Multiple Time Series Priors

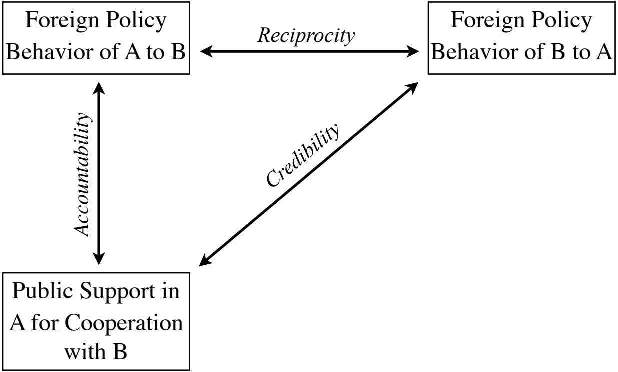
(Or how we learned to stop worrying, and love Bayesian time series)

Patrick T. Brandt
University of Texas, Dallas
pbrandt@utdallas.edu
John R. Freeman
University of Minnesota
freeman@umn.edu

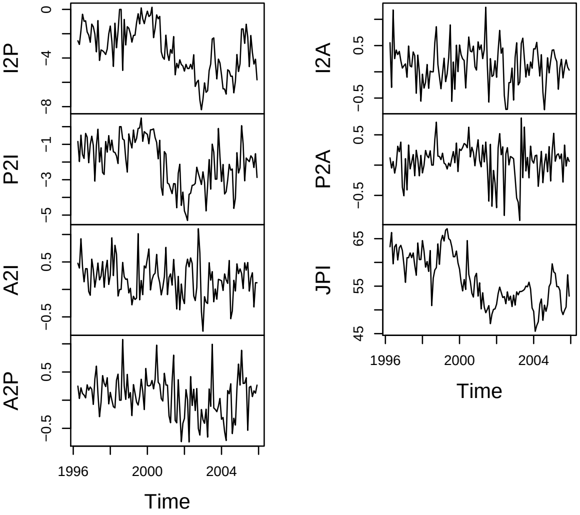
April 18, 2007

Example of "structure" and "dynamics"

Example of structure and dynamics in a model of international conflict with an audience (Brandt, Colaresi and Freeman, forthcoming, JCR):

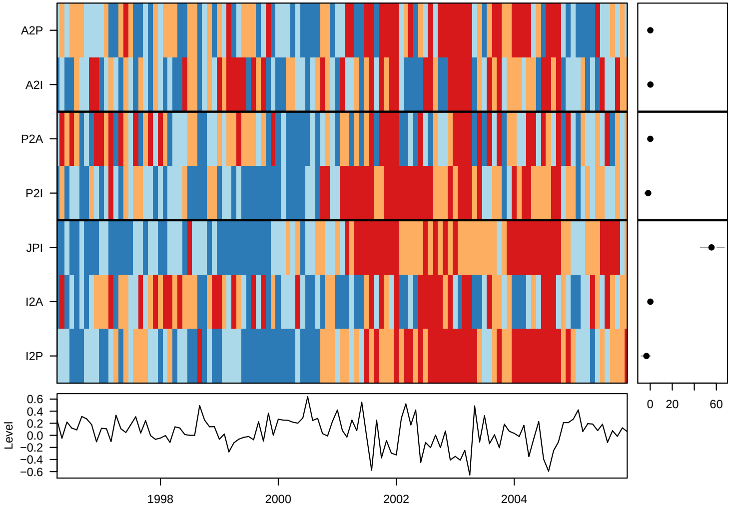


Boring version of the data



Better version of the data

Multiple time series plotting method from Peng (2008 JSS):



Why do we want to impose restrictions / structure?

Time series (regression) analysis includes a whole series of parameters that describe the cycles, trends, and deterministic components of the data. These raise issues of

- Model Scale and Complexity
- Dynamics
- Specification uncertainty
- Endogeneity

But there is high uncertainty, since these are large models with MANY parameters. We need some "loose" restrictions on the parameters — so we use Bayesian priors.

What should we worry about?

- What are the dynamic implications of a prior and how are these related to the dynamics of our data in computing the posterior?
- How can we elucidate or elicit prior beliefs about dynamics?
- How are prior beliefs correlated?
- What contemporaneous structure should be imposed on multiple time series data?

