CPB Netherlands Bureau for Economic Policy Analysis

CPB Background Document | February, 2015

Technical background document for BVAR models used at CPB

Joris de Wind

Contents

1 Introduction

Thanks to the seminal work of Sims (1980), vector autoregressions (VARs) have a long tradition in applied macroeconomics, and are widely used for policy analysis and forecasting. They are flexible time series models that can capture complex dynamic interrelationships among macroeconomic variables. The VAR framework imposes no restrictions on the lag structure of the system and is therefore not parsimonious at all, i.e. all variables are allowed to respond to all variables at all lags. The generality of the VAR model brings along a large number of parameters even for systems of moderate size. Bayesian shrinkage makes it nevertheless possible to estimate VAR models with a *large* set of variables, which without Bayesian shrinkage would lead to overfitting and poor out-of-sample forecasting performance. For a general discussion of Bayesian VARs, see the excellent survey paper of Koop and Korobilis (2010) and the forecasting applications of Bańbura, Giannone, and Reichlin (2010) and Giannone, Lenza, and Primiceri (2013) as well as the references therein. Here, we only would like to emphasize that Bayesian VARs are among the frontier forecasting models and are well suited to include a *large* set of variables, just like dynamic factor models.

This document consists of two parts that are self-contained. First, we set out some notational conventions for Bayesian VARs and derive the formulae for the posterior distribution under various prior settings. In general, the posterior distribution is a weighted combination of prior distribution and likelihood function. The weights are (partly) endogenous, which is the key rationale behind Bayesian shrinkage, as will be explained in this document. This document is rather technical and all the results are derived from scratch.

Second, we present details on the two Bayesian VAR instruments used at the CPB, i.e. one instrument for forecasting the Dutch economy and another instrument for forecasting world trade. Both instruments contain a wide range of variables that cover the full spectrum of the economy, only the relevant set of variables is different for the two forecasting applications. For both instruments, we describe the data used, discuss which Bayesian VAR settings are chosen, and explain what type of output we look at. The description of the two applications is self-contained so that readers may decide to skip the derivations of the Bayesian VAR formulae.

The organization of the rest of this document is as follows. In section 2, we set out the notational conventions used in this document. In section 3, we derive from scratch the formulae for the posterior distribution under various prior settings and give some intuition behind the weighting between prior and data. In section 4, we describe the various ways how one can set the hyperparameters of the prior distribution. In

section 5, we explain how the Bayesian VAR framework can be used for forecasting under missing observations and scenario analysis. In section 6, we give the Gibbs sampling procedure that can be used to estimate Bayesian VARs with priors that are not conjugate to the likelihood function. Finally, in section 7, we present details on the two Bayesian VAR instruments used at the CPB.

2 Notational conventions

Consider a VAR model with n variables and p lags

$$y_t = c + \sum_{i=1}^{p} B_i y_{t-i} + u_t \tag{1}$$

where y_t is an $n \times 1$ vector of observed endogenous variables, c is an $n \times 1$ vector of constants, $\{B_i\}_{i=1}^{p}$ are $n \times n$ matrices of autoregressive parameters, and u_t is an $n \times 1$ vector of shocks. Shocks are assumed to be distributed according to the normal distribution $u_t \sim N(0, \Sigma)$.

It is more convenient to rewrite the VAR in concise matrix form. Let T denote the number of time periods in the sample after accounting for initialization and write

$$Y = XB + U \tag{2}$$

where Y is a $T \times n$ matrix of regressands, X is a $T \times k$ matrix of regressors, U is a $T \times n$ matrix of shocks, and B is a $k \times n$ matrix of regression parameters, with $k = 1 + pn$ is the number of regression parameters per VAR equation. The matrix of regressands is constructed as $Y = [\{y_t\}_{t=1}^{T}]'$ and the matrix of regressors as $X = [\{x_t\}_{t=1}^{T}]'$ with $x_t = [1, \{y'_{t-i}\}_{i=1}^{p}]'$. This notation implies that the i^{th} column of the matrix B consists of the regression parameters of the i^{th} VAR equation.

To derive all the results in section 3, we also need a vectorized version of the above matrix equation. Let $\otimes$ denote the Kronecker product and $\text{vec}(\cdot)$ the column stacking operator. Using the vectorization rule that $\text{vec}(ABC) = (C' \otimes A)\,\text{vec}(B)$, we can derive that

$$y = (I_n \otimes X)\,\beta + u \tag{3}$$

where $y = \text{vec}(Y)$ is a $Tn \times 1$ vector of regressands and $\beta = \text{vec}(B)$ is a $kn \times 1$ vector of regression parameters. Furthermore, $u = \text{vec}(U)$ is a $Tn \times 1$ vector of shocks that are distributed according to the normal distribution $u \sim N(0, \Sigma \otimes I_T)$. Note that y stacks the $T \times 1$ vectors $\{y_i\}_{i=1}^{n}$ below each other, that is $y = [\{y'_i\}_{i=1}^{n}]'$.

3 Bayesian VAR formulae

We first describe the likelihood function and after that we set out the prior and posterior distributions.

3.1 Likelihood function

The likelihood function is simply the sampling distribution $p(y|\beta,\Sigma)$ but then viewed as a function of the parameters. The multivariate normal distribution of the shocks in the vectorized form (3) implies that

$$p(y|\beta,\Sigma) \propto |\Sigma \otimes I_T|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}u'(\Sigma \otimes I_T)^{-1}u\right) \tag{4a}$$

$$\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left((Y-XB)'(Y-XB)\Sigma^{-1}\right)\right) \tag{4b}$$

where equation (4b) can be derived from equation (4a) with several steps of algebra. Note that equation (4b) can also be derived directly from the concise matrix form (2) using the matrix normal distribution (as opposed to the multivariate normal distribution). We have left out the constant-of-proportionality since we do not need it anyway.

