WP/16/231

IMF Working Paper

System Priors for Econometric Time Series

by Michal Andrle and Miroslav Plašil

INTERNATIONAL MONETARY FUND

IMF Working Paper

Research Department

System Priors for Econometric Time Series[1]

Prepared by Michal Andrle and Miroslav Plašil

Authorized for distribution by Benjamin Hunt

November 2016

Abstract

The paper introduces "system priors", their use in Bayesian analysis of econometric time series, and provides a simple and illustrative application. System priors were devised by Andrle and Benes (2013) as a tool to incorporate prior knowledge into an economic model. Unlike priors about individual parameters, system priors offer a simple and efficient way of formulating well-defined and economically-meaningful priors about high-level model properties. The generality of system priors are illustrated using an AR(2) process with a prior that most of its dynamics comes from business-cycle frequencies.

JEL Classification Numbers: C11, C51, C18, C22

Keywords: system priors, Bayesian analysis, time series

Author's E-Mail Address: mandrle@imf.org, miroslav.plasil@cnb.cz

[1] We would like to thank Ben Hunt, Jan Brůha, Aleš Bulíř, Mika Kortelainen, and Joris de Wind for useful comments or discussions. The authors are solely responsible for all errors and omissions.

CONTENTS

FIGURES

I. INTRODUCTION

This paper introduces "system priors", their use in Bayesian analysis of econometric time series, and provides a simple and illustrative application. System priors were devised by Andrle and Benes (2013) as a tool to incorporate prior knowledge into an economic model. Although system priors were originally proposed within the DSGE context, they are by no means specific to DSGE modeling and can be used equally well in any other domain. Unlike priors about individual parameters, system priors offer a simple and efficient way of formulating well-defined and economically-meaningful priors about high-level model properties. In more technical terms, eliciting prior beliefs about selected system properties of the model introduces restrictions on the joint prior distribution of individual model coefficients. To see this, it is enough to realize that many features of a model are determined by non-trivial functions of its individual parameters. An example of system priors is when researchers have views about the characteristics of a model's impulse-response function. In principle, views about all meaningful and computable model properties can be expressed as system priors.

This paper illustrates the generality of system priors using a simple but relevant example of an AR(2) process. As the exposition in Andrle and Benes (2013) may be less accessible to those not familiar with the literature on Dynamic Stochastic General Equilibrium (DSGE) models, we provide a more nuanced one highlighting the generality of the approach. In particular, we demonstrate how researchers' economically-meaningful priors about high-level model properties can be easily implemented into model estimation and inference and how these a priori beliefs restrict the parameter space of individual coefficients. In our application, we assume a stationary process and incorporate a belief that a significant share of its variance comes from business-cycle frequencies. Such a prior might be an advantage when an AR(2) process is used for modeling cyclical components of economic variables and researchers need to confine the parameter space to regions they find economically plausible.

We keep the illustration as simple as possible for the ease of exposition and conceptual clarity of general principles, however the application of system priors to state-space models, Bayesian Vector Autoregressive (BVAR) models, or other type of linear or non-linear models is just a straightforward extension of the basic principles introduced in this paper. See Andrle and Benes (2013) for examples of system priors applied to structural DSGE models (priors about the sacrifice ratio and impulse-response functions, for instance) and Andrle and Plašil (2016) for in-depth discussion of system priors for forecasting and structural BVARs.[2]

Implementing system priors into standard Bayesian computations is fairly straightforward. Similar to traditional Bayesian inference, the initial (most commonly marginal independent) priors on parameters are updated using the likelihood function of the model. However, they are also updated using the information contained in the system priors, i.e. by prior views about the aggregate behavior of the model. Priors on individual parameters and system priors constitute a composite prior that reflects all available prior information. In some sense, the procedure is related to "dummy observation" priors (Theil and Goldberger, 1961), and allows to directly formulate priors on general nonlinear functions of parameters. From the non-Bayesian point of view, system priors can simply be interpreted as another penalty in the criterion function, along with the likelihood and marginal prior distribution penalties. The formal discussion below will make the computational implementation clear.