Outline

Distinctiveness of Bayesian Time Series

Compare discussions in Gill, (2004, PA) and Jackman (2004, Annals). These address value of Bayesian models in political science. But,

- Flat prior equivalence to MLE breaks down in regions of non-stationarity; but under a suitable prior the posterior can be a known pdf
- "Modern" Bayesian time series analysis often uses historical or cross-unit time series data as basis for the benchmark or baseline prior.
- As will be explained, for B-SVAR models, estimation (sampling) is a guided or normalized "random tour" of the parameter space.

Dynamics and beliefs

Beliefs about dynamic data:

- "Past performance is not a predictor of future results": the prior your retirement fund / broker wants you to have to minimize *their* risk.
- "Best prediction of tomorrow is *today* with a random shock": makes sense, but what is the scale of the shock?
- "Some weighted function of past results": this is the moving average or autoregressive process of the past values, but "What are the weights?"

Parameterizing a prior for dynamics

Consider a simple example:

$$y_t = \phi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$

This is an AR(1) process model. Prior beliefs about ϕ, are an important component of any Bayesian model of this process.

Importance of prior in dynamic models

Consider simulated data for

$$y_t = 0.5y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \quad t = 1, \ldots 200,$$

with the following priors for ϕ:

Diffuse Case $\phi \sim N(0, 1)$

Calibration or Empirical $\phi \sim N(0.5, 0.01)$

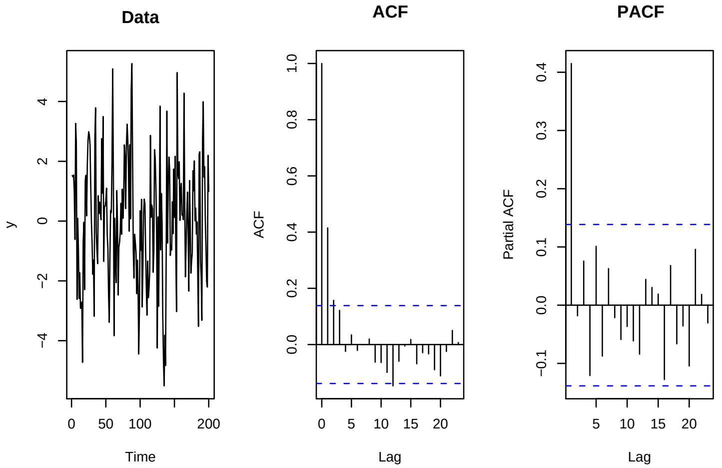
Informed, random walk $\phi \sim N(1, 0.01)$

Simulation

We can do this in Wash U's favorite `MCMCpack`:

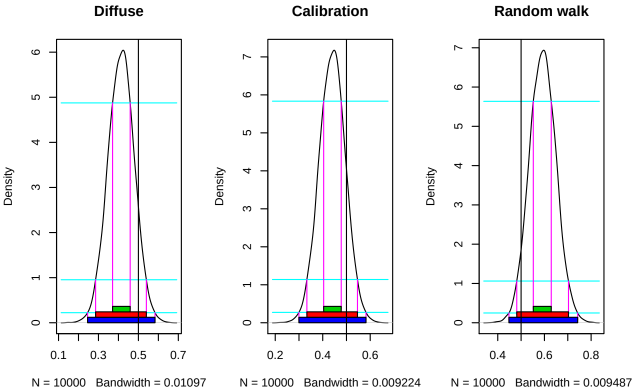
```
library(MCMCpack); library(hdrcde)

set.seed(123)

N <- 200
y <- arima.sim(n=N, list(ar=0.5), sd=2)

pdf(file="AR1example.pdf", width=6, height=4)
par(mfrow=c(1,3))
plot(y, main="Data");  acf(y, main="ACF"); pacf(y, main="PACF")
dev.off()

M1 <- MCMCregress(y[2:N] ~ lag(y, 1)[0:(N-1)],
                  b0=0, B0=1, c0=1, d0=1)

M2 <- MCMCregress(y[2:N] ~ lag(y, 1)[0:(N-1)],
                  b0=c(0, 0.5), B0=100, c0=1, d0=1)

M3 <- MCMCregress(y[2:N] ~ lag(y, 1)[0:(N-1)],
                  b0=c(0,1), B0=100, c0=1, d0=1)

pdf(file="AR1posteriors.pdf", width=6, height=4)
par(mfrow=c(1,3))
hdr.den(M1[,2], main="Diffuse"); abline(v=0.5)
hdr.den(M2[,2], main="Calibration"); abline(v=0.5)
hdr.den(M3[,2], main="Random walk"); abline(v=0.5)
dev.off()
```

Example data



So what do the posteriors look like?

