MAKING MACRO MODELS BEHAVE REASONABLY

CHRISTOPHER A. SIMS

ABSTRACT. Using the idea of generalized dummy observations, we extend the methods of Del Negro and Schorfheide, who have proposed a way to use a dynamic stochastic general equilibrium (DSGE) model to generate a prior distribution for a structural vector autoregression (SVAR). The method proposed here is more explicit and systematic about the prior's assertions about the SVAR identification, and it provides a mechanism for varying the tightness of the prior across frequencies, so that for example the long run properties of the DSGE can be asserted more confidently than its short-run behavior.

In every large scale macro modeling project we make compromises. Models whose properties we understand and can interpret behaviorally are generally not rich enough, either in number of variables or in dynamics, to fit the data well. Models that are rich enough to fit well become complex and can end up having implications that we believe are implausible. Practical macro modelers who face real-time demands for forecasts and policy projections have struggled continuously with these tradeoffs.

Date: November 8, 2006.

Key words and phrases. DSGE,macroeconomic model,Bayesian.

Forecasts and short-term policy projections from models are often modified with "add factors". Longer term policy analysis exercises with models often leave estimated equations aside altogether, or impose properties on the model's behavior that the data are not allowed to alter. These measures reflect a fact about modeling: often we construct a model, find it does not have quite the properties we would like, and then intervene by one means or another to push its results in a more reasonable direction. In principle of course, the right course would be to reformulate the model so that it stops displaying the properties we find unreasonable, but there are two problems with this suggestion. One is that reformulating the model may be difficult; if we clearly understood the mapping from the model structure to the model implications that bother us, we probably would not have built the model to produce these results in the first place. Another is that repeated, unsystematic, alterations of the model to fix one unwanted property or result after another can increase the complexity of the model and introduce new types of unwanted behavior faster than it fixes old ones.

Bayesian procedures hold the promise of letting us make the interaction of our beliefs about desirable model properties with the model formulation process more explicit and systematic. There are two ways to proceed here, both of which show up even in some of the earliest papers in the literature [DeJong et al., 1996, 2000, Ingram and Whiteman, 1994]. One can use Bayesian methods to allow use of more densely parameterized behavioral models, or one can use Bayesian methods to import beliefs based on behavioral models (perhaps not so densely parameterized)

into densely parameterized models without a complete behavioral interpretation. Bayesian methods are important for the first approach because they allow us, using modern Bayesian computational methods, to handle inference on models with many parameters. Also, with models that are richly enough parameterized to fit the data well, use of prior distributions is essential to any reasonable inference.

The first approach has advantages. It provides us with more complete stories about what behavioral mechanisms produce a given forecast, forecast error, or policy scenario. As we develop more models of this type, we will learn what aspects of a behavioral model the macroeconomic data does or does not pin down. It has disadvantages also, however. At the current stage, it seems that these models more often than not fail to fit as well as models with little or no behavioral structure. Furthermore, as they try to come close to fitting as well as descriptive time series models, the behavioral models tend to introduce frictions and lags that, while essential to the fit, have weak foundations in economic theory.

The second approach has been pursued recently by DelNegro and Schorfheide [2004].[1] It is computationally practical, and as they have implemented it results in a structural VAR (SVAR) model with a prior informed by a dynamic general equilibrium (DSGE) model. A model formulated this way is likely to fit quite a bit better than a purely behavioral model, and since it is a structural VAR, it allows for

[1] A recent paper [DelNegro et al., 2006] uses the same methods, but discusses results as if it is using the less structured model to inform estimation of a behavioral model. In fact, in both papers the data is modeled as generated by a structural VAR. The behavioral model is only a means to generation of a prior.

a substantial part of the "story-telling" that goes into discussing policy scenarios and the reasons for forecast failures. In SVAR's, each event or forecast is generated by independent, behaviorally interpretable disturbances. What is missing is only the possibility of using the model itself to evaluate welfare effects of policies or the effects of drastic policy interventions.