System priors do not require linearity of the model, Gaussian structural shocks, or error terms, or any particular form of the prior distribution on coefficients or model system properties. This naturally comes at some cost as they are computationally more expensive than natural-conjugate priors and their variations available in the literature. In this day and age, however, our view is that having economically meaningful models should be favored over convenience of having lightning-fast estimates of the models. Note also, that that computations with system priors can be sped up considerably by using modern parallel architectures.

[2] Most of the BVAR priors are only vaguely motivated in terms of economic theory and rather serve as elementary parameter-shrinkage method. There are several noticeable exceptions: the steady-state priors of Villani (2008), sums-of-coefficient priors of Sims and Zha (1998), or recent 'prior for the long run' by Giannone, Lenza, and Primiceri (2014). However, these top-down priors are designed for ad-hoc problems and do not provide a general framework for implementing priors on high-level properties of the VAR model.

II. System Priors

Estimating models with system priors closely follows general principles of Bayesian inference. The difference rests in the form of the prior distribution formulation. To demonstrate this, let us start with a traditional Bayesian setup: we assume that joint prior beliefs about a $(k \times 1)$ vector of individual parameters, θ, of a model M are expressed using independent marginal probability distributions, i.e. as: $p_m(\theta) = p_m(\theta_1) \times \ldots \times p_m(\theta_k)$. Other setups of priors are possible with no loss of generality. We further assume that given the observed data, Y, it is possible to evaluate the likelihood function of the model, $L(Y \mid \theta;M)$ for different values of parameters. Applying the Bayes law, it is well-known that the posterior distribution of parameters is proportional to a product of the likelihood and the prior distribution:

$$p(\theta \mid Y;M) \propto L(Y \mid \theta;M) \times p_m(\theta). \tag{1}$$

Now, let us incorporate a priori views about the model's system properties. To proceed, let us define a statistic, $r = h(\theta;M)$, which is a function of the model structure and its individual coefficients, and can be easily evaluated for different values of parameters. Such a function can describe impulse-response function characteristics or frequency-domain properties of the model, for instance. As in the case of individual parameters, prior beliefs about the values of the statistic r can be summarized by a feasible functional form, by a probability distribution. We will call it the system prior and denote it as $p_s(r \mid \theta;h,M) \equiv p_s(h(\theta);M)$. Putting together the effects of the marginal prior, system prior, and the likelihood function, the posterior distribution of the parameters emerges as

$$p(\theta \mid Y;M) \propto L(Y \mid \theta;M) \times [\, p_s(h(\theta);M) \times p_m(\theta)]. \tag{2}$$

The form of the posterior kernel in (2) is an intuitive one. For a given value of the parameter, θ, the posterior distribution is based on a two-step updating process. In the first step the marginal prior, p_m, is updated with the system priors, p_s, resulting in the composite prior distribution. As system priors operate on functions of parameters, the composite prior implies some restrictions on individual coefficients but generally not in a unique or invertible way. In the second step, composite prior beliefs are updated with information contained in the data using a likelihood function of the model.

To help with the intuition, it is useful to think of system priors as an artificial likelihood function[3] summarizing information contained in the artificial data on r, which are put into an auxiliary probabilistic model with a structure corresponding to function $r = h(\theta;M)$. In other words, system priors can be interpreted as measuring how likely the values of individual parameters are given the "observed" distribution of r. As such, they penalize parameter values not conforming to prior beliefs about system properties of the model. The shape of this likelihood depends on the distributional assumptions about system priors, expressed by the functional form and the hyper-parameters. Combining the prior distribution of individual parameters both with the artificial and with the conventional likelihood function results in posterior distribution of parameters expressed in (2).[4]

Essentially, estimation with system priors is just applying the Bayes law twice: first with the artificial likelihood function to obtain the composite prior and second with the conventional likelihood function of the underlying model to obtain the posterior distribution of the model parameters.

