Econometrics with System Priors

Michal Andrle

IMF, Research Dept.

IMF ICD Seminar,
Nov 7, 2017

Joint work with J. Benes and M. Plasil

Disclaimer:

The views expressed herein are those of the authors and should not be attributed to the International Monetary Fund, its Executive Board, or its management.

System Priors

System priors:

Prior views about **system properties** of the model.
These may be complex functions of all underlying parameters.

System priors are very explicit, transparent, economically meaningful, and can relate to any of the model's properties.

Your boss and colleagues will understand your priors.
And can disagree.

Aren't My Priors Obious?

Parameters	Prior			Posterior maximisation		Metropolis-Hastings sampling			
	Type	Mean	Stdev	Mode	Stdev	Mean	Mode	5%	95%
σ_l	invg	2.00	0.50	1.81	0.39	2.08	1.81	1.29	2.85
h	beta	0.70	0.20	0.79	0.04	0.78	0.79	0.72	0.84
μ^w	invg	1.25	0.20	1.38	0.11	1.38	1.38	1.21	1.56
μ^m	invg	1.20	0.20	1.23	0.20	1.35	1.23	0.97	1.72
μ^c	invg	1.05	0.05	1.06	0.10	1.08	1.06	0.91	1.24
μ^i	invg	1.05	0.05	1.02	0.09	1.05	1.02	0.89	1.21
μ^g	invg	1.05	0.05	1.01	0.09	1.04	1.01	0.88	1.19
μ^x	invg	1.05	0.05	1.03	0.09	1.06	1.03	0.90	1.22
S	norm	7.69	1.50	9.73	1.27	9.90	9.73	7.80	11.97
ξ_w	beta	0.83	0.10	0.86	0.03	0.83	0.86	0.77	0.89
ξ_d	beta	0.75	0.10	0.74	0.04	0.75	0.74	0.69	0.82
ξ_m	beta	0.50	0.10	0.30	0.06	0.30	0.30	0.21	0.40
ξ_c	beta	0.60	0.10	0.84	0.02	0.84	0.84	0.80	0.88
ξ_i	beta	0.60	0.10	0.73	0.04	0.74	0.73	0.68	0.80
ξ_g	beta	0.60	0.10	0.74	0.06	0.73	0.74	0.64	0.84
ξ_x	beta	0.60	0.10	0.73	0.05	0.73	0.73	0.65	0.81
κ_w	beta	0.50	0.10	0.59	0.06	0.61	0.59	0.51	0.71
κ_d	beta	0.50	0.10	0.47	0.08	0.47	0.47	0.33	0.60

ROUND 1:
Examples and Intuition

"Invert, always invert. (Carl Jacobi) "

AR(2) Example

Assume an AR(2), say a model of an *output gap*

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2)$$

What are plausible priors for ϕ_1, ϕ_2?

Is choosing $\phi_1 \sim N(0, \sigma_{\phi_1})$ and $\phi_2 \sim N(0, \sigma_{\phi_2})$ reasonable?

AR(2) Example

Assume an AR(2), say a model of an *output gap*

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2)$$

What are plausible priors for ϕ_1, ϕ_2?

Is choosing $\phi_1 \sim N(0, \sigma_{\phi_1})$ and $\phi_2 \sim N(0, \sigma_{\phi_2})$ reasonable?

Let's use a system prior:

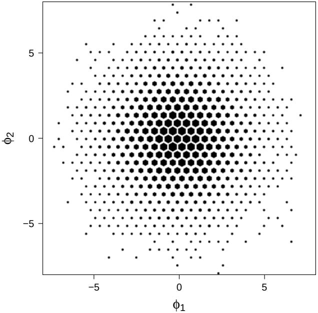
stationarity

+

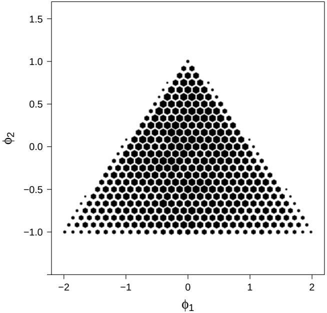
"around" 60% of variance from cyclical frequencies

It's an output gap, right?