Posteriors for ϕ for the three priors:



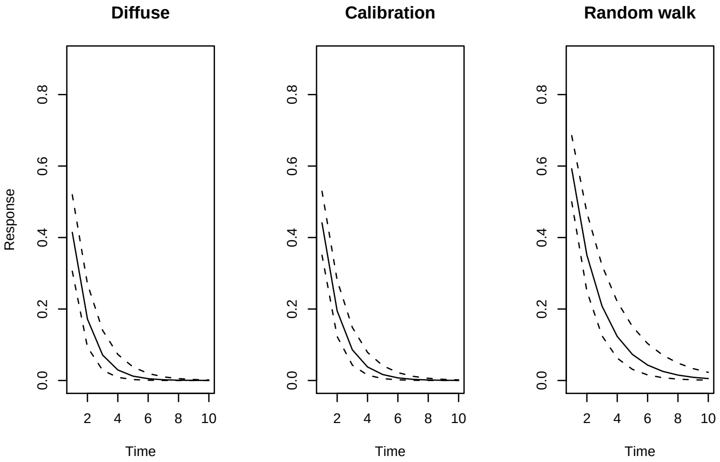
Influence of the prior is large!

So what?

- So I made up an example where I can cook the prior? (cue Homer Simpson)
- The bigger point is the sensitivity, since the priors imply very different dynamics (this is coming on the next slide).
- The informed priors beliefs imply much more persistent dynamics (this too is coming on the next slide).
- This means we are saying different things about how shocks decay across these priors / posteriors (yep, this is also on the next slide).

Associated impulse responses

These trace out the impacts of a one unit change or a shock to the residuals over time, with 90% pointwise credible intervals:



So what is the difference?

- The impulse responses for diffuse and calibration priors reach zero after 6 periods.
- Random walk prior impulse response reaches zero after 10 periods.
- So the prior is generating the story about the dynamic effects.
- This also impacts the estimation of the 90% credible intervals (more on this later).

Further Bayesian Perspectives in AR(1) Model

- Choose uniform prior for ϕ (Sims 1988, Sims and Uhlig 1991, DeJong and Whiteman 1991)
- Choose Jeffreys-type prior for ϕ (Phillips 1991 a,b)
- Use predictive elicitation allowing a family of piecewise conjugate (normal or normal-inverse gamma) prior distributions that permit different opinions when ϕ is less than, equal to, or greater than 1 (Kadane et al 1996)

Multiple time series models

We generally implement these ideas on a vector autoregression (VAR) modeling framework.

- VAR models have one equation for each of m endogenous variables in the system.
- Equations right hand side variables include p lags of all of the endogenous variables in the system.
- Model has $m^2 p + m$ regression parameters, plus contemporaneous or error covariance terms (at most $\frac{m(m+1)}{2}$).

(S)VAR specification

Model

$$\underset{1\times m}{y_t}\,\underset{m\times m}{A_0} + \sum_{\ell=1}^{p} \underset{1\times m}{y_{t-\ell}}\,\underset{m\times m}{A_\ell} = \underset{1\times k}{Z_t}\,\underset{k\times m}{D} + \underset{1\times m}{\epsilon_t}, \quad t = 1, 2, \ldots, T,$$

Structural innovations

$$E[\epsilon_t | y_{t-s}, s > 0] = \underset{1\times m}{0}, \quad \text{and} \quad E[\epsilon_t' \epsilon_t | y_{t-s}, s > 0] = \underset{m\times m}{I}.$$

Notation:

$$\begin{aligned} A_+ &= (m^2 p + k) \times m \text{ stacking of the coefficients } A_\ell \\ a_+ &= \text{column major stacking of } A_+ \\ a_0 &= \text{column major stacking of } A_0 \end{aligned}$$

Litterman (and related) priors

Litterman (1980, 1986) presents a prior for multivariate time series models that encompasses a few simple ideas:

- Random walk prior for the dynamic coefficients: first lag coefficients are 1 with some standard deviation.
- Discounting of sample error covariances
- Variances of higher order lags are smaller
- Own-lags matter more than other-lags

Sims-Zha prior

Problem: Litterman prior is equation-by-equation rather than for the full system. This leads to an unknown and possibly intractable posterior sampling problem.