From a non-Bayesian perspective, the criterion function (2) is simply a penalized likelihood problem with two types of penalties. As such, it can be subject to standard or ad-hoc designed optimization routines to estimate the parameters and carry out the inference. The equivalence between the literature on parameter shrinkage in statistics and a suite of selected priors in Bayesian analysis is a good example of dual interpretation.[5] If feasible, the criterion function can be optimized numerically with respect to θ to find the posterior mode. The

[3] This brings it close to the idea of "dummy observation" priors. As pointed out in Sims (2005), "The prior takes the form of the likelihood function for the dummy observations." Yet, the system priors are not equivalent to dummy-observation priors. It would be the case if for the distribution of coefficients θ, for which the statistic $r = h(\theta)$ has a distribution $p_s(r)$ with chosen hyper-parameters, one would draw samples of the data using the model and use those in the inference.

[4] Regardless of whether Jacobian terms are involved or not, the resulting prior distribution of aggregate model properties is key to understanding all the consequences of the prior specification used, namely in non-trivial models.

[5] For instance, the popular ridge regression can be recast as a Bayesian problem with Gaussian priors.

inverse Hessian matrix evaluated at the posterior mode can then serve directly for (non-Bayesian) inference or as an important ingredient of the Markov Chain Monte Carlo (MCMC) procedures.

To analyze the composite joint prior distribution in greater detail, computations analogous to posterior sampling in (2) are needed with the evaluation of the conventional likelihood function switched off. Such analysis and associated prior predictive analysis of model's properties is highly recommended, to check if the formulation of the priors lead to desired or plausible properties of the model, see Geweke (2010).

Existing Bayesian computations and computer code can stay almost unchanged when system priors are employed—see the pseudo-code for the posterior kernel in the Appendix. The only difference is that for a particular *j*-th draw of the parameter vector, θ_j, three, instead of two components need to be evaluated – with the system prior component adding to the overhead.[6] Given the computational progress in the last decade and years to come, there is no need for the system priors to have closed-form solutions or conjugate forms.

III. Example – System Priors for an AR(2) Process

After the theoretical exposition of system priors, let us proceed with an illustration using an AR(2) process and the thought process that goes along the system priors. Admittedly, the process by itself may not be particularly useful for macroeconomic time series. However, it can be an important part of richer structural time series models. For instance, it is not uncommon to use the second-order autoregressive model for modeling a cyclical component of output and other economic variables (see e.g. Watson, 1986, Clark, 1987, and Kuttner, 1994).[7] Let us then consider a zero-mean AR(2) process:

[6] All three components, however, can be evaluated independently and thus in parallel for a particular draw of parameter vector or resources can be re-used in multiple components, as both the likelihood function and system priors make use of a model solution for a new vector of parameters.

[7] There are other common specifications of the cyclical components in the literature, for instance the trigonometric form in Harvey and Trimbur (2003), which in its univariate form corresponds to a restricted ARMA(2,1) process.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0,\sigma_\varepsilon^2) \tag{3}$$

What would be reasonable priors for the two auto-regressive coefficients ϕ_1 and ϕ_2? A common point of departure would be to start with normally distributed independent marginal priors for the individual coefficients, that is $\phi_1 \sim N(0,\sigma^2_{\phi_1})$ and $\phi_2 \sim N(0,\sigma^2_{\phi_2})$. However, this hardly sounds right if the prior is supposed to convey some relevant, economically-meaningful information. When the coefficients can vary independently and the joint distribution is spread out, it implies a wide array of model dynamics, including wild oscillations or unstable non-stationary impulse-response functions. Researchers have been aware of this issue for a long time and have been striving to come up with better ways of formulating priors, even in the particular, and simple, example of the second-order autoregressive process (see, e.g. Planas et al., 2008).[8]

In the case of the AR(2) process, a polar-form specification of the cycle was proposed as one of the solutions as it helps incorporate prior beliefs about cyclical behavior more efficiently. In the polar-form specification, the coefficients are analytically re-parameterized such that the priors are imposed on the periodicity and the amplitude of the cycle.[9] However, such re-parameterization may still be too vague for other types of a priori views about the business cycle. Further, analytical re-parameterization is not generally feasible except for very simple models. Luckily, there is absolutely no need for it. System priors usually will not have closed-form solutions. Not having a closed-form solution may add some computing overhead but does not affect the general principles.

