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 (DS), who have proposed a way to use a dynamic stochastic general equilibrium (DSGE) model to generate a prior distribution for a structural time series model that relaxes the tight theoretical restrictions of the DSGE. The advantages of this paper's approach over that of DS are that the time series model strictly nests the DSGE specification, that the prior information is more able to resolve weak identification, that uncertainty about identification is treated more like other forms of uncertainty, and that giving greater weight to the prior at particular frequencies is more straightforward.

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

---

Date: May 22, 2008.

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

believe are implausible. Practical macro modelers who face real-time demands for forecasts and policy projections have struggled continuously with these tradeoffs.

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 parametrized behavioral models, or one can use Bayesian methods to import beliefs based on behavioral models (perhaps not so densely parametrized) into densely parametrized 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 parametrized 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

[1] A recent paper [DelNegro et al., 2007] 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

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 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

is modeled as generated by a structural VAR. The behavioral model is only a means to generation of a prior.

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 responses 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 a structural time series model 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

$$y_t = Ay_{t-1} + H\varepsilon_t \,, \tag{1}$$

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. Constant terms are represented as elements of y that always have the value 1. Any exogenous variables are assumed to be included in y. Their exogeneity is reflected in block triangularity of A. Not all the elements of y are necessarily observable. While this model has the form of a first-order structural vector autoregression (SVAR), it differs from SVAR's in existing 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 and solved to eliminate expectational terms, usually can be cast this form, with and A and H both functions of the model's behavioral parameters. Since only some components of the y_t vector are taken to correspond to observable data, this model implies a structural vector autoregressive-moving-average (SVARMA) time series model for the observed data. It is "structural" because it explains observed data as functions of the interpretable disturbances ε_t, whose interpretations

are carried over from the DSGE. Thus it can correctly be used to assess the effects of hypothetical time paths to the disturbances that represent changes in policy.

We will 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 and solved to take the form

$$y_t = A^*(\theta) + H^*(\theta)\varepsilon_t \,. \tag{2}$$

We assume that, like most DSGE's in the literature, this model is too tightly parametrized to fit the data in detail. Nonetheless we expect it to be a fairly good approximation, and we want to use it 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. According to its own internal logic, the DSGE is "structural" relative to a wider class of interventions than is the SVARMA. The DSGE purports to be able to correctly predict the effects of permanent, unanticipated, known-never-to-be-repeated shifts in policy behavior. However, since such policy changes by construction must never, or almost never, occur, and since actual DSGE models are heavily laden with tight restrictions that no one believes to be literally true, we may not be giving up a great deal in making the SVARMA the central policy model

We use the DSGE together with a prior distribution for its parameters θ to generate a prior distribution for A and H, the VARMA parameters. We postulate that conditional on θ, (A, H) has a distribution that is centered on $\big(A^*(\theta), H^*(\theta)\big)$, the linearized DSGE coefficients. It is likely that the DSGE has very low order dynamics,

so the length of the y vector, with all lagged values included, is likely to be higher than implied by the DSGE. This means that the $A^*(\theta)$ matrix we take as a mean for A has some blocks of zeros that reflect the absence of higher order lags in the DSGE. Also, as discussed in more detail below, The ε vector could have more elements than are present in the DSGE shock vector, so that $H^*(\theta)$ has columns of zeros. More generally, zero elements in $H^*(\theta)$ may or may not be treated as zero with certainty in H.

We express our beliefs about the connection of A, H to $A^*(\theta), H^*(\theta)$ via a set of dummy observations. There are several types of dummy observation. In the first two types, the j'th dummy observation consists of a pair $\bar{y}_1^j, \bar{y}_0^j$ of current and lagged values for the y vector.

