CENTRE FOR CENTRAL BANKING STUDIES

ECONOMIC MODELLING AND FORECASTING

Estimating Bayesian VAR models

by

Ole Rummel
Centre for Central Banking Studies
Bank of England

12 February 2015

ole.rummel@bankofengland.co.uk

1

Contents

1   Introduction

To illustrate Bayesian VAR methods in EViews using a selection of the priors and Bayesian methods described in the presentation, we replicate some of the results in Koop and Korobilis' (2009, Section 2.5.5) empirical illustration of Bayesian VAR methods. I have used the Matlab code that the authors make available on their website to calculate results for three prior distributions: the non-informative prior, the Minnesota prior and the natural conjugate (Normal-(inverse) Wishart) prior.[1] These three Bayesian approaches were chosen because they have analytical results for the posterior distribution and because they can be reproduced in EViews. These estimation results are shown in Table 1 below. Note that Table 1 is arranged by lags, which is not the way EViews usually displays estimation output.[2] Using the original data from Koop and Korobilis, we will try to replicate the results using the existing priors and estimation approaches in the BVAR add-in available in EViews.

Table 1: Posterior means of VAR(4) coefficients for three prior distributions

	Non-informative prior			Minnesota prior			Natural conjugate prior		
	$\Delta\pi_t$	u_t	r_t	$\Delta\pi_t$	u_t	r_t	$\Delta\pi_t$	u_t	r_t
Intercept	0.3059	0.4203	-0.0043	0.3100	0.4222	0.0179	0.3041	0.4181	-0.0028
	(0.1105)	(0.1161)	(0.2692)	(0.1086)	(0.1141)	(0.2644)	(0.1090)	(0.1140)	(0.2585)
$\Delta\pi_{t-1}$	1.5035	0.0038	0.5504	1.4620	0.0228	0.4774	1.4923	0.0058	0.5386
	(0.0728)	(0.0765)	(0.1772)	(0.0681)	(0.0714)	(0.1656)	(0.0715)	(0.0748)	(0.1697)
u_{t-1}	-0.2664	1.2911	-0.7198	-0.2505	1.2566	-0.6570	-0.2630	1.2833	-0.7119
	(0.0732)	(0.0769)	(0.1782)	(0.0692)	(0.0726)	(0.1684)	(0.0721)	(0.0755)	(0.1711)
r_{t-1}	-0.0590	-0.0243	0.7763	-0.0543	-0.0277	0.7717	-0.0580	-0.0254	0.7761
	(0.0315)	(0.0331)	(0.0767)	(0.0303)	(0.0318)	(0.0738)	(0.0312)	(0.0327)	(0.0741)
$\Delta\pi_{t-2}$	-0.4653	0.0632	-0.7820	-0.3970	0.0016	-0.5376	-0.4468	0.0597	-0.7517
	(0.1309)	(0.1375)	(0.3187)	(0.1142)	(0.1195)	(0.2777)	(0.1279)	(0.1339)	(0.3036)
u_{t-2}	0.1940	-0.3250	0.7886	0.1537	-0.3024	0.6398	0.1867	-0.3408	0.7692
	(0.1132)	(0.1190)	(0.2757)	(0.1008)	(0.1055)	(0.2450)	(0.1111)	(0.1163)	(0.2637)
r_{t-2}	0.0620	-0.0219	-0.0289	0.0524	-0.0136	-0.0256	0.0605	-0.0208	-0.0298
	(0.0390)	(0.0410)	(0.0951)	(0.0362)	(0.0379)	(0.0881)	(0.0387)	(0.0405)	(0.0918)
$\Delta\pi_{t-3}$	-0.0734	-0.1597	0.8250	-0.0965	-0.0728	0.5043	-0.0805	-0.1556	0.7952
	(0.1316)	(0.1382)	(0.3204)	(0.1061)	(0.1107)	(0.2579)	(0.1286)	(0.1346)	(0.3052)
u_{t-3}	-0.0178	-0.1310	-0.3556	0.0134	-0.1470	-0.1953	-0.0133	-0.1342	-0.3370
	(0.1148)	(0.1206)	(0.2796)	(0.0944)	(0.0986)	(0.2295)	(0.1127)	(0.1179)	(0.2673)
r_{t-3}	-0.0072	0.0963	0.0997	-0.0005	0.0798	0.1243	-0.0060	0.0957	0.1026
	(0.0389)	(0.0409)	(0.0947)	(0.0338)	(0.0353)	(0.0821)	(0.0385)	(0.0403)	(0.0914)
$\Delta\pi_{t-4}$	0.0360	0.1270	-0.4862	0.0316	0.0822	-0.3355	0.0355	0.1245	-0.4751
	(0.0743)	(0.0781)	(0.1809)	(0.0626)	(0.0654)	(0.1522)	(0.0730)	(0.0764)	(0.1733)
u_{t-4}	0.0411	0.0771	0.3109	0.0341	0.0767	0.2337	0.0409	0.0771	0.3038
	(0.0691)	(0.0726)	(0.1682)	(0.0584)	(0.0611)	(0.1420)	(0.0680)	(0.0712)	(0.1614)
r_{t-4}	0.0008	-0.0253	0.0570	-0.0010	-0.0125	0.0317	0.0002	-0.0245	0.0550
	(0.0317)	(0.0333)	(0.0771)	(0.0275)	(0.0287)	(0.0668)	(0.0313)	(0.0328)	(0.0744)

2   The basic VAR model and the curse of dimensionality

Consider the basic VAR(p) model of $t = 1, \dots, T$ observations of a set of n endogenous (macroeconomic) variables $y_t = (y_{1t}, \dots, y_{nt})'$ with a maximum lag order of p:

---

[1] Indeed, the EViews program `kokomodel.prg` that can be found in the BVAR add-in folder is a direct conversion of the Matlab code entitled `BVAR_ANALYT.m` from Gary Koop's website accompanying the Koop and Korobilis (2009) monograph.

[2] Table 1 shows estimated coefficients by variable, holding the lag length constant, while the standard EViews output shows estimated coefficients by lag length, holding the variable constant.

$$y_t = c + B_1 y_{t-1} + \dots + B_p y_{t-p} + \varepsilon_t \qquad (1)$$

where $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})'$ is an unobservable, n-dimensional vector error process with $E(\varepsilon_t) = 0$ and time-invariant (i.e., constant), positive-definite variance-covariance matrix $\text{cov}(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = \Sigma$ (white noise); c is a $(n \times 1)$ vector of intercepts; and B_j is a $(n \times n)$ matrix of regression coefficients for the j-th lag ($j = 1, \dots, p$).

VAR models with many variables and long lags contain many parameters. As such, the number of coefficients in the basic unrestricted (reduced-form) VAR model in equation (1) easily proliferates, meaning that:

- there are lots of coefficients to estimate – the total number of parameters to be estimated in c and the B_j's ($j = 1, \dots, p$) equals $n(np + 1)$;
- unrestricted OLS estimates of the $n(np + 1)$ coefficient values are often not very well determined (in the sense of imprecise) in a finite set of T sample points; and
- if the $B_1, \dots, B_p$ are imprecisely estimated because of limited data, the impulse response functions, forecasts and forecast error variance decompositions that are based on them will also be imprecisely estimated

As a result, unrestricted estimation of these models requires lots of data. In fact, it will be the case that if the estimated parameters are not very precisely estimated, the results are hard to interpret and forecasts may appear more precise than they really are because standard error bands do not account for parameter uncertainty.

Alternative methods for estimating the coefficients have therefore been developed, the most prominent of which is Bayesian estimation.

The BVAR add-in in EViews allows us to work with (useful) Bayesian VAR models that have analytical results for posterior distributions and predictive densities. Other models that do not have such results require intensive computational (simulation) methods such as the Gibbs sampler or the Metropolis-Hastings algorithm. These methods, generally referred to as Markov chain Monte Carlo (MCMC) methods, are not as readily available in EViews, and generally require bespoke EViews programs for the particular estimation problem at hand.

3   The Bayesian approach to VAR estimation

The Bayesian approach suggests a solution to this curse of dimensionality by introducing **prior** distributions. The estimates and forecasts can be improved upon if one has prior information about the structure of the model, possible values of the parameters or even functions of the parameters.

In a classical estimation framework, it is difficult to incorporate non-sample information into the estimation. But using Bayes' law, non-sample information is easily incorporated into a Bayesian framework. The latter also naturally incorporates parameter uncertainty into common measures of precision. Details of the underlying Bayesian methods are given in, *inter alia*, Sims and Zha (1998) and Zha (1998).

In essence, the analysis of VAR models using Bayesian methods requires knowledge of the distributional properties of the estimates of the parameters. In the following, we collect the parameters of interest, which are the vectorised model coefficients, $\beta = \text{vec}(B)$, and the

variance-covariance matrix of the residuals, Σ, as one in $\theta = (\beta, \Sigma)$. We denote the prior distribution of the parameters of interest by $p(\theta)$ and the corresponding posterior distribution by $p(\theta|Y)$.

Using Bayes' theorem, the posterior distribution is summarised as the product of the likelihood $L(Y|\theta)$ and the prior $p(\theta)$:

$$p(\theta|Y) \propto L(Y|\theta)\, p(\theta) \tag{2}$$

On this basis, the main aim of Bayesian estimation is to find the moments of the posterior distribution of the parameters of interest. Just as in classical (frequentist) estimation, we can interpret location and dispersion of $p(\beta|Y)$ and $p(\Sigma|Y)$ as point estimates and precisions of the parameters.

In order to do this, the VAR(p) can be rewritten in matrix form in different ways. One approach expresses the model in terms of the multivariate Normal and the other in terms of the matric-variate Normal distribution.

The former arises if we use an ($nT \times 1$) vector y ($y = \mathrm{vec}(y_t)$), which stacks all T observations on the first dependent variable and then all T observations on the second dependent variable, etc.