[8] Planas et al. (2008) have recently commented on the problem with a business cycle modeled as an AR(2) [p. 19]: "Indeed, assuming a normal prior distribution on parameters ϕ_1, ϕ_2, we found it difficult to reproduce our prior knowledge by tuning the mean and the covariance matrix of the autoregressive parameters… In some cases, the implied distribution for the periodicity and amplitude can be counterintuitive…Putting the prior on the AR coefficients […in traditional way…] is probably inadequate for cyclical analysis."

[9] The polar-form specification is as follows, (1-2Acos(2pi/tau)L + A^2L^2)y(t)=epsilon(t), where A is the amplitude and tau is the periodicity. The amplitude is given by $A=\sqrt{-\phi_2}$ and the periodicity by $\tau = 2\pi/\mathrm{acos}\{\phi_1/2A\}$.

In our application example, we incorporate a prior view that the second-order AR process is stationary and more than 60% of its variance comes from business-cycle frequencies (i.e. frequencies of 8-32 periods in a quarterly model).

The process in (3) is stationary only if $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ and $|\phi_2| < 1$. These assumptions restrict the parameter space but they do not restrict the oscillatory properties of the model in a sensible way, as illustrated below. More disciplined behavior of the model can be achieved through a prior assumption about spectral characteristic of the process. Spectral density of the model can be interpreted as a distribution of variance across frequencies and is thus a natural starting point for formulating a system prior in this case.

The spectral density of y_t, denoted $S_y(w)$, can be computed as follows:

$$S_y(w) = \frac{\sigma_\varepsilon^2}{2\pi\left[1+\phi_1^2+\phi_2^2+2\phi_1\left(\phi_2-1\right)\cos(w)-2\phi_2\cos(2w)\right]}, \tag{4}$$

where $w \in [0,\pi)$ is the angular frequency. A brief inspection shows that spectral density is a non-linear function of both auto-regressive parameters. The variance of the error term, σ_ε^2, determines the level of the spectrum but not its shape.[10] Therefore, any prior exploiting a spectral restriction would result in nontrivial joint prior distribution for individual regression parameters. To introduce the system prior outlined above, we define the total variance of the process y_t as the integral of the spectrum (4) over the full frequency range and business cycle variance as the integral limited to the range of business cycle frequencies (a,b). Specifying the business-to-total variance ratio as:

$$r = \int_a^b S_y(w)\,\mathrm{d}w \Big/ \int S_y(w)\,\mathrm{d}w, \tag{5}$$

[10] Although for the AR(2) process the spectrum can be expressed in a closed form, nothing would change if the closed-form was not available.

results in a statistic that is univariate, has clear units, and has clear interpretation. The ratio (5) can only take values within the interval [0,1]. A change in the shock variance, σ_{ε}^{2}, shifts the spectrum up or down but never affects the ratio. As such, the spectral prior is completely uninformative about the coefficient σ_{ε}^{2}. In general, system priors are not equally informative about all coefficients.

Now, let us present two complementary examples of implementing a system prior that reflects prior beliefs about the cyclical component of the output. First, one may consider a condition that at least 60% of variance of y_t originates from business cycle frequencies. Second, prior beliefs about the ratio can be expressed using a statistical distribution. Given the range of admissible values for r, a Beta-distributed prior is a feasible option, as its support is in the [0,1] interval (non-Gaussian or non-conjugate prior poses no difficulty here). In our example, $r \sim Be(15,5)$ is used, which places a large portion of the probability mass of the variance of y_t as coming from business cycle frequencies. Other hyper-parameter settings are possible and used values only serve for illustration.