Each these dummy observations generates a factor in the log prior density as if $\bar{y}_1^j$ were an observation on a $N(A\bar{y}_0^j, HH')$ random variable. H is likely to be of less than full row rank, implying singularity of the covariance matrix. Therefore to write the log-pdf element for y corresponding to the dummy observation we can use the singular value decomposition $H = UDV'$, where $V'V = U'U = I$ and D is a square diagonal matrix with no zeros on the diagonal:

$$-\log|D| - \tfrac{1}{2}(\bar{y}_1^j - A\bar{y}_0^j)'UD^{-2}U'(\bar{y}_1^j - A\bar{y}_0^j)\,. \tag{3}$$

The two types of dummy observation in this form are Type 1, in which $\bar{y}_1^j = A^*(\theta)\bar{y}_0^j$, and Type 2, in which $\bar{y}_0^j = 0$ and $\bar{y}_1^j \neq 0$. The former type of dummy observation expresses a belief that A is close to $A^*(\theta)$. The latter has no implications for beliefs

about A, but implies beliefs about H. In order for the Type 2 observations to imply H is close to $H^*(\theta)$, the corresponding $\bar{y}_1^j$ should be of the form $\bar{y}_1^j = H^*(\theta)e_i$, where e_i is a unit vector.

These two types of dummy observation are convenient, because they yield terms in the log likelihood that are of the same form as the pdf of y in the sample period. (I.e., they lead to a conjugate prior.) They have two limitations that may be important, however. One is that they convey prior beliefs only about HH', not directly about H. Since for any $F = HQ$ with $QQ' = I$, $FF' = HH'$, such prior beliefs can leave some dimensions of variation in H unrestricted. This is possible even when $H^*(\theta)$ would be fully identified from observations on y, because H generally relaxes some of the restrictions on $H^*(\theta)$. On the other hand, in some cases there may be zero restrictions on H itself that are treated as non-stochastic, so that a distribution on HH' implies a proper distribution on H itself. This can happen even in cases where H would not be identified from observations on observable components of y alone. In that case a proper prior can be formed from observations of these first two types alone. Nonetheless, even when the two conjugate types of dummy observations can provide a proper prior, it is probably a good idea to use non-conjugate dummy observations that are capable of forcing H to match $H^*(\theta)$ when the prior is

sufficiently tightened. A simple way to do this is to add to the prior a factor proportional to a $N(H^*(\theta), \Omega)$ pdf in H, where Ω is fairly small[2]. While this makes the prior non-conjugate, only the H component becomes non-conjugate, so the extra computational burden or MCMC simulation is modest. We label this form of direct dummy observation on H Type 3.

A second limitation of these two conjugate dummy observation types is that the Type 1 observations force the prior to imply the same pattern of uncertainty about coefficients across elements of the lagged y vector in all of the model's equations. This limitation is familiar from the Minnesota prior for VAR's, where avoiding it takes us out of the realm of conjugate priors and makes computations much harder. Here, however, we already expect to be treating only some of the elements of y as observable, and therefore to be generating the likelihood function with the Kalman filter. It is therefore no additional computational burden to allow for dummy observations on $S^j\bar{y}_1^j$, treating it as $N\left(S_j\bar{y}_0^j, S^jHH'(S^j)'\right)$, where S^j is a weighting matrix (e.g. a simple unit vector) that characterizes which parts or linear combinations of y^j are observable. This type of dummy observation allows for introducing equation-specific patterns of covariance in prior beliefs about deviations of coefficients from those in the linearized DSGE. We label this form of dummy observation Type 4.

[2]Some elements of H will be constrained deterministically to be zero, so this pdf will actually be a Normal pdf in the unconstrained elements of H, centered at the corresponding elements of $H^*(\theta)$.