The likelihood function can be decomposed into two components, namely a normal (conditional) distribution for β given Σ and an inverse-Wishart (marginal) distribution for Σ. For this purpose, use the well-known decomposition rule that $(Y-XB)'(Y-XB) = (Y-X\hat{B})'(Y-X\hat{B}) + (\hat{B}-B)'X'X(\hat{B}-B)$, where $\hat{B} = (X'X)^{-1}X'Y$ is the ordinary least squares estimate of B, giving us

$$p(y|\beta,\Sigma) \propto |\Sigma|^{-\frac{k}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left((\hat{B}-B)'X'X(\hat{B}-B)\Sigma^{-1}\right)\right) \tag{5a}$$

$$|\Sigma|^{-\frac{T-k}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left((Y-X\hat{B})'(Y-X\hat{B})\Sigma^{-1}\right)\right) \tag{5b}$$

The first line (5a) is the kernel of the matrix normal distribution and the second line (5b) is the kernel of the inverse-Wishart distribution. It is more convenient to rewrite the matrix normal distribution in terms of the multivariate normal distribution, so that

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

$$\Sigma|y \sim IW\left(\hat{S},T-k-n-1\right) \tag{6b}$$

where $\hat{\beta} = \operatorname{vec}(\hat{B})$ and $\hat{S} = (Y-X\hat{B})'(Y-X\hat{B})$. Note that we can equally get $\hat{\beta} = (x'x)^{-1}x'y$ with $x = (I_n \otimes X)$ from the vectorized form (3).

Altogether, under a flat prior, we have a conditional normal distribution for β given Σ and a marginal inverse-Wishart distribution for Σ. The joint posterior mode of the model parameters is given by

$$\beta^* = \hat{\beta} \tag{7a}$$

$$\Sigma^* = \frac{\hat{S}}{T} \tag{7b}$$

implying that the posterior distribution (under a flat prior) is centered around the ordinary least squares estimates.[1] Finally, note that under the non-informative Jeffreys prior only the degrees-of-freedom parameter of the inverse-Wishart distribution increases by $n + 1$.

3.2 Prior and posterior distributions

It turns out to be possible to specify the prior distribution in such a way that the posterior distribution is analytically available (as in the flat prior case discussed above). We distinguish three cases:

- Σ known and normal prior distribution for β
- Σ unknown and normal inverse-Wishart prior distribution for $\{\Sigma, \beta\}$, with Kronecker structure
- Σ unknown and normal inverse-Wishart prior distribution for $\{\Sigma, \beta\}$, no restrictions

The posterior distribution is analytically available in the first two cases, but not in the third case. Whenever the posterior distribution is analytically available and in the same family as the prior distribution, the prior distribution is said to be conjugate to the likelihood function.

The three cases are discussed in the remainder of this section. Moreover, at the end of the discussion of the first case, we elaborate on the rationale behind Bayesian shrinkage.

3.2.1 Σ known and normal prior distribution for β

The prior discussed here has come to be known as the Minnesota prior. The Minnesota prior is based on the assumption that Σ is known, so that in practice it should be pre-estimated.[2] This is not a good Bayesian treatment, but it makes the computations

[1] Note that the mode of the marginal distribution for Σ is actually $\frac{\hat{S}}{T-k}$. However, the degrees-of-freedom correction drops out when accounting for the feedback from the conditional distribution for β.

[2] In the original Minnesota prior Σ is assumed to be a diagonal matrix, but this is actually not needed and relaxed in later work.

much easier in that the posterior distribution is analytically available. With a proper Bayesian treatment of Σ, we would need to impose some restrictions on the prior in order to have an analytical posterior. It should be noted that the latter approach, which is discussed in section 3.2.2, is more common nowadays and we elaborate on the Minnesota prior just for completeness and to give a quick insight in the history of Bayesian VARs.

When Σ is known (or pre-estimated), we only have to specify a prior distribution for β. The Minnesota prior is based on the normal prior distribution

$$\beta \sim N\left(\underline{\beta}, \underline{V}\right) \tag{8}$$

where $\underline{\beta}$ is a $kn \times 1$ vector with the prior mean and $\underline{V}$ is a $kn \times kn$ matrix with the prior covariance. The Minnesota prior puts a particular structure on $\underline{\beta}$ and $\underline{V}$. The idea behind the Minnesota prior is to shrink the model towards the random walk, with stronger shrinkage for coefficients on longer lags and across variables.

The random walk (without drift) implies that $\underline{\beta} = \text{vec}\left([O_{n\times 1} \;\; I_{n\times n} \;\; O_{n\times n(p-1)}]'\right)$, i.e. $\underline{\beta}$ is simply a vector with all zeros except for the elements corresponding to the first own lags. Besides, the ones can be replaced by zeros or any number if the data enter the model in growth rates as opposed to levels.

The prior covariance matrix $\underline{V}$ is assumed to be diagonal. Let $\underline{v}_i$ denote the $k \times 1$ vector with the diagonal elements associated with the parameters of the i^{th} VAR equation, so that $\underline{V} = \text{diag}\left(\{\underline{v}_i\}_{i=1}^{n}\right)$. Furthermore, let $\underline{v}_{iq}$ denote the q^{th} element of $\underline{v}_i$. Note that $\underline{v}_{i1}$ is associated with the constant term, while $\underline{v}_{iq}$, $q \neq 1$, is associated with the r^{th} lag of variable j with $r = \lfloor \frac{q-2}{n} \rfloor + 1$ and $j = q - 1 - \lfloor \frac{q-2}{n} \rfloor n$. A common implementation of the Minnesota prior chooses the elements of $\underline{v}_i$ according to

$$\underline{v}_{iq} = \begin{cases} \frac{\pi_1}{r^2} & \text{if } q \neq 1 \text{ and } j = i \text{ (for coefficients on the } r^{\text{th}} \text{ own lags)} \\ \frac{\pi_2}{r^2}\frac{\sigma_{ii}}{\sigma_{jj}} & \text{if } q \neq 1 \text{ and } j \neq i \text{ (for coefficients on the } r^{\text{th}} \text{ foreign lags)} \\ \pi_3 \sigma_{ii} & \text{if } q = 1 \text{ (for coefficients on deterministic terms)} \end{cases} \tag{9}$$

where σ_{ii} is the i^{th} diagonal element of Σ. The σ_{ii} terms are used to account for differences in the scaling of the various variables in the model. The above rule implies that shrinkage is stronger for longer lags and across variables (provided that $\pi_1 > \pi_2$). Typically, π_3 is chosen to be large implying very weak shrinkage for deterministic terms.

When Σ is known (or pre-estimated), we can drop several terms from the likelihood function (5) that was derived in section 3.1 for the general case with unknown Σ. In fact, we only need the $\exp(\cdot)$ term on line (5a), which is the kernel of the matrix normal distribution. It is more convenient to rewrite the matrix normal kernel in terms of the

multivariate normal kernel, giving us

$$p\left(y|\beta\right) \quad \propto \quad \exp\left(-\frac{1}{2}\left(\beta-\hat{\beta}\right)'\hat{V}^{-1}\left(\beta-\hat{\beta}\right)\right) \tag{10}$$

where $\hat{V}=\Sigma\otimes\left(X'X\right)^{-1}$.