Bayesian methods in the abstract seem to require that one proceed directly to specifying prior beliefs about model parameters. As we have already noted, in practice economists tend to have ideas about how a model should behave that are not easy to translate into properties of a probability distribution for the parameters. It can therefore be helpful to note that priors can be built up from "dummy observations". This is an idea that perhaps began with Theil's adding mental observations to the data set to resolve multicollinearity in linear regression models. But the idea is more general than that. For example, suppose we have prior beliefs about a model's impulse responses. The full set of impulse responses in a SVAR model is a large 3-dimensional array, larger than the vector of parameters in the SVAR, usually. This means that any prior on the impulse responses that could be translated into a prior on the parameters of the SVAR would be singular, concentrated on a low-dimensional manifold in the full space of impulse responses. Furthermore, the mapping between parameters and impulse responses is highly nonlinear, so that computationally and intuitively demanding Jacobian terms are involved in connecting a pdf for impulse responsess to one for the model parameters.

But mental observations on the impulse responses need involve no Jacobian terms, and there is no need for them to be limited in dimension to the number of parameters in the model. They provide a much more intuitively clear route to building up a prior distribution in high-dimensional models than a direct attack on formulating a pdf for parameters. The idea is simple — independent mental observations (dummy observations) are combined as a product of functions of parameters (like, e.g., e to minus the mean square deviations of some impulse response from an a priori likely form) that penalize implausible behavior of the model. Technical details, and a discussion of how a prior formulated this way affects model comparisons, appear in Sims [2005]

This paper presents a proposed extension of the DelNegro/Schorfheide methodology. It attempts to be more consistent and explicit about the connection of the DSGE to the equations of the SVAR and to develop posterior inference directly on the parameters of those equations. It also uses an approach that allows giving different emphasis to prior beliefs at different frequencies. And it does these things with an approach based on the idea of generalized dummy observations.

I. USING A DSGE AS A STRUCTURAL PRIOR

We take the model that generates the data and that allows our structural interpretation of the sources of variation in the data to be in the form

$$A(L)y_t = c + \varepsilon_t\,, \tag{1}$$

where ε is a vector of independent normal shocks whose variances are one or zero. In other words, $\varepsilon_t \sim N(0, \Gamma)$, where Γ is a diagonal matrix with ones and zeros on the diagonal. Not all the elements of y are necessarily observable. While this model has the form of a structural vector autoregression (SVAR), it differs from SVAR's in exiting literature in that it allows for zero-variance elements of ε_t and postulates that the y_t vector is not necessarily the observed data vector. Dynamic stochastic general equilibrium models (DSGE's), when linearized, usually have this form, with c and $A(L)$ both functions of the model's behavioral parameters.

We will in fact assume we have available a DSGE model that considers the same y_t vector and ε_t vector and can be linearized around a (possibly non-unique) steady state to take the form

$$A(L;\theta)y_t = c(\theta) + \varepsilon_t \,. \tag{2}$$

We assume that, like most DSGE's in the literature, this model is too tightly parameterized to fit the data in detail. Nonetheless we expect it to be a fairly good approximation, and it is all we have available as a source of identifying restrictions to let us interpret historical variation and project the effects of possible current and future events or policy actions.

We use the DSGE together with a prior distribution for its parameters θ to generate a prior distribution for $A(L)$, the SVAR parameters. We postulate that conditional on θ, $(A(L), c)$ has a distribution that is centered on $\big(A(L;\theta), c(\theta)\big)$, the linearized DSGE coefficients. It is likely that the DSGE has very low order dynamics, so the order of our $A(L)$ polynomial for the SVAR model is likely to be higher

than implied by the DSGE. This means that the $A(L;\theta)$ polynomial we take as a mean for $A(L)$ has some zero coefficient matrices at higher order lags.

We express our beliefs about the connection of $A(L), c$ to $A(L,\theta), c(\theta)$ via a set of dummy observations, with the j'th dummy observation taking the form

$$A(L)\bar{y}^j - c\bar{x}^j = A(L;\theta)\bar{y}^j - c(\theta)\bar{x}^j - \bar{\varepsilon}^j \,. \tag{3}$$

The dummy observation shocks $\bar{\varepsilon}^j$ have a diagonal covariance matrix Δ that does not depend on j, though not necessarily the same covariance matrix Γ as the SVAR shocks. We have to scale the dummy observations to reflect the relative strength of our beliefs in them. The $\bar{x}^j$'s that multiply c in the dummy observations in place of the usual unit vector can arise from this scaling; they can also arise from dummy observations that make assertions about coefficients in $A(L)$ without reference to c.