The latter arises if we define Y to be a ($T \times n$) matrix which stacks the T observations on each dependent variable in columns next to each other. In particular, by denoting:

$$x_t = \begin{bmatrix} 1 & y'_{t-1} & \cdots & y'_{t-p} \end{bmatrix} \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_T \end{bmatrix} \quad B = \begin{bmatrix} c \\ B_1 \\ \vdots \\ B_p \end{bmatrix} \tag{3}$$

and $\beta = \mathrm{vec}(B)$ and $\varepsilon = \mathrm{vec}(\varepsilon_t)$, one can rewrite equation (1) either as:

$$Y_{(nT \times 1)} = (I_n \otimes X)\beta + \varepsilon \qquad \varepsilon \sim N(0, \Sigma \otimes I_T) \tag{4}$$

or:

$$Y_{(T \times n)} = X_{(T \times (np+1))}B_{((np+1) \times n)} + E_{(T \times n)} \qquad E \sim N(0, \Sigma) \tag{5}$$

If $K = 1 + np$ is the number of coefficients in each equation of the VAR, X in equations (3), (4) and (5) above will be a ($T \times k$) matrix. While β is a ($kn \times 1$) vector of VAR coefficients in the VAR given by equation (4), B in equation (5) is a ($k \times n$) matrix of VAR coefficients, with the relationship between the two determined by the fact that $b = \mathrm{vec}(B)$.

The approach in equation (4) in particular lends itself quite easily to producing a (non-standard) posterior distribution. As shown in Canova (2007) and Koop and Korobilis (2009), equation (4) is useful for decomposing the likelihood function of a VAR(p) into the product of a Normal density for β, conditional on the OLS estimates of the VAR parameters ($\hat{\beta}$) as well as Σ, and a Wishart (inverse Wishart) density for Σ^{-1} (Σ).

In particular:

$$p(\beta|\Sigma, y) \sim \mathrm{N}(\hat{\beta}, \Sigma \otimes (X'X)^{-1}) \tag{6}$$

and:

$$p(\Sigma^{-1}|y) \sim \mathrm{W}(S^{-1}, T - K - n - 1)) \tag{7}$$

where $K = 1 + np$, $\hat{B} = (X'X)^{-1}X'Y$ is the OLS estimate of B, $\hat{\beta} = vec(\hat{B})$ and:

$$S = (Y - X\hat{B})'(Y - X\hat{B}) \tag{8}$$

4   The data

The authors use a quarterly US data set which includes the inflation rate, $\Delta\pi_t$, defined as the annual percentage change in a chain-weighted GDP price index; the unemployment rate, u_t, defined as the seasonally-adjusted civilian unemployment rate for all workers over the age of 16; and the interest rate, r_t, defined as the yield on the three-month Treasury bill rate. While all the data can be obtained from the Federal Reserve Bank of St. Louis's FRED website (http://research.stlouisfed.org/fred2/), I found some differences between the website data and the underlying dataset that can be downloaded from Gary Koop's website at http://personal.strath.ac.uk/gary.koop/bayes_matlab_code_by_koop_and_korobilis.html.

Data downloaded by myself from the FRED website either have the suffix _ave or _end, depending on whether the underlying monthly data has been averaged to yield the quarterly number or the final month of each quarter has been chosen to yield the corresponding quarterly figure. The corresponding data from Gary Koop's website has the suffix _koko to differentiate it from the other data.

The underlying data set for the Bayesian VAR therefore is $y_t = (\Delta\pi_t, u_t, r_t)'$ and can be found in the EViews workfile called **koko_bvars.wf1**. To open the EViews workfile from within EViews, choose **File**, **Open, EViews Workfile…**, select **koko_bvars.wf1** from the appropriate folder and click on **Open**. Alternatively, you can double-click on the workfile icon outside of EViews, which will open EViews automatically.

The sample runs from 1953 Q1 to 2006 Q3. The dataset includes the inflation rate (`d12pgdp_koko`), the unemployment rate (`unrate_koko`), and the three-month Treasury bill rate (`r3_koko`). The dataset is interesting as these three variables can frequently be found in new Keynesian VARs. Examples of papers that use these, or similar variables, include Cogley and Sargent (2005), Primiceri (2005) and Koop *et al.* (2009).

The underlying data are plotted in Figure 1.

Figure 1: Data used in the empirical exercise

UNRATE_KOKO

D12PGDP_KOKO

R3_KOKO
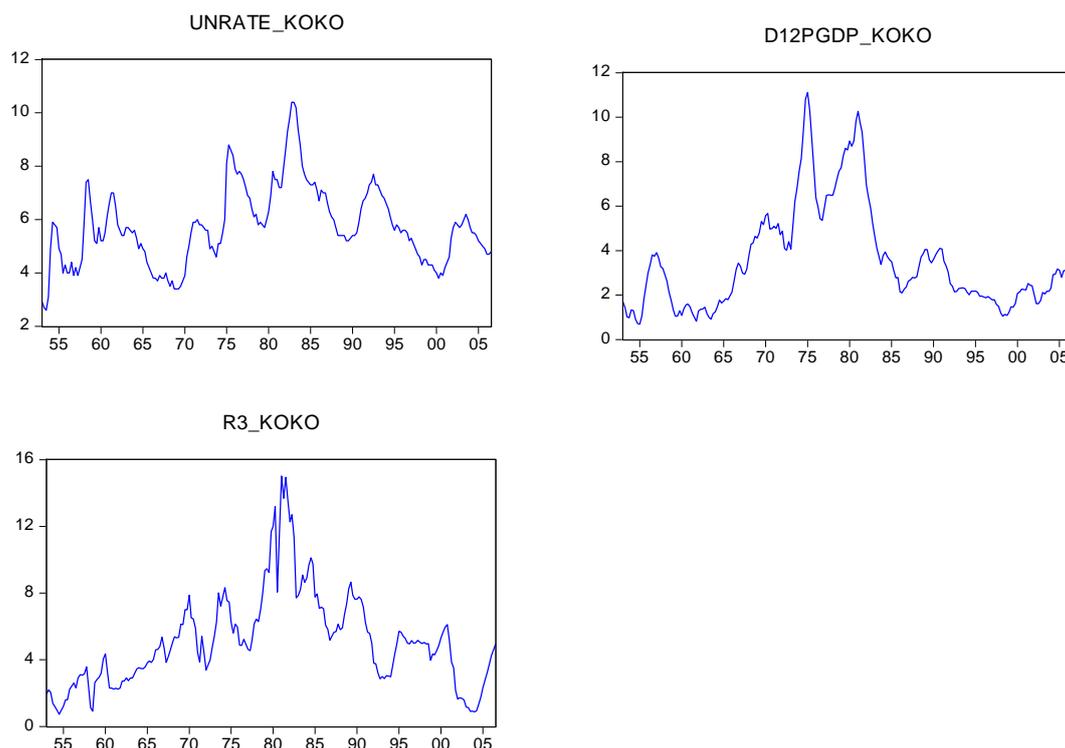
5 The BVAR add-in and Bayesian estimation in EViews

To estimate a Bayesian vector autoregression (BVAR) model in EViews, we use the **Bayesian VAR** add-in for EViews 7.2, which can be downloaded for free from:[3]

http://www.eviews.com/cgi/ai_download.cgi?ID=BVAR.aipz

To illustrate Bayesian VAR analysis in EViews using this data, we follow Koop and Korobilis and work with an unrestricted VAR including an intercept and four lags of all variables included in every equation and consider the following five priors available in EViews:

- Sims-Zha Normal-Wishart prior (which is the default option, as the natural joint prior for normally distributed data is the Normal-Wishart distribution if the assumption of a fixed and diagonal variance-covariance matrix is relaxed) (`prior = 0`);
- Sims-Zha Normal-flat prior (`prior = 1`);
- Koop and Korobilis (Ko-Ko) Minnesota prior (`prior = 2`);
- Ko-Ko Normal-Wishart prior (`prior = 3`); and a
- diffuse (flat-flat) prior (`prior = 4`)

---

[3] Starting with EViews 8, Bayesian estimation of VAR models is now part of the specification menu for VAR objects. Generally speaking, EViews provides commands for estimating BVARs and forecasting without simulation; if simulation is required, the researcher has to do a significant amount of programming.

5.1 The unrestricted VAR(4) benchmark model

Before analysing the different Bayesian options, let us start with a plain VAR(4) in the three variables, estimated using classical techniques, as a benchmark for comparison.

To estimate an unrestricted VAR(4) using all three variables over the sample period from 1953 Q1 to 2006 Q3, we select **Quick**, **Estimate VAR…**, enter `d12pgdp_koko`, `unrate_koko` and `r3_koko` in that order as endogenous variables and set `1 4` for the lag interval. An equivalent way of getting EViews to do the estimation is to type the following command in the command window:

```
var var4.ls 1 4 d12pgdp_koko unrate_koko r3_koko @ c
```

This command creates and estimates a VAR with the name `var4` and a lag length of four. If you do not fancy typing the three variable names, you can copy and paste them across from the `vartext` EViews text file. Note that you have to copy them before opening the VAR object – and then paste them into the **Endogenous Variables** box. After estimation using the **Quick**, **Estimate VAR…** route, we should give the VAR object a name by clicking on **Name** and calling it `var4`.