Finally, the likelihood function can be combined with the prior distribution. In particular, multiplying the likelihood kernel with the prior kernel yields the posterior kernel, giving us

$$p\left(\beta|y\right) \quad \propto \quad \exp\left(-\frac{1}{2}\left(\beta-\hat{\beta}\right)'\hat{V}^{-1}\left(\beta-\hat{\beta}\right)\right)\exp\left(-\frac{1}{2}\left(\beta-\underline{\beta}\right)'\underline{V}^{-1}\left(\beta-\underline{\beta}\right)\right) \tag{11}$$

It should be noted that the sum of two quadratic expressions is just another quadratic expression. Therefore, the result of the multiplication of the two $\exp\left(\cdot\right)$ terms simply yields another $\exp\left(\cdot\right)$ term or in other words a multivariate normal kernel

$$p\left(\beta|y\right) \quad \propto \quad \exp\left(-\frac{1}{2}\left(\beta-\overline{\beta}\right)'\overline{V}^{-1}\left(\beta-\overline{\beta}\right)\right) \tag{12}$$

By matching coefficients, which is a very powerful algebra trick, we can find $\overline{\beta}$ and $\overline{V}$. This yields the posterior distribution

$$\beta|y \quad \sim \quad N\left(\overline{\beta},\overline{V}\right) \tag{13a}$$

with

$$\overline{\beta} \quad = \quad \left(\hat{V}^{-1}+\underline{V}^{-1}\right)^{-1}\left(\hat{V}^{-1}\hat{\beta}+\underline{V}^{-1}\underline{\beta}\right) \tag{13b}$$

$$\overline{V} \quad = \quad \left(\hat{V}^{-1}+\underline{V}^{-1}\right)^{-1} \tag{13c}$$

Summing up, the Minnesota prior yields a normal posterior distribution for β with mean and variance equal to (13b) and (13c), respectively.

One can recognize that the posterior mean for β is a weighted combination of the prior mean $\underline{\beta}$ and maximum likelihood estimator $\hat{\beta}$, with weights $\overline{V}\underline{V}^{-1}$ and $\overline{V}\hat{V}^{-1}$, respectively. The weights are (partly) controlled by the user who sets the hyperparameters that enter $\underline{V}$. Nevertheless, the weights are also endogenous since the data enter $\hat{V}$, i.e. the covariance matrix with parameter uncertainty. In fact, the data is informative about which (combinations of) parameters are robust and which are not. The weighting scheme takes this information into account, with relatively large data weights for dimensions that are found to be robust and relatively large prior weights for dimensions that are found to be unstable. The Bayesian shrinkage approach effectively recognizes the robust relationships in the data and replaces the unstable relationships by a parsimonious benchmark that is known to have decent forecasting performance.[3]

[3] The parsimonious benchmark used in the Minnesota prior is the random walk with drift, yet it is also

3.2.2 Σ unknown and normal inverse-Wishart prior distribution for $\{\Sigma, \beta\}$, with Kronecker structure

The prior discussed here is called the natural conjugate prior as it is from the same family of distributions as the likelihood and posterior. The natural conjugate prior is widely used since it is the only prior that yields an analytical posterior (without assuming that Σ is known). The natural conjugate prior has the exact same normal inverse-Wishart structure as the likelihood function, that is

$$\begin{aligned} \beta|\Sigma &\sim N\left(\underline{\beta}, \Sigma \otimes \underline{V}\right) && \text{(14a)} \\ \Sigma &\sim IW\left(\underline{S}, \underline{\upsilon}\right) && \text{(14b)} \end{aligned}$$

where we would like to emphasize that the Kronecker structure is strictly necessary to get an analytical posterior. With the natural conjugate prior, also the posterior satisfies the exact same functional form, that is

$$\begin{aligned} \beta|\Sigma, y &\sim N\left(\overline{\beta}, \Sigma \otimes \overline{V}\right) && \text{(15a)} \\ \Sigma|y &\sim IW\left(\overline{S}, \overline{\upsilon}\right) && \text{(15b)} \end{aligned}$$

or in terms of the probability density function

$$p\left(\beta, \Sigma|y\right) \propto |\Sigma|^{-\frac{k}{2}} \exp\left(-\frac{1}{2}(\beta-\overline{\beta})'\left(\Sigma^{-1}\otimes\overline{V}^{-1}\right)(\beta-\overline{\beta})\right)|\Sigma|^{-\frac{\overline{\upsilon}+n+1}{2}}\exp\left(-\frac{1}{2}\operatorname{tr}\left(\overline{S}\Sigma^{-1}\right)\right) \tag{16}$$

In order to show that the posterior indeed follows the functional form (16), multiply the likelihood function (5) with the prior distribution (14), giving us

$$\begin{aligned} p\left(\beta, \Sigma|y\right) \propto &|\Sigma|^{-\frac{k}{2}} \exp\left(-\frac{1}{2}(\beta-\hat{\beta})'\left(\Sigma^{-1}\otimes X'X\right)(\beta-\hat{\beta})\right)|\Sigma|^{-\frac{T-k}{2}}\exp\left(-\frac{1}{2}\operatorname{tr}\left(\hat{S}\Sigma^{-1}\right)\right) \\ &|\Sigma|^{-\frac{k}{2}} \exp\left(-\frac{1}{2}(\beta-\underline{\beta})'\left(\Sigma^{-1}\otimes\underline{V}^{-1}\right)(\beta-\underline{\beta})\right)|\Sigma|^{-\frac{\underline{\upsilon}+n+1}{2}}\exp\left(-\frac{1}{2}\operatorname{tr}\left(\underline{S}\Sigma^{-1}\right)\right) \end{aligned} \tag{17}$$

where the first line is the likelihood function (rewritten in multivariate normal form) and the second line is the prior distribution.

We can now find the unknown quantities by matching coefficients. First, from the quadratic and linear terms in β we can find $\overline{\beta}$ and $\overline{V}$. This yields the (conditional)

---

possible to use a different parsimonious benchmark. For example, the so-called combination prior is based on more general unit root processes, see section 3.2.2 for details.

posterior distribution for β

$$\begin{aligned}
\beta|\Sigma, y &\sim N\left(\overline{\beta}, \Sigma \otimes \overline{V}\right) && \text{(18a)}\\
&\text{with} \\
\overline{\beta} &= \text{vec}\left(\overline{B}\right) && \text{(18b)}\\
\overline{B} &= \left(X'X + \underline{V}^{-1}\right)^{-1}\left(X'Y + \underline{V}^{-1}\underline{B}\right) && \text{(18c)}\\
\overline{V} &= \left(X'X + \underline{V}^{-1}\right)^{-1} && \text{(18d)}
\end{aligned}$$

where $\overline{V}$ followed directly from matching the quadratic terms and $\overline{\beta}$ followed from matching the linear terms. For the derivation of the latter, we have also used the reverse of the vectorization rule that $\text{vec}\,(ABC) = (C' \otimes A)\,\text{vec}\,(B)$ and the fact that $X'X\hat{B} = X'Y$.[4]

It should be noted that after matching the quadratic and linear terms in β, there is a discrepancy between the constant terms. The discrepancy, which is below denoted by $\tilde{S}$, as well as one of the two $|\Sigma|^{-\frac{k}{2}}$ terms should be absorbed by the inverse-Wishart posterior distribution for Σ.