There is no uniquely best way to specify $\bar{y}_1^j$ and $\bar{y}_0^j$ in Type 1 dummy observations. The dynamics of the system are controlled by the eigenvectors and eigenvalues of A. A natural way to proceed, then, is to use Type 1 dummy observations of the form $\bar{y}_1 = \lambda x$, $\bar{y}_0 = x$, where x and λ are an eigenvector and corresponding eigenvalue of $A^*(\theta)$. This approach has the advantage that it allows us to weight dummy observations more heavily at frequencies where we give more credence to the DSGE model. Observations corresponding to eigenvalues near 1 or eigenvalues with half-lives or oscillation frequencies in the business cycle range could be more heavily weighted, for example. A variant of this idea that would be more numerically stable is to form the Schur decomposition $QTQ' = A^*(\theta)$, where $Q'Q = I$ and T is upper triangular with the eigenvalues of A down the diagonal. The first k columns of Q span the same space as the first k eigenvectors of $A^*(\theta)$. The advantage of the Schur decomposition approach is that when there are repeated eigenvalues, they may correspond to an invariant space of dimension higher than one. The columns of Q will span the invariant spaces, while standard eigenvector calculations will not recover them. Then the Type 1 dummy observations would use $\bar{y}_0^j = Q_{\cdot j}$, $\bar{y}_1^j = QT_{\cdot j}$, where (e.g.) $T_{\cdot j}$ is the j'th column of T.

The precision of prior beliefs about the connection between $A^*(\theta)$, $H^*(\theta)$ and A, H can be controlled two ways for Type 1 dummy observations. One can scale up or down the size of $\bar{y}_1^j, \bar{y}_0^j$, with larger dummy observations then getting more weight and therefore implying greater precision of prior beliefs, or one can in effect repeat

the dummy observations. In the expression (3) for the log pdf term corresponding to a dummy observation, scales up the second, quadratic term, while leaving the $-\log R$ term unaffected. Repeating the dummy observation amounts to scaling both the $\log R$ and the quadratic term by the same constant. When both terms are scaled, the implications of the dummy observation for H are tightened along with the implications for A, while scaling the quadratic term only mainly increases the precision of the prior for A. Since the Type 1 dummy observations are chosen to fit perfectly at the prior mean, they will tend to shrink the prior mean of HH', the reduced form covariance matrix. It is therefore probably wise to repeat the Type 2 and/or 3 dummy observations whenever the repetition weighting is applied to Type 1 observations.

The eigenvectors of $A^*(\theta)$ are of course determined only up to a scale factor. In a stationary model, a natural way to set the scale is to set the length of the $\bar{y}_0^j$ vector at the standard deviation of the projection of y onto the $\bar{y}_0^j$ direction, using the unconditional distribution of y_t implied by $A^*(\theta), H^*(\theta)$. The appeal of this approach is that it controls the weight of the prior relative to the data. The central form of the prior, before scaling to tighten or loosen, is calibrated to have similar weight to a corresponding number of observations drawn from the distribution for y_t implied by $A^*(\theta), H^*(\theta)$. But this is not always desirable. For example, the DSGE may imply a tight cointegrating relation between two variables and we may regard this implication as robust, so we want to impose it fairly precisely in our prior. But the tight cointegrating relation implies that the corresponding linear combination of variables

will have low unconditional variance. An eigenvector of $A^*(\theta)$ corresponding to the cointegrating vector would therefore get low, not high, weight, if we simply scaled the corresponding dummy observation by the standard deviation of y_t in that direction.

In fact, the DSGE is likely to imply a singular joint distribution for y_t, so that certain directions of variation have zero variance. One would not want to put zero weight on dummy observations in these directions — this would in effect ignore the information about the y_t distribution about which the DSGE has the sharpest implications. The mirror image of this pitfall is that if A has any eigenvalues greater than or equal to one in absolute value, there will be no unconditional distribution for y_t. This latter is in fact a problem even for stationary models if they have roots close to one in absolute value. The variance of y_t in the direction of eigenvectors corresponding to near unit roots will be extremely large, producing dummy observations at those eigenvectors with extremely large weight in the prior.

