

**ONLINE APPENDIX TO: MARGINALIZED PREDICTIVE LIKELIHOOD
COMPARISONS WITH APPLICATIONS TO DSGE, DSGE-VAR, AND VAR
MODELS**

ANDERS WARNE, GÜNTER COENEN AND KAI CHRISTOFFEL*

APRIL 3, 2013

APPENDIX A: POSTERIOR PROPERTIES OF THE RANDOM WALK MODEL

The purpose of this Appendix is to provide technical details on the predictive density of the random walk model with a standard diffuse prior on the residual covariance matrix. An analytical expression of the predictive density is derived and its mean vector and covariance matrix are also determined.

To these ends, let

$$y_t = y_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \quad (\text{A.1})$$

where the residuals ε_t are assumed to be i.i.d. $N(0, \Omega)$ with Ω positive definite and y_0 is fixed. The diffuse prior is given by

$$p(\Omega) \propto |\Omega|^{-(n+1)/2}. \quad (\text{A.2})$$

Stacking the model in (A.1) into $n \times T$ matrices $y = [y_1 \cdots y_T]$ and $\varepsilon = [\varepsilon_1 \cdots \varepsilon_T]$, with the realized values, for convenience, being denoted the same way, the posterior distribution is proportional to the prior times the likelihood, which in natural logarithms can be expressed as

$$\log p(y|y_0, \Omega) + \log p(\Omega) = -\frac{nT}{2} \log(2\pi) - \frac{T+n+1}{2} \log |\Omega| - \frac{1}{2} \text{tr}[\Omega^{-1} \varepsilon \varepsilon']. \quad (\text{A.3})$$

Recognizing that the last two terms on the right hand side of (A.3) form the log of the kernel of the n -dimensional inverted Wishart distribution with location matrix $\varepsilon \varepsilon'$ and T degrees of freedom, we obtain

$$\begin{aligned} \log p(\Omega|y, y_0) = & -\frac{nT}{2} \log(2) - \frac{n(n-1)}{4} \log(\pi) - \log \Gamma_n(T) + \frac{T}{2} \log |\varepsilon \varepsilon'| \\ & - \frac{T+n+1}{2} \log |\Omega| - \frac{1}{2} \text{tr}[\Omega^{-1} \varepsilon \varepsilon'], \end{aligned} \quad (\text{A.4})$$

where

$$\Gamma_n(T) = \prod_{i=1}^n \Gamma([T-i+1]/2),$$

* All authors: Directorate General Research, European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany; e-mail: anders.warne@ecb.europa.eu, gunter.coenen@ecb.europa.eu, kai.christoffel@ecb.europa.eu.

for $T \geq n > 0$ with $\Gamma(\cdot)$ being the gamma function. From Bayes theorem it therefore follows that the log marginal likelihood is given by (A.3) minus (A.4), i.e.

$$\log p(y|y_0) = -\frac{n(2T-n+1)}{4} \log(\pi) + \log \Gamma_n(T) - \frac{T}{2} \log |\varepsilon\varepsilon'|. \quad (\text{A.5})$$

NORMAL APPROXIMATION OF MARGINAL PREDICTIVE LIKELIHOOD

When forecasting with the random walk model it holds that

$$E[y_{T+h}|y, y_0, \Omega] = y_T, \quad h = 1, \dots, h^*. \quad (\text{A.6})$$

The forecast error is therefore equal to the accumulation of ε_{T+i} over $i = 1, \dots, h$, while the forecast error covariance matrix given Ω is

$$C(y_{T+h}|y, y_0, \Omega) = h\Omega, \quad h = 1, \dots, h^*. \quad (\text{A.7})$$

From Rao-Blackwellization we know that the covariance matrix $C(y_{T+h}|y, y_0)$ is equal to the mean of the covariance matrix in (A.7) with respect to the posterior of Ω plus the covariance matrix of the deviation of the mean in (A.6) and its population mean $E[y_{T+h}|y, y_0]$. The latter term is zero since the population mean is also y_T , while the former term is given by h times the mean of the posterior of Ω .¹ That is,

$$C(y_{T+h}|y, y_0) = \frac{h}{T-n-1} \varepsilon\varepsilon'. \quad (\text{A.8})$$

When computing the marginal predictive likelihood with a normal approximation for the full system we therefore make use of the realized forecast errors $y_{T+h} - y_T$ and the covariance matrix in (A.8).

When forecasting only a subset of the variables we need to take into account how the posterior distribution for the covariance matrix of the corresponding subset of residuals is related to the posterior $p(\Omega|y, y_0)$. Let S be an $n \times n_s$ matrix of columns from I_n which selects $y_{s,t} = S'y_t$. Similarly, let S_\perp the the $n \times (n - n_s)$ matrix which selects the remaining variables from the y_t vector. Define

$$M = [S \ S_\perp], \quad (\text{A.9})$$

i.e. M is an $n \times n$ matrix made up of all the columns of the identity matrix and therefore has a unit determinant while $M^{-1} = M'$. The posterior distribution of $\Omega_M = M'\Omega M$ is an n -dimensional inverted Wishart with location matrix $M'\varepsilon\varepsilon'M$ and T degrees of freedom. Letting $\Omega_S = S'\Omega S$, it follows from, e.g., Bauwens, Lubrano, and Richard (1999, Theorem A.17) that the posterior of Ω_S is an n_s -dimensional inverted Wishart with location matrix $S'\varepsilon\varepsilon'S$ and $T - n + n_s$ degrees of freedom.