AR(2) Example – Joint Prior

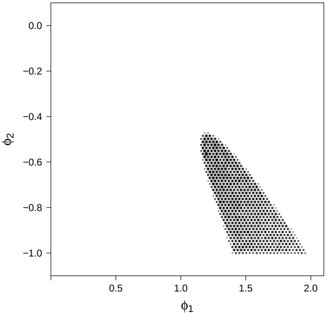


Independent Normal

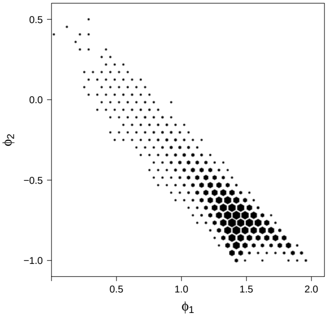


Independent Normal + Stationary

AR(2) Example – Joint Prior

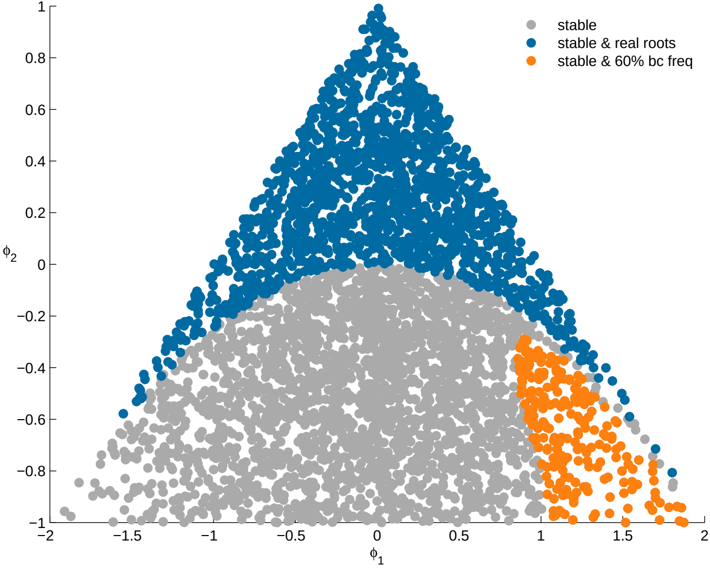


Stationary
+
At least 60% at BC freqs...

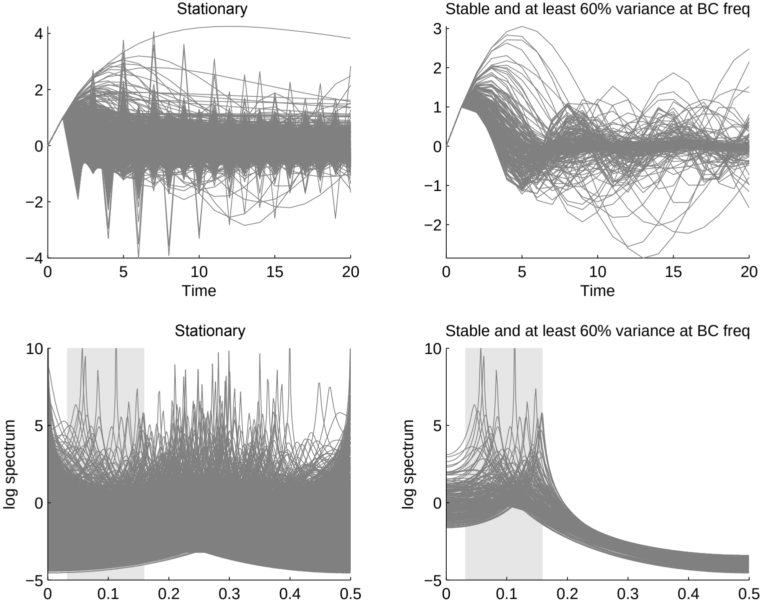


Stationary
+
BC freq. share $\sim$ *Beta*$(15, 5)$
("around 60%")

AR(2) Example – Smaller Sub-Space



AR(2) Example: Implications of Priors...



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

Simple DNK Model Example

Simple New-Keynesian "gap" model used for illustration:

$$
\begin{aligned}
y_t &= \alpha_1 y_{t+1|t} + \alpha_2 y_{t-1} + \alpha_3 (rr_t - \overline{rr}_t) + \varepsilon_t^y & (1)\\
\pi_t^c &= \lambda_1 \pi_{t+1|t}^c + (1-\lambda_1)\pi_{t-1}^c + \lambda_2 \hat{y}_t + \varepsilon_t^\pi & (2)\\
i_t &= \gamma_1 i_{t-1} + (1-\gamma_1) \times \Big[(\overline{rr}_t + \overline{\pi}_t) + \gamma_2(\pi_{c,t+3|t}^{y/y} - \overline{\pi}_{t+3}) + \gamma_3 y_t\Big] + \varepsilon_t^i & (3)\\
rr_t &= i_t - \pi_{t+1|t} & (4)\\
\overline{\pi}_t &= \overline{\pi}_{t-1} + \varepsilon_t^{\overline{\pi}} & (5)
\end{aligned}
$$

Standard practice:
Specify marginal independent priors for $\{\alpha_1, \alpha_2, \alpha_3, \lambda_1, \ldots\}$

Economists Have Views on Sacrifice Ratio...