These dummy observations are interpreted as "mental observations" on A, c, not as observations on data satisfying the model (1). That is, we regard each dummy observation as generating a factor in the log prior density of the form

$$-\tfrac{1}{2}\log|\Delta| - \tfrac{1}{2}\sum_{\Delta_{ii}\neq 0}\frac{(\bar{\varepsilon}_i^j)^2}{\Delta_{ii}^2}\,, \tag{4}$$

i.e. a standard Gaussian density for $\bar{\varepsilon}^j$, with $\bar{\varepsilon}^j$ defined by solving (3).

If there are n elements of y and k lags in the system, any set of $n(k+1)+1$ or more dummy observations of this form with linearly independent $\bar{y}^j, \bar{x}^j$ vectors will define a proper Gaussian prior for $A(L), c$ with mean $A(L;\theta), c(\theta)$. (Here of course we are, to be more precise, putting a distribution on the coefficients in $A(L)$

and $A(L;\theta)$, not the matrix polynomials in the lag operator themselves.) A prior formulated this way is not a completely general Gaussian prior on these coefficients. If we use A^* to refer to the $n \times (nk+1)$ matrix $[A_0, A_1, \ldots, c]$ and $\bar{y}^{j*}$ to refer to the column vector obtained by stacking up the current and lagged variable vectors in $\bar{y}^j$ and $\bar{x}^j$, then we are using dummy observations only of the form $A^*\bar{y}^{k*}$. In other words, the $\bar{y}^j$ weights applied to coefficients in one equation (row of A^*) also apply to every other equation. This means that, while we can make the means of coefficients vary across equations, their covariance matrix has the same form, up to a scalar multiple, in every equation. The form of the distribution implied by the dummy observations is Gaussian with

$$E[A^*] = A^*(\theta) \tag{5}$$

$$\text{Var}(\vec{A}^*) = \left(\sum \bar{y}^{j*}\bar{y}^{j*\prime}\right)^{-1} \otimes \Delta\,. \tag{6}$$

There is a one-dimensional redundancy in this parameterization, which can be resolved for example by setting one diagonal element of Δ to one as a normalizing convention.

Δ may contain zeros on the diagonal, for example for equations that are identities. The corresponding equations in the SVAR are then constrained by the prior to exactly match their counterparts in the DSGE. If all the behavioral equations in the DSGE have error terms, then all the zeros on the diagonal of Δ will correspond to those on the diagonal of Γ in the DSGE. More generally, though, the DSGE might contain fewer shocks than the SVAR, so that Γ has more zeros than Δ.

Allowing for more general Gaussian priors on A^* is straightforward, but can lead to much higher-dimensional matrix algebra in evaluating the posterior. In particular applications, with particular motives for a more general specification of the covariance matrix, problem-specific coding could keep the dimensionality increase from being too great; and the computations may be feasible in general, if somewhat slower than with pure dummy observation priors of the form described above. In the application considered in this paper we stick to these pure dummy observation priors.

If we use this prior as it stands, there is an undesirable byproduct. The inverse of the model's implied covariance matrix of one-step-ahead forecast errors is $A_0'A_0$. Our Gaussian density puts positive density on singular A_0 matrices, hence on unbounded covariance matrices of prediction errors. It turns out that with A^* normal, some elements of the reduced form coefficient matrix $B^* = A_0^{-1}A^*$ will be distributed with Cauchy tails, i.e. have no finite moments of integer order. This can be easily seen for the trivial special case where A^* is 1×2 with mean zero and identity covariance matrix, in which case the single reduced form coefficient is $b = a_1/a_0$, well known to be exactly a Cauchy random variable with pdf proportional to $1/(1+b^2)$.

If $\bar{y}^j$ were data generated by the model, the first column of $\bar{y}_j$, representing current data, would be thought of as the realization of a random variable correlated with $\bar{\varepsilon}^j$, and conversion of the pdf for $\bar{\varepsilon}^j$ to one for $\bar{y}^j$ would require inclusion of $|A_0|$ as a Jacobian term. The natural conjugate prior for this model — a prior that takes

the same form, as a function of the parameters as does the likelihood — therefore includes an $|A_0|^q$ Jacobian term, where q in the likelihood would correspond to the number of observations. This enforces zero density at points where A_0 is singular. Such a term in the prior to downweight singularities in A_0 is reasonable. Because the determinant is just the product of eigenvalues, it is invariant to orthogonal rotations of A_0 and thus in itself puts no restrictions on the location of zeros in A_0, for example. In a sense it expresses beliefs about the overall scale of A_0, not its form. Note that unlike the case where a prior is placed on B^*, the reduced form coefficient matrix, here we do not need the $|A_0|^q$ term to make the prior integrable in $A(L), c$. It will be integrable even with $q = 0$. It is a good idea to choose $q \geq 1$, however, in order to avoid giving much weight to singular A_0's.

