand lowers resources pushing down consumption and output, and on the other hand induces the pooer households to work harder. While on impact, output rises, because the estimated persistence of the government spending disturbance is quite high, output wuickly turns negative. Depending on the horizon taken, the cumulative government spending multiplier is between -0.02 and 0, even though Ramey (2008), among others, has found multipliers close to 1.

We now turn to the unrestricted VAR(1). The second panels of table 2 and figure 2 summarize the posterior distributions. The Frisch elasticity of labor supply is similar, 0.74,

[10] The coincidence of prior and posterior could be due to θ being weakly identified or just to random chance. We explored this by shifting the prior to the right and repeating the estimation. TO ADD

but the elasticity of intertemporal substitution is much lower. Its median is now 0.36, while its 5% lower bound is 0.25, bringing the DSGE estimates in line with the single-equation Euler equation estimates. The non-diagonal terms in the coefficient matrix Φ are both significantly different from zero, and the lagged technology term in the law of motion for government spending is quite large. According to these estimates, when productivity falls, there is a lagged increase in government spending, matching what we would expect from the automatic and discretionary stabilizers in U.S. fiscal policy.

Figure 4 plots impulse responses and variance decompositions in the VAR(1) case, identified following the lead provided by the results in Evans (1992) by assuming that productivity affects government spending with a one-quarter lag.[11] There are three striking differences relative to the AR(1) case. First, the response of output to a productivity disturbance is now significantly more delayed, partially overcoming the Cogley and Nason puzzle. The reason is that an increase in productivity leads to a subsequent fall in government spending. While this initially makes the impact on output smaller, after a few periods, it boost output up. Second, an improvement in technology now lowers hours, although only slightly, matching the results from the literature that followed Gali (1998). Third, technology now accounts for a much larger fraction of the business cycle in line with the evidence: three quarters for output and one half for hours, versus one quarter and 10% in the independent AR(1) case.

Allowing for correlated disturbances therefore resolves three long-standing puzzles associated with this model.[12] But is the VAR(1) model a better fit to the data than the independent AR(1)s? In term of model fit, the answer is a clear yes. The log marginal density of the VAR(1) model is 1,374, whereas that of the AR(1) model is 1,355, so the posterior odds ratio is an overwhelming 178 million in favor of the former.[13]

[11] With contemporaneously correlated disturbances, identification is an unavoidable issue. We have also tried two alternatives to the one in the main text: ordering productivity first in the Choleski decomposition of Ω, and estimating a restricted VAR(1) model where we impose that Ω is diagonal. These led to very similar results and the restricted VAR(1) case fit the data almost as well as the unrestricted case. This irrelevance of the identification strategy is a result of the posterior estimates of the non-diagonal element of Ω being close to zero.

[12] With regard to the fourth puzzle, the fiscal multiplier is higher with correlated disturbances, but it is still close to zero, very far from the target of one.

[13] TO ADD: OUT-OF-SAMPLE FORECASTING PERFORMANCE.

2 The general theory of the conjugate-conditionals algorithm

This section starts by laying out the set of models on which the algorithm may be used. It then presents the two conjugate conditional distribution, followed by the Gibbs algorithm together with alternative algorithms that exploit the conjugate conditionals. Finally, we discuss alternative priors and complications that may arise from restrictions on the statistical parameters. All proofs are in the appendix.

2.1 The model

We consider an economic model that relates the following vectors of variables:

- y_t : endogenous economic variables, of dimension n_y;
- s_t : exogenous disturbances, of dimension n_s;
- e_t : exogenous innovations to the disturbances, of dimension n_s;
- x_t : observables, of dimension n_x;
- ε : economic parameters, of dimension n_ε,
- σ : statistical parameters, of dimension n_σ,

in a sample $t = 1, ..., T$ with the convention that a variable dated t is determined at that date. The sample realization of a variable, say x_t, from $t = 1$ to date j is denoted by $x^j \equiv \{x_t\}_{t=1}^{j}$. We use $p(.)$ to denote a general posterior distribution and $q(.)$ to denote a prior distribution.

As first noted by Blanchard and Kahn (1981), and more recently by Sims (2001), the dynamics of many linear (or linearized) economic models are described by a system of equations:

$$\Psi_0(\varepsilon)y_t = \Psi_1(\varepsilon)y_{t-1} + \Psi_2(\varepsilon)s_t + \Psi_3(\varepsilon)w_t, \tag{9}$$

together with a set of boundary conditions, coming from initial states and transversality conditions. The Γ_i matrices typically have many zero elements and their size is higher than n_ε, embodying the cross-equation restrictions that come from optimal behavior, technologies and other constraints. The endogenous errors w_t arise from replacing terms involving expectations by $w_{i,t+1} = y_{i,t+1} - E_t(y_{i,t+1})$ so that $E_t w_{t+1} = 0$.[14]

The disturbances, s_t, are assumed to be linear covariance-stationary processes that are well approximated by a vector autoregression of finite order k:

$$s_t = \Phi(\sigma)(L)\, s_{t-1} + e_t, \text{ with } e_t \sim N(0, \Omega(\sigma)) \tag{10}$$

where $\Phi(\sigma)(L) = \sum_{j=1}^{k} \Phi_j(\sigma) L^j$, a matrix lag polynomial. If these matrices are left unrestricted, then the number of statistical parameter $n_\sigma = kn_s + n_s(n_s+1)/2$, which could be quite large. If there are restrictions on the disturbance processes, n_σ may be smaller and its components may have to lie in restricted sets.