Here is a compromise between the desire to pin down the ratio of precision in the prior and in the sample data and the desire to allow the prior to contribute information where the data are weakly informative. Experience may prove that better compromises are possible. First, using an extremely dispersed initial distribution for y_0 and the algebra of the Kalman filter, calculate $\hat{y}_0 = E_\theta[y_0 \mid z_0]$ and $M_0 = \mathrm{Var}_\theta(y_0 \mid z_0)$. Here z_0 is the initial vector of observations on components of y_0, and the θ subscript refers to the distribution implied by the $A^*(\theta), H^*(\theta)$ stochastic process for y. If

there is to be any hope of estimating the model, the observables should pin down all the non-stationary dimensions of variation, so the end result should not be so dispersed in any dimension as the initial distribution for y_0. Find the eigenvector decomposition $W'\Lambda W = \hat{M}_0$ (where W is orthonormal and Λ is diagonal with eigenvalues of $\hat{M}$ down the diagonal), and form $\hat{\Lambda}$ by increasing any eigenvalue smaller in size than some critical fraction κ of the largest eigenvalue λ_{max} to $\kappa\lambda_{max}$. Define $\hat{M}_S = W'\hat{\Lambda}W$, and scale the Type 1 dummy observation vector $\bar{y}_0^j$ to have length $\left(x_j'\hat{M}_S x_j\right)^{1/2}$, where $x_j = \bar{y}_0^j / \left\|\bar{y}_0^j\right\|$ is the unit vector in the $\bar{y}_0^j$ direction. Of course then, since in this type of observation $\bar{y}_1^j = A^*(\theta)\bar{y}_0^j$, $\bar{y}_1^j$ is then automatically scaled as well.

Note that some elements of A and H that are zero in the DSGE we will want to constrain also to be zero in the less restricted model. For example, some equations in the DSGE may be accounting identities. The corresponding rows of $H^*(\theta)$ will be zero, and we will want to set those rows of H to be zero also. Some rows of $A^*(\theta)$ will be unit vectors, defining the j'th lag in the y_t vector to be the $(j-1)$'th lag in the y_{t-1} vector. These rows will take the same form in A and will contain no free parameters.

We can weight frequencies by applying a function $\omega(\rho, \phi)$, where ρ is the absolute value of the eigenvalue associated with a given Type 1 dummy observation and ϕ is the phase angle in radians. The use of the long run covariance matrix to scale the dummy observations will tend to emphasize low frequency components in the

prior automatically, but we may want to increase this emphasis by making ω a decreasing function of $(1-\rho)^2+\phi^2$, the distance of the eigenvalue from one in polar coordinates. This of course also downweights explosive roots that might appear in the DSGE. It also might be reasonable to make ω especially small for negative real roots or complex roots with, say, $|\phi| > .8\pi$ (assuming the convention is $\phi \in (-\pi, \pi)$), especially if they have large ρ. This expresses skepticism that a DSGE's implication of rapid oscillations would be correct.

II. 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, H given θ (the conditional prior laid out in the previous section), and a conditional pdf for the data given A, H (the standard SVARMA 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, H, C given to obtain the posterior pdf value for $\{z_t\}$ given A, H, C.

However, while constructing the conditional pdf of $A, H \mid \theta, \{y_t\}$ is straightforward, constructing that for $A^* \mid \theta, \{z_t\}$ is not, as it would require integrating out the unobserved components of y_t to obtain a z process that is a vector ARMA 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, H,

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 and A, H. 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. The posterior pdf of H given $\{y_t\}$ is not standard, because of our Type 3 dummy observations, but it has an explicit analytic form.

Our expanded parameter set $A, H, \theta, \{y_t\}$ thus has two high-dimensional components, A and $\{y_t\}$, and two of moderate dimension, θ and H. 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 at least at first use an alternate-directions-search algorithm, varying one of the four 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, but we can also include a finishing phase in which maximization is joint across all four blocks.

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 and H blocks, 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, H can be arbitrarily strong. It may therefore be necessary to use an occasional independence Metropolis-Hastings MCMC step[3] on θ and A, H jointly to break the dependence.

