2018-06-22 BEAR Frequently Asked Questions

AR coefficient

Is it possible to individually set the Auto-regressive coefficient in the BVAR for all the variables separately?

So far it is unfortunately not possible to set individual values for different variables in BEAR. It is a decision we made to keep the design of the toolbox simple, but this might change at some point.

Dummy observations

For the Bayesian Model with dummy observations, following the original paper "LARGE BAYESIAN VARs" (2008) Could you tell me if the Autoregressive coefficient in the BVAR settings corresponds to δ(delta) in the paper.

The AR coefficient in BEAR indeed corresponds to delta in the Banbura paper.

Cholesky and triangular factorisation

What is the difference between Cholesky and triangular factorisation?

With Cholesky factorisation, the variance-covariance matrix for the structural shocks is normalised to identity. This is a simplifying but sometimes undesirable assumption when the aim of the model precisely consists in evaluating the variance of the shocks. In this case, triangular factorisation should be used, and the individual variance values and be recovered from the Gamma matrix.
A consequence of these differences also arises in terms of interpretation of the impulse response functions. With triangular factorisation, the IRFs provide the response to a unit structural shock, while with Cholesky factorisation, the response is that to a one standard deviation shock.
For impulse response functions, the difference between Cholesky and triangular factorisation is thus a simple matter of scaling, the response being otherwise similar, but this is not necessarily the case. For certain applications (e.g. conditional forecasts), the two settings will produce different results. In this case, which setting is best to employ is a decision left to the user.

Normal -Wishart prior – what is Lambha 3 and Lambha 4?

lambha 3 - relative influence of first lags compared to past lags
Lambha 4 - tightness around exogenous

Data treatment:

How do I choose if I estimate a model in growth rates or levels? Does the BEAR provide criteria to select that?

This is matter of personal choice, driven by economic theory and interpretation. BEAR is not capable of telling you which choice is best and does not provide any criterion, nor does any other software. This is a consideration external to estimation methods.

Dummy extensions:

What is the difference between the dummy observation prior and the dummy initial observation extension? How do I choose in general when to use the dummy observation extensions?

The sums-of-coefficients extension is used to drive the estimates on unit-root behaviours (root=1) rather than on explosive behaviour (root>1). Due to unit roots, the model still implies non-stationarity.
For this reason, this application is often supplemented with the dummy-initial-observation extension, which drives the estimate towards cointegration. The data still follows a random walk, but may have stationary linear combinations.
The two applications are complimentary and typically used together. Sums-of-coefficients typically requires moderate shrinkage (lambda6=0.1), while dummy-initial observation requires more shrinkage (lambda7=0.001).

Block exogeneity

What do we get different results after estimation when applying Block exogeneity?

Choleski structural VARs happens after estimation….means can get different results even though have block exogeneity.

Grid search

Is it possible to also do the grid search over lags?

Not at the moment.

Unconditional forecasts

What value do unconditional forecasts converge to?

Unconditional forecasts will converge to posterior median, i.e. the forecastable part (which is not same as steady-state).

Sign restrictions

How is this implemented?

We follow the approach of Arias, Rubio-Ramirez and Waggoner. The idea of the underlying algorithm is to compute regular IRFs from the Gibbs sampling algorithm, apply a first rotation to them by the way of cholesky factorisation, then a second random rotation. If the restrictions are satisfied after the random rotation the draw is retained, otherwise it is rejected and the procedure is repeated until acceptance is obtained.

Relative entropy methodology - Titling

When does Titling work?

Tilting needs two requirements to be satisfied to work:

-the conditions need to be close to the regular forecast values

- the conditions must be few (less than a dozen)

Please refer to the user guide where this is explained in details!

Long-run priors:

What is the difference with Dummy initial observation

Dummy initial observation implies a general cointegration framework, by forcing the model to be stationary despite the presence of a unit root. Long run priors are much more flexible, they allow the user to specify their own cointegration relations within the variables composing the model.

Normal diffuse prior and overall tightness:

Is the ND prior similar to an OLS estimates, given that it is a non-informative prior? If I increase the parameter on the overall tightness, do I get closer to OLS estimates?

Normal diffuse is not equivalent to OLS estimates. It is diffuse only for the variance-covariance matrix of VAR residuals. The VAR coefficients remain defined by the standard Minnesota scheme. An increase in the overall tightness does however push the prior on the coefficients closer to a diffuse setting. If the coefficient is large enough, then the prior on both the residual variance-covariance matrix and the VAR coefficients gets diffuse, and OLS estimates are approached.