The reduced-form solution to the system in equations (9)-(10) is:

$$\begin{pmatrix} y_t \\ \tilde{s}_t \end{pmatrix} = \begin{pmatrix} \Lambda_1(\varepsilon,\sigma) & \Lambda_2(\varepsilon,\sigma) \\ 0 & \tilde{\Phi}(\sigma) \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \tilde{s}_{t-1} \end{pmatrix} + \begin{pmatrix} \Lambda_3(\varepsilon,\sigma) \\ \Upsilon \end{pmatrix} e_t. \tag{11}$$

The first row has the rational expectations solution of the model, where the coefficient matrices Λ_i are typically complicated non-linear functions of the economic and shock parameters.[15] The second row re-writes the VAR in state-space form, where $\tilde{s}_t = \left(s'_t, ..., s'_{t-k+1}\right)'$, the large matrix $\tilde{\Phi}(\sigma)$ includes all the elements of $\Phi(\sigma)(L)$, and Υ is a selection matrix of zeros and ones. Letting $z_t \equiv (y'_t, \tilde{s}'_t)'$ be the state vector, we write this more compactly as:

$$z_t = \Gamma_1(\varepsilon,\sigma) z_{t-1} + \Gamma_2(\varepsilon,\sigma) e_t \tag{12}$$

The data is for a vector of variables x_t, which is a linear combination of the elements of z_t that may or may not involve the economic parameters:

$$x_t = H(\varepsilon)\, z_t. \tag{13}$$

In some cases, $n_x \geq n_z$ and the matrix H has full rank, so we can directly observe z_t. In most cases, this is not possible, and we deal with this general case. We abstract from

[14]We will treat y_t as deviations from a steady-state so we ommit constants from the system, but it is straightforward to include these.

[15]By the principle of certainty equivalence, the Λ_i might depend on Γ_i and Φ_i, but will not depend on Ω.

measurement error in these observations to avoid confusion with the economic disturbances specified in the model, but including measurement error does not change our conclusions significantly.

Given this setup, the estimation problem is the following: we observe x^T and we know its dynamics are determined by the state-space system in equation (12)-(13). Starting from a prior distribution $q(\varepsilon, \sigma)$, the goal is to characterize the posterior $p(\varepsilon, \sigma | \, x^T)$, and this is done numerically by simulating J draws. Note already that this is a different problem form having an unrestricted state-space model because the economic theory puts tight constraints on how the parameters in (ε, σ) affect $\Gamma_1(\varepsilon, \sigma)$ and $\Gamma_2(\varepsilon, \sigma)$, and because the $\tilde{s}_t$ component of the z_t vector involves only σ, not ε. It is this last feature that leads to the conjugate-conditionals method.

2.2 The conjugate conditional distributions

Our first result is the following.

Proposition 1 *A sufficient statistic for the conditional* $p\left(\sigma | \, x^T, \varepsilon, z^T\right)$ *is* s^T*:*

$$p\left(\sigma | \, x^T, \varepsilon, z^T\right) = p(\sigma | \, s^T).$$

Since the elements of σ *are* $\Phi(L)$ *and* Ω*, from the definition of a conditional probability:*

$$p\left(\sigma | \, x^T, \varepsilon, z^T\right) = p(\tilde{\Phi} \Big| \, s^T, \Omega) p(\Omega | \, s^T).$$

Then, if $q(\Omega)$ *is an inverse-Wishart with* ν_0 *degrees of freedom and scale matrix* Ξ_0*, and* $q(\bar{\Phi} \big| \, \Omega)$ *is a multivariate normal, where* $\bar{\Phi} = [vec(\Phi_1')', ..., vec(\Phi_k')']'$ *a* $n_s^2 k$ *vector with mean* Φ_0 *and variance* $\Omega \otimes \Theta_0$*, the posterior distributions belong to the same families, with* *posterior parameters:*

$$\begin{aligned}
\nu_1 &= \nu_0 + T - k && (14)\\
\Xi_1 &= \Xi_0 + S + \breve{\Phi}_0' \Theta_0^{-1} \breve{\Phi}_0 + \breve{\Phi}_{OLS}' X' X \breve{\Phi}_{OLS} - \breve{\Phi}_1' \Theta_1^{-1} \breve{\Phi}_1 && (15)\\
\breve{\Phi}_1 &= \Theta_1 \Theta_0^{-1} \breve{\Phi}_0 + \Theta_1 X' X \breve{\Phi}_{OLS} && (16)\\
\Theta_1 &= \left(\Theta_0^{-1} + X'X\right)^{-1} && (17)
\end{aligned}$$

where the VAR is written as $Y = X\breve{\Phi} + E$, with Y a $(T-k)xn_s$ matrix, X a $(T-k)xn_s k$ matrix, and $\breve{\Phi} = [\Phi_1, ..., \Phi_k]$. The auxiliary matrices are: $\breve{\Phi}_{OLS} = (X'X)^{-1}X'Y$ and $S = \left(Y - X\breve{\Phi}_{OLS}\right)' \left(Y - X\breve{\Phi}_{OLS}'\right)$.

The first part in the proposition notes that once we know z^T and extract s^T, then the second block of the reduced-form state-space system is simply a VAR. The second part shows that the family of normal-inverse-Wishart priors are conjugate for this problem (Kadiyala and Karlsson, 1997).