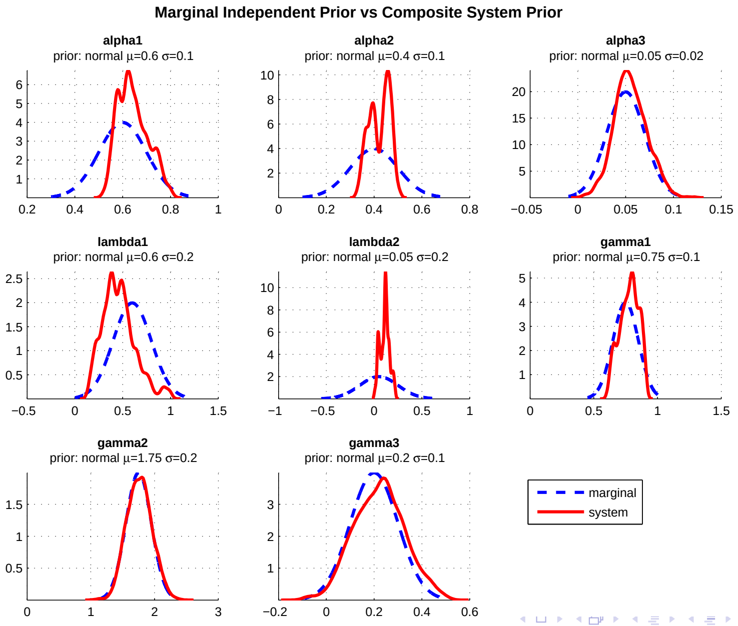
Sacrifice ratio:
Cumulative loss of output after a permanent disinflation by 1 percentage point.

System prior:
Assume the sacrifice ratio to be distributed as $N(-0.8, 0.05)$.

Note:

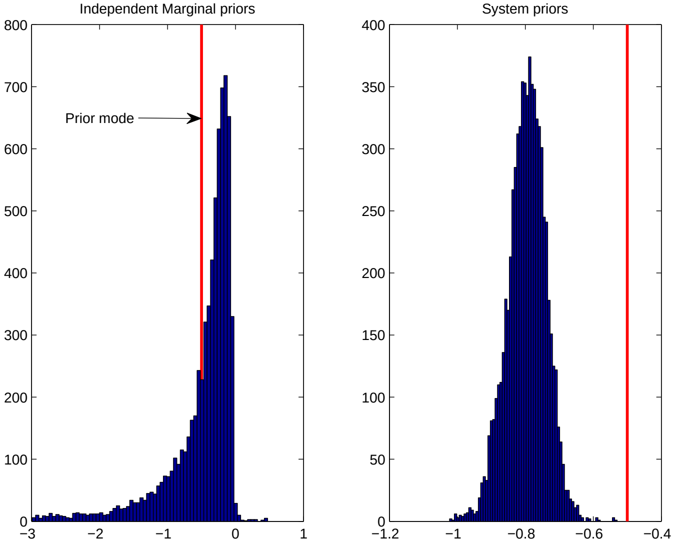
- Individual-country data samples uninformative about sacrifice rat.
- Cross-country evidence on disinflation is often available...
- Sacrifice-ratio prior does not violate the likelihood principle

System Prior Ties Down Parameters. . . (Some)



System Priors Do What They're Supposed To

Prior distribution of the sacrifice ratio…



DSGE: Some Priors We Have Used…

- **Properties of IRFs**
  (signs, speed of convergence to SS, tertiary cycles, …)
- **Spectral properties of the model**
  (measurement errors, shock contrib. at freq. bands)
- **Spectral properties of the filter transfer function**
  (gains, cut-offs, …)
- **Shock-decomposition priors**
  (in year 200X shocks did this and that)
- **Policy scenarios**
  (sacrifice ratio, delayed MP response, …)
- …

TC-VAR: Examples of SPriors We Have Used...