Table 2: Unrestricted VAR(4) estimation results
(1954Q1 - 2006Q3)

Vector Autoregression Estimates
Sample (adjusted): 1954Q1 2006Q3
Included observations: 211 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	D12PGDP_KO KO	UNRATE_KOK O	R3_KOKO
D12PGDP_KOKO(-1)	1.503497	0.003804	0.550374
	(0.07203)	(0.07569)	(0.17541)
	[20.8726]	[0.05026]	[3.13758]
D12PGDP_KOKO(-2)	-0.465324	0.063162	-0.781993
	(0.12952)	(0.13610)	(0.31542)
	[-3.59255]	[0.46409]	[-2.47921]
D12PGDP_KOKO(-3)	-0.073420	-0.159654	0.824951
	(0.13023)	(0.13684)	(0.31715)
	[-0.56376]	[-1.16668]	[2.60116]
D12PGDP_KOKO(-4)	0.035952	0.126990	-0.486216
	(0.07353)	(0.07727)	(0.17907)
	[0.48892]	[1.64355]	[-2.71525]
UNRATE_KOKO(-1)	-0.266419	1.291132	-0.719760
	(0.07245)	(0.07613)	(0.17644)
	[-3.67719]	[16.9597]	[-4.07944]
UNRATE_KOKO(-2)	0.193963	-0.352004	0.788593
	(0.11207)	(0.11776)	(0.27292)
	[1.73066]	[-2.98909]	[2.88941]

UNRATE_KOKO(-3)	-0.017793	-0.130997	-0.355610
	(0.11364)	(0.11941)	(0.27674)
	[-0.15658]	[-1.09705]	[-1.28500]
UNRATE_KOKO(-4)	0.041103	0.077067	0.310941
	(0.06835)	(0.07182)	(0.16645)
	[0.60136]	[1.07305]	[1.86808]
R3_KOKO(-1)	-0.058956	-0.024297	0.776289
	(0.03120)	(0.03278)	(0.07597)
	[-1.88989]	[-0.74124]	[10.2187]
R3_KOKO(-2)	0.062048	-0.021909	-0.028936
	(0.03865)	(0.04061)	(0.09411)
	[1.60555]	[-0.53953]	[-0.30747]
R3_KOKO(-3)	-0.007206	0.096263	0.099745
	(0.03850)	(0.04045)	(0.09375)
	[-0.18719]	[2.37976]	[1.06398]
R3_KOKO(-4)	0.000794	-0.025306	0.056982
	(0.03133)	(0.03292)	(0.07629)
	[0.02533]	[-0.76873]	[0.74687]
C	0.305916	0.420308	-0.004312
	(0.10942)	(0.11497)	(0.26645)
	[2.79590]	[3.65582]	[-0.01618]
R-squared	0.983403	0.950288	0.931588
Adj. R-squared	0.982397	0.947275	0.927442
Sum sq. resids	19.12495	21.11564	113.4162
S.E. equation	0.310790	0.326565	0.756841
F-statistic	977.6458	315.4093	224.6857
Log likelihood	-46.10485	-56.55149	-233.9022
Akaike AIC	0.560236	0.659256	2.340306
Schwarz SC	0.766748	0.865768	2.546818
Mean dependent	3.583876	5.791943	5.160995
S.D. dependent	2.342468	1.422199	2.809711
Determinant resid covariance (dof adj.)		0.004909	
Determinant resid covariance		0.004057	
Log likelihood		-317.1567	
Akaike information criterion		3.375893	
Schwarz criterion		3.995430	

As you can see from Table 2, even estimating a small three-variable VAR(4) generates a lot of output. Each column in the table corresponds to the equation for one endogenous variable in the VAR. For each right-hand side variable, EViews reports the coefficient point estimate, the estimated coefficient standard error (in round brackets) and the *t*-statistic (in square brackets). At first glance, the point estimates in Table 2 appear to be equivalent to the point estimates in Table 1 above, and we will return to this point in the next Section.

5.2 The non-informative, diffuse or flat-flat prior

We can compare the above model with the EViews option of imposing a diffuse (or flat-flat) prior. EViews automates the most popular diffuse prior, which is also referred to as the non-informative, Jeffrey's or flat-flat prior. This prior takes the form:

$$p(\beta, \Sigma) \propto |\Sigma|^{-(n+1)/2} \tag{9}$$

where n is the number of variables in the VAR. Using Bayes' Theorem, we can derive the conditional posteriors as:

$$p(\beta \mid Y) \propto \exp[-0.5(\beta - \bar{\beta})' \bar{V}^{-1}(\beta - \bar{\beta})] \tag{10}$$

where β = vec(B), $\bar{\beta} = vec(\bar{B})$ and the posterior mean and covariance are given by $\bar{B} = (X'X)^{-1}X'Y$ and $\bar{V} = (X'X)^{-1}$. In this case, the posterior results for Σ are $\bar{S} = (Y - X\hat{B})'(Y - X\hat{B})$, where $\hat{B} = (X'X)^{-1}X'Y$. We can see that the posterior results are based on the well-known OLS quantities.

To estimate a Bayesian vector autoregression (BVAR) model in EViews, we use the **Bayesian VAR** add-in for EViews 7.2. After installation, go to **Add-ins** and choose **Bayesian VAR**. Upon opening the EViews Bayesian VAR add-in from the **Add-ins** menu, a dialog box – which has been pre-populated with some settings – will appear:

This first dialog box lets you specify the *n* endogenous variables in the VAR model, while the second box lets you specify any exogenous variables that enter into the VAR.[4] The next couple of dialog and/or check boxes, entitled 'Enter an estimation sample', 'Enter a number of lags', 'Calculate impulse responses' and 'Enter a number of periods for IRFs' are pretty self-explanatory. The first sets the estimation sample (which, in our case, is 1953q1 2006q3, which has been chosen automatically); the second lets you enter the number of lags, *p*, in the VAR; and the fourth – in conjunction with the box to 'Calculate impulse responses' – allows you to choose the calculation of impulse responses and the periods for which impulse responses will be calculated.

Note how the Bayesian VAR estimation window includes an option to store the VAR to be estimated as an EViews **Model** object, which will then be available for post-estimation purposes. The **Model Object Name** dialog box lets you specify the name for the EViews model object, which can be used for later use, such as forecasting with the estimated BVAR model. To keep the model permanently in the workfile, we give the model a name by entering a name such as `bvar_noninformative` into the **Model Object Name** textbox. After we finish setting all the estimation options on the **Bayesian VAR estimation** screen, we click **OK**.[5] These options are as follows:

---

[4] Other exogenous variables can be added to the estimation formula, just as for OLS and the conventional VAR model. The BVAR estimation will automatically take this into consideration and return the coefficient estimates for those variables. Note that this option is only available for models estimated using the Sims and Zha (1998) prior.

[5] The model object `bvar_noninformative` can then be used to produce a forecast for the endogenous variables.

After selecting all the different options as required, click **OK** and a display of the estimation results will appear. These are in the form of a **Spool** object, showing the estimation results for the BVAR. If you selected to perform impulse responses, a table of responses, along with corresponding graphs, will also be shown in the spool.

We should find that the posterior results shown in Table 3 below for point estimates are identical to the classical VAR estimation results in Table 2 above.

Table 3: Bayesian estimation results using a non-informative prior

Bayesian Vector Autoregression
Sample: 1953Q1 2006Q3
Included observations: 211 (after adjustment)
Standard errors in ()
Priors: Diffuse (Flat-flat) prior

	D12PGDP_KOKO	UNRATE_KOKO	R3_KOKO
D12PGDP_KOKO(-1)	1.503497	0.003804	0.550374
	(0.07277)	(0.07646)	(0.17721)
D12PGDP_KOKO(-2)	-0.465324	0.063162	-0.781993
	(0.13085)	(0.13749)	(0.31866)
D12PGDP_KOKO(-3)	-0.073420	-0.159654	0.824951
	(0.13157)	(0.13825)	(0.32040)
D12PGDP_KOKO(-4)	0.035952	0.126990	-0.486216
	(0.07429)	(0.07806)	(0.18091)
UNRATE_KOKO(-1)	-0.266419	1.291132	-0.719760
	(0.07320)	(0.07691)	(0.17825)
UNRATE_KOKO(-2)	0.193963	-0.352004	0.788593
	(0.11322)	(0.11897)	(0.27572)
UNRATE_KOKO(-3)	-0.017793	-0.130997	-0.355610
	(0.11481)	(0.12063)	(0.27958)
UNRATE_KOKO(-4)	0.041103	0.077067	0.310941
	(0.06905)	(0.07256)	(0.16816)
R3_KOKO(-1)	-0.058956	-0.024297	0.776289
	(0.03152)	(0.03311)	(0.07675)
R3_KOKO(-2)	0.062048	-0.021909	-0.028936
	(0.03904)	(0.04102)	(0.09508)
R3_KOKO(-3)	-0.007206	0.096263	0.099745
	(0.03889)	(0.04087)	(0.09471)
R3_KOKO(-4)	0.000794	-0.025306	0.056982
	(0.03165)	(0.03326)	(0.07708)
C	0.305916	0.420308	-0.004312
	(0.11054)	(0.11615)	(0.26918)

While point estimates are unchanged across the unrestricted VAR estimation results and the non-informative prior, we note that the diffuse prior has larger standard errors. Why should that be? In essence, a diffuse prior means that we know 'nothing', i.e., we are incorporating an enormous amount of (potential) uncertainty. In fact, there is more uncertainty than in the OLS case, where we are, roughly speaking, averaging over the prediction errors only. In the diffuse prior Bayesian case, we are averaging over the prediction errors **plus** the extra uncertainty that comes from admitting that we know nothing about the parameter estimates.

We should note that we can also estimate the BVAR without having the additional dialog boxes. For example, the above BVAR(4) with a diffuse or flat-flat prior could equally well have been estimated by typing:

```
bvar(prior=4) 4 d12pgdp_koko unrate_koko r3_koko
```

in the command window. Without a specification of the respective hyperparameters of the specific priors, EViews will assume the default settings. In case you wanted to change hyperparameter values, you can do so as part of the options associated with the `bvar(·)` command, such as:

```
bvar(prior=2,a1=0.5,a2=0.5,a3=0.1) 4 lm2 rl rs ly pi
```

which would set the three hyperparameters a_1, a_2 and a_3 in the Koop and Korobilis version of the Minnesota prior in a BVAR(4) involving the five variables `lm2`, `rl`, `rs`, `ly` and `pi`.

5.3   The Koop and Korobilis (Ko-Ko) Minnesota prior

The next BVAR that we consider uses the Minnesota prior. The Minnesota prior is based on an approximation that leads to great simplification in prior elicitation and computation. The simplicity of the Minnesota prior comes from the fact that Σ is assumed to be known (and replaced by its estimate, $\hat{\Sigma}$), meaning that no prior distribution for Σ is required. The original Minnesota prior simplifies even further by assuming Σ to be a diagonal matrix. When Σ is replaced by an estimate, we only have to worry about a prior for β and the prior distribution for β under the Minnesota prior is given by:

$$p(\beta|\Sigma) \sim \mathrm{N}(\beta_0, V_0) \tag{11}$$

The default setting is $\beta_0 = 0_{(mp \times 1)}$ and $V_0 \neq 0$. In other words, the Minnesota prior involves setting most or all of its elements to zero, thus ensuring shrinkage of the VAR coefficients towards zero and lessening the risk of over-fitting.