If the disturbances were observed in the data, this first proposition would suffice. In that case, from x^T we could back out s^T, and the joint posterior that we are after would be:

$$p(\varepsilon, \sigma | x^T) = p(\varepsilon | x^T, \sigma) p(\sigma | x^T) = p(\varepsilon | x^T, \sigma) p(\sigma | s^T).$$

The first equality follows from the definition of a conditional distribution, and the second equality uses the result that s^T is a sufficient statistic. Proposition 1 characterizes $p(\sigma | s^T)$ analytically, so in a first step, we could take the J draws for σ quickly and easily even when n_σ is very large. These draws for the statistical parameters can then be combined with the data to numerically characterize the unknown conditional for the smaller number n_ε of economic parameters $p(\varepsilon | x^T, \sigma)$.[16]

[16] To give one example where this would work, consider estimating a Calvo New Keynesian Phillips curve:

$$\pi_t = \beta E(\pi_{t+1}) + \chi^{-1}(1-\chi)(1-\beta\chi) mc_t + \varepsilon_t,$$

where π_t is endogenous inflation, mc_t is the observed exogenous marginal cost, ε_t is a measurement error, and the two economic parameters are the discount factor, β, and the fraction of fixed prices, χ. The disturbance in this model is mc_t, so allow it to follow an AR(k), with $k+1$ parameters. The conventional Metropolis random-walk algorithm would sample from the $k+3$ parameters in one block. Our method instead would

Most of the times (including the business-cycle models in this paper), we do not observe the disturbances directly. Proposition 1 can still be used as long as we do a parameter expansion and change the problem to drawing from the joint posterior $p(\varepsilon, \sigma, z^T | x^T)$. Since:

$$p(\varepsilon, \sigma, z^T | x^T) = p(\varepsilon, \sigma | x^T) p(z^T | x^T, \varepsilon, \sigma),$$

if we know the analytical form of $p(z^T | x^T, \varepsilon, \sigma)$, then sampling from this expanded posterior is not harder than the original distribution and we can recover the distribution of interest because $p(\varepsilon, \sigma | x^T) = \int p(\varepsilon, \sigma, z^T | x^T) dz^T$. In practice, retaining the (ε, σ) components of the draws gives a valid draw from their joint posterior.

The next result derives the analytical form for $p(z^T | x^T, \varepsilon, \sigma)$, by noting that this is a standard filtering problem (Kim and Nelson, 1999, Durbin and Koopman, 2001) on the state-space system in equations (12)-(13).

Proposition 2 *The state vector realizations z^T are normally distributed, conditionally on the observations x^T and the parameters (ε, σ). Their density can be factored into a product of normal densities:*

$$p\left(z^T \middle| x^T, \varepsilon, \sigma\right) = f\left(z_T | x^T, \varepsilon, \sigma\right) \prod_{t=1}^{T-1} f\left(z_t | z_{t+1}, x^t, \varepsilon, \sigma\right). \tag{18}$$

with means and covariances denoted by:

$$\begin{array}{ll} z_{t|t} \equiv E(z_t | x^t, \varepsilon, \sigma), & P_{t|t} \equiv E\left((z_t - z_{t|t}) (z_t - z_{t|t})' \middle| x^t, \varepsilon, \sigma \right), \\ z_{t|t+1} \equiv E(z_t | z_{t+1}, x^t, \varepsilon, \sigma), & P_{t|t+1} \equiv E\left((z_t - z_{t|t+1}) (z_t - z_{t|t+1})' \middle| z_{t+1}, x^t, \varepsilon, \sigma \right), \end{array} xx \tag{19}$$

These means and covariances are given by the Kalman filter recursions:

$$\begin{array}{ll} z_{t|t-1} = \Gamma_1 z_{t-1|t-1}, & x_{t|t-1} = H z_{t|t-1}, \\ P_{t|t-1} = \Gamma_1 P_{t-1|t-1} \Gamma_1' + \Gamma_2 \Omega \Gamma_2', & K_t = P_{t|t-1} H' \left(H P_{t|t-1} H' \right)^{-1}, \\ z_{t|t} = z_{t|t-1} + K_t \left(x_t - x_{t|t-1} \right), & P_{t|t} = \left(I - K_t H \right) P_{t|t-1}. \end{array} \tag{20}$$

---

instead first obtain J draws for the $k+1$ statistical parameters quickly from the known conditional for an AR(k) fit to the observed data on mc_t. Then it would use each of these draws as regressors in the Phillips curve to characterize the distribution of β and χ.

followed by a second set of recursions:

$$M_t = P_{t|t}\Gamma_1 P_{t+1|t}^{-1}, \quad z_{t|t+1} = z_{t|t} + M_t\left(z_{t+1} - \Gamma_1 z_{t|t}\right), \quad P_{t|t+1} = P_{t|t} - M_t P_{t+1|t} M_t' \tag{21}$$

Sampling from $p\left(z^T \middle| x^T, \varepsilon, \sigma\right)$ is an easy matter, as even for very large n_z, most software programs can take draws from the multivariate normal quickly. Moreover, while the Kalman filter recursions can take some time, they were required anyway in order to calculate the likelihood function. Given a draw for z^T, we can read off directly the draw for the state vector s^T or its extended counterpart $\tilde{s}^T$, and apply the result in the previous proposition.

2.3 Hybrid algorithms to exploit the conjugate conditionals

In the previous sub-section we already discussed one algorithm that exploits the conjugate conditionals in the first proposition when the disturbances are directly observed. We now describe algorithms, sometimes referred to as hybrid, Metropolis-within-Gibbs, or block-Metropolis, that exploit both propositions when the disturbances are not directly observed. These algorithms rely on the following result (Tierney, 1994).