- Want a prior that is consistent across the equations, or for the whole system (so drop the own v. other lag) .
- Initial conditions for the likelihood matter when modeling trending data: add a prior on initial conditions.
- Presence of cointegration implies a sum of coefficients prior: add a prior on the sum of autoregressive coefficients.
- Scale of prior for intercepts should be separate from other parameters because these describe variability around the mean or trends.

Sims-Zha Prior Hyperparmeters

Need to specify the prior variances or standard deviations on MANY dynamic parameter coefficients (at least 10^2 or 10^3+):

Parameter	Range	Interpretation
λ_0	[0,1]	Scale of the error covariance matrix
λ_1	> 0	Standard deviation about A_1 (persistence)
λ_2	$= 1$	Weight of own lag versus other lags
λ_3	> 0	Decay of variance of coefficients on lags
λ_4	≥ 0	Scale of standard deviation of intercept
λ_5	≥ 0	Scale of standard deviation of exogenous variables coefficients
μ_5	≥ 0	Sum of autoregressive coefficients component
μ_6	≥ 0	Dummy initial observations component

Structural VAR

Structure is typically the contemporaneous relationships:

- Endogeneity
- Within-period effects v. lagged effects
- (Zero) restrictions in A_0.
- Decomposition of the contemporaneous residual variances. Establishes how impulse response functions (IRFs) are interpreted
- Can also add structure by restricting lagged relationships (but we do not do that, cf. Waggoner and Zha 2003a)

SVAR $\rightarrow$ VAR

SVAR's become reduced form VARs if we post-multiply by A_0^{-1}:

$$
\begin{aligned}
y_t A_0 + \sum_{\ell=1}^{p} y_{t-\ell} A_\ell &= Z_t D + \epsilon_t \\
y_t A_0 A_0^{-1} + \sum_{\ell=1}^{p} y_{t-\ell} A_\ell A_0^{-1} &= Z_t D A_0^{-1} + \epsilon_t A_0^{-1} \\
y_t + \sum_{\ell=1}^{p} y_{t-\ell} A_\ell A_0^{-1} &= Z_t D A_0^{-1} + \epsilon_t A_0^{-1} \\
y_t &= -\sum_{\ell=1}^{p} y_{t-\ell} A_\ell A_0^{-1} + Z_t D A_0^{-1} + \epsilon_t A_0^{-1} \\
y_t &= \sum_{\ell=1}^{p} y_{t-\ell} B_\ell + Z_t C + u_t
\end{aligned}
$$

where

$$C = DA_0^{-1} \quad B_\ell = -A_\ell A_0^{-1}, \quad \ell = 1, 2, \ldots, p, \quad u_t = \epsilon_t A_0^{-1}$$

Posterior for B-SVAR model

$$\underbrace{q(A)}_{Posterior} \propto \underbrace{L(Y|A)}_{Likelihood} \cdot \underbrace{\pi(a_0)}_{Prior\ of\ A_0} \cdot \underbrace{\phi(\widetilde{a_+}, \Psi|A_0)}_{Prior\ of\ A_+|A_0}$$

$$\propto \pi(a_0)|A_0|^T|\Psi|^{-0.5} \times exp[-0.5(a_0'(I \otimes Y'Y)a_0$$
$$-2a_+'(I \otimes X'Y)a_0 + a_+'(I \otimes X'X)a_+ + \widetilde{a_+}'\Psi\widetilde{a_+})].$$

- Let Y be left-hand side and X be right-hand (vectorized).
- This posterior with have 2^m modes because it is invariant to changing the signs of structural innovations / coefficients.
- Need to map $2^m - 1$ modes back to one for posterior convergence and interpretation. *This is likelihood normalization.*
- We can use Gibbs sampling since the kernel is normal. See Waggoner and Zha (2003a,b) for details and normalization.