In EViews, the Minnesota prior is imposed using the options for the Ko-Ko Minnesota prior. In Koop and Korobilis (2009), the authors set $\beta_{Mn} = 0$, except for the first own lag of each variable, for which it is set to 0.9. Σ is diagonal with elements s_i^2 obtained from univariate regressions of each dependent variable on an intercept and the chosen number of lags of all variables. Koop and Korobilis (2009) specify their prior covariance matrix V_0 as a diagonal matrix with elements $v_{ij,l}$ ($l = 1, \ldots, p$) as:

$$
v_{ij,l} = \begin{cases} a_1 / p^2 & \text{for coefficients on own lags} \\ (a_2\sigma_i)/(p^2\sigma_j) & \text{for coefficients on lags of variable } i \neq j \\ a_3\sigma_i & \text{for coefficients on exogenous variables} \end{cases}
$$

where σ_j^2 is the i-th diagonal element of Σ. The prior simplifies the complicated choice of fully specifying all the elements of V_0 to choosing three scalars: a_1, a_2 and a_3. We can see that Koop and Korobilis (2009) introduce their own hyperparameters. But it should be noted that these hyperparameters are related to the ones given for the Sims-Zha prior in Table 5 below. In particular:

- $a_1 = \lambda_1$;
- $a_2 = \lambda_1\lambda_2$;
- $a_3 = (\lambda_0\lambda_4)^2$;
- $0 = \mu_5$; and
- $0 = \mu_6$

This form of the prior captures the sensible property that, as the lag length p increases, coefficients are increasingly shrunk toward zero. Moreover, by setting $a_1 > a_2$, own lags are more likely to be important predictors than lags of other variables. A big advantage of the Minnesota prior is that it leads to simple posterior inference involving only the Normal distribution. The resulting posterior distribution for β has the form:

$$
p(\beta \mid \hat{\Sigma}, Y) \sim N(\overline{\beta}, \overline{V}) \tag{12}
$$

where:

$$
\overline{V} = (V_0^{-1} + (\hat{\Sigma}^{-1} \otimes X'X))^{-1} \tag{13}
$$

$$
\overline{\beta} = \overline{V}(V_0^{-1}B_0 + (\hat{\Sigma}^{-1} \otimes X)'Y) \tag{14}
$$

where $\hat{B}$ is the OLS estimate of B. If it were not for the $\overline{V}^{-1}$ term in equation (14), the estimator for β looks like the multivariate least-squares estimator. We also note that a disadvantage of the Minnesota prior is that it does not provide a full Bayesian treatment of Σ as an unknown parameter. Instead, it simply replaces Σ by $\hat{\Sigma}$, ignoring any uncertainty in this parameter. Other Bayesian methods discussed in this exercise treat Σ as an unknown parameter.

The options for the Minnesota prior are as follows:

Pressing **OK** at the end of the first dialog box then opens up another, which allows us to choose the specific setting for the Minnesota prior as formulated by Koop and Korobilis (2009).

We keep EViews' default options as they are ($a_1 = 0.5$, $a_2 = 0.5$, $a_3 = 100$ and $\beta_0 = 0$), and press **OK** to estimate the BVAR. The results are given in Table 4.

Table 4: Bayesian estimation results using Koop and Korobilis' Minnesota prior

Bayesian Vector Autoregression
Sample: 1953Q1 2006Q3
Included observations: 211 (after adjustment)
Standard errors in ()
Priors: KoKo Minnesota / Litterman prior
(beta0, a1, a2, a3) = (0, 0.5, 0.5, 100)

	D12PGDP_KOKO	UNRATE_KOKO	R3_KOKO
D12PGDP_KOKO(-1)	1.250893	0.019548	0.391605
	(0.06058)	(0.06535)	(0.15498)
D12PGDP_KOKO(-2)	-0.117837	0.022975	-0.300950
	(0.09644)	(0.10404)	(0.24672)
D12PGDP_KOKO(-3)	-0.137222	-0.088908	0.335469
	(0.09667)	(0.10428)	(0.24731)
D12PGDP_KOKO(-4)	-0.000231	0.088441	-0.303809
	(0.06182)	(0.06669)	(0.15815)
UNRATE_KOKO(-1)	-0.200559	1.062069	-0.613166
	(0.06178)	(0.06664)	(0.15805)
UNRATE_KOKO(-2)	0.055816	-0.094126	0.437514
	(0.08630)	(0.09310)	(0.22078)
UNRATE_KOKO(-3)	0.035601	-0.158551	-0.017917
	(0.08699)	(0.09384)	(0.22255)
UNRATE_KOKO(-4)	0.055407	0.050585	0.211563
	(0.05798)	(0.06254)	(0.14832)
R3_KOKO(-1)	-0.027766	-0.052554	0.665425
	(0.02775)	(0.02993)	(0.07099)
R3_KOKO(-2)	0.020522	0.004148	0.050480
	(0.03266)	(0.03523)	(0.08355)
R3_KOKO(-3)	0.008676	0.071238	0.120685
	(0.03235)	(0.03490)	(0.08277)
R3_KOKO(-4)	-0.004721	0.006967	0.053655
	(0.02777)	(0.02996)	(0.07105)
C	0.351917	0.514383	0.050479
	(0.11232)	(0.12116)	(0.28735)

Interestingly, we find some major differences with the results given in Table 1 above, even though we are using the **same** setting of the hyperparameters as in Koop and Korobilis' Matlab code.

4.4 The natural conjugate (Normal-Wishart) prior due to Sims and Zha (1998)

The Normal-Wishart prior is an example of **natural conjugate priors**. These are priors where the prior, likelihood and posterior distribution come from the same family of distributions. Kadiyala and Karlsson (1997) were able to show that the likelihood $L(Y|\beta, \Sigma)$ from the VAR model can be decomposed into two parts. One is a distribution for β given Σ and another is a distribution for Σ^{-1} (Σ). The former is assumed to be Normal, while the latter is a Wishart (inverse Wishart) distribution.

If the assumption of a fixed and diagonal variance-covariance matrix is relaxed, the natural joint prior for normal data is the Normal-Wishart distribution:

$$p(B|\Sigma) \sim N(B_0, V_0) \qquad p(\Sigma^{-1}) \sim IW(v_0, S_0^{-1}) \qquad (15)$$

where B_0 and V_0 in the prior distribution for B given Σ are specified as:

$$B_0 = \begin{pmatrix} I_m \\ 0_{(mp \times m)} \end{pmatrix}$$

and each (diagonal) element of V_0 is calculated as:

$$V_{0l,ij} = \left(\frac{\lambda_o \lambda_1}{\sigma_j l^{\lambda_3}} \right)^2 \text{ for } i, j = 1, \dots, m$$

where σ_j^2 is the j-th diagonal element of Σ for the l-th lag of variable i in equation j and λ_1, λ_2 and λ_3 are three hyperparameters set by the researcher.

The shrinkage parameter λ_0 measures the overall tightness of the prior: when $\lambda_0 \rightarrow 0$, the prior is imposed exactly and the data do not influence the estimates, while as $\lambda_0 \rightarrow \infty$, the prior becomes loose and prior information does not influence the estimates, which will therefore approach the standard OLS estimates. The parameter λ_2 implements additional shrinkage on lags of other variables than on lags of the dependent variable. We refer to this as the cross-shrinkage parameter. In Sims and Zha's (1998) benchmark specification, this parameter is set to $\lambda_2 = 1$, which implies that no cross-variable shrinkage takes place, as required for the Normal-(inverse) Wishart case.

The prior parameters for the variance-covariance matrix Σ are set to $v_0 = T + 1$ and $S_0 = \lambda_0^{-2} SSE$, where $SSE = (Y - X\hat{B})'(Y - X\hat{B})$ and $\hat{B} = (X'X)^{-1} X'Y$, i.e., the OLS estimate of β. The resulting posterior distributions of B and Σ are then:

$$p(B|Y) \sim N(\overline{B}, \overline{V}) \qquad p(\Sigma|Y) \sim IW(v_0 + T - k - n - 1, \overline{S})$$

where:

$$\overline{V} = (V_0^{-1} + X'X)^{-1}$$

$$\overline{B} = \overline{V}(X'Y + V_0^{-1}B_0)$$

$$\overline{S} = Y'Y + S_0 - \overline{B}'V_0^{-1}\overline{B} + B_0'V_0^{-1}B_0 \qquad (16)$$

The set of hyperparameters in the case of the Sims and Zha (1998) prior are given in Table 5.

Table 5: Hyperparameters of the Sims and Zha (1998) prior

Hyperparameter	Purpose
λ_0	Overall tightness of the prior on the covariance matrix. Smaller values mean a tighter prior.
λ_1	Standard deviation or tightness of the prior on the AR(1) coefficient. Smaller values mean a tighter prior.
λ_2	=1 by assumption under Sims and Zha.
λ_3	Controls how fast higher lags go to zero. Larger values mean lags decay faster.
λ_4	Standard deviation or tightness of the prior on constant. Smaller values mean a tighter prior.
λ_5	Tightness of prior on exogenous variables. Smaller values mean a tighter prior.
μ_5	Higher values imply that coefficients on own lags add up to one.
μ_6	Higher values imply that there are fewer than N stochastic trends in the VAR.

The Sims and Zha (1998) prior is also readily available in the BVAR add-in. We note in passing the additional two hyperparameters μ_5 and μ_6 that deal with any nonstationarity and cointegration of the variables appearing in the VAR. This issue is discussed in more detail in the Appendix. The upshot is that a VAR model in first differences (growth rates) does not need the cointegration/unit root prior by construction (because it is stationary), while these priors do help and should be included in models estimated on the levels of potentially nonstationary variables, consistent with evidence in Banbura *et al.* (2010).

Pressing **OK** at the end of the first dialog box opens up another, which allows us to choose the specific settings for the Sims and Zha (natural-conjugate) prior. To begin with, we use EViews' default settings, which are shown in the following screenshot.

We keep EViews' default options as they are and press **OK** to estimate the BVAR. The results are given in Table 6.