The Kalman filter, when we apply it to evaluating the pdf for $\{y_t\}$ conditional on A, H, requires that we initialize the filter with a pre-observation distribution for y_1. If the A, H 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 values that imply non-stationarity. In most cases, the DSGE model will imply that the components of the y_t vector that are not observable are stationary, conditional on z_t — that is, that the unobservable elements of y are cointegrated with the observable elements. (Arguably, if this were not true we should not be trying to estimate the model from observations on z alone.) In that case, we can put zero prior probability on A, H values that fail to preserve this cointegration, and make

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

the initial distribution for y_1 the conditional distribution for y_1 given z_1 under the A, H process.

Evaluation of the marginal data density — the integrated posterior density — is required for making Bayesian model comparisons. Here the availability of analytic integration over $\{y_t\} \mid \theta, A, H, \{z_t\}$ or $\{A \mid H, \theta, \{y_t\}\}$ will increase accuracy for the usual modified harmonic mean method of evaluating the marginal data density.

III. 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,c,b] 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. [2007] 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 H matrix. They do produce a prior 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 H and θ, is conjugate, but in their published and circulated work they do not consider applying different weights by frequency.[4]

The most important potential advantage of this paper's approach, however, is that it provides greater scope for letting prior information help resolve weak identification. In the DS approach, except for the somewhat ad hoc method used to connect the SVAR H to the DSGE H, the DSGE generates prior information for the SVAR only in the form of artificial posterior distributions for A, H, constructed as if data from the DSGE had been observed. But the DSGE may imply tight cointegrating relations or exact dependencies among variables that imply singularity or near-singularity in the reduced form VAR's right-hand-side variable matrix. This in turn would imply that the data from the DSGE would contain little or no information about these tight relationships, because there would be no observed variation in them. This paper's approach would automatically take account of and use such implied relationships, instead of implicitly discounting them. This opens up the possibility, for example,

[4]Del Negro et al. [2008] propose a way to weight a prior by frequency in the Del Negro/Schorfheide SVAR framework.

of generating a prior from a DSGE, say a simple RBC model, that implies a singular joint distribution for the observable z's. The DS approach cannot be applied to such a case.

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.

IV. An example application

No complete application is as yet available at the time of writing this draft. It may nonetheless be useful to work through the mechanics of setting up the Type 1 dummy observations in an extremely simple rational expectations model in this framework. The model is a naive pricing model for an asset paying an exogenous, serially correlated, dividend:

$$\beta E_t[P_{t+1} + \delta_{t+1}] = P_t \tag{4}$$

$$\delta_t = \gamma + \rho\delta_{t-1} + \varepsilon_t \,. \tag{5}$$

We consider generating the prior at a particular setting $(\beta, \gamma, \rho) = (.95, .2, .8)$ of the structural parameter vector. The y_t vector is $y_t = (P_t, \delta_t, 1)'$. We consider both the case where $z_t = P_t$ and where $z_t = (P_t, \delta_t)$. The solved model has

$$A^*(\theta) = \begin{bmatrix} 0 & 2.53 & 16.47 \\ 0 & 0.80 & 0.20 \\ 0 & 0 & 1 \end{bmatrix}, \qquad H^*(\theta) = \begin{bmatrix} 3.17 \\ 1.00 \\ 0 \end{bmatrix} . \tag{6}$$

The third rows of $A^*(\theta)$ and $H^*(\theta)$ are just defining the constant term as a variable that is constant and subject to no shocks. These elements of A and H are therefore treated as known constants, not unknown parameters. The first two zeros in the first column of $A^*(\theta)$ reflect the fact that in the "DSGE" P_t and δ_t are deterministically related, so that only one of them is needed on the right-hand side of the system. The coefficients on δ_t (the second column of $A^*(\theta)$ in the first and second equation are in the same ratio as the coefficients on ε_t in $H^*(\theta)$, so that a linear combination of the two equations implies that a linear combination of P_t and δ_t is a constant. These are constraints that we probably do *not* want to enforce deterministically in the A, H model. We would like the A, H model to recognize the possibility that the theoretical model is misspecified in these dimensions, through perhaps not allowing enough lags in δ and/or not recognizing that current and past δ are not the only sources of information available to the market about future δ. (In fact, we might in practice want at least to expand the y_t vector to include additional lags of δ in the A, H model, with corresponding zeros in the $A^*(\theta)$ matrix.)