With $q > 0$ the mode of A_0 is shifted away from $A_0(\theta)$ in the direction of larger absolute values of its eigenvalues, though the amount of shift is modest if the dummy observations, prior to the premultiplication by $|A_0|^q$, were fairly tightly concentrated around $A_0(\theta)$. A rough idea of the effect can be obtained from the scalar case, where the derivative of the log of the mode of the A_0 distribution with respect to q is $\Delta / \left(\sum(\bar{y}^j)^2 A_0(\theta)^2\right)$. In other words, the effect of a unit increase in q on the prior mode is small if the ratio of the standard deviation of the initial Gaussian form of the prior to that distribution's prior mean is small. If the prior standard error of the Gaussian form is one quarter of its mean, for example, a unit increase in q increases the prior mode of A_0 about 6%. Of course this gives only

a rough idea of the effect for multivariate cases, and it therefore is reasonable to check sensitivity of results to the choice of q.

II. Choosing the dummy observation weights

The most straightforward way to set the prior would be to have one independent dummy observation setting the prior mean for each variable and lag, plus one for the constant term. Such a dummy observation would have $\bar{y}^j = 0$ except at one lag for one variable. Indeed in the first version of the Minnesota prior, as described in Litterman [1986], the prior took exactly this form (except that it was on reduced form VAR, rather than on SVAR, coefficients). Users of Bayesian VAR's (BVAR's) soon realized, though, that they got better performance from a more realistic prior that was more assertive about low-frequency than high-frequency variation in the data. Dummy observations that asserted independent normal priors for sums of coefficients, as in Doan et al. [1984], proved useful. Such dummy observations imply negative cross-correlation in beliefs about coefficients at different lags and focus on the implied low-frequency behavior of the data.

Economists seem to put more credence in the implications of DSGE's at low frequencies than at higher ones, at least compared to linear time series models. This is evident in the practice at central banks of using models with a "core" that is more or less insulated from influence by data. The notion of a core model is most explicit in the recent Bank of England Quarterly model [Harrison et al., 2005], where the core is a DSGE, but it is also present in the US Federal Reserve Board's

FRBUS and in a number of central bank models that have followed the lead of the Bank of Canada's QPM.

We can mimic the behavior of the Minnesota prior and, roughly, the revealed priors of policy modelers, by using $\bar{y}^j$'s with different frequency characteristics and weighting them differently. To be specific we can use, where k is the number of lags (and hence $k+1$ the numbers of columns in A^* corresponding to y's rather than x's)

$$\bar{y}^{ji}_{i\ell} = \begin{cases} \alpha\sigma_i \dfrac{1}{\sqrt{2}} & j = 1 \\ \alpha\sigma_i \cos\left(\dfrac{\pi j(\ell-1)}{k+1}\right) \ell^\gamma \left(\dfrac{j}{2}\right)^{-\omega} & j \text{ even, } 1 < j < \frac{k+1}{2} \\ \alpha\sigma_i \sin\left(\dfrac{\pi (j-1)(\ell-1)}{k+1}\right) \ell^\gamma \left(\dfrac{j-1}{2}\right)^{-\omega} & j \text{ odd, } 1 < j < \frac{k+1}{2} \\ \alpha\sigma_i \dfrac{(-1)^{\ell-1}}{\sqrt{2}} & j \text{ even, } j = k+1\,. \end{cases} \tag{7}$$

$$\bar{y}^{ji}_{h\ell} = 0,\ h \neq i \tag{8}$$

Each dummy observation asserts, with more or less confidence determined by the weighting, that a pattern of variation in the data that produces a given residual vector in the DSGE should produce a similar residual vector in the SVAR. Here j, i indexes the dummy observation, i indexes variables, and ℓ indexes the lag. The γ parameter, when positive, makes the prior tighter for more distant lags. The ω parameter, when positive, makes the prior bind less tightly at higher frequencies. The σ_i parameters are prior estimates of the relative scales of variation in the variables. α is an overall scale parameter determining the tightness of the prior. The most natural interpretation of the dummy observations in this group are that they

make assertions about the effects of deviations form the steady state or trend path, and the $\bar{x}^{ji}$'s, the weights on the constant term are therefore set to zero.