Proposition 3 *A Markov Chain Monte Carlo that at step j has $(\varepsilon^{(j-1)}, \sigma^{(j-1)}, z^{T(j-1)})$ and then (i) draws $z^{T(j)}$ from the conditional, $p(z^{T(j)} \middle| x^T, \varepsilon^{(j-1)}, \sigma^{(j-1)})$; (ii) draws $\sigma^{(j)}$ from the conditional $p(\sigma^{(j)} \middle| x^T, \varepsilon^{(j-1)}, z^{T(j)})$, and (iii) draws $\varepsilon^{(j)}$ from the conditional $p(\varepsilon^{(j)} \middle| x^T, \sigma^{(j)}, z^{T(j)})$, will produce a set of J draws that converge to draws from the posterior distribution $p(\varepsilon, \sigma, z^T \middle| x^T)$.*

This is the Gibbs-sampling result that drawing in turn from conditionals converges to the joint. It cannot be directly applied to our problem because we do not know the conditional distribution for the economic parameters, $p(\varepsilon | x^T, \sigma, z^T)$. However, we can draw from a proposal distribution that will approximate it. There are a few approaches to do so:

1) *The random-walk Metropolis.* Starting with a draw $\varepsilon^{(j-1)}$, at step j (i) draw a candidate ε^* from a proposal density, $\pi(\varepsilon | \varepsilon^{(j-1)})$, (ii) compute the ratio $r = p\left(\varepsilon^* | x^T, \sigma, z^T\right) / p\left(\varepsilon^{(j-1)} \middle| x^T, \sigma, z^T\right)$, (iii) with probability $\min(r, 1)$ set $\varepsilon^{(j)} = \varepsilon^*$ and otherwise set $\varepsilon^{(j-1)} = \varepsilon^*$. In practice, we follow convention and use for proposal a normal distribution with mean $\varepsilon^{(j-1)}$ and covariance $c\mathcal{H}_{mo,\varepsilon}^{-1}$ where $\mathcal{H}_{mo,\varepsilon}$ is the Hessian for ε at the mode of the posterior distribution, and

c is a scaling factor adjusted to control the rate of convergence. As for the ratio r, using Bayes theorem,

$$r = \frac{\mathcal{L}\left(x^T|\varepsilon^*, \sigma^{(j)}, z^{T(j)}\right) q(\varepsilon^*)}{\mathcal{L}\left(x^T|\varepsilon^{(j-1)}, \sigma^{(j)}, z^{T(j)}\right) q(\varepsilon^{(j-1)})},$$

where $\mathcal{L}(.)$ is the likelihood function and $q(.)$ the prior.[17]

2) *The independent Metropolis.* The algorithm is identical to the previous one with the exception of the proposal density $\pi(.)$, which is now independent of $\varepsilon^{(j-1)}$. One strategy that often leads to good results is to look for local maxima of the posterior distribution. If there is only one, pick $\pi(.)$ to be a normal centered at the mode with covariance $c\mathcal{H}_{mo,\varepsilon}^{-1}$ where c is a factor well above 1. If there are many, pick $\pi(.)$ to be a mixture of normal centered at these modes, with mixing weights proportional to the mode's posterior densities.

3) *Rejection sampling.* In this algorithm, we pick an auxiliary density $\pi(\varepsilon)$, from which we sample a candidate ε^*, which is accepted with probability $\min(r', 1)$ where

$$r' = \mathcal{L}\left(x^T|\varepsilon^*, \sigma^{(j)}, z^{T(j)}\right) q(\varepsilon^*)/k\pi(\varepsilon^*).$$

The scalar k is set to ensure that r' is close to 1, and a typical choice for the proposal density is a normal centered at the mode, or a mixture of normals around local maxima.

We have not found that any of these algorithms clearly dominates the others. It depends on the model and he data. Finally, in our experience, it improved inference considerably to importance re-sample the J draws for (ε, σ), that is to draw from the original set of J draws, with probabilities proportional to their posterior density.

2.4 Alternative priors and parameter restrictions

Proposition 1 dealt with the case where we put no restrictions on the VAR matrices in (10). This may not be the case. For instance, in section 2 we wished to consider the independent AR(1)s case. In this section we summarize the conjugate families for restricted VARs, and in the appendix are the formulae for the posterior means and variances.

[17] The analytical for for the likelihood function is derived in the appendix.

If we assume that disturbances follow independent AR(k)s, then the Φ_j and Ω matrices are all diagonal. In that case, the family of conjugate priors and posteriors is: $i = 1, ..., n_s$ independent normals for $[\Phi_j(i)]_{j=1}^k$, and $i = 1, ...n_s$ independent inverse-gamma squared for each of $\Omega(i)$. The means and variances of the prior distributions can be chosen freely.

More generally, we may wish to impose that some of the elements of Φ_j and Ω are zero, or appear more than once. In this case, the system in (10) is a system of seemingly unrelated regressions (SUR). Collecting the disturbances into the vector $\bar{s}$ of size $n_s(T-k)$, it is written as

$$\bar{s} = Z\beta + \varepsilon, \quad \text{with } \varepsilon \sim N(0, \Omega \oplus I_{t-k}),$$

where Z contains the lagged states as well as blocks of zeros allowing for a rich set of restrictions on the VAR. The coefficients β include the elements of Φ; for instance, with an unrestricted VAR, the $\bar{s}$ are stacked along the diagonal blocks of the Z matrix. In this general SUR case, as long as $\beta|\,\Omega$ is normal and Ω^{-1} has a Wishart distribution, then so will the posteriors (Zellner, 1962).