Second, from the three remaining determinant terms we can find the degrees-of-freedom parameter $\overline{v}$ and from the trace terms and the discrepancy $\tilde{S}$ (rewritten in matrix form) we can find the scale matrix $\overline{S}$. This yields the (marginal) posterior distribution for Σ

$$\begin{aligned}
\Sigma|y &\sim IW\left(\overline{S}, \overline{v}\right) && \text{(19a)}\\
&\text{with} \\
\overline{S} &= \hat{S} + \underline{S} + \tilde{S} && \text{(19b)}\\
\tilde{S} &= \hat{B}'X'X\hat{B} + \underline{B}'\underline{V}^{-1}\underline{B} - \overline{B}'\overline{V}^{-1}\overline{B} && \text{(19c)}\\
\overline{v} &= T + \underline{v} && \text{(19d)}
\end{aligned}$$

Finally, there are no remaining terms, which implies that the posterior distribution indeed satisfies the normal inverse-Wishart structure given by equations (18) and (19).

Combination prior Now that we know the prior structure (and associated posterior), how do we set the prior details? At the CPB, we are using a combination of three priors

[4] The (conditional) posterior mean for β has the same 'endogenous weighting of prior and data' interpretation as with the Minnesota prior, i.e. the Bayesian shrinkage approach effectively recognizes the robust relationships and replaces the unstable ones. See the last paragraph of section 3.2.1 for a more extensive explanation.

as in Sims and Zha (1998) and Giannone, Lenza, and Primiceri (2013) among others. The various prior components are discussed in turn.

The *baseline prior* is a version of the Minnesota prior, so as to shrink the model towards the random walk. Relative to the original Minnesota prior, we also need to specify a prior for the covariance matrix Σ and furthermore the prior for the vector of regression parameters β has to satisfy the same Kronecker structure as the likelihood function. The priors for Σ and β are discussed in turn.

Regarding Σ, we set the degrees-of-freedom parameter of the inverse-Wishart prior distribution equal to $\underline{\upsilon} = n + 2$, which is the minimum value under which the prior mean exists, implying that the prior for Σ is rather uninformative. The prior mean is equal to $\frac{\underline{S}}{\underline{\upsilon}-n-1}$, which under our choice for $\underline{\upsilon}$ simplifies to $\underline{S}$. We set the scale matrix $\underline{S}$ to a diagonal matrix with on the main diagonal the variances of separately pre-estimated AR(1) models, if necessary rescaled with $\underline{\upsilon} - n - 1$.[5]

Regarding β, the idea behind the Minnesota prior is to shrink the model towards the random walk. Just as in section 3.2.1 we set the prior mean equal to $\underline{\beta} = \text{vec}\left([O_{n\times 1} \;\; I_{n\times n} \;\; O_{n\times n(p-1)}]'\right)$, i.e. $\underline{\beta}$ is simply a vector with all zeros except for the elements corresponding to the first own lags. Note again that the ones can be replaced by zeros or any number if the data enter the model in growth rates as opposed to levels.

With the original Minnesota prior, the prior covariance matrix was chosen so as to imply stronger shrinkage for coefficients on longer lags and across variables. For convenience, here is the rule again

$$\underline{v}_{iq} = \begin{cases} \frac{\pi_1}{r^2} & \text{if } q \neq 1 \text{ and } j = i \text{ (for coefficients on the } r^{\text{th}} \text{ own lags)} \\ \frac{\pi_2}{r^2}\frac{\sigma_{ii}}{\sigma_{jj}} & \text{if } q \neq 1 \text{ and } j \neq i \text{ (for coefficients on the } r^{\text{th}} \text{ foreign lags)} \\ \pi_3\sigma_{ii} & \text{if } q = 1 \text{ (for coefficients on deterministic terms)} \end{cases} \tag{20}$$

where $r = \lfloor \frac{q-2}{n} \rfloor + 1$ and $j = q - 1 - \lfloor \frac{q-2}{n} \rfloor n$. With the natural conjugate prior it is not possible to distinguish between own and foreign lags, because of the Kronecker structure. Therefore, we have to set $\pi_1 = \pi_2$ and choose the same $\underline{v}_q = \underline{v}_{iq}$ for all i. Furthermore, the Kronecker product with Σ already accounts for the σ_{ii} term in the numerator. Yet, we still need to account for the σ_{jj} term in the denominator, for which we use the diagonal elements of the prior scale matrix $\underline{S}_{jj}$. After some reparameterizations, we have the following rule

$$\underline{v}_q = \begin{cases} \frac{\lambda^2}{r^2}\frac{1}{\underline{S}_{jj}/(\underline{\upsilon}-n-1)} & \text{if } q \neq 1 \text{ (for coefficients on the } r^{\text{th}} \text{ lags)} \\ \kappa & \text{if } q = 1 \text{ (for coefficients on deterministic terms)} \end{cases} \tag{21}$$

[5] This is common practice. Although this is not a good Bayesian treatment, it will hardly influence the results given that the prior for Σ is rather uninformative.

where the hyperparameter λ controls the overall tightness of the prior and κ is chosen to be large implying very weak shrinkage for deterministic terms.[6] In section 4, we discuss several approaches how one can choose a good value for the hyperparameter λ.

In addition to the baseline prior, we use two additional priors. The first one favors more general unit root processes and the second one favors cointegration. Both priors can be implemented by using dummy observations (and are therefore conjugate by construction).