Note that these $\bar{y}^{ij}$'s, as functions of ℓ, are the standard entries in a finite Fourier transform matrix, except that the term in γ makes them increase with lag length and the term in ω makes the rows smaller at higher frequencies. They are also closely analogous to the dummy observations that form the Minnesota prior. The $j = 1$, or zero-frequency, components are the "sums of coefficients" dummy observations used in the Minnesota prior, except for the γ term and the fact that they shrink toward the DSGE rather than toward independent random walks. The Minnesota prior includes separate dummy observations for each variable and lag in addition to the sums-of-coefficients dummy observations. Here instead we include the higher frequency dummy observations. The Minnesota prior approach is equivalent to giving the dummy observations on different frequencies equal weights at all frequencies other than zero.

We have so far just $(k+1)n$ dummy observations and need one more to have a proper prior over all of the A_j's and c. We can add one, call it $\bar{y}^{1,n+1}$, in which all variables take their $j = 1$ values from (7) with $\gamma = 0$ and $\bar{x}^{ji} \neq 0$. We choose $\bar{x}^{ji}$ so that this dummy observation asserts that the SVAR has approximately the same steady state as the DSGE. To do this, in case the DSGE is stationary, we set $\bar{y}^{1,n+1} = \lambda \bar{y}_0(\theta)$ at all lags, where $\bar{y}_0(\theta)$ is the steady state, and $\bar{x}^{1,n+1} = \lambda$. If the DSGE implies some unit roots then $\bar{y}_0(\theta)$ should be chosen consistent with all

the stationary linear combinations of y being at their steady states and the non-stationary ones being at the values implied by the initial conditions in the data.

III. INFERENCE

To characterize the joint pdf of parameters and $\{y_t\}$ now requires just specification of a prior on θ, which will be model-specific. Then we will have a marginal pdf for θ (the prior on θ), a conditional pdf for A^* given θ (the conditional prior laid out in the previous section), and a conditional pdf for the data given A^* (the standard SVAR pdf). Their product is the joint pdf. We do not assume, though, that the full y vector is observed. Observations are $z_t = Cy_t$, where z_t is a shorter vector than y_t. The Kalman filter gives us an algorithm for proceeding recursively through the sample with $A(L), c$ given to obtain the posterior pdf value for $\{z_t\}$ given $A(L), c$.

However, while constructing the conditional pdf of $A^* \mid \theta, \{y_t\}$ is straightforward, constructing that for $A^* \mid \theta, \{z_t\}$ is not. The z process will be an ARIMA process with coefficients nonlinear functions of A^*. It is therefore convenient to treat the unobserved dimensions of y_t explicitly and conduct inference jointly on them, A^*, and θ. A recursive algorithm similar to the Kalman filter will allow us also to generate a sample from the conditional distribution of the unobserved components of y_t conditional on the observed z_t's. Conditional on those y values and θ, the posterior pdf of A^* is in a standard form and can be evaluated analytically or sampled from directly.

Our expanded parameter vector $A^*, \theta, \{y_t\}$ thus has two high-dimensional components, A^* and $\{y_t\}$, and one of moderate dimension, θ. For each high-dimensional, component we know how to maximize analytically or sample from directly the posterior conditional on the other parameters. In the maximization phase of inference, therefore, we will use an alternate-directions-search algorithm, maximizing in one of the three parameter blocks at a time. Such methods generally work well in the initial phases of optimization, but can become very slow near the optimum if there is strong dependence across the alternating directions. Since we are aiming at using the optimization mainly to generate a good starting point for MCMC posterior sampling, the possibility that it slows down near the optimum is not necessarily a serious problem.

For MCMC sampling from the posterior, alternating directions corresponds to Gibbs sampling. In the two high-dimension blocks, we can sample directly from the conditional posterior. For the θ block, some version of Metropolis-Hastings sampling will be required. As is well known, sampling sequentially from alternate blocks of parameters leads to a legitimate MCMC sampler. However here too there could be difficulties if there is strong dependence across the dimensions. It seems likely that uncertainty about unobserved states is not strongly dependent on the values of other parameters. When the prior is tight, though, dependence between θ and A^* can be arbitrarily strong. It may therefore be necessary to use an occasional

independence Metropolis-Hastings MCMC step[2] on θ and A^* jointly to break the dependence.