Table 6: Bayesian estimation results using Sims and Zha's (1998) Normal-Wishart natural conjugate prior ($\lambda_1 = 1$)

Bayesian Vector Autoregression
Sample: 1953Q1 2006Q3
Included observations: 211 (after adjustment)
Standard errors in ()
Priors: Sims-Zha Normal-Wishart prior
(L0, L1, L3, L4, mu5, mu6, qm) = (1, 1, 1, 1, 0.1, 0.1, 4)

	D12PGDP_KOKO	UNRATE_KOKO	R3_KOKO
D12PGDP_KOKO(-1)	1.490779	0.016860	0.510245
	(0.06808)	(0.07168)	(0.16603)
D12PGDP_KOKO(-2)	-0.440252	0.035089	-0.637859
	(0.11772)	(0.12396)	(0.28711)
D12PGDP_KOKO(-3)	-0.078573	-0.121168	0.627529
	(0.11308)	(0.11907)	(0.27580)
D12PGDP_KOKO(-4)	0.027563	0.101950	-0.392796
	(0.06536)	(0.06882)	(0.15940)
UNRATE_KOKO(-1)	-0.252402	1.289261	-0.691783
	(0.06875)	(0.07239)	(0.16767)

UNRATE_KOKO(-2)	0.173261	-0.340515	0.710510
	(0.10326)	(0.10873)	(0.25184)
UNRATE_KOKO(-3)	-0.002331	-0.129665	-0.246367
	(0.10008)	(0.10538)	(0.24410)
UNRATE_KOKO(-4)	0.038521	0.074012	0.249480
	(0.06111)	(0.06435)	(0.14905)
R3_KOKO(-1)	-0.055490	-0.024417	0.776494
	(0.03001)	(0.03160)	(0.07318)
R3_KOKO(-2)	0.056928	-0.016076	-0.025326
	(0.03647)	(0.03840)	(0.08895)
R3_KOKO(-3)	-0.003247	0.086542	0.110225
	(0.03493)	(0.03678)	(0.08520)
R3_KOKO(-4)	-0.001435	-0.020522	0.042489
	(0.02839)	(0.02989)	(0.06924)
C	0.272587	0.374309	0.009755
	(0.10049)	(0.10581)	(0.24508)

We then repeat the exercise, making sure to change the tightness of the prior on the AR(1) coefficient, controlled by λ_1, to $\lambda_1 = 0.1$ (rather than its default value of 1). Leave the other hyperparameters at their default settings. The hyperparameter λ_1 controls the standard deviation or tightness of the prior on the AR(1) coefficients in the VAR, and thus how closely the random walk approximation is to be imposed. The default setting for λ_1 was 1, so our new value of $\lambda_1 = 0.1$ assigns a much tighter prior on the coefficients, i.e., we shrink the diagonal elements of the B_1 coefficient matrix in the VAR (1) toward one (and all other coefficients toward zero). We can then compare the results to the Bayesian estimation with the original value of $\lambda_1 = 1$ to investigate the differences between the two settings of the hyperparameter. The results of this experiment are shown in Table 7.

Table 7: Bayesian estimation results using Sims and Zha's (1998)
Normal-Wishart natural conjugate prior ($\lambda_1 = 0.1$)

Bayesian Vector Autoregression
Sample: 1953Q1 2006Q3
Included observations: 211 (after adjustment)
Standard errors in ()
Priors: Sims-Zha Normal-Wishart prior
(L0, L1, L3, L4, mu5, mu6, qm) = (1, 0.1, 1, 1, 0.1, 0.1, 4)

	D12PGDP_KOKO	UNRATE_KOKO	R3_KOKO
D12PGDP_KOKO(-1)	1.203390	0.020236	0.221814
	(0.03956)	(0.04076)	(0.09316)
D12PGDP_KOKO(-2)	-0.112073	0.001784	-0.078974
	(0.04305)	(0.04435)	(0.10138)
D12PGDP_KOKO(-3)	-0.063939	-9.01E-06	-0.010224

	(0.02959)	(0.03049)	(0.06969)
D12PGDP_KOKO(-4)	-0.031541	0.011627	-0.044707
	(0.02188)	(0.02255)	(0.05154)
UNRATE_KOKO(-1)	-0.146192	1.086126	-0.238095
	(0.04102)	(0.04226)	(0.09660)
UNRATE_KOKO(-2)	0.027268	-0.110282	0.137648
	(0.03996)	(0.04117)	(0.09412)
UNRATE_KOKO(-3)	0.032174	-0.061660	0.059340
	(0.02770)	(0.02854)	(0.06525)
UNRATE_KOKO(-4)	0.030342	-0.020322	0.059505
	(0.02060)	(0.02123)	(0.04852)
R3_KOKO(-1)	-0.004078	-0.017626	0.897687
	(0.01932)	(0.01991)	(0.04551)
R3_KOKO(-2)	0.003941	0.017071	-0.009956
	(0.01658)	(0.01708)	(0.03905)
R3_KOKO(-3)	-0.000874	0.019822	0.016580
	(0.01200)	(0.01236)	(0.02825)
R3_KOKO(-4)	-0.002849	0.006849	0.000298
	(0.00936)	(0.00964)	(0.02204)
C	0.366943	0.363909	0.088870
	(0.10224)	(0.10533)	(0.24078)

Log marginal density	502.1759
Predictive marginal density	538.8118
Coef log-posterior	3308.476

As expected, we find that the coefficients on the first lag of each variable in the BVAR(4) in Table 7 (d12pgdp_koko(-1), unrate_koko(-1) and r3_koko(-1)) are closer to unity than the corresponding entries in Table 6, which is a result of us imposing the random walk assumption more tightly. That being said, many of the other coefficients have now also changed as a result of us imposing a different prior from the EViews default case. Additional output at the bottom of the results table includes the log marginal density. The specification of the BVAR can be optimised by maximising the log marginal density. This could be useful, say, when selecting the optimal lag length, as is done in Carriero *et al.* (2015), or the settings of some of the hyperparameters.

We should note that we can also estimate this BVAR without having the additional dialog boxes. For example, the above BVAR(4) with the Sims and Zha (1998) prior could equally well have been estimated by using:

```
szbvar(options) lags endogenous variables @ exogenous
variables
```

in the command window. Without a specification of the respective hyperparameters of the specific priors under `options`, EViews will assume the default settings.

5.5 The Sims and Zha (1998) Normal-flat prior

The Normal-flat prior due to Sims and Zha (1998) assumes independence between B and Σ, where:

$$p(B) \sim \mathrm{N}(B_0, V_0) \qquad p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$$

Note that, because B and Σ are now assumed to be independent, Σ no longer enters into the prior distribution for B, which is written simply as $p(B)$ rather than $p(B|\Sigma)$. The resulting posterior distributions of B and Σ are then:

$$p(B\,|\,Y) \sim N(\overline{B}, \overline{V}) \qquad p(\Sigma^{-1}\,|\,Y) \sim IW(T, \overline{S}^{-1})$$

where:

$$\overline{V} = (V_0^{-1} + X'X)^{-1}$$

$$\overline{B} = \overline{V}(X'Y + V_0^{-1}B_0)$$

$$\overline{S} = Y'Y + \overline{B}'V_0^{-1}\overline{B} + B_0'V_0^{-1}B_0$$

These equations for the posterior distribution are just as in the case of the Sims-Zha Normal-Wishart prior. The last equation is just as equation (16), except that S_0 is missing. This is because our prior assumption about Σ does not depend on S_0.

The Sims and Zha (1998) Normal-flat prior is also readily available in the BVAR add-in. Pressing **OK** at the end of the first dialog box opens up another, which allows us to choose the specific settings for the Sims and Zha (natural-conjugate) prior. At this point, we use EViews' default settings.

We note that this second screen is just as in the case of the Sims and Zha Normal-inverse Wishart prior. The estimation results are given in Table 8.

Table 8: Bayesian estimation results using Sims and Zha's (1998) Normal-flat prior

Bayesian Vector Autoregression
Sample: 1953Q1 2006Q3
Included observations: 211 (after adjustment)
Standard errors in ()
Priors: Sims-Zha Normal-flat prior
(L0, L1, L3, L4, mu5, mu6, qm) = (1, 1, 1, 1, 0.1, 0.1, 4)

	D12PGDP_KOKO	UNRATE_KOKO	R3_KOKO
D12PGDP_KOKO(-1)	1.490779	0.016860	0.510245
	(0.06790)	(0.07149)	(0.16555)
D12PGDP_KOKO(-2)	-0.440252	0.035089	-0.637859
	(0.11742)	(0.12362)	(0.28629)
D12PGDP_KOKO(-3)	-0.078573	-0.121168	0.627529
	(0.11279)	(0.11875)	(0.27501)
D12PGDP_KOKO(-4)	0.027563	0.101950	-0.392796
	(0.06519)	(0.06863)	(0.15894)
UNRATE_KOKO(-1)	-0.252402	1.289261	-0.691783
	(0.06857)	(0.07220)	(0.16719)

UNRATE_KOKO(-2)	0.173261	-0.340515	0.710510
	(0.10299)	(0.10844)	(0.25112)
UNRATE_KOKO(-3)	-0.002331	-0.129665	-0.246367
	(0.09982)	(0.10510)	(0.24340)
UNRATE_KOKO(-4)	0.038521	0.074012	0.249480
	(0.06095)	(0.06418)	(0.14862)
R3_KOKO(-1)	-0.055490	-0.024417	0.776494
	(0.02993)	(0.03151)	(0.07297)
R3_KOKO(-2)	0.056928	-0.016076	-0.025326
	(0.03638)	(0.03830)	(0.08870)
R3_KOKO(-3)	-0.003247	0.086542	0.110225
	(0.03484)	(0.03669)	(0.08496)
R3_KOKO(-4)	-0.001435	-0.020522	0.042489
	(0.02831)	(0.02981)	(0.06904)
C	0.272587	0.374309	0.009755
	(0.10023)	(0.10553)	(0.24438)

Comparing Table 6 and Table 8, we can see that the point estimates have rea\mined unchanged, while the standard errors have changed. In general, they are marginally smaller than for the Sims and Zha (1998) Normal-Wishart natural conjugate prior (keeping constant the EViews' default settings for the hyperparameters).

5.6 The Koop and Korobilis (Ko-Ko) natural-conjugate (Normal-Wishart) prior

Just as in the Sims and Zha case above, if the assumption of a fixed and diagonal variance-covariance matrix is relaxed, the natural joint prior for normal data is the Normal-Wishart distribution:

$$p(\beta|\Sigma) \sim N(\beta_0, \Sigma \otimes V_0) \qquad p(\Sigma^{-1}) \sim W(v_0, S_0^{-1}) \qquad (17)$$

where the default values for β and Σ^{-1} in the prior distribution for β given Σ and Σ^{-1} are specified as:

$$p(\beta|\Sigma) \sim N(0, 10I_m) \qquad p(\Sigma^{-1}) \sim W(m, I_m) \qquad (18)$$

such that $\beta_0 = 0$, $V_0 = 10I_m$, $v_0 = m$ and $S_0 = I_m$.