- **Steady-state priors**
  (intercept is not the slope...)
- **Properties of IRFs**
  (signs, speed of convergence to SS, tertiary cycles, statonarity,...)
- **Spectral properties of the model**
  (measurement errors, variance at BC freqs.)
- **Spectral properties of the filter transfer function**
  (output-gap gains, cut-offs, ...)
- **Shock-decomposition priors**
  (in year 200X shocks did this and that)
- ...

System Priors (References)

Details on Methodology:

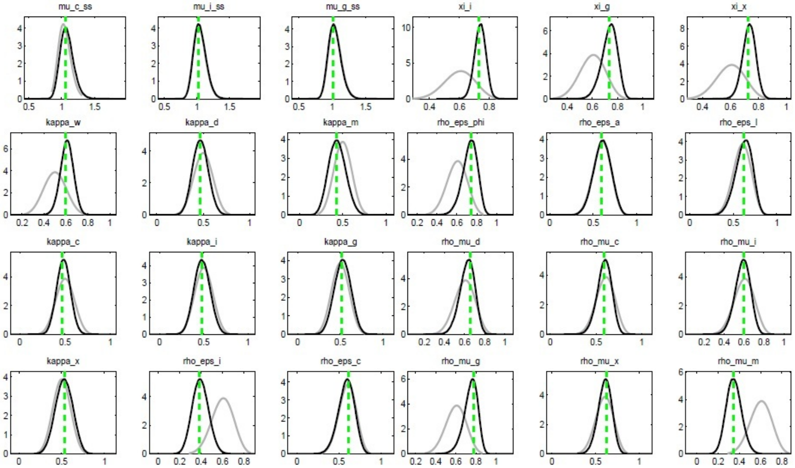
- Andrle and Beneš, IMF WP **2013** (DSGE Applications)
- Andrle and Plašil, IMF WP **2016** (TSeries Applications)

Implementations:

- ECB's New Area Model (NAWM) toolbox YADA
- Codes for IRIS Toolbox
- Codes for Dynare Toolbox
- AR(2) Example code for R

ROUND 2:
Issues with Current Practice

. . . is this enough information?



Does this substitute for prior predictive tests. . . ?

Parameters	Prior			Posterior maximisation		Metropolis-Hastings sampling			
	Type	Mean	Stdev	Mode	Stdev	Mean	Mode	5%	95%
σ_l	invg	2.00	0.50	1.81	0.39	2.08	1.81	1.29	2.85
h	beta	0.70	0.20	0.79	0.04	0.78	0.79	0.72	0.84
μ^w	invg	1.25	0.20	1.38	0.11	1.38	1.38	1.21	1.56
μ^m	invg	1.20	0.20	1.23	0.20	1.35	1.23	0.97	1.72
μ^c	invg	1.05	0.05	1.06	0.10	1.08	1.06	0.91	1.24
μ^i	invg	1.05	0.05	1.02	0.09	1.05	1.02	0.89	1.21
μ^g	invg	1.05	0.05	1.01	0.09	1.04	1.01	0.88	1.19
μ^x	invg	1.05	0.05	1.03	0.09	1.06	1.03	0.90	1.22
S	norm	7.69	1.50	9.73	1.27	9.90	9.73	7.80	11.97
ξ_w	beta	0.83	0.10	0.86	0.03	0.83	0.86	0.77	0.89
ξ_d	beta	0.75	0.10	0.74	0.04	0.75	0.74	0.69	0.82
ξ_m	beta	0.50	0.10	0.30	0.06	0.30	0.30	0.21	0.40
ξ_c	beta	0.60	0.10	0.84	0.02	0.84	0.84	0.80	0.88
ξ_i	beta	0.60	0.10	0.73	0.04	0.74	0.73	0.68	0.80
ξ_g	beta	0.60	0.10	0.74	0.06	0.73	0.74	0.64	0.84
ξ_x	beta	0.60	0.10	0.73	0.05	0.73	0.73	0.65	0.81
κ_w	beta	0.50	0.10	0.59	0.06	0.61	0.59	0.51	0.71
κ_d	beta	0.50	0.10	0.47	0.08	0.47	0.47	0.33	0.60

Issues with ‘standard DSGE’ priors (1)

- Assumption of independent marginal priors is often **unrealistic**
- Reporting only marginal parameter prior and posterior distributions is **not informative enough**
- Independent priors induce **unintended consequences** for the prior distribution of model features (IRFs, moments, etc.)
- Independent marginal priors are **not always transparent** enough; looking at them gives you often no clue...
- Very little or **no economics** of “adjustment-costs” coefficients or similar priors...