Forecasts

I noticed that I am getting slightly different forecasts every time I run my model, with both the Minnesota and the Normal-Wishart priors. My understanding is that these have closed form

posteriors so there shouldn't be any randomness in the posterior. The discrepancy becomes smaller once I increase the number of iterations and the number of burn-in observations.

Yes, there is actually randomness in the production of forecasts, whatever the prior you are using. True, the Minnesota and Normal-Wishart priors have analytical posteriors for the VAR coefficients, but the forecasts don't. Producing forecasts necessarily involves simulations, even if the distribution for the VAR coefficients is known, hence the small differences you observe from one run to another. Typically, these differences should be small, and become negligible by increasing the number of iterations. To be absolutely clear about 'randomness': forecasts in a Bayesian context are just like any other parameter: random variables, characterised by a posterior distribution. There exists most of the time analytical formulas for the posterior distribution of one-period ahead forecasts in a Bayesian context, though those formulas may be already difficult to recover. For anything beyond one period, it is not possible anymore to recover the posterior distribution analytically so that the distribution must be obtained by simulation methods. If you want a good start on these subject, you can take a look at the BEAR user guide, or at this paper:
http://www.sciencedirect.com/science/article/pii/B9780444627315000154

We have the same variables and the same specifications with Minnesota (full VAR estimates) prior. However, now we change the order of the endogenous variables in the second run. The two consecutive runs give different RMSE's in the unconditional GDP forecasts. It is not only in this example but also in many other runs ordering changes the forecasts, hence the different RMSE. In general is this so because we have a Choleski decomposition?

This is most likely due to the randomness of the forecasts, and not to the ordering. Because forecasts are slightly different at every run, the RMSE is slightly affected. The difference is usually tiny, except if the forecasts are fairly close to the actual. Then a small change in the forecasts can result in a potentially large change in RMSE.

<u>Conditional forecasts</u>

For conditional forecasts if you need to impose more than one condition the toolbox doesn`t do it with sign restrictions?

Conditional forecasts in general may work or not, depending on how you implement your restrictions. This is not an issue specific to sign restrictions, though it may be harder to find a consistent setting with sign restrictions due to the fact that the structural matrices are typically not triangular.

Also, note that conditional forecasts are invariant only to orthogonal transformations of SVAR that have a unit variance-covariance matrix. This implies that they are different if you use a Cholesky or triangular factorisation scheme.

<u>Historical Decompositions</u>

The historical decompositions are largely explained by an 'exogenous shock', despite the SVAR being exactly identified. I have run the SVAR in levels, but first differencing or using log data does not get rid of this exogenous shock either?

The "exogenous" part you can find in the historical decomposition is basically the determinist part of the model and is not really an exogenous shock. In the beginning of the period this is mainly driven by the initial conditions, as the model does not have shocks to start with. If you do not have exogenous variables in your model, you can think of it as an approximation of the posterior median (or in non-Bayesian terms simply a kind of average or trend component) of your time series. You can also think of it in terms of a "steady-state", "long-term" or sometimes in finance "fundamental value" of the respective time series. For inflation, for instance, you would assume, that the exogenous part would give you a value of close to 2, in absence of all shocks. So what people usually do, is to decompose the actual time series into shocks minus the exogenous part (you take deviations from the steady-state). The idea is that without shock you are exactly in steady-state or at a trend value and that the structural shocks drive the time series away from its steady state. Another suggestion: if you are interested in the shock decomposition, manage to run the model in stationary form. With the new version of BEAR you will also be able to better determine trends and deal better with non-stationary time series, but for now please treat the data carefully and make sure it is more or less stationary.

The median shock contributions + the exogenous part do not add up to the actual time series. Is this normal and how do I deal with this?

This is normal and related to the Gibbs sampling algorithm involved in the computation. To explain simply, the posterior distributions are recovered by simulating a large number of models, then putting the results from all the models together to form an empirical distribution. While for each individual model the decomposition adds up to the actual series, the point estimates are obtained from many different models in the simulation, hence the discrepancies. There is no perfect solution to this, but common practice is to take the shock contribution, and subtract it from the actual series to obtain the deterministic component. The leads to leave apart the "exogenous" component.

How can I interpret the initial conditions and the exogenous part on the historical decomposition? Is it a kind of steady-state?

The historical decomposition is made of two components : a determinist component which can be indeed interpreted as a steady-state, and a fluctuation component represented by the shocks. The exogenous part represents the first component, but it also includes the impact of the initial conditions at the beginning of the sample. The impact of these initial conditions normally fades away relatively quickly as the sample gets further away from its starting point.