The resulting posterior distributions for β and Σ are then:

$$p(\beta | \Sigma, Y) \sim N(\bar{\beta}, \Sigma \otimes \bar{V}) \qquad p(\Sigma^{-1} | Y) \sim W(\bar{v}, \bar{S}^{-1})$$

where:

$$\overline{V} = (V_0^{-1} + X'X)^{-1} \tag{19}$$

$$\overline{\beta} = \overline{V}(V_0^{-1}B_0 + X'X\hat{B}) \tag{20}$$

$$\overline{v} = v_0 + T \tag{21}$$

$$\overline{S} = SSE + S_0 - \hat{B}'X'X\hat{B} + B_0'V_0^{-1}B_0 - \overline{B}'(V_0^{-1} + X'X)\overline{B} \tag{22}$$

$$SSE = (Y - X\hat{B})')Y - X\hat{B}) \tag{23}$$

In equations (19) to (23), $B_0 = \text{unvec}(\beta_0)$ and $\hat{B}$ is the OLS estimate of B. This once again illustrates the role of familiar OLS quantities in the derivation of posterior results with informative priors.

We note two properties of this prior that can be undesirable in some circumstances. To begin with, the ($I_m \otimes X$) form of the explanatory variables in equation (17) means that every equation must have the same set of explanatory variables. This may be appropriate for an unrestricted VAR, but may not be suitable if the researcher wishes to impose restrictions. Zero restrictions therefore cannot be imposed with the natural conjugate prior described above.

The second – possibly undesirable – property of this prior requires us to denote individual elements of Σ as σ_{ij}. The fact that the prior variance-covariance matrix has the form ($\Sigma \otimes V_0$), which is necessary to ensure natural conjugacy of the prior, also implies that the prior covariance of the coefficients in equation i is $\sigma_i V_0$. This means that the prior covariance of the coefficients in any two equations must be proportional to one another, which is another possibly restrictive feature.

Pressing **OK** at the end of the first dialog box opens up another, which allows us to choose the specific settings for the Sims and Zha (natural-conjugate) prior. To begin with, we use EViews' default settings.
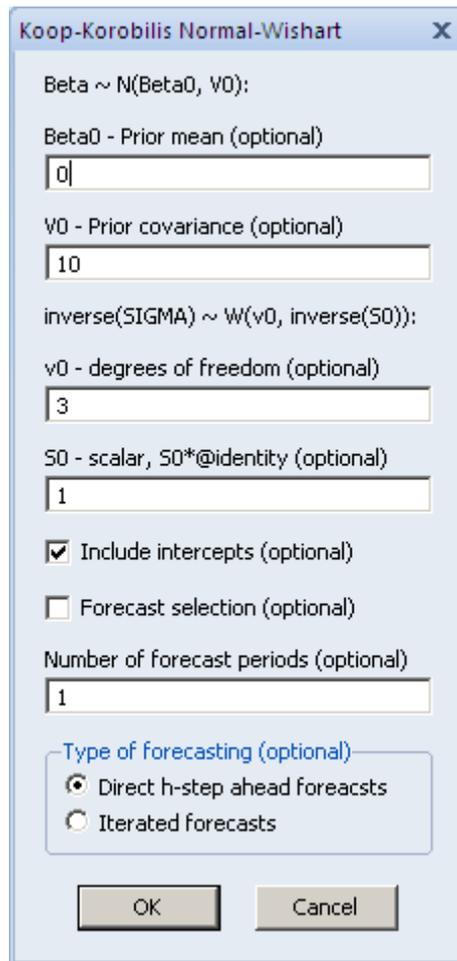
We keep EViews' default options as they are and press **OK** to estimate the BVAR. Much to my surprise, the BVAR cannot be calculated in EViews and the add-in fails to solve. This seems to be the case regardless of the options chosen, i.e., any setting of the hyperparameters yields the same results.

One experiment in this vein is the following. Koop and Korobilis (2009, p. 291) also state that we can set up a non-informative version of the natural conjugate prior under which $\beta_0 = 0_{(mn \times 1)}$, $V_0 = 100I_{(n \times n)}$, $v_0 = 0$ and $S_0 = 0_{(n \times n)}$. These options are shown in the following screenhot.
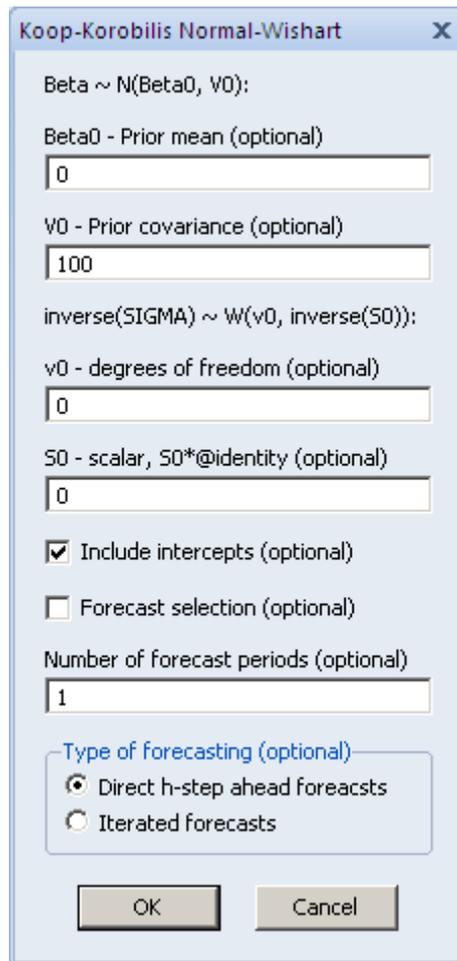
Again, we find that EViews is unable to estimate the BVAR with these prior settings, primarily because v_0, the (optional) degree of freedom parameter, even though it appears as optional in the parameter settings, has to be set equal to at least the number of variables in the model.

6   Impulse response functions

The EViews BVAR add-in offers two additional tools. The first is the ability to calculate impulse response functions based on the estimated BVAR model. For illustrative purposes, let us re-do the estimation of the BVAR model with the Sims and Zha (1998), this time also calculating impulse response functions over 20 periods (five years). These can then be compared to the unrestricted VAR impulse response functions.

In order to have EViews generate impulse response functions for us, we simply tick the check box for **Calculate impulse responses** and select the required number of periods (20) in the very next box. The options are reproduced in the following screenshot.
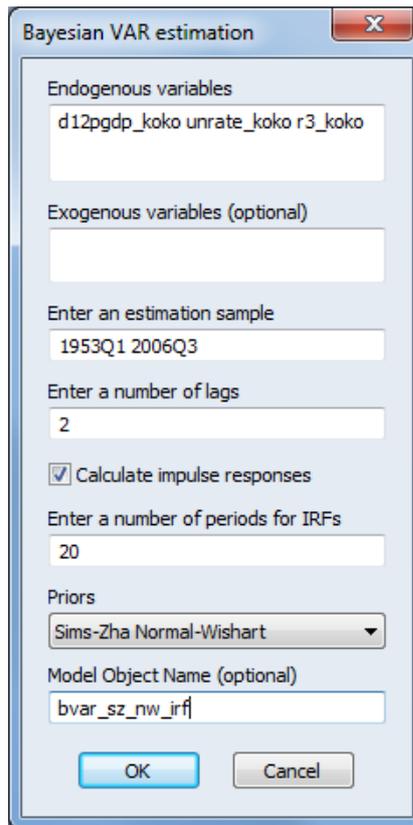
We leave the settings of the hyperparameters on the second dialog box at EViews' default settings and press OK to estimate the BVAR. The resulting output, i.e., the impulse response functions, will be part of the spool object that we have already come across. This is illustrated in the following screenshot.