Estimation of B-(S)VAR models

Several steps

- Specification of the structure of A_0 (non-Bayesian).
- Specification of the prior hyperparameters (very Bayesian).

Estimation or Sampling

- Find peak of posterior numerically.
- Use a Gibbs sampler to draw A_0 for the model. Normalize A_0 draws at each iteration.
- Conditional on A_0, sample other coefficients and quantities $(A_+, B_\ell, \text{impulse responses})$.
- Summarize results.

Outline of B-SVAR Gibbs sampler

- Basic idea: recursively sample the parameters of the posterior distribution using conditional marginal distributions of the parameters and data.

$$q(A) \propto Pr(A_+|A_0)Pr(A_0)$$

- Hard part: Sampling A_0 is expensive. Waggoner and Zha (2003a) present a Gibbs sampler that improves on the importance sampler of Sims and Zha (1998). This allows us to sample A_0 for all (over-) identified B-SVAR models (hard part is because pdf is non-standard).
- Implementation: We have implemented this sampler in the `R` package `MSBVAR`.

Inference and Reporting Results

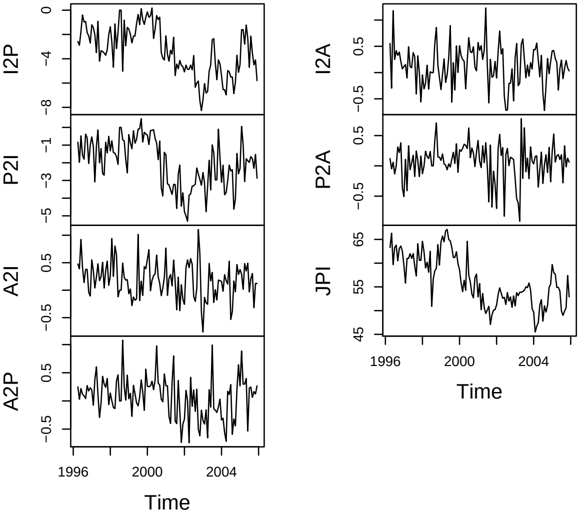
- Model selection: need to report prior hyperparameters and A_0 structure. Can use Bayes factors, BICs and LPDs to summarize and evaluate models. This is an active area: see Sims, Waggoner and Zha (forthcoming), *Journal of Econometrics*.
- Impulse responses and forecasts must have error bands, preferably Bayesian shape error bands.

Example I: Testing Structures

Brandt, Colaresi and Freeman (forthcoming, JCR) looks at

1. How to test competing contemporaneous structures. This can be accomplished by computing log marginal data densities and other posterior quantities.
2. Forecasting conflict to provide an early warning indicator / model.
3. Presence or structures of reciprocity, credibility and accountability dynamics in Israeli-Palestinian conflict.
4. Impact of Jewish support for the peace process (JPI) on dyadic conflict is important for forecasts.

Monthly mean Israeli-Palestinian conflict, 1996:4-2005:3



Goal of this paper

- Determining the structure, or contemporaneous relationships among the series that are consistent with the reciprocity, accountability, and credibility dynamics.
- Showing that models with explicit structure are far superior (i.e., have higher posterior probability) than those with recursive or agnostic contemporaneous structures.
- Forecasts are vastly improved by including Jewish public opinion dynamics.

Different Models

- Recursive: responses depend on an arbitrary ordering of equations.
- Bystander: no reaction to or from the public.
- Follower: public opinion follows leaders' actions.
- Accountability: public holds leaders responsible.
- Credibility: accountability and adversary monitors public's ability to hold leaders responsible

Credibility Model

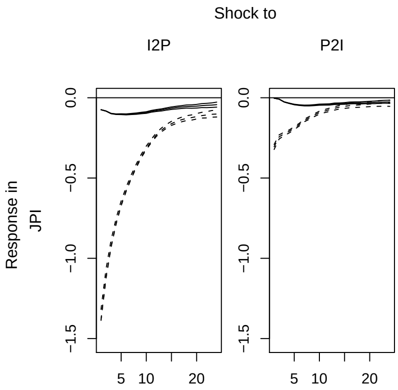
Variable	$I2P_t$	$P2I_t$	$A2I_t$	$A2P_t$	$I2A_t$	$P2A_t$	JPI_t
$I2P_t$	X	X	X				C
$P2I_t$	X	X		X			C
$A2I_t$	X		X		X		C
$A2P_t$		X		X		X	C
$I2A_t$			X		X		
$P2A_t$				X		X	
JPI_t	C	C					X

Note: other structural models come from zero restrictions on the "C" terms.