The first additional prior is called the *sum-of-coefficients prior* and centers the sum of the coefficients $\sum_{i=1}^{p} B_i$ around the identity matrix I_n. This prior implies a unit root in each VAR equation and is thus inconsistent with cointegration. That is why in the literature the sum-of-coefficients prior is also referred to as the *no cointegration prior*. The prior is implemented with a set of n dummy observations

$$y^+ = \text{diag}\left(\frac{\bar{y}_0}{\mu}\right) \tag{22a}$$

$$x^+ = [O_{n\times 1} \;\; y^+ \;\; \cdots \;\; y^+] \tag{22b}$$

where $\bar{y}_0$ is an $n \times 1$ vector containing the averages of the first p observations with which the VAR model is initialized. For posterior inference, the $n \times n$ matrix y^+ and the $n \times k$ matrix x^+ are added on top of the data matrices Y and X, respectively. The hyperparameter μ controls the prior tightness and if $\mu \to \infty$, the prior becomes uninformative. In section 4, we discuss several approaches how one can choose a good value for the hyperparameter μ.

The second additional prior is called the *single-unit-root prior* and pushes the VAR either in the direction of (i) at least one unit root and zero constant or (ii) no unit root and $\bar{y}_0$ equal to its unconditional mean. The single-unit-root prior is thus consistent with cointegration. In the literature, this prior is also known as the *dummy-initial-observation prior*, although this name is not quite informative. The prior is implemented with a single dummy observation

$$y^{++} = \frac{\bar{y}_0'}{\delta} \tag{23a}$$

$$x^{++} = [\frac{1}{\delta} \;\; y^{++} \;\; \cdots \;\; y^{++}] \tag{23b}$$

For posterior inference, the $1 \times n$ vector y^{++} and the $1 \times k$ vector x^{++} are added on top of the data matrices Y and X, respectively. The hyperparameter δ controls the prior tightness and if $\delta \to \infty$, the prior becomes uninformative. In section 4, we discuss several approaches how one can choose a good value for the hyperparameter δ.

[6] It is quite common to choose $\kappa = 10^7$.

3.2.3 Σ unknown and normal inverse-Wishart prior distribution for $\{\Sigma, \beta\}$, no restrictions

The prior discussed here is called the independent normal inverse-Wishart prior. This prior has a normal inverse-Wishart structure like in section 3.2.2, yet without the Kronecker structure and the conditioning on Σ in the prior for β. The independent normal inverse-Wishart prior is given by

$$\beta \sim N\left(\underline{\beta}, \underline{V}\right) \tag{24a}$$

$$\Sigma \sim IW\left(\underline{S}, \underline{\upsilon}\right) \tag{24b}$$

The joint posterior distribution $p\left(\beta, \Sigma | y\right)$ is not analytically available without the exact same Kronecker structure as in the likelihood function.[7] However, the conditional posterior distributions $p\left(\beta | \Sigma, y\right)$ and $p\left(\Sigma | \beta, y\right)$ are both analytically available, so that we can use Gibbs sampling for posterior inference. It is not hard to implement the Gibbs sampling procedure but it is computationally demanding, which can be cumbersome for e.g. forecast competitions. In section 6, we set out the Gibbs sampling procedure for Bayesian VARs with independent normal inverse-Wishart priors.

4 Choosing the hyperparameters

We are left with the task of choosing good values for the hyperparameters. There are several ways how to do this:

- Use standard values such as the ones recommended by Sims and Zha (1998), namely $\lambda = 0.2$, $\mu = 1$, and $\delta = 1$. This often works well in practice, even though the number of time periods as well as the system size vary considerably between applications.
- Choose the hyperparameters by maximizing the marginal likelihood as proposed by Giannone, Lenza, and Primiceri (2013). We will elaborate on the marginal likelihood in section 4.1.
- Choose the hyperparameters to avoid overfitting as in Bańbura, Giannone, and Reichlin (2010).
- Choose the hyperparameters by running a forecast competition.

[7] An educated guess would be that β follows a normal posterior distribution even without the Kronecker structure in the prior. However, when trying to match coefficients it turns out that it is not possible to match the determinant term of the presumed normal posterior distribution, so that the educated guess must be wrong.

4.1 Marginal likelihood

The marginal likelihood is analytically available for Bayesian VARs with conjugate priors. In this subsection, we only give the formula and refer to Giannone, Lenza, and Primiceri (2013) for the tedious but straightforward derivation.[8] For the case without dummy observations, we have

$$p(y) = \pi^{-\frac{nT}{2}} \frac{\Gamma_n\left(\frac{T+\underline{v}}{2}\right)}{\Gamma_n\left(\frac{\underline{v}}{2}\right)} |\underline{V}|^{-\frac{n}{2}} \left|X'X + \underline{V}^{-1}\right|^{-\frac{n}{2}} |\underline{S}|^{\frac{\underline{v}}{2}} \left|\hat{S} + \underline{S} + \tilde{S}\right|^{-\frac{T+\underline{v}}{2}} \tag{25}$$

where $\Gamma_n(\cdot)$ denotes the multivariate Gamma function. It should be noted that equation (25) is numerically unstable for large systems and one should work with the equivalent but stable expression

$$p(y) = \pi^{-\frac{nT}{2}} \frac{\Gamma_n\left(\frac{T+\underline{v}}{2}\right)}{\Gamma_n\left(\frac{\underline{v}}{2}\right)} \left|I_k + D_{\underline{V}}' X'X D_{\underline{V}}\right|^{-\frac{n}{2}} |\underline{S}|^{-\frac{T}{2}} \left|I_n + D_{\underline{S}}'(\hat{S} + \tilde{S}) D_{\underline{S}}\right|^{-\frac{T+\underline{v}}{2}} \tag{26}$$

where $D_{\underline{V}}$ and $D_{\underline{S}}$ are the lower-triangular parts of the Cholesky decompositions of $\underline{V}$ and $\underline{S}$, respectively. To compute the first determinant term, calculate the eigenvalues of $D_{\underline{V}}' X'X D_{\underline{V}}$, add one to each eigenvalue, and take the product. The third determinant term can be computed similarly.

Finally, we can also use expression (26) for the case with dummy observations. Let y_{ext} denote the extended dataset including both the real and artificial data and let y_{art} denote the dataset including only the artificial data. The marginal likelihood is then simply given by

$$p(y) = \frac{p(y_{\text{ext}})}{p(y_{\text{art}})} \tag{27}$$

5 Forecasting and scenario analysis

Yet another representation is preferred when using Bayesian VARs for forecasting and scenario analysis. In particular, it is convenient to rewrite the VAR in state space form and exploit the extensive toolkit that is available for this representation. The key tool of the state space toolkit is the Kalman filter, which can be used for forecasting under missing observations inter alia.[9]

[8] It should be noted that the notation and representation of the marginal likelihood is slightly different here than in Giannone, Lenza, and Primiceri (2013).