TO ADD: DISCUSSION OF SIMS-ZHA (1998) PRIORS.

Finally, in some models, the transversality conditions impose the constraint that the VAR in (10) is stationary. There is no conjugate family for both Φ and Ω in this case, but there is a conjugate family for Φ conditional on Ω, the truncated normal in the region of stationarity. In this case, the conditional distribution of Ω is not known must be characterized using the hybrid algorithms described in the previous sub-section. Still, we at least know the conditional distribution for Φ and since this is where the curse of dimensionality bites as k increases, even in this worst-scenario case, our conjugate-conditionals should offer some efficiency improvements.

3 Monetary business cycles with correlated disturbances

TO BE ADDED

4 Conclusion

TO BE ADDED

Appendix

A.1. Section 1. TO BE ADDED

References

Christiano, Lawrence J. and Martin Eichenbaum (1992). "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations." *American Economic Review*, 82 (3), 430-50.

Del Negro, Marco and Frank Schorfheide, forthoming. "Monetary Policy Analysis with Potentially Misspecified Models. " *American Economic Review.*

Evans, Charles L. (1992). "Productivity shocks and real business cycles." *Journal of Monetary Economics*, 29 (2), 191-208.

Fair, Ray S. (2004). *Estimating How the Macroeconomy Works.* Harvard University Press.

Ireland, Peter N. (2004). "A method for taking models to the data." *Journal of Economic Dynamics and Control*, 28 (6), 1205-26.

Smets, Frank and Rafael Wouters (2007). "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach." *American Economic Review*, 97 (3), 586-606.

Zellner, Arnold (1962). "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association*, 57, 348-68.

Table 1. Prior Distribution

Parameter	*Density*[a]	*Mode*	*Percentile* 5	*Percentile* 50	*Percentile* 95
Economic					
γ	1 + *G*	1.500	1.1730	2.0108	4.1444
θ	*G*	2.500	1.6151	2.6336	4.0113
Statistical					
Panel A. Independent AR(1)s					
Φ_A	*N*	0.5960	0.2956	0.5964	0.8994
Φ_G	*N*	0.5278	0.2290	0.5276	0.8298
Ω_A	IG^2	0.0002	0.0001	0.0004	0.0018
Ω_G	IG^2	0.6990	0.4421	1.2495	5.9013
Panel B. Unrestricted VAR(1)					
Φ_{AA}	*N*	0.5960	0.2949	0.5960	0.8975
Φ_{AG}	*N*	0.0000	-0.0053	0.0000	0.0053
Φ_{GA}	*N*	0.0000	-17.353	-0.0232	17.312
Φ_{GG}	*N*	0.5278	0.2236	0.5273	0.8262
Ω_{AA}	*IW*	0.0002	0.0002	0.0005	0.0024
Ω_{AG}	*IW*	0.0000	-0.0433	0.0001	0.0438
Ω_{GG}	*IW*	0.6990	0.5873	1.6621	7.8314

a. The densities are the gamma (*G*), normal (*N*) and the inverse-Wishart (*IW*).

Table 2. Posterior Distributions

			Percentile		
Parameter	*Mean*	*Mode*	*5*	*50*	*95*
Panel A. Independent AR(1)s					
Economic					
γ	1.7638	1.3192	1.1425	1.6710	2.7097
θ	1.1845	0.8995	0.7504	1.1226	1.8221
Statistical					
Φ_A	0.8039	0.7947	0.7299	0.8035	0.8780
Φ_G	0.7429	0.7413	0.6703	0.7454	0.8074
Ω_A	0.0002	0.0001	0.0001	0.0002	0.0002
Ω_G	0.3292	0.2982	0.2375	0.3212	0.4456
Panel B. Unrestricted VAR(1)					
Economic					
γ	2.8243	2.8389	1.9133	2.7546	3.9637
θ	2.6287	2.7177	1.5467	2.5625	3.9266
Statistical					
Φ_{AA}	0.9244	0.9411	0.8237	0.9374	0.9690
Φ_{AG}	0.0050	0.0048	0.0042	0.0050	0.0058
Φ_{GA}	-7.2108	-7.1836	-9.6423	-7.3625	-4.4326
Φ_{GG}	0.8411	0.8826	0.6107	0.8720	0.9148
Ω_{AA}	0.0002	0.0001	0.0001	0.0002	0.0002
Ω_{AG}	0.0071	0.0069	-0.0013	0.0070	0.0153
Ω_{GG}	2.8654	1.8943	0.8045	1.9932	7.9015