Computationally, the inference is based on simulation techniques with the conventional likelihood function omitted to learn only about the composite prior. In the case of the minimum of 60% of the variability coming from business cycle frequencies we employ rejection sampling with normally distributed marginal priors used as the proposal distribution. In the latter and more general case, our results are based on the sequential Monte Carlo sampling (SMC, see e.g. Herbst and Schorfheide, 2014) which can be seen as an alternative to the traditional Metropolis-Hasting random walk algorithm, which in our simple case might suit as well. For simple models both algorithms should provide almost identical results, however sequential Monte Carlo sampling is strongly preferred if complex models (containing dozens of parameters) are estimated.[11]

Fig. 1 shows the combinations of parameters that correspond to Normally distributed marginal priors, $\phi_1, \phi_2 \sim N(0,2)$, in the upper-left panel and combinations that conform with

[11] The R code for the examples presented is available upon request.

the stationarity restriction in the upper-right panel. The bottom panels illustrate the combinations in line with the requirement on sufficient variance of y_t coming from business cycle frequencies. Clearly, the considered system prior is fairly informative and leads to a non-Normal joint distribution of parameters. Both computational ways of implementing system priors reflect similar prior beliefs, hence they lead to similar results. System priors pose few restrictions on the actual technical design of the prior – it is the meaningfulness of the prior for the analysts and their audience that matters.[12]

Knowing just the combinations and full joint prior distribution of individual parameters that satisfy the constraints is not enough to evaluate the role of priors. The key knowledge is the understanding of how these priors translate into the behavior of the model in as many aspects as relevant. The analyst should investigate if there are any unintended consequences of the chosen priors. For this purpose, the prior-predictive distribution of the models' properties must be analyzed. In our case, the prior-implied distribution of the impulse-response function alongside spectral characteristics are natural candidates for closer inspection.

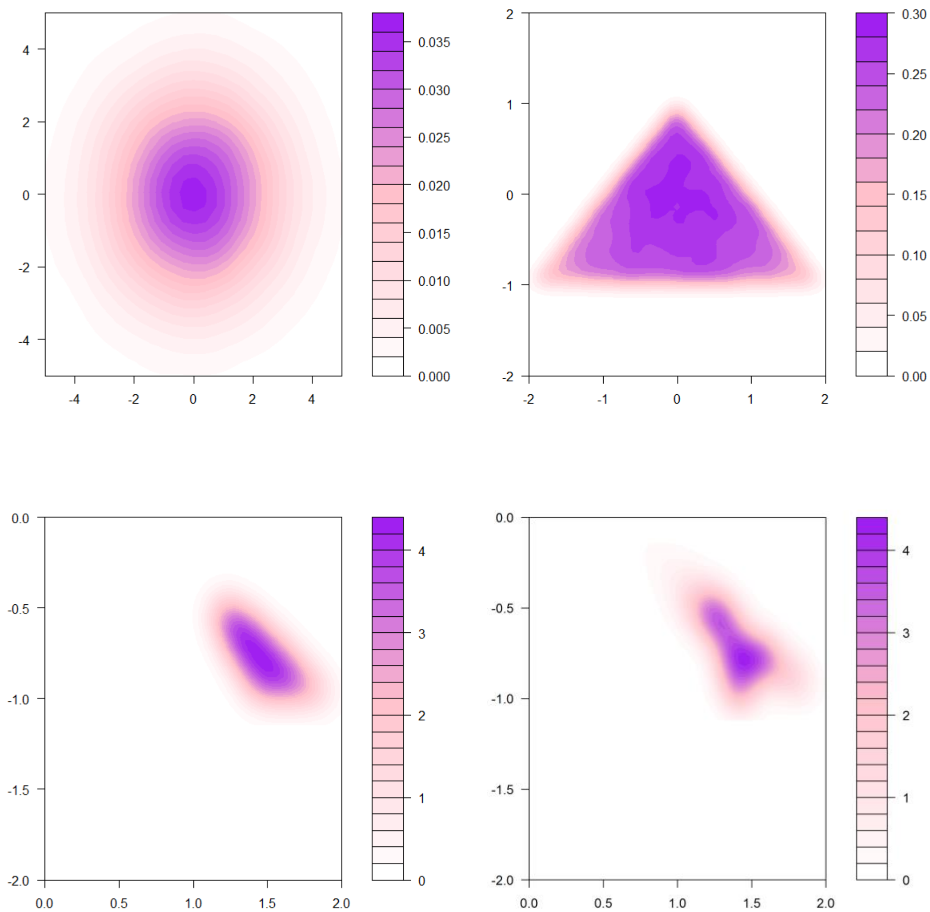
Fig. 2 depicts the spectral densities and impulse response functions for parameters in regions complying with the requirement of stationarity and sufficient variance coming from business cycle frequencies. It is apparent that the stationarity condition itself does not restrict the process in an economically-meaningful way, while the system priors do. Our system prior is not diffuse, it is fairly informative. However, it is also very transparent, simple to implement, and easy for others to agree or disagree with, should they wish to do so.