[9] It should be noted that this section takes as given a point estimate of the VAR parameters, estimated on the largest rectangular block of data without missing observations. Note furthermore that

In subsection 5.1, we derive the VAR's measurement and state equation, which together characterize the state space form. In subsection 5.2, we describe the Kalman filter equations and show how to apply the Kalman filter equations for forecasting under missing observations (which is crucial for practical work where datasets quite often have a ragged-edge pattern).

5.1 State space representation for VAR models

For a general treatment of state space models, see, for example, the excellent book by Durbin and Koopman (2012).[10] Here, we only give a (non-unique) state space representation for VAR models. The VAR's measurement and state equation are given by

$$y_t = Z\alpha_t \tag{28a}$$

$$\alpha_{t+1} = C + T\alpha_t + R\eta_t \qquad \text{with } \eta_t \sim N(0, Q) \tag{28b}$$

The first equation is the measurement equation and the second equation is the state equation. The state equation is simply the VAR's companion form.[11] The $np \times 1$ state vector is constructed as $\alpha_t = [\{y'_{t-i}\}_{i=0}^{p-1}]'$. In order to conform to the notation of Durbin and Koopman (2012), the $n \times 1$ vector with state innovations is time shifted by one period, that is $\eta_t = u_{t+1}$. Below we describe how the various system matrices look like for the general VAR(p) case, while in appendix A we give an illustration for the VAR(3) case.

Regarding the state equation, the $np \times 1$ vector with constants is constructed as $C = [C_1' \;\; C_2']'$ where $C_1 = c$ and $C_2 = O_{n(p-1)\times 1}$ and the $np \times np$ matrix with autoregressive parameters is constructed as $T = [T_1' \;\; T_2']'$ where $T_1 = [\{B_i\}_{i=1}^{p}]$ and $T_2 = [I_{n(p-1)\times n(p-1)} \;\; O_{n(p-1)\times n}]$. Note that c and $[\{B_i\}_{i=1}^{p}]$ are simply the first and remaining columns of B', respectively. Furthermore, the $np \times n$ selection matrix R is constructed as $R = [I_{n\times n} \;\; O_{n\times n(p-1)}]'$ and the covariance matrix of the state innovations is simply $Q = \Sigma$. Regarding the measurement equation, the $n \times np$ matrix Z is constructed as $Z = [I_{n\times n} \;\; O_{n\times n(p-1)}]$.

---

it is straightforward to run a Monte Carlo simulation based on the posterior distribution of the VAR parameters.

[10] Since we wanted to follow the notation of Durbin and Koopman (2012) as much as possible, we have recycled some notation in this section.

[11] It is always possible to rewrite a VAR(p) model with n variables as a restricted VAR(1) model with np variables. The latter representation is called the companion form.

5.2 Missing observations and the Kalman filter

For the state space representation given in equation (28), the Kalman filter recursions are given by

$$\begin{aligned} \upsilon_t &= y_t - Z a_t && \text{(29a)} \\ F_t &= Z P_t Z' && \text{(29b)} \\ K_t &= T P_t Z' F_t^{-1} && \text{(29c)} \\ a_{t+1} &= C + T a_t + K_t \upsilon_t && \text{(29d)} \\ P_{t+1} &= T P_t T' + RQR' - K_t F_t K_t' && \text{(29e)} \end{aligned}$$

where υ_t is the prediction error, F_t is the prediction error variance, K_t is the Kalman gain, $a_t = \mathrm{E}\left(\alpha_t | Y^{t-1}\right)$ is the filtered state, and $P_t = \mathrm{Var}\left(\alpha_t | Y^{t-1}\right)$ is the filtered state variance. The notation Y^τ is used to denote the history of y_t up to and including time period τ. The Kalman filter recursions are initialized by

$$\begin{aligned} a_1 &= C + T\alpha_0 && \text{(30a)} \\ P_1 &= RQR' && \text{(30b)} \end{aligned}$$

where α_0 consists of the first p observations with which the VAR model is initialized.

In practical forecasting applications, we often need to work with a ragged-edge dataset and/or with other types of missing observations. In the presence of missing observations, the measurement equation needs to be adjusted, that is

$$W_t y_t = W_t Z \alpha_t \qquad \text{(31)}$$

where W_t is a known matrix that depends on the pattern of the missing observations. In particular, the rows of W_t are a subset of the rows of I_n, i.e. the i^{th} row of I_n is included in W_t if and only if the i^{th} element of y_t is not missing. So $W_t = I_n$ if y_t is fully observed, whereas W_t is an empty $0 \times n$ matrix if y_t is fully missing.[12] In the presence of missing observations, the Kalman filter recursions are given by

$$\begin{aligned} \upsilon_t &= W_t y_t - W_t Z a_t && \text{(32a)} \\ F_t &= W_t Z P_t Z' W_t' && \text{(32b)} \\ K_t &= T P_t Z' W_t' F_t^{-1} && \text{(32c)} \\ a_{t+1} &= C + T a_t + K_t \upsilon_t && \text{(32d)} \\ P_{t+1} &= T P_t T' + RQR' - K_t F_t K_t' && \text{(32e)} \end{aligned}$$

[12] Standard matrix algebra can also be used for empty matrices.

which is simply expression (29) yet with Z replaced by $W_t Z$ and y_t replaced by $W_t y_t$. Apart from this, the formulae are unchanged as well as the initialization.

After running the forward recursions of the Kalman filter, we also need to run the *backward* recursions of the Kalman smoother so as to back and nowcast the missing observations. The Kalman smoother delivers the smoothed state $\hat{\alpha}_t = \mathrm{E}(\alpha_t|Y)$ as well as the smoothed state variance $V_t = \mathrm{Var}(\alpha_t|Y)$.[13] The backward recursions of the Kalman smoother are given by

$$r_{t-1} = Z' W_t' F_t^{-1} \upsilon_t + (T - K_t W_t Z)' r_t \tag{33a}$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \tag{33b}$$

$$N_{t-1} = Z' W_t' F_t^{-1} W_t Z + (T - K_t W_t Z)' N_t (T - K_t W_t Z) \tag{33c}$$

$$V_t = P_t - P_t N_{t-1} P_t \tag{33d}$$

where r_t and N_t are auxiliary variables. The backward recursions of the Kalman smoother are initialized by $r_T = 0$ and $N_T = 0$.

It is straightforward to use the Kalman filter and smoother for forecasting by treating future observations as missing. In particular, the point forecast for y_τ is simply equal to $Z\hat{\alpha}_\tau$. It should be noted that the Kalman filter and smoother approach delivers the exact same forecasts as the standard iterative approach when there are no missing observations.