Figure 1. Convergence: Random-walk Metropolis versus conjugate conditionals

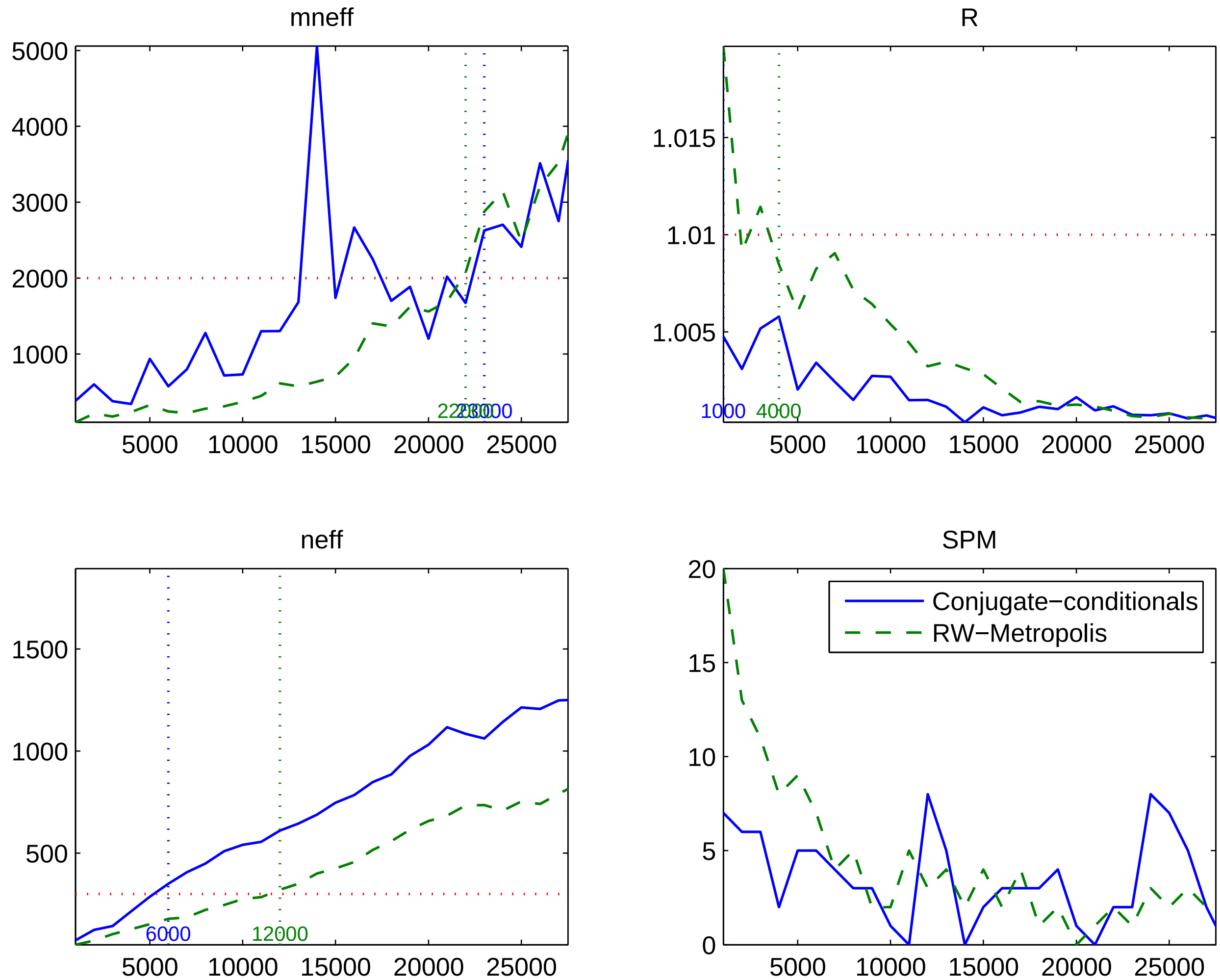
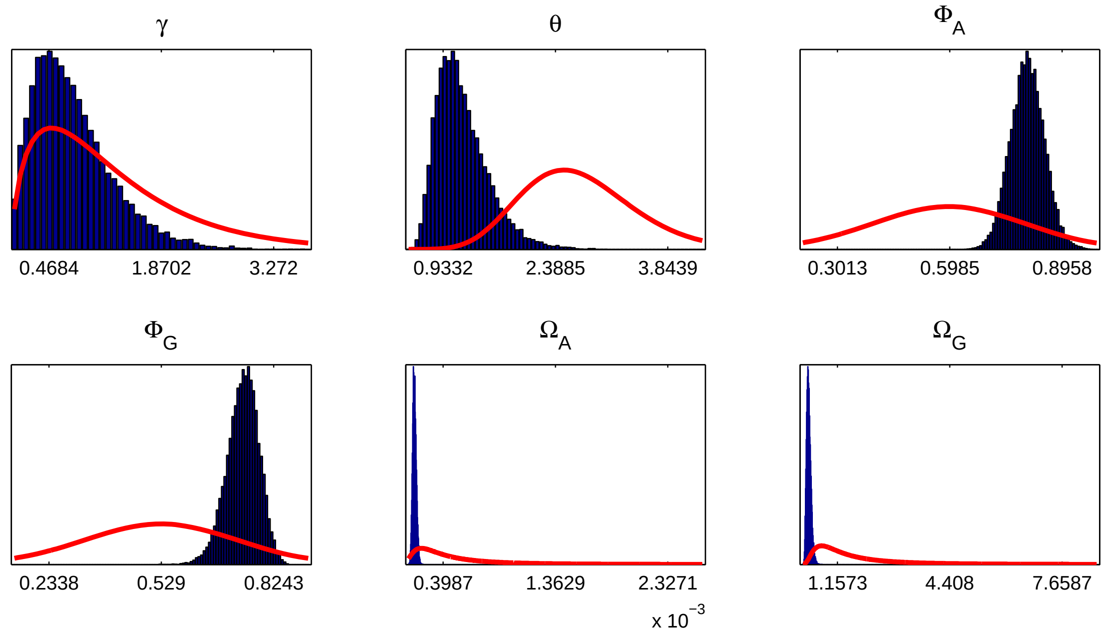


Figure 2. Prior and posterior distributions for the parameters

Panel A. Independent AR(1)s case



Panel B. Unrestricted VAR(1)