We could have specified other meaningful priors, for example directly in terms of the impulse-response function of the model. The scope of system priors is wide. System priors are a flexible tool, which easily extends to any other type of econometric and statistical models, including the state-space models (Andrle and Benes, 2013) or Bayesian VARs

[12] The computer-code implementation of system priors differs from standard Bayesian analysis in that the prior restrictions are not off-the-shelf functions and users are expected to specify their own. Once a clear interface is established and documented, users only pass their function or function object with clear inputs and outputs to the system.

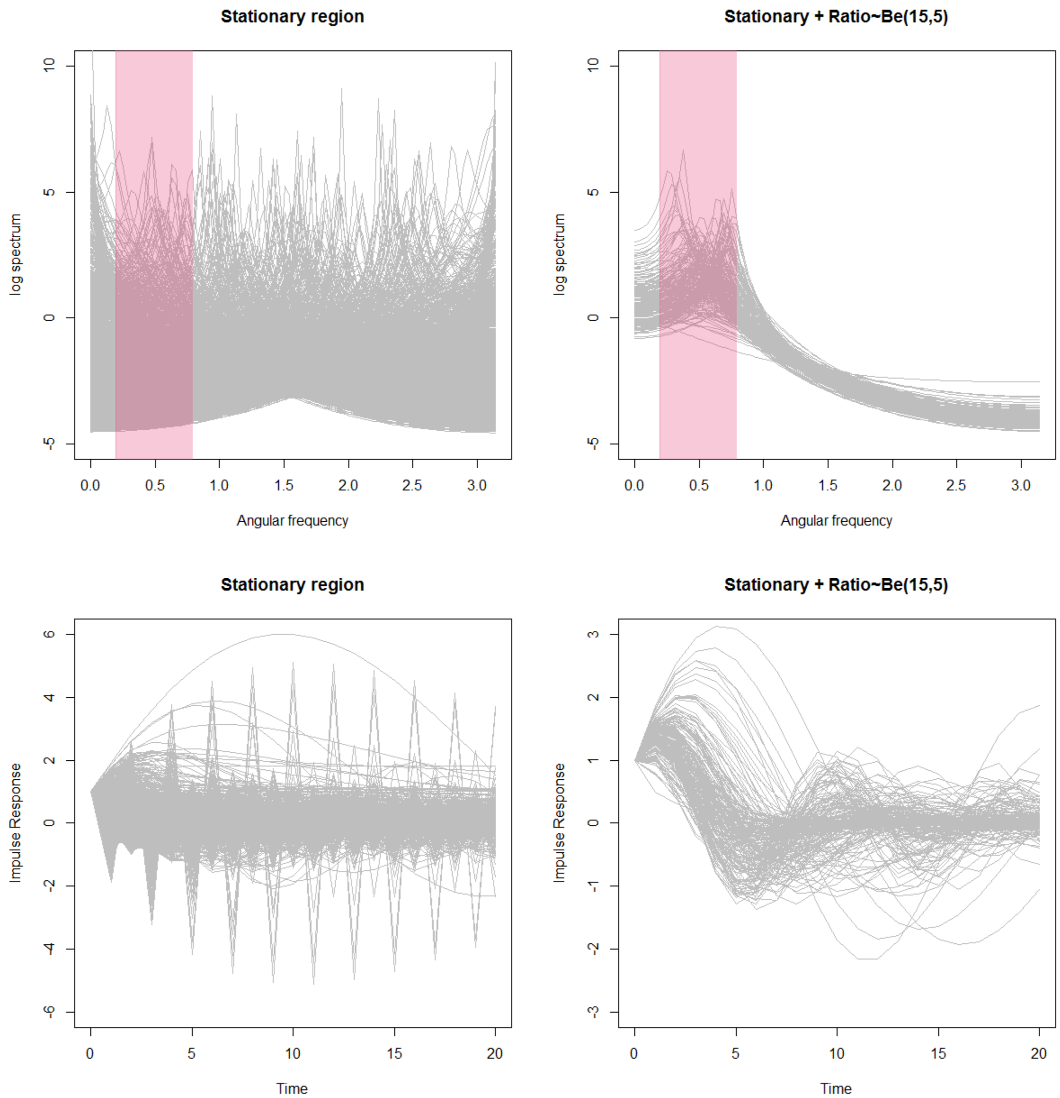
(Andrle and Plašil, 2016). Recall also, that non-Bayesian analysis can embrace the penalized loss-function approach to inference as well.