Stochastic Volality

How do I know that I have to account for stochastic volatility? Is there a test for this?

There is no test for this. Stochastic volatility is an assumption that must be taken by the user.

BEAR prints out numbers for the following things:
AR coefficient on residual variance γ:
IG shape on residual variance $\alpha 0$:
IG scale on residual variance $\delta 0$:
Prior mean on inertia $\gamma 0$:
Prior variance on inertia $\zeta 0$:
I am assuming the first of these is the gamma in equation 5.2.3 of your technical guide and that BEAR does not allow the user to estimate it (it must be fixed at some value specified by the user --- I cannot see any discussion of this in section 5.2 of the guide). But I thought gamma was the "inertia" and hence was puzzled as to what the last two of these are. But I think they are not relevant (using the standard model), but these intended for a different model (the one is section 5.3) so can just be ignored. right?

The terms $\gamma 0$ and $\zeta 0$ can indeed be ignored as they are only relevant for the random inertia model. They are nevertheless reported (even if not used) since the Estimation info reports all the hyperparameters involved in the stochastic volatility interface.

I cannot find any prior for f (see equation 5.2.8), but from the discussion there, BEAR does not give the user the possibility of using a prior for this, but rather just uses some relatively non-informative choice (the pedantic might want to know exactly what this is). right?

For f, BEAR indeed does not let you choose a value for the prior, and implements a diffuse prior (normal with mean 0 and large variance). In theory, this could be let to be specified by the user, but in practice there is no real point in that. Every f represents a vector of the inverse of a covariance matrix. I don't see how any relevant informative prior could be designed on part of an inverse matrix, so a diffuse looks pretty much like the only choice.

Structural VARs

A SVAR always produces N orthogonal shocks – N being the number of (endogenous) variables – regardless of the identification scheme. This is my interpretation of what Fry/Pagan (2011) call the "multiple shocks problem" and where they say that "if only a single shock is to be isolated (more generally any number less than N) some information will need to be provided on what strategy was used to deal with this issue. At the moment little mention is made in many published articles using sign restrictions." (p.952). I want to identify 2 structural shocks (demand + supply) in a model of 3 endogenous variables. Consequently, 1 shock would remain, which is orthogonal to either of them and, hence, a "neither-demand-nor-supply" shock…If this is true so far, I find it very good that BEAR "does not allow" for less than n shocks, since this would only hide the remaining shocks which exist by construction.

An SVAR always produces N orthogonal shocks. If you set restrictions on less than N shocks, there will be at least one that will be undetermined. By undetermined, I mean a shock that will be orthogonal to the others, but for which no meaningful economic interpretation can be obtained (in your case, this is the neither supply nor demand shock). This is actually not necessarily a problem, as long as you keep the following 2 points in mind: 1) not determining one (or more) shocks does have an influence on the determined shocks. Of course, if you had instead set restrictions on all your shocks, you would have obtained a different overall set of restrictions, which would have produced different IRFs , both restricted and unrestricted. So someone here could come and say: your setting is not robust to the addition of supplementary restrictions. This is true, but represents a more general problem with sign restrictions: different restrictions will produce (potentially very) different responses. 2) as long as you really don't care about those undetermined shocks, you are still fine about those which are determined. Naturally, it is not possible to interpret the IRFs for the undetermined shocks in any way, so that any comment on those undetermined shocks would be irrelevant.

Time-varying BVARs

What is the advantage of sparse matrix approach to Kalman filtering approach?

It is faster, more stable, and less sensitive to initial conditions as it uses a diffuse prior for the initial period.

Is sparse matrix sensitive to starting values?

No , again since it uses a diffuse prior.

Will you get different results to Kalman filtering approach?

Chan and Jeliazkov test for similarity and obtain indeed similar results.

Matlab Output

Where can I find the whole (simulated) posterior distribution of IRFs from a BVAR

Draws from the Gibbs sampler are stored in the cell irf_record. Every matrix in the cell contains one response, the number of rows corresponds to the number of simulations, and the number of columns to the number of periods for you IRFs.

What are variables names for posterior distribution of VAR Coefficients and Covariance Matrices?

In general, the variables just take the same name as in the technical guide. For instance:
For VAR coefficients: betabar and Omegabar,
for Sigma: Sbar and alphabar, and so on.
It should therefore be easy to recover them by just referring to the notations in the guide.