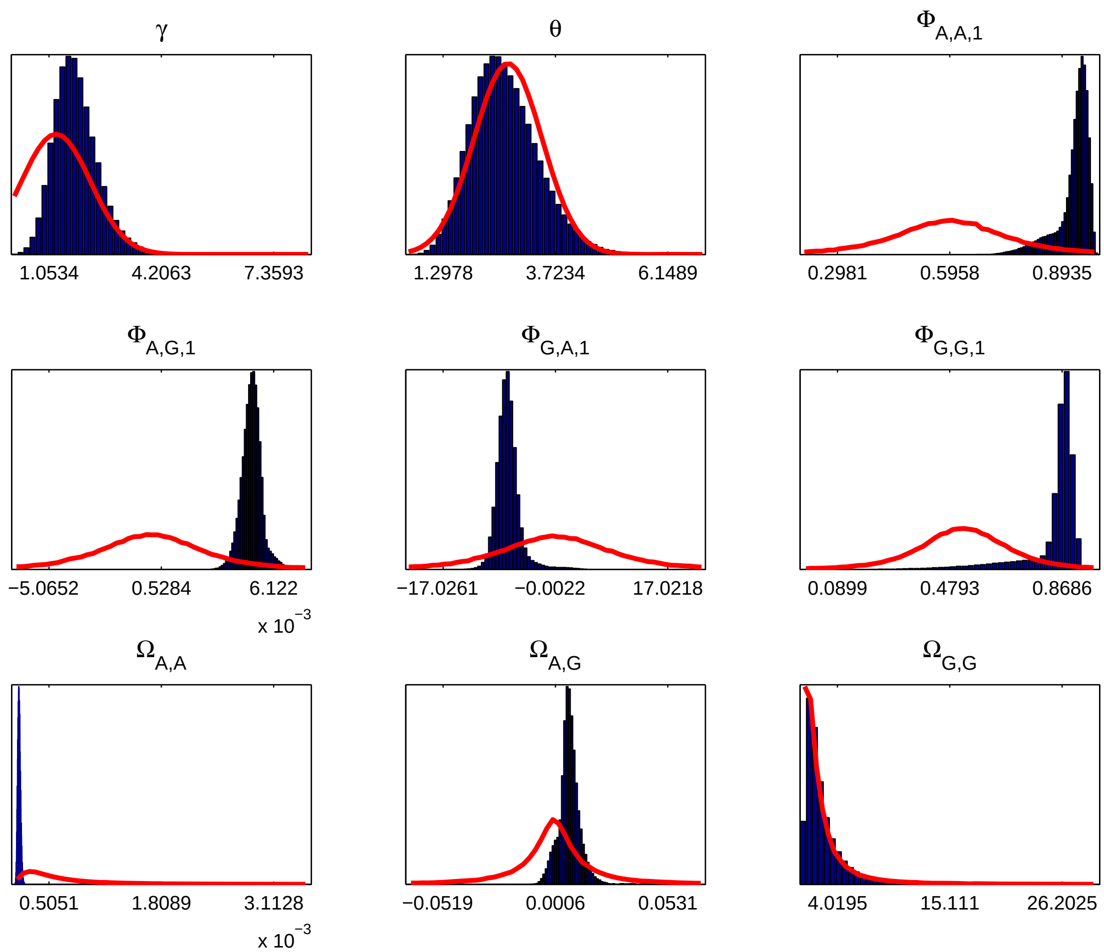


Figure 3. Impulse response functions with independent AR(1)s

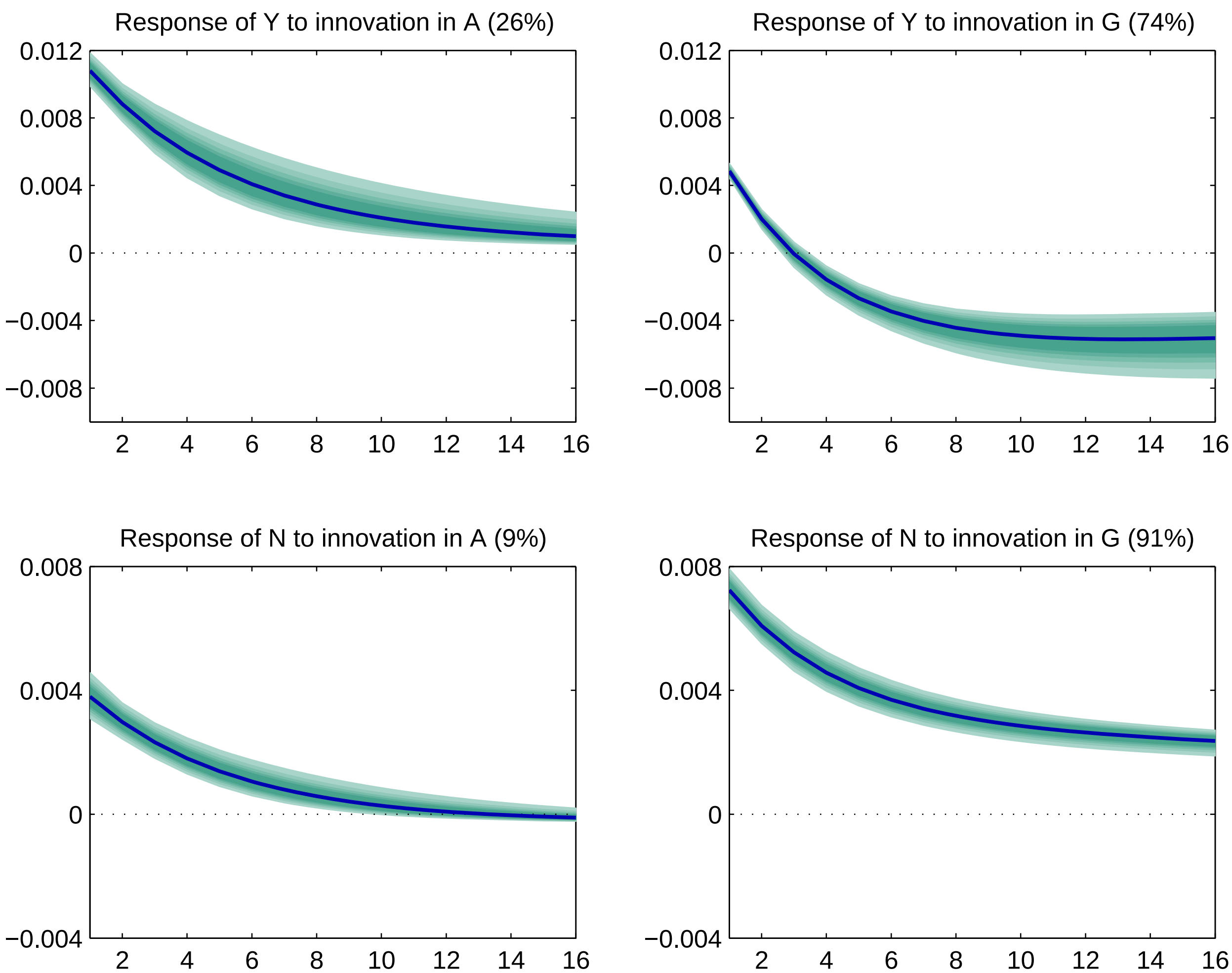