Figure 1: Parameter regions for different priors



Note: Kernel estimates of the joint prior density. Left upper panel: normally distributed independent marginal priors for ϕ_1 and ϕ_2, upper right: identical priors restricted to the stationarity region, bottom left: stationarity + at least 60% of variability comes from business cycle frequencies, bottom right: stationarity + the share of business cycle frequencies given by $Be(15,5)$.

Figure 2: Model properties for admissible regions



Note: The business cycle frequencies are denoted by the shaded region.

IV. CONCLUSION

Building on Andrle and Benes (2013), we provided a background theory of system priors accompanied by an illustrative example, placing emphasis on the elements and mechanics of system priors' application. System priors bring on board views about high level features of models, not necessarily just individual coefficients. As such, they provide a more refined way of incorporating prior information on complex functions of parameters, like impulse-responses or frequency-response functions.

The specification and implementation of system priors was illustrated using a second-order autoregressive process, which, despite its simplicity, can display nontrivial dynamics. Gaussian independent priors on the autoregressive coefficients do not restrain the model dynamics in a meaningful way when it comes to cyclical properties of the process. The polar re-parameterization suggested in the literature is a specific modification with only a modest improvement. However, it was illustrated that imposing a restriction that more than 60% of the model's dynamics comes from business cycle frequencies allows the parameters to be estimated only in a region with plausible cyclical dynamics of the impulse-response function. Other economically relevant priors could have been chosen due to the generality of system priors and options are virtually unlimited in more sophisticated models.

We believe that system priors are a useful top-down approach to eliciting priors –possibly hierarchical– about model characteristics as long as these are computable functions of the underlying coefficients. System priors allow researchers to work with informative and economically-meaningful priors in econometric and structural economic models, be it state-space models, Bayesian vector auto-regressions, or others. Importantly, system priors are easy to incorporate within the existing Bayesian toolkits with only little computational overhead.

REFERENCES

Andrle, M., and Benes, J. (2013): "System Priors: Formulating Priors about DSGE Models System Properties," *Working Paper WP/13/257*, International Monetary Fund, Washington DC.

Andrle, M., and Plašil, M. (2016): "System Priors for VARs," *mimeo*.