Posterior fit of A_0's

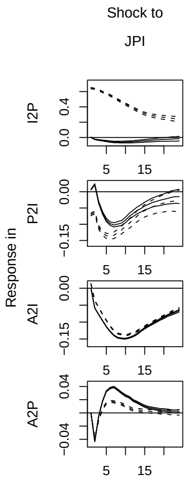
Model	$log(Pr(A_0 \| Model))$	$Pr(A_0 \| Model)$
Recursive	-22.29	2.1×10^{-10}
Bystander	-2.29	0.10
Follower	-3.84	0.02
Accountability	-2.49	0.08
Credibility	-0.56	0.57

Opinion responses



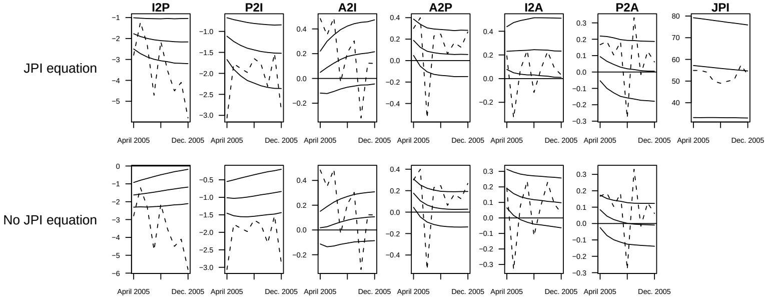
Follower = solid lines; Credibility = dashed lines; 68% credible intervals

Accountability / Credibility responses



Accountability = solid lines; Credibility = dashed lines; 68% credible intervals

Forecasts


I2P
P2I
A2I
A2P
I2A
P2A
JPI
JPI equation
No JPI equation
April 2005
Dec. 2005

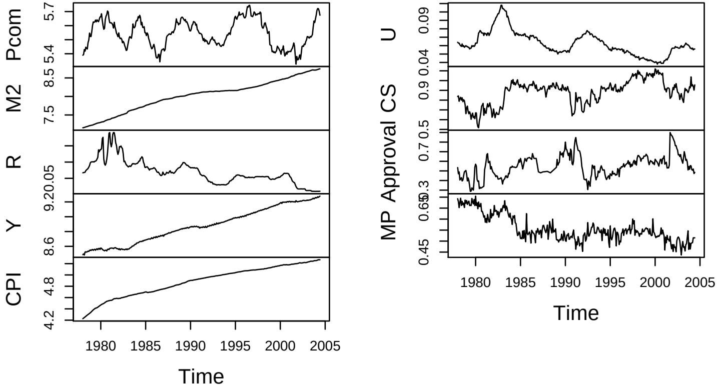
Example II: Evaluating Competing Priors

Our work on macro-political dynamics looks at the interactions of the real economy and the polity. The variables in our analysis include:

- The economy: commodity prices, money, interest rates, output, inflation, and unemployment.
- The polity: consumer sentiment, presidential approval, macropartisanship

Monthly series from 1978-2004 cover both a standard macroeconomic model (Sims and Zha 1998) and a standard public opinion dynamics model (Erikson, MacKuen, Stimson 2002)

Macropolity Data



What needs to go into this dynamic model?

Three things are needed in specifying this model:

- Prior beliefs about the dynamics of the 9 series.
- Use 13 lags: so there are $9^2 \times 13 + 9 = 1062$ dynamic parameters (plus others).
- How do we specify the *contemporaneous relationships* for the 45 possible parameters in the contemporaneous A_0 matrix?
- Remember this multiplies out across the data and dynamics!
- The latter are crucial for causal relationships while the former prior matters for the speed of adjustment of the polity and economy.