What is different from our earlier estimation and the associated spool objects is the list appearing on the left-hand side, consisting of `est_results`, `irf_table` and `irf_graph`. These three labels allow us to toggle between the estimation results (by clicking on `est_results`), a table of the values of the impulse response functions (by clicking on `irf_table`) and the nine graphs of the impulse response functions (by clicking on `irf_graph`). The latter are reproduced below in Figure 2. In order to do so, we need to select it within the spool object, which is done by right-clicking on the selected object. You then select **Extract**, which brings up an untitled graph in EViews.[6]
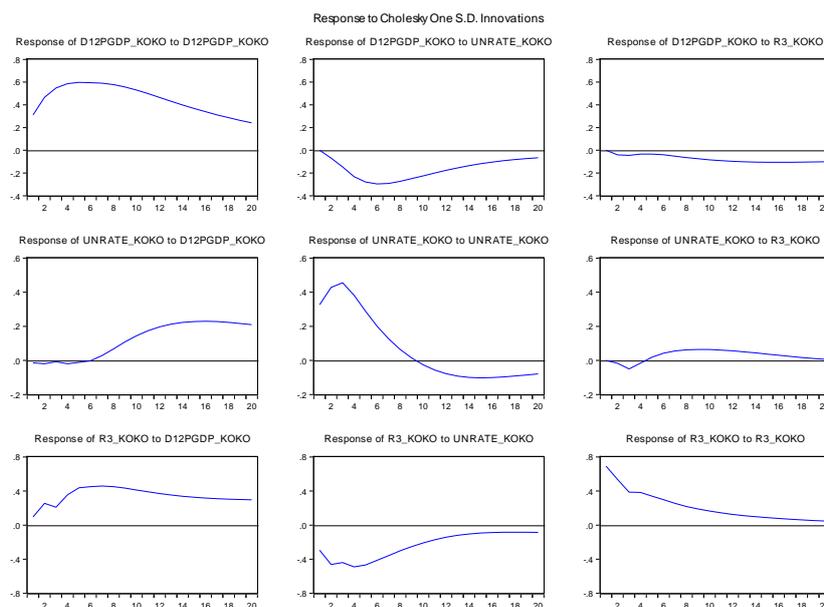
Figure 2: Impulse response functions of the BVAR(4) using Sims and Zha's (1998) Normal-Wishart natural conjugate prior
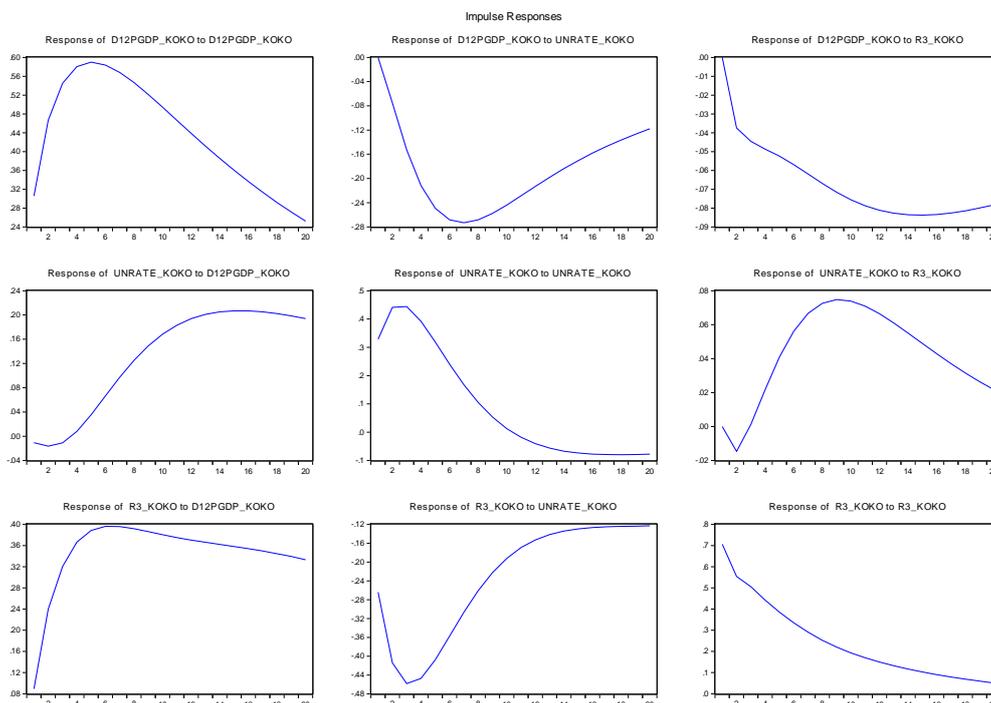
These impulse response functions can then be compared to those arising from the unrestricted VAR(4) in Section 4.1, which are shown in Figure 3.

---

[6] A similar operation on the `irf_table` portion of the spool object generates an EViews table.

Figure 3: Impulse response functions of the unrestricted VAR(4) model

Response to Cholesky One S.D. Innovations

While the shapes of the impulse response functions in Figures 2 and 3 are broadly comparable, the exact magnitudes differ somewhat, although this is conditional on the exact chart. Koop and Korobilis (2009, Fig. 2.2, p. 295) produce a chart reproduced as Figure 4 below, which shows the impulse response functions of the posterior for the non-informative prior in the case of a unit shock. Note that the elements are arranged by shock, rather than variables, which is the case in EViews, such that Figure 4 will have to be transposed to make it comparable to Figure 3 above. In addition, the impulse response functions in Figure 3 are in response to a one standard-deviation shock, such that the axes will be different across figures. The similarity is nonetheless quite pleasing.

Figure 4: Impulse response functions of the diffuse BVAR prior
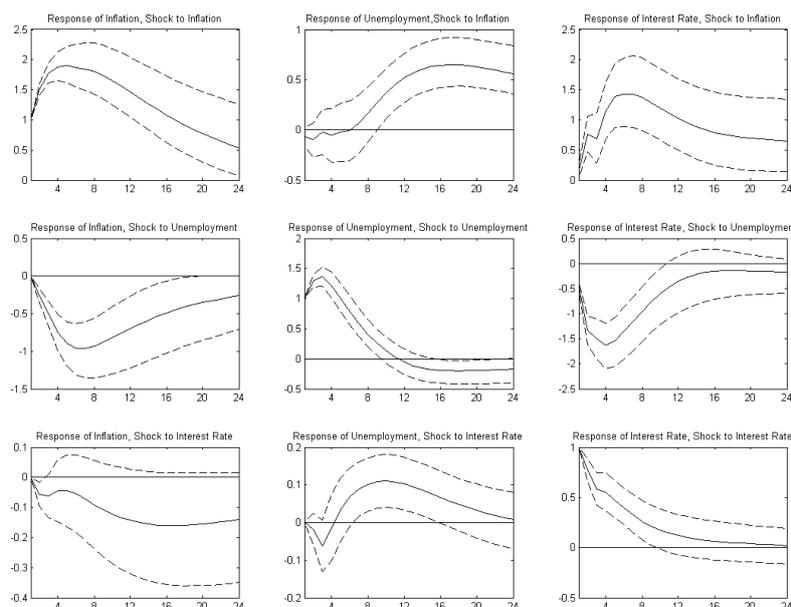
7   Forecasting

The other additional tool is the ability to forecast with the estimated BVAR model. In many forecasting situations, the forecast horizon, h, is longer than the frequency used in collecting data and estimating the forecasting model. In these situations, the forecaster encounters a multi-period forecasting problem and possible approaches to multi-period forecasting are either **direct** or **iterated** forecasts. **Iterated** multi-period-ahead time-series forecasts are made using a one-period-ahead model, iterated forward for the desired number of periods. Assume the case of a generic univariate AR(1) model given by:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t \tag{24}$$

The one-step-ahead forecast is generated as:

$$y_{t+1|t} = \hat{\alpha} + \hat{\beta} y_t \tag{25}$$

while the forecasts for h-step-ahead horizons are obtained as:

$$y_{t+2|t} = \hat{\alpha} + \hat{\beta} y_{t+1|t} = \hat{\alpha} + \hat{\beta}(\hat{\alpha} + \hat{\beta} y_t)$$
$$y_{t+3|t} = \hat{\alpha} + \hat{\beta} y_{t+2|t} = \hat{\alpha} + \hat{\beta}(\hat{\alpha} + \hat{\beta} y_{t+1|t}) = \hat{\alpha} + \hat{\beta}(\hat{\alpha} + \hat{\beta}(\hat{\alpha} + \hat{\beta} y_t))$$
$$\vdots$$
$$y_{t+h|t} = \hat{\alpha} + \hat{\beta} y_{t+h-1} \tag{26}$$

In contrast, **direct** forecasts are made using a horizon-specific estimated model, where the dependent variable is the multi-period ahead value being forecasted (this approach directly optimises the relevant loss function, i.e., the h-step ahead mean squared forecast error). The estimated model is then:

$$y_t = \alpha_h + \beta_h y_{t-h} + \varepsilon_t \tag{27}$$

where the variable at time t is projected directly onto its past value in period $t - h$, which means that the estimated coefficients summarise the h-step ahead effect. Note that a different α_h and β_h are obtained for each forecast horizon. The direct forecasts are then derived as:

$$y_{t+h} = \alpha_h + \beta_h y_t \tag{28}$$

Note that for the one-step-ahead horizon, the two methods produce the same results. Which approach is better is an empirical matter. In theory, iterated forecasts are more efficient if the one-period-ahead model is correctly specified, while direct forecasts are more robust to model misspecification (Marcellino *et al.* (2006)).

Let us start with an illustrative one-step-ahead forecast, for which the two forecasting methods should yield the same result. We once again use the BVAR(4) with the Sims and Zha (1998) Normal-Wishart natural conjugate prior. This time, though, we operate with the model that we have created when doing the estimation. The model corresponding to the Sims

and Zha (1998) reference prior is called **bvar_sims_zha1**. Open the model by double-clicking on its icon in the workfile.

To construct the unconditional forecast (within the **Model** object), either go to **Proc**, **Solve Model...** or click on the **Solve** button on the model toolbar. EViews will display a tabbed dialog containing the solution options.[7] For the current forecasting exercise, we will choose the following options: on the **Basic Option** tab, select **Deterministic** under **Simulation type**, **Dynamic solution** in the **Dynamics** box and select 2006q4 2006q4 as the **Solution sample**. Leave everything else unchanged and click on **OK** to start solving, i.e., forecasting, the model. All the required options are shown in the screenshot below.
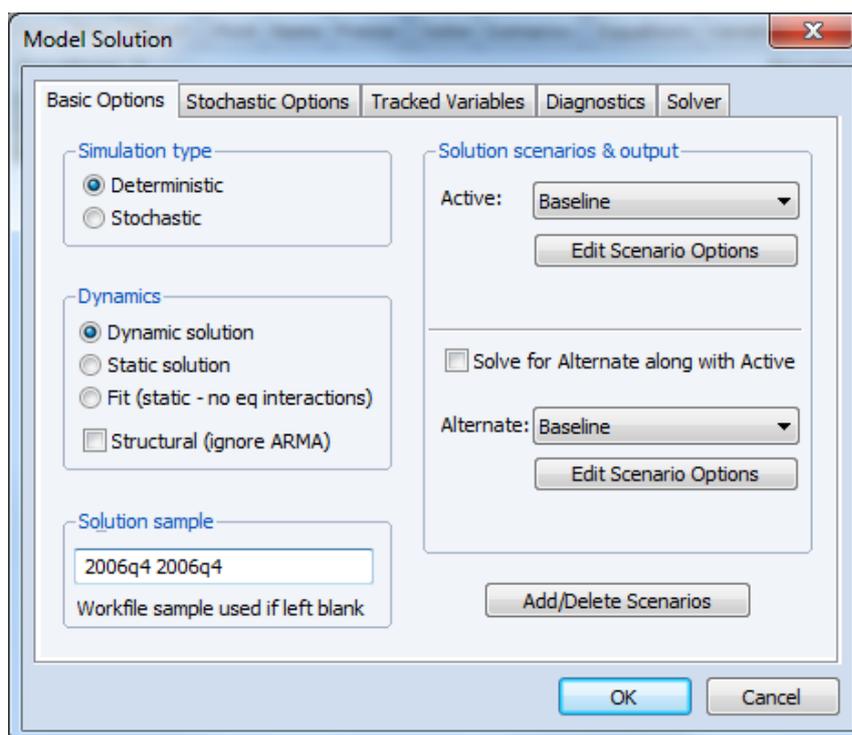
In a **deterministic** simulation, all equations in the model are solved so that they hold without error during the solution sample, all coefficients are held fixed at their point estimates and all exogenous variables, if present, are held constant at their historical values in the sample. This results in a **single** path for the endogenous variables which can be evaluated by solving the model once.