Furthermore, the missing observation approach can also be used for scenario analysis, e.g. for conditioning on a future path of anticipated observations on one of the variables of the VAR model. Similarly, by leaving out the last data point of a variable it is possible to determine the contribution of that variable to the forecasts for other variables.

6 Gibbs sampling procedure for non-conjugate Bayesian VARs

For a general treatment of Gibbs sampling, see, for example, Casella and George (1992). Here, we only explain how to implement the Gibbs sampling procedure for Bayesian VARs with independent normal inverse-Wishart priors (see section 3.2.3 for details about this prior). The idea behind the Gibbs sampler is to iterate between the conditional posterior distributions, that is

0. Initialize the Gibbs sampler with β^0 and set i to 1.

[13] Note that the smoothed estimate of α_τ is based on the entire history of y_t, whereas the filtered estimate of α_τ is just based on the history of y_t up to and including time period $\tau - 1$.

1. *Draw a covariance matrix* Σ^i *from* $p(\Sigma|\beta^{i-1}, y)$.

2. *Draw a vector of regression parameters* β^i *from* $p(\beta|\Sigma^i, y)$.

3. *Increment* i *by* 1 *and go back to step 1 until* $i > g$.

4. *Throw away the first* b *draws as burn-in and keep the remaining* $g - b$ *draws for posterior inference.*

Subject to regularity conditions (see, for example, Roberts and Smith, 1994), the Gibbs sampler generates a Markov chain with $p(\beta, \Sigma|y)$ as stationary distribution. It is very important to check convergence of the Markov chain, e.g. by comparing the results of various independent Markov chains that have started from different initial conditions.

In the remainder of this section, we will derive the conditional posterior distributions $p(\Sigma|\beta, y)$ and $p(\beta|\Sigma, y)$, which are the key ingredients for the Gibbs sampling procedure.

6.1 Conditional posterior distribution $p(\Sigma|\beta, y)$

We need to combine the likelihood function with the prior distribution and collect the relevant terms for the conditional posterior distribution $p(\Sigma|\beta, y)$. Regarding the likelihood function, it is most convenient to start off from the matrix normal form, i.e. equation (4b), which for convenience is repeated here

$$p(y|\beta, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \operatorname{tr}\left((Y - XB)'(Y - XB)\Sigma^{-1}\right)\right) \tag{34}$$

It should be noted that in the current Gibbs step $\beta = \operatorname{vec}(B)$ is considered to be known, so that the above likelihood function has already an inverse-Wishart structure for Σ.

In case of the independent normal inverse-Wishart prior, the prior distribution for Σ is given by

$$\Sigma \sim IW(\underline{S}, \underline{\upsilon}) \tag{35}$$

or in terms of the probability density function

$$p(\Sigma) \propto |\Sigma|^{-\frac{\underline{\upsilon}+n+1}{2}} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\underline{S}\Sigma^{-1}\right)\right) \tag{36}$$

Multiplying the likelihood function (34) with the prior distribution (36), yields the conditional posterior distribution

$$p(\Sigma|\beta, y) \propto |\Sigma|^{-\frac{T+\underline{\upsilon}+n+1}{2}} \exp\left(-\frac{1}{2} \operatorname{tr}\left((S + \underline{S})\Sigma^{-1}\right)\right) \tag{37}$$

where $S = (Y - XB)'(Y - XB)$. This is simply an inverse-Wishart distribution with degrees-of-freedom parameter $T + \underline{\upsilon}$ and scale matrix $S + \underline{S}$, that is

$$\Sigma|\beta, y \sim IW\left(S + \underline{S}, T + \underline{\upsilon}\right) \tag{38}$$

6.2 Conditional posterior distribution $p(\beta|\Sigma, y)$

The Gibbs step discussed here follows exactly the case with the original Minnesota prior in which Σ is assumed to be known. Therefore, we just include the conditional posterior distribution $p(\beta|\Sigma, y)$ and refer to section 3.2.1 for the derivation. Analogue to expression (13), we have the conditional posterior distribution

$$\beta|\Sigma, y \sim N\left(\overline{\beta}, \overline{V}\right) \tag{39a}$$

with

$$\overline{\beta} = \left(\hat{V}^{-1} + \underline{V}^{-1}\right)^{-1}\left(\hat{V}^{-1}\hat{\beta} + \underline{V}^{-1}\underline{\beta}\right) \tag{39b}$$

$$\overline{V} = \left(\hat{V}^{-1} + \underline{V}^{-1}\right)^{-1} \tag{39c}$$

where $\hat{\beta}$ is the ordinary least squares estimate of β and $\hat{V} = \Sigma \otimes (X'X)^{-1}$.

7 Applications

7.1 Dutch economy

At the CPB, we have developed a Bayesian VAR model for the Dutch economy. This instrument is used to benchmark the forecasts that come from CPB's macro-econometric model Saffier II and serves as input for the expert opinion that is part of the official CPB quarterly projections (i.e. the Central Economic Plan and Macro Economic Outlook as well as the June and December projections).

The Bayesian VAR instrument is used to generate forecasts for a wide range of variables, including GDP, macro aggregates, and prices. These forecasts are compared to the forecasts that come from the macro-econometric model Saffier II. We also run various decomposition exercises in which we determine the contribution of (blocks of) variables to the forecasts for other variables. To determine the contribution of a block of variables we leave out the last data point of each variable in that block and look how this impacts the forecasts for mainly GDP and the macro aggregates. This is implemented using the missing observation approach, see section 5 for details. Furthermore, we also analyze which variables contribute to the changes in the forecasts

relative to the forecasts made in the previous quarter. The insights from these exercises are used as input for expert opinion.

We use quarterly data from 2001-Q1 onwards and, as discussed in section 5, we can handle ragged-edge patterns at the end of the sample so that we can exploit the most recent information of various indicators.[14] Our dataset consists of a wide range of variables that cover the full spectrum of the economy, i.e. GDP and macro aggregates (in real terms), prices and wages, indicators on consumer confidence, indicators on business confidence, international variables, and financial variables. The full list of 22 variables is given in Table 1. We would like to emphasize that Bayesian VARs are well suited to include a large set of variables. As explained in section 3.2.1, the Bayesian shrinkage approach effectively recognizes the robust relationships in the data and replaces the unstable relationships by a parsimonious benchmark that is known to have decent forecasting performance (e.g. the random walk with drift). This makes the Bayesian VAR framework very appealing since we do not have to worry about a variable selection procedure.[15]

We include four lags in our quarterly model. The data enter the model in levels, as is standard for Bayesian VARs, and the non-stationary variables are log-transformed. We use the so-called combination prior, which is superior in terms of forecasting performance relative to the Minnesota prior.[16] The key difference between these priors is that the Minnesota prior shrinks the VAR towards univariate random walks, whereas the combination prior shrinks towards more flexible unit root processes and favors cointegration (which likely is important for macroeconomic forecasting).[17] Finally, for the hyperparameters of the prior distribution we use the values that are recommended by Sims and Zha (1998), that is $\lambda = 0.2$, $\mu = 1$, and $\delta = 1$.[18]

---

[14] Due to the 2014 revision of the national accounts data by Statistics Netherlands, the sample period only starts in 2001-Q1, but data from 1995-Q1 onwards will be published shortly.