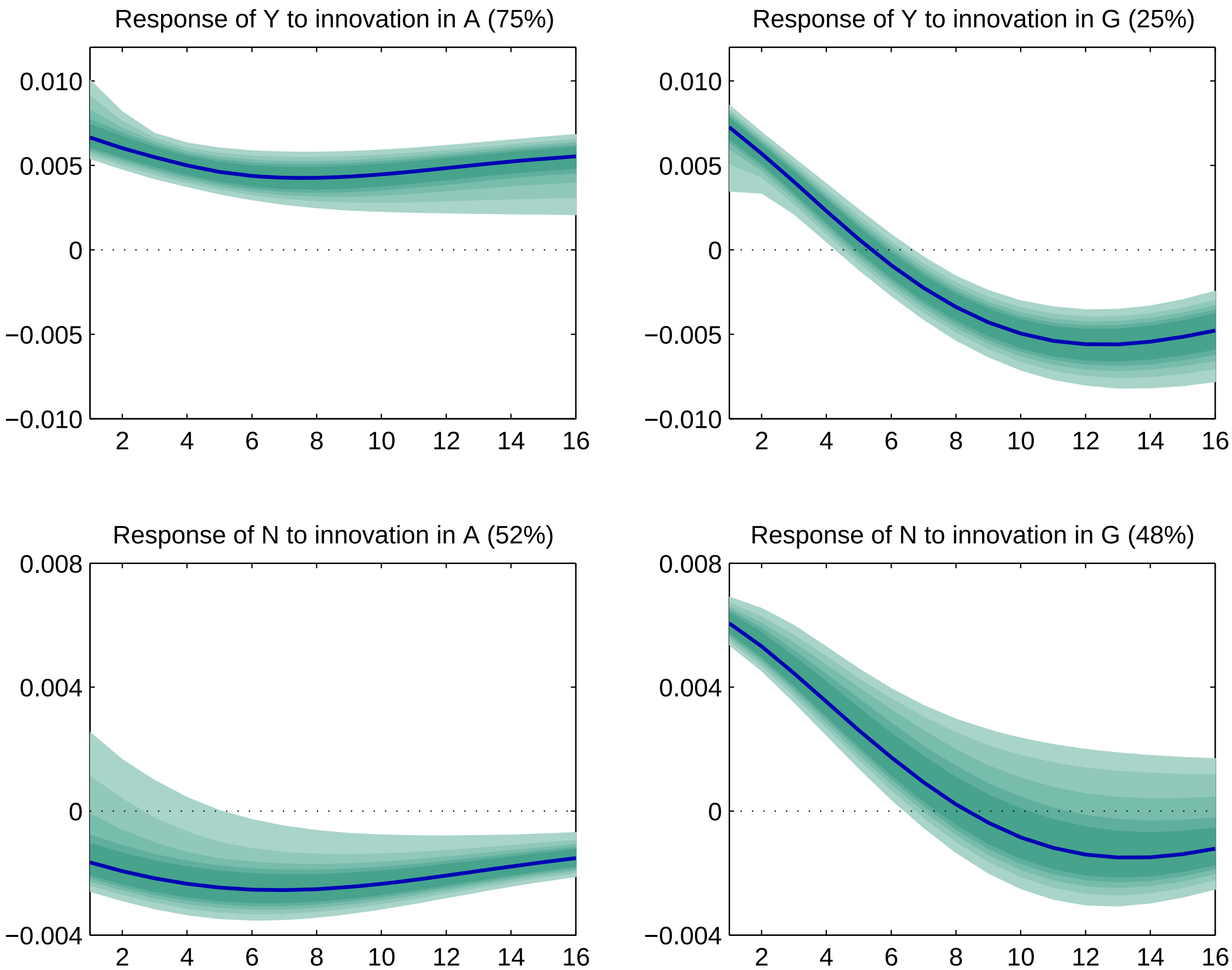


Figure 4. Impulse response functions with unrestricted VAR(1)



a