Choosing the hyperparameters

Choice of hyperparameter values is critical because it implies beliefs about the dynamic paths of the model.

- EMS-SZ tight prior: assumes benchmark or baseline prior with stochastic trends and limited drift
- EMS-SZ loose prior: greater uncertainty
- Diffuse prior: little information since prior variances are large.

Prior

Hyperparameter	EMS-SZ Tight	EMS-SZ Loose	Diffuse
Error covariance matrix (λ_0)	0.6	0.6	1
Standard deviation of A_1 (λ_1)	0.1	0.15	10
Decay of lag variances (λ_3)	1	1	0
Standard deviation of intercept (λ_4)	0.1	0.15	10
Standard deviation of exogenous vars. (λ_5)	0.07	0.07	10
Sum of AR coefficients component(μ_5)	5	2	0
Initial condition component (μ_6)	5	2	0

Specifying A_0

Sector	Variables	Pcom	M2	R	Y	CPI	U	CS	A	MP
Information	Pcom	X	X	X	X	X	X	X	X	X
Monetary Policy	M2		X	X					X	
Money Demand	R		X	X	X	X			X	
Production	Y				X					
Production	CPI				X	X				
Production	U				X	X	X			
Macropolity	CS				X	X	X	X		
Macropolity	A				X	X	X	X	X	
Macropolity	MP				X	X	X	X	X	X

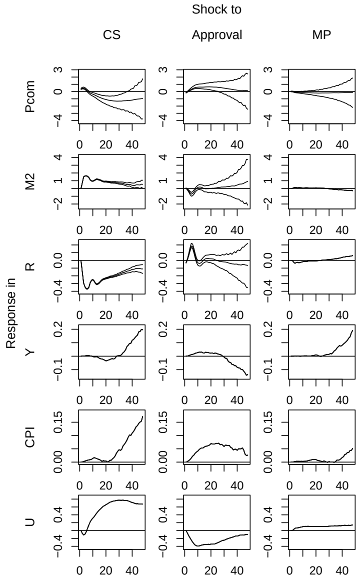
Results

What happens when we look at the competing priors for the American macro-political economy?

	EMS-SZ Tight	EMS-SZ Loose	Diffuse
$log(m(Y))$	4636	5482	12580
$log(Pr(A_0, A_+\|Y))$	2365	1527	-4507

- So the *Loose* prior generates a better fit than the *Tight* one.
- The *Diffuse* prior overfits and has $log(Pr(A_0, A_+|Y))$ that is too low.
- *Tight* versus *Loose* dynamic inferences differ sensibly.

IRFs for the Economy



Summary

What we have learned to love about Bayesian time series:

- With help of folks like Phil Schrodt, we learned how to modify SZ prior to make it more consistent with what WE have learned about IR and other kinds of political data/processes in the last several decades.
- We have learned how to use this (modified) prior to begin to make useful forecasts
- Have developed the `MSBVAR` `R` package which provides a general implementation of the Gibbs sampler for B-SVAR models, posterior impulse response summaries, and forecasts.
- Implementation of Gibbs sampler and posterior inferences requires likelihood normalization to address 2^m modes problem.

Challenges

Challenges ahead:

- Learning how best to assess fit of these models: marginal data densities, posterior probability calculations, posterior predictive checks, and / or forecast validation.
- Time-variation in the parameters: need to consider Markov-switching and changepoint processes in BVARs.
- More complex / efficient sampling: employing Markov-switching greatly expands the size of the parameter space, the need for storage / memory, and complicates posterior reporting.

Recent papers / learning more

- Brandt and Freeman (2006 PA): Error bands for IRFs in BVAR models, forecasting, and counterfactual (policy) analysis for a six equation model of the Levant.
- Brandt and Freeman (2008): "Modeling Macro-political Dynamics," the macropartisanship debate as contending priors in seven equation B-SVAR models
- Sattler, Freeman, and Brandt (2008 CPS) and Brandt, Colaresi, Freeman (2008 JCR): Theoretical debates in IPE and conflict studies as contending specifications of A_0; eleven and seven equation B-SVAR models, respectively.
- Web: `http://www.utdallas.edu/~pbrandt` or `http://yule.utdallas.edu`