[15] For example, Bańbura, Giannone, and Reichlin (2010) have shown that a Bayesian VAR with 131 variables outperforms a VAR with 3 variables, although the same forecasting performance is already achieved with 20 variables.

[16] We use the mean squared prediction error to evaluate the out-of-sample forecasting performance, yet the same conclusion would be drawn based on the mean absolute error.

[17] See section 3.2 for details about the various priors.

[18] We have also experimented with the maximum marginal likelihood procedure of Giannone, Lenza, and Primiceri (2013) with similar forecasting performance as a result.

Table 1 List of variables used in Dutch application

Variable description	Block of economy	Data source
GDP	Macro	CBS
Private consumption	Macro	CBS
Government consumption	Macro	CBS
Investment (private and government)	Macro	CBS
Exports of goods and services	Macro	CBS
Imports of goods and services	Macro	CBS
CPI	Prices	CBS
Wages	Prices	CBS
Consumer confidence	Consumer	CBS
Economic climate (part of consumer confidence)	Consumer	CBS
Willingness to buy (part of consumer confidence)	Consumer	CBS
Expected business activity industry	Business	CBS
Actual business activity industry	Business	CBS
Utilization rate industry	Business	CBS
Bankruptcies	Business	CBS
World trade (trade weighted for Netherlands)	International	CPB-WTM
OECD composite leading indicator euro area	International	OECD
OECD composite leading indicator US	International	OECD
AEX stock index	Financial	DNB
M1 money supply	Financial	DNB
Short interest rate	Financial	DNB
House prices	Financial	DNB

Data sources: Statistics Netherlands (CBS), CPB World Trade Monitor (CPB-WTM), OECD, and De Nederlandsche Bank (DNB). Where applicable, the data are seasonally adjusted at their source.

7.2 World trade

At the CPB, we have also developed a Bayesian VAR model for world trade. It plays a similar role as the model for the Dutch economy. In particular, the Bayesian VAR forecasts serve as a source of expert opinion for the international projections made by the CPB.

We currently use the model only for forecasting world trade, although potentially we could look at a wider range of variables. Moreover, we run the same decomposition exercises as with the model for the Dutch economy, i.e. horizontal analyses to determine

the contribution of blocks of variables to the world trade forecast and vertical analyses to understand why the world trade forecast has changed relative to the previous forecast.

The data for the world trade application are available on a monthly basis from 1991-M1 onwards. The dataset suffers from a ragged-edge pattern at the end of the sample, but with the missing observation approach, as discussed in section 5, we can exploit the most recent information. Our dataset consists of a wide range of variables including goods trade volumes and prices, industrial production, retail trade volumes, composite leading indicators, purchasing managers' indices, and various other variables. The full list of 22 variables is given in Table 2.

We include six lags in our monthly model. We follow the exact same Bayesian VAR settings as with the model for the Dutch economy, see the last paragraph of section 7.1 for details.

A State space representation for VAR(3) model

Regarding the measurement equation, we have

$$Z = \begin{pmatrix} I_{n\times n} & O_{n\times n} & O_{n\times n} \end{pmatrix} \tag{40}$$

and regarding the state equation, we have

$$C = \begin{pmatrix} c \\ O_{n\times 1} \\ O_{n\times 1} \end{pmatrix} \qquad T = \begin{pmatrix} B_1 & B_2 & B_3 \\ I_{n\times n} & O_{n\times n} & O_{n\times n} \\ O_{n\times n} & I_{n\times n} & O_{n\times n} \end{pmatrix} \qquad R = \begin{pmatrix} I_{n\times n} \\ O_{n\times n} \\ O_{n\times n} \end{pmatrix} \qquad Q = \Sigma \tag{41}$$

Table 2 List of variables used in world trade application

Variable description	Block of economy	Data source
World trade	Goods trade volumes	CPB-WTM
Imports (advanced economies)	Goods trade volumes	CPB-WTM
World trade price	Goods trade prices	CPB-WTM
Import price (world)	Goods trade prices	CPB-WTM
Import price (advanced economies)	Goods trade prices	CPB-WTM
Manufactures	Goods trade prices	CPB-WTM
Fuels (HWWI)	Goods trade prices	CPB-WTM
Primary commodities excl. fuels (HWWI)	Goods trade prices	CPB-WTM
Industrial production excl. construction (world)	Industrial production	CPB-WTM
Industrial production excl. construction (adv. economies)	Industrial production	CPB-WTM
Retail trade (OECD countries)	Retail trade volumes	OECD
Retail trade (euro area)	Retail trade volumes	OECD
Composite leading indicator (OECD countries)	Composite leading indicators	OECD
Composite leading indicator (China)	Composite leading indicators	OECD
Ifo business climate (Germany)	Purchasing managers' indices	CESifo
ISM manufacturing (US)	Purchasing managers' indices	FRED
Brent oil price	Various	Datastream
Baltic exchange dry index	Various	Datastream
MSCI world index	Various	Datastream
World steel production	Various	Datastream
World semiconductor billings	Various	WSTS
Tech pulse index	Various	FRBSF

Data sources: CPB World Trade Monitor (CPB-WTM), OECD, CESifo Group Munich (CESifo), Federal Reserve Economic Data (FRED), Datastream, World Semiconductor Trade Statistics (WSTS), and Federal Reserve Bank of San Francisco (FRBSF). Where applicable, the data are seasonally adjusted at their source.

References

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press.

Giannone, D., Lenza, M., and Primiceri, G. E. (2013). Prior selection for vector autoregressions. *Review of Economics and Statistics*, forthcoming.

Koop, G. and Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends(R) in Econometrics*, 3(4):267–358.

Roberts, G. O. and Smith, A. F. M. (1994). Simple conditions for the convergence of the gibbs sampler and metropolis-hastings algorithms. *Stochastic Processes and their Applications*, 49(2):207–216.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968.

Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T +31 70 3383 380

info@cpb.nl | www.cpb.nl

February 2015