Issues with ‘standard DSGE’ priors (2)

Marginal independent priors give little control over priors!

- Overly **diffuse** marginals imply loose control over a particular system feature of the model...
- Overly **tight** marginals give little chance for data to speak
- Marginal priors are **too blunt** for economically meaningful priors

Prior-predictive analysis often absent (but badly needed)

- What is the prior distribution of your monetary policy shock IRF?
- Could the response of labor to a TFP shock be positive in your model at all? Do priors tilt it that way?

System Priors: Motivation

- **Economically meaningful. . .**
- **Top-Down vs. Bottom-Up Specification. . .**
  - Calibrated models use[d] top down specification
  - Top down **priors** on system behavior of the model
  - Top down approach allows you to implement **priors that make sense** and that other economists would understand
- **System priors induce cross-dependence among parameters**
  - A prior on a model feature is consistent with a set of parameterizations (iso-parametric path)
  - Just one system prior is enough to induce a joint distribution prior across multiple structural parameters

ROUND 3 – FINAL: Theory and Computation

“Premature optimization is the root of all evil. . . ” (D. Knuth)

Bayesian Updating $\equiv$ Inverse Probability

Method of Inverse Probability:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{6}$$

Bayesian Updating

A **model** M with parameters $\theta \in \Theta$. Observed data, Y^o.

Likelihood function $L(Y|\theta; M)$

Marginal independent priors:
$p_m(\theta) = p(\theta_1) \times \ldots \times p(\theta_N)$

Bayesian learning – use data to update your priors:

$$p(\theta|Y^o; M) \propto L(Y^o|\theta; M) \times p_m(\theta). \tag{7}$$

BUT the prior $p(\theta)$ can be anything, as long it's a distribution...

System Priors

Let's keep marginal independent priors on parameters: $p_m(\theta)$.

Specify a **system feature** $r = h(\theta; M)$, for $\theta \in \Theta$.
Example: IRFs, sacrifice ratio, ...

System prior about feature $r = h(\theta; M)$:

$$p_s(r\,|\theta; M) \equiv p_s(h(\theta; M)|\theta; M) \tag{8}$$

Bayesian update – use the r-likelihood to update $p_m(\theta)$:

$$p_c(\theta|r; M) \propto p_s(h(\theta)|\theta; M) \times p_m(\theta). \tag{9}$$

System Priors – Putting it All Together

Given the marginal priors, system priors, and the data:

Bayesian updating:

$$p(\theta | Y^o, M) = L(Y^o|\theta; M) \times [p_s(h(\theta); M) \times p_m(\theta)]$$

System Priors – Putting it All Together

Given the marginal priors, system priors, and the data:

Bayesian updating:

$$p(\theta | Y^o, M) = L(Y^o | \theta; M) \times [p_s(h(\theta); M) \times p_m(\theta)]$$

This is not a short-cut.
Not a trick.

Please, try this at home!

System Priors: Computations

Loss function with three components:

(i) likelihood function (or other criterion function), $L(Y^o|\theta, M)$

(ii) marginal independent priors, $p_m(\theta|M)$

(iii) **system priors**, $p_S(r|M)$, with $r = h(\theta)$

Posterior sampling:

- Simple extension of standard MCMC, e.g. RW-Metropolis
- To sample from spriors, 'switch-off' the likelihood!
- Due to the global nature of composite prior, adaptive **Sequential Monte-Carlo** is preferred (massively parallel)

Available for: YADA (ECB Tbx), DYNARE, IRIS, ...

System Priors: 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
```

Relationship to the Literature

- Faust (2009) and Gupta and Faust (2011) point out unintended consequences of ‘standard’ marginal independent priors using **prior-predictive analysis**
- Geweke (2010) discusses prior-predictive analysis at length
- Canova and Sala (2010) point out identification problems of DSGE models
- Fernandez-Villaverde and Rubio-Ramirez (2008): How Structural are Structural Parameters?
- Work of E.T. Jaynes on priors and max-ent priors and ‘moment approach’ to prior selection in J.O. Berger (1985)

Conclusions

- System priors are the economically meaningful way of using priors
- System priors may solve many problems of marginal independent priors
- System priors induce individual parameter priors
- System priors encompass ‘standard’ way of doing things

Thank you for your patience...