Options in the **Dynamics** box determine how EViews uses historical data for the endogenous variables when solving the model:

- when **Dynamic solution** is chosen, only values of the endogenous variables preceding the solution sample are used when forming the forecast – a dynamic solution is typically the correct method to use when forecasting values several periods into the future (a multi-step-ahead forecast) or for evaluating how a multi-step-ahead forecast would have performed historically; and

---

[7] This **Basic Option** page contains the most important options for the simulations. While the options on other tabs, such as **Stochastic Options**, can often be left at their default values, the options on this page will definitely need to be set appropriately for the task we are trying to perform.

- when **Static solution** is chosen, values of the endogenous variables up to the previous period are used each time the model is solved – a static solution is typically used to produce a set of one-step-ahead forecasts over the historical data to examine the historical fit of the model within the sample period.[8]

Once the calculations are finished, the model window will switch to the **Solutions Messages** view, thereby indicating that the model has been solved successfully. If that should not be the case, an error message will appear in this view, which may (or may not) be helpful in indicating what has gone wrong in the model solution.

Now that we have solved the model, we would like to look at the results. We find that EViews has created three new series for us: `d12pgdp_koko_0`, `unrate_koko_0` and `r3_koko_0`, which contain the forecasts from the baseline scenario, i.e., the BVAR(4) we have just estimated. Since the newly created series are standard EViews objects, we could use the workfile window to open the series and examine them directly. The forecasted series can now be compared to the actual values in 2006 Q4. For `d12pgdp_koko`, the forecast value is 2.763, while the actual value that quarter is 3.275; for `unrate_koko`, the forecast value is 4.922, while the actual value that quarter is 4.700 and for `r3_koko`, the forecast value is 4.537, while the actual value that quarter is 4.600.

Longer-term forecasts, using either the iterated or the direct method can also be produced using the EViews model object, as can standard error bands around the forecasts.

8   Conclusions

As we have seen, putting BVARs to practical use raises a lot of detailed questions about model specification, estimation, inference and forecast construction. Carriero *et al.* (2015) address some of the pertinent issues having to do with model specification, such as:

- the choice of tightness and lag length of the BVAR;
- the treatment of the error variance and the imposition of cross-variable shrinkage;
- whether or not to transform the variables to get stationarity; and
- whether to complement this choice with the imposition of prior favouring cointegration and unit roots

Using the Sims and Zha (1998) prior as their benchmark, their results firmly support simpler (in their words 'quick and easy') specifications of BVARs over more complicated ones. In particular, for researchers interested in point forecasts only, there are no drawbacks from using methods that do not involve simulation. According to the authors, an approach that works well is to specify a Normal-(inverse) Wishart prior along the lines of Sims and Zha (1998) on the VAR in levels and optimise both the overall tightness and the lag length. In particular, they advise that optimising over the lag length of the BVAR is more 'helpful' than optimising the tightness of the prior. Finally, they strongly advise against using the Minnesota prior that treats the error variance-covariance matrix as fixed and diagonal.

---

[8] As such, a static solution cannot be used to predict more than one observation into the future at a time.

References and further reading

Bańbura, M, Giannone, D and Reichlin, L (2010), 'Large Bayesian vector auto regressions', *Journal of Applied Econometrics*, Vol. 25, No. 1, pages 71-92.

Blake, A P and Mumtaz, H (2012), *Applied Bayesian econometrics for central bankers*, Centre for Central Banking Studies Technical Handbook – No. 4. http://www.bankofengland.co.uk/education/Documents/ccbs/technical_handbooks/pdf/techbook4.pdf.

Canova, F (2007), *Methods for applied macroeconomic research*, Princeton, Princeton University Press.

Carriero, A, Clark, T E and Marcellino, M (2015), 'Bayesian VARs: specification choices and forecast accuracy', *Journal of Applied Econometrics*, Vol. 30, No. 1, pages 46-73.

Clements, M P and Hendry, D F (1996), 'Intercept corrections and structural change', *Journal of Applied Econometrics*, Vol. 11, No. 5, pages 475-94.

Cogley, T and Sargent, T (2005), 'Drifts and volatilities: monetary policies and outcomes in the post WWII US', *Review of Economic Dynamics*, Vol. 8, No. 2, pages 262-302.

Diebold, F X and Kilian, L (2000), 'Unit-root tests are useful for selecting forecasting models', *Journal of Business & Economic Statistics*, Vol. 18, No. 3, pages 265-73.

Doan, T, Litterman, R and Sims, C A (1984), 'Forecasting and conditional projection using realistic prior distributions', *Econometric Reviews*, Vol. 3, No. 1, pages 1-100.

Giannone, D, Lenza, M and Primiceri, G E (2012), 'Prior selection for vector autoregressions', *NBER Working Paper No. 18467*. http://www.nber.org/papers/w18467.pdf.

Kadiyala, K R and Karlsson, S (1997), 'Numerical methods for estimation and inference in Bayesian VAR-models', *Journal of Applied Econometrics*, Vol. 12, No. 2, pages 99-132.

Kim, C-J and Nelson, C R (1999), *State-space models with regime switching*, Cambridge, Mass, MIT Press.

Koop, G and Korobilis, D (2009), 'Bayesian multivariate time series methods for empirical macroeconomics', *Foundations and Trends in Econometrics*, Vol. 3, No. 4, pages 267-358. http://personal.strath.ac.uk/gary.koop/koop_korobilis_Foundations_and_Trends_2010.pdf.

Koop, G, Leon-Gonzalez, R and Strachan, R (2009), 'On the evolution of the monetary policy transmission mechanism', *Journal of Economic Dynamics and Control*, Vol. 33, No. 4, pages 997-1017.

Marcellino, M, Stock, J H and Watson, M W (2006), 'A comparison of direct and iterated multistep AR methods for forecasting economic time series', *Journal of Econometrics*, Vol. 127, Nos. 1-2, pages 499-526.

Primiceri, G (2005), 'Time varying structural vector autoregressions and monetary policy', *Review of Economic Studies*, Vol. 72, No. 3, pages 821-52.

Sims, C A (1993), 'A nine-variable probabilistic macroeconomic forecasting model', Chapter 4 in Stock, J H and Watson, M E (eds), *Business cycles, indicators and forecasting*, Chicago, University of Chicago Press, pages 179-212 (with comment). http://www.nber.org/chapters/c7192.pdf.

Sims, C A and Zha, T A (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review*, Vol. 39, No. 4, pages 949-968.

Waggoner, D F and Zha, T A (1999), 'Conditional forecasts in dynamic multivariate models', *Review of Economics and Statistics*, Vol. 80, No. 4, pages 639-51.

Zha, T A (1998), 'A dynamic multivariate model for use in formulating policy', *Federal Reserve Bank of Atlanta Economic Review*, Vol. 83, No. 1, pages 16-29.

Appendix: The sum-of-coefficients and dummy initial observation priors

It is in principle unclear whether transforming variables by taking the first difference (i.e., growth rates) can enhance the forecasting performance of BVARs. For example, some researchers and practitioners prefer to leave variables in log levels and impose prior means of unit roots with additional priors to be discussed below (Banbura *et al.* (2010); Giannone *et al.* (2012)). One reason is that such a specification can better take into consideration the existence of long-run (cointegrating) relationships across the variables, which are omitted in a VAR in differences. In contrast, Clements and Hendry (1996) show that in a classical framework differencing can improve the forecasting performance in the presence of instability. Furthermore, Diebold and Kilian (2000) show that, for variables with unit roots, forecasting accuracy can be improved by differencing.

When specifying a VAR with potentially nonstationary variables in levels, it often makes sense to address the issues of unit roots and/or cointegration by imposing the so-called sum-of-coefficients and dummy initial observation priors due to Doan *et al.* (1984) and Sims (1993). Both of these prior can be implemented by augmenting the system with dummy observations.

The first is motivated by the frequent practice of specifying a VAR model involving data that contain unit roots in first differences of the data. This specification corresponds to the restrictions that the sums of coefficients on the lags of the dependent variable in each equation of the VAR equal one, while coefficients on lags of other variables sum to zero. One way of implementing this prior is to add a set of n initial dummy observations to the data. For example, if there are $n = 2$ variables then there are two sum-of-coefficients dummy observations for each equation. An alternative way of interpreting this prior is that it expresses a belief that when the average of lagged values of a variable is at some level, $\bar{y}_i$, that same value $\bar{y}_i$ is likely to be a good forecast of y_{it}. It also implies that knowing the average of lagged values of variable i does not help in predicting a variable $i \neq j$. This set-up is facilitated by the extra hyperparameter, μ_5 in the Sims and Zha (1998) prior. A weight of $\mu_5 \geq 0$ is then attached to these dummy observations and, as $\mu_5 \to \infty$, the estimated VAR will increasingly satisfy the sum of coefficients restriction. Note also that as $\mu_5 \to \infty$, the forecast growth rates will eventually converge to their sample average values.

The dummy initial observation component of a prior, proposed by Sims (1993), introduces a single dummy observation in which, up to a scaling factor, all values of all variables are set equal to the corresponding averages of initial conditions (i.e., y_{i0}), and the last column of the data matrix is set at the value of 1. We designate the scale factor for this single dummy observation as μ_6. If we assume that each series in the VAR contains a unit root, then, as $\mu_5 \to \infty$, the sum of coefficients restriction implies that there are as many unit roots in the VAR as there are variables in y_t. On the other hand, it might be reasonable to suppose that there are fewer than n unit roots in the VAR, as would be the case if there were stable long-run (cointegration) relations between the trending series. Sims (1993) observed that introducing an additional dummy observation could make some allowance for this possibility. A weight of $\mu_6 \geq 0$ is then attached to these dummy observations for each equation. If the series individually contain unit roots, then, as $\mu_6 \to \infty$, increasingly more weight will be put on a VAR that can be written in a form in which all series share a single unit root and the intercept will be close to zero.