The eigenvalues of this $A^*(\theta)$ are 1.0 0.8 and 0.0, with corresponding eigenvectors being the columns of

$$\begin{bmatrix} 0.9972 & .9536 & 1.00 \\ 0.0525 & .3011 & 0 \\ 0.0525 & 0 & 0 \end{bmatrix}.$$

There are no repeated roots here, so we could use the eigenvectors themselves as the dummy observations. The Schur decomposition here is trivial, with Q the identity

and T being $A^*(\theta)$ itself, since that is already in upper triangular form. If we used the columns of Q, then, the dummy observations would simply perturb one element of y on the right-hand side at a time.

The model is stationary, despite the unit root, because the unit root applies only to the constant term, which is unaffected by the disturbance. The unconditional second moment matrix implied by the theoretical model is

$$\begin{bmatrix} 388.9 & 27.8 & 19.00 \\ 27.80 & 3.778 & 1.000 \\ 19.00 & 1.000 & 1.000 \end{bmatrix} .$$

If we pick a κ of, say, .01, the corresponding $\hat{\hat{M}}$ emerges as

$$\begin{bmatrix} 388.9 & 27.63 & 18.79 \\ 27.63 & 5.900 & 1.348 \\ 18.79 & 1.348 & 4.835 \end{bmatrix}$$

Since $Q = I$, the corresponding weights on the columns of Q are just the square roots of the diagonals of this $\hat{\hat{M}}$ matrix, or $(19.72, 2.43, 2.20)$. The three Type 1 dummy observations, arranged in columns, are then

$$\bar{y}_1 = \begin{bmatrix} 0 & 6.15 & 36.2 \\ 0 & 1.94 & 0.440 \\ 0 & 0 & 2.20 \end{bmatrix} \qquad \bar{y}_0 = \begin{bmatrix} 19.7 & 0 & 0 \\ 0 & 2.43 & 0 \\ 0 & 0 & 2.20 \end{bmatrix} .$$

Notice that the first dummy observation asserts that the coefficient of P_{t-1} has expected value zero in every equation, but since the first two equations have non-zero error variance, this assertion is not deterministic for those two equations.

Here there appears to be no particular reason to think the theoretical model is more accurate at certain frequencies, so there is no need to consider a non-constant $\omega()$ function. It would be natural to introduce as a hyperparameter the number of implicit repetitions of the dummy observations, since it is not obvious how tight we would want to make the prior. And of course to apply the model we would need to proceed to putting a prior on δ, β, γ and generate Type 3 dummy observations to capture the theoretical model's assertion of a tight relation between the disturbances in the first two equations. In fact, we would in practice want to add a column to H, to recognize that in reality the forecast errors in P and δ would not be perfectly correlated. As for initial conditions, since the theoretical model is stationary we could simply use the unconditional joint distribution of y to form it.

If we could observe both P and δ, this approach would give us a model, based on a simple theory in which P and δ are perfectly correlated, in which P and δ are asserted to behave approximately as the theory predicts, but with a non-singular joint distribution. If we could observe only P, β and γ would not be separately identified from the likelihood function alone. We might have fairly strong prior beliefs about reasonable values for β, however, which would allow the model to generate corresponding posterior implications for γ.

V. Conclusion

[Omitted until these ideas have been implemented with a serious DSGE model.]

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 Del Negro, Francis X. Diebold, and Frank Schorfheide. Priors from Frequency-Domain Dummy Observations. Technical report, Federal Reserve Bank of New York and University of Pennsylvania, 12 February 2008.

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 and Forecasting Performance of New Keynesian Models. *Journal of Business and Economic Statistics*, 25(2):123–162, 2007.

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.

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

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

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 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 2003b.

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 2003c.

DEPARTMENT OF ECONOMICS, PRINCETON UNIVERSITY

E-mail address: `sims@princeton.edu`