The Kalman filter requires that we initialize the filter with a pre-observation distribution for y_1. If the $A(L), c$ model is stationary, the obvious choice is the ergodic distribution for y_1 implied by the model. However it is common for DSGE's to imply unit roots, and even if a given θ implies only roots near 1, the conditional posterior will put some probability on $A(L), c$ values that imply non-stationarity. The common practice in VAR and SVAR models is to use the pdf for the data from $k+1$ onwards, conditional on the initial conditions for $t = 1, \ldots, k$. That will generally be possible here, also, since we can expect that z_t captures all the dimensions of non-stationarity in the model. In that case the conditional distribution of y_1 given z_1 will be well-defined, even if there is no unconditional distribution for y implied by the model, and we can initiate the Kalman filter with this conditional distribution. Conditioning on initial conditions this way wastes information, as the stationary dimensions of variation in initial z will carry information about the parameters that is ignored. Also, as explained in Sims [revised 1996, 2000], conditioning on initial conditions tends to lead to model fits that attribute too much explanatory power to deterministic components dependent on initial conditions. The use of the DSGE prior probably mitigates this latter tendency, but the world awaits a better systematic approach to handling initial conditions than conditioning on them.

---

[2]See Robert and Casella [2004] for a discussion of Metropolis-Hastings algorithms.

Evaluation of the marginal data density — the integrated posterior density — is required for making Bayesian model comparisons. Here the availability of analytic integration over $\{y_t\} \mid \theta, A^*, \{z_t\}$ or $\{A^* \mid \theta, \{y_t\}\}$ will increase accuracy for the usual modified harmonic mean method of evaluating the marginal data density. It seems that only one of these simplifications is usable at a time, though one could try each.

It is worth mentioning here that standard DSGE solution programs do not produce results in the form (2). The program `gensys`, for example. produces a solution in the form

$$y_t = G(\theta)y_{t-1} + H(\theta)\varepsilon_t \,. \tag{9}$$

Here $H(\theta)$, called the impact matrix in `gensys`, is generally non-square and less than full rank, so premultiplying (9) by $H(\theta)^{-1'}$ to achieve the SVAR-like form (2) is not an option. In fact, with $H(\theta)$ singular there is not in general a unique $A_0(\theta)$. Nonetheless it is possible to define an algorithm that maps $I - G(\theta)L$ and $H(\theta)$ into a unique $A(L;\theta)$ for every theta. The normalization that delivers a unique θ has no effect on the equations defining ε or on the model's implied behavior of the data. The mechanics of the translation are described in the Appendix.

IV. Relation to previous work

There has been continuing interest in using Bayesian methods to connect behaviorally interpretable macroeconomic models with statistical models that fit well, going back at least to Ingram and Whiteman [1994] and DeJong et al. [1996, 2000]. In the first of these papers a DSGE model is used to generate a prior for a reduced

form VAR. In the last a prior is placed on parameters of a simple linearized DSGE, which is then compared with a BVAR in a forecasting exercise. More recently Smets and Wouters [2003a,b,c] demonstrated that this latter approach could be extended to models more closely approaching the type and scale of those used in central bank policy analysis.

The most closely related previous work is that of DelNegro and Schorfheide [2004], DelNegro et al. [2006] and Sims and Zha [1998]. Like Ingram and Whiteman [1994], Del Negro and Schorfheide use a DSGE to develop a prior for a VAR. They go beyond the earlier work in making the model for the data a structural VAR, but their approach does not admit any direct specification of the degree of uncertainty about the DSGE's A_0 matrix. They do produce a prior on A_0 (actually, directly on H) based on θ but they do so via a mapping that depends on an arbitrary ordering of variables. The result is that some identifying restrictions from the DSGE are imposed deterministically and others stochastically, with the exact nature of the prior depending on the arbitrary ordering in ways that are difficult to grasp. As in this paper, they construct a prior that, conditional on A_0 and θ, is conjugate, but in their published and circulated work they do not consider applying different weights by frequency.[3] Finally, their approach does not produce distributions for $A(L), c$, only for the reduced form parameters and H. An advantage of producing posteriors directly on $A(L;\theta)$ is that uncertainty about the coefficients in a given

[3]From conversation with them, I understand that they have research underway that does so, though through a different approach than that taken here.

equation in this form corresponds to uncertainty about whether the shock is well-determined by the prior and the data. For example, if the data and prior allow fairly precise identification of the monetary policy shock, but weak identification of the distinction between technology shocks and labor supply shocks, this would be evident from distributions for the equations defining these shocks.