Clark, P. (1987): "The Cyclical Comovement of U.S. Economic Activity," *Quarterly Journal of Economics*, 102, pp. 797-814.

Geweke, J. (2010): *Complete and Incomplete Econometric Models*, Princeton University Press.

Giannone, D., M. Lenza, and G.E. Primiceri (2014): "Priors for the Long Run", mimeo

Harvey, A.C., and Koopman, S.J. (1997): *Multivariate structural time series models*, In: Heij, C., et al. (Eds.), System Dynamics in Economic and Financial Models. Wiley, Chichester.

Herbst, E., and Schorfheide, F. (2014): "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, Vol. 29, Issue 7, pp.1073-1098.

Kuttner, K. (1994): "Estimating potential output as a latent variable," *Journal of Business and Economic Statistics*, 12, 3, 361-368.

Planas, C., Rossi, A., and Fiorentini, G. (2008): "Bayesian Analysis of the Output Gap," *Journal of Business & Economic Statistics*, 26 (1), pp. 18-32.

Sims, Ch. (2005): "Conjugate Dummy Observations for VARs," *Princeton University mimeo*.

Sims, Ch. And T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models", International Economic Review, 39, 949—968

Theil, H., and Goldberger, A. (1961): "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, Vol. 2, No. 1, pp. 65-78.

Villani, M. (2009): "Steady-State Priors for Vector Autoregressions", Journal of Applied Econometrics, Vol 24, 630—650

Watson, M. (1986): "Univariate detrending methods with stochastic trends," *Journal of Monetary Economics*, 18, pp. 49-75.

V. APPENDIX: PSEUDO CODE FOR THE POSTERIOR KERNEL

The following is a simplified pseudo-code for implementing the computations to evaluate the formula (2) in the main body of the text, restated here for convenience:

$$p(\theta \mid Y;M) \propto L(Y \mid \theta;M) \times [\, p_s(h(\theta);M) \times p_m(\theta)].$$

The function evaluating all three components of $p(\theta \mid Y;M)$ takes as inputs the vector of coefficients, θ, to evaluate the criterion function for, the model (either already solved for θ or to be solved for θ), observed data required for evaluation of the log-likelihood and possibly also for evaluation of the system priors.

A crucial input is the user-defined function, `logsprior_user_fun`, that can evaluate the system priors for a given coefficient vector, θ. The function handle, or a function object[13], needs to follow a pre-specified application programming interface (API) to be used with a general toolbox.

The evaluation of the function can be efficient with solving the model with a new vector of coefficients only once or evaluating all three components in parallel.

The switches allow to switch between Bayesian estimation with System Priors, Bayesian estimation without system priors, maximum likelihood estimation with no explicit priors, or investigation of the compound prior by switching off the likelihood component. Although the "do_xx" switches are not shown explicitly as inputs, they are included in the function (or function object).

[13] For illustration of function objects in multiple programming languages, see https://en.wikipedia.org/wiki/Function_object. Function object is an object that can be called like a function, yet can do more, for instance "remember" a lot of data, its previous state, etc.

PSEUDO CODE:

```
[crit]  = function(theta, Model, Data, logsprior_user_fun, … )
BEGIN


      /* Evaluate the marginal priors:  p_m(θ)]. */
      IF (do_mprior == TRUE)
            Log_mprior = evalMarginalPriors(theta, hyperParameters);
      ELSE
            Log_mprior = 0;
      END

      /* Evaluate the SYSTEM priors:  p_s(h(θ);M)*/
      IF (do_sprior == TRUE)
            Log_sprior = call(@logsprior_user_fun(theta, Model, Data);
      ELSE
            Log_sprior = 0;
      END

      /* Evaluate the likelihood or other criterion function: L(Y|θ;M) */
      IF (do_loglik == TRUE)
            Log_lik  = evalLoglikelihood(theta, Data, Model);
      ELSE
            Log_lik  = 0;
      END


      /* Assemble and return the posterior value */
      crit = Log_lik + Log_sprior + Log_mprior

END
```