The paper by Zha and myself takes an approach similar to that in this paper, but focuses entirely on SVAR's, shrinking toward independent random walks.

V. An example application

TO BE WRITTEN

Appendix A. From H to A_0

Suppose we are given a system of the form

$$\underset{n\times 1}{y_t} = G y_{t-1} + \underset{n\times m}{H}\,\varepsilon\,. \tag{10}$$

We can find a singular value decomposition of H as

$$UDV' = H\,, \tag{11}$$

where U and V are orthonormal and square and D has the dimensions of H and is diagonal (meaning its only non-xero elements are on the diagonal of its upper $m \times m$ matrix). If D is singular, some linear combinations of the shocks have no influence at all on y. This could happen, for example if only the sum of two demand shocks matters, or if the total number of behvioral shocks exceeds the length of the y vector. In such cases it is likely to be true that we can redefine ε to be of lower

dimension, but still mutually uncorrelated, so that D is full rank. We will not lay out a general method for dealing with such cases here, however.

With D non-singular, we can set

$$A_0 = \begin{bmatrix} V & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} D^{-1} & 0 \\ 0 & I \end{bmatrix} U' \tag{12}$$

$$A_1 = A_0 G \tag{13}$$

to obtain

$$A_0 y_t = A_1 y_{t-1} + \begin{bmatrix} \varepsilon \\ 0 \end{bmatrix} \tag{14}$$

REFERENCES

David DeJong, Beth Ingram, and Charles H. Whiteman. A Bayesian Approach to Calibration. *Journal of Business and Economic Statistics*, 14:1–10, January 1996.

David N. DeJong, Beth F. Ingram, and Charles H. Whiteman. A Bayesian Approach to Dynamic Macroeconomics. *Journal of Econometrics*, 15:311–320, May 2000.

Marco DelNegro and Frank Schorfheide. Priors from general equilibrium models for vars. *International Economic Review*, 45:643–673, May 2004.

Marco DelNegro, Frank Schorfheide, Frank Smets, and Raf Wouters. On the Fit of New-Keynesian Models. Technical report, Federal Reserve Bank of Atlanta, March 2006.

Thomas Doan, Robert Litterman, and Christopher A. Sims. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1): 1–100, 1984.

Richard Harrison, Kalin Nikolov, Meghan Quinn, Gareth Ramsay, Alasdair Scott, and Ryland Thomas. *The Bank of England Quarterly Model*. Bank of England, 2005.

Beth F. Ingram and Charles H. Whiteman. Supplanting the "minnesota" prior: Forecasting macroeconomic time series using real business cycle model priors. *Journal of Monetary Economics*, 34(3):497–510, 1994.

Robert B. Litterman. Forecasting with bayesian vector autoregressions — five years of experience. *Journal of Business and Economic Statistics*, 4:25–38, 1986.

Christian R. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, 2nd edition, 2004.

Christopher A. Sims. Inference for multivariate time series with trend. Technical report, presented at the 1992 American Statistical Association Meetings, `http://sims.princeton.edu/yftp/trends/ASAPAPER.pdf`, revised 1996.

Christopher A. Sims. Dummy Observation Priors Revisited. Technical report, Princeton University, April 2005. URL `http://sims.princeton.edu/yftp/DummyObs/`.

Christopher A. Sims. Using a likelihood perspective to sharpen econometric discourse: Three examples. *Journal of Econometrics*, 95(2):443–462, 2000. `http://www.princeton.edu/~sims/`.

Christopher A. Sims and Tao Zha. Bayesian methods for dynamic multivariate models. *International Economic Review*, 1998.

Frank Smets and Raf Wouters. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1: 1123–1175, September 2003a.

Frank Smets and Raf Wouters. Shocks and frictions in us business cycles: a Bayesian DSGE approach. Discussion paper, European Central Bank and National Bank of Belgium, Frank.Smets@ecb.int,Rafael.Wouters@nbb.be, May 2003b.

Frank Smets and Raf Wouters. Shocks and frictions in us and euro area business cycles: a Bayesian DSGE approach. Technical report, European Central Bank and National Bank of Belgium, Frank.Smets@ecb.int,Rafael.Wouters@nbb.be, February 2003c.

DEPARTMENT OF ECONOMICS, PRINCETON UNIVERSITY

E-mail address: sims@princeton.edu