With this in mind, the normal approximation of the marginal predictive likelihood for the subset of variables is based on the mean forecast error $y_{s,T+h} - y_{s,T}$ and the population covariance

¹ More generally, the posterior distribution of $h\Omega$ is inverted Wishart with location parameter $h\varepsilon\varepsilon'$ and T degrees of freedom.

matrix

$$C(y_{s,T+h}|y, y_0) = \frac{h}{T-n-1} S' \varepsilon \varepsilon' S. \quad (\text{A.10})$$

ANALYTICAL FORM OF THE MARGINAL PREDICTIVE LIKELIHOOD

The determination of the marginal predictive likelihood requires an expression for the conditional likelihood function $p(y_{s,T+h}|y, y_0, \Omega)$. From equation (20) and using $y_{T+h|T} = y_T$ and $\Sigma_{y,T+h|T} = h\Omega$ we find that the conditional log-likelihood for the random walk model is given by

$$\log p(y_{s,T+h}|y, y_0; \Omega) = -\frac{n_s}{2} \log(2\pi h) - \frac{1}{2} \log |\Omega_S| - \frac{1}{2h} \text{tr} [\Omega_S^{-1} \varepsilon_{s,T,h} \varepsilon'_{s,T,h}], \quad (\text{A.11})$$

where $\varepsilon_{s,T,h} = y_{s,T+h} - y_{s,T}$, and the term involving $\log(h)$ is due to $|h\Omega_S| = h^{n_s} |\Omega_S|$.

The product of the conditional likelihood of $y_{s,T+h}$ and the posterior of Ω_S is given by:

$$\begin{aligned} p(y_{s,T+h}, \Omega_S | y, y_0) &= \frac{|S' \varepsilon \varepsilon' S|^{(T-n+n_s)/2}}{(2\pi h)^{n_s/2} 2^{(T-n+n_s)n_s/2} \pi^{n_s(n_s-1)/4} \Gamma_{n_s}(T-n+n_s)} \times \\ &\times |\Omega_S|^{-(T-n+2n_s+2)/2} \exp \left[-\frac{1}{2} \text{tr} (\Omega_S^{-1} [S' \varepsilon \varepsilon' S + h^{-1} \varepsilon_{s,T,h} \varepsilon'_{s,T,h}]) \right]. \end{aligned} \quad (\text{A.12})$$

Recognizing that the two terms involving Ω_S is the kernel of an n_s -dimensional inverted Wishart distribution with location matrix $S' \varepsilon \varepsilon' S + h^{-1} \varepsilon_{s,T,h} \varepsilon'_{s,T,h}$ and $T-n+n_s+1$ degrees of freedom, it follows that the integral of the density $p(y_{s,T+h}, \Omega_S | y, y_0)$ with respect to Ω_S is equal to the expression in the first term on the right hand side of equation (A.12) times the inverse of the integration constant of the $IW_{n_s}(S' \varepsilon \varepsilon' S + h^{-1} \varepsilon_{s,T,h} \varepsilon'_{s,T,h}, T-n+n_s+1)$ distribution. We therefore find that

$$p(y_{s,T+h}|y, y_0) = \frac{\Gamma_{n_s}(T-n+n_s+1) |h S' \varepsilon \varepsilon' S|^{-1/2}}{\pi^{n_s/2} \Gamma_{n_s}(T-n+n_s) |I_{n_s} + (h S' \varepsilon \varepsilon' S)^{-1} \varepsilon_{s,T,h} \varepsilon'_{s,T,h}|^{(T-n+n_s+1)/2}}. \quad (\text{A.13})$$

In other words (and as expected), the density of $y_{s,T+h}|y, y_0$ is an n_s -dimensional t -distribution with mean $y_{s,T}$, covariance matrix given in (A.10), and $T-n+n_s$ degrees of freedom; see, e.g., Bauwens et al. (1999, Appendix A) for details.²

APPENDIX B: POSTERIOR PROPERTIES OF THE LARGE BVAR MODEL

The large BVAR is estimated with the methodology suggested in Bańbura, Giannone, and Reichlin (2010) and therefore relies on using dummy observations when implementing the normal-inverted Wishart version of the Minnesota prior. Below we will first present the prior and posterior distribution and thereafter show the relation between the prior parameters and the T_d dummy observations; see also Lubik and Schorfheide (2006).

² Notice also that $|I_{n_s} + (h S' \varepsilon \varepsilon' S)^{-1} \varepsilon_{s,T,h} \varepsilon'_{s,T,h}| = 1 + \varepsilon'_{s,T,h} (h S' \varepsilon \varepsilon' S)^{-1} \varepsilon_{s,T,h}$; see, e.g., Magnus and Neudecker (1988, Proof of Theorem 1.9).

The VAR representation of y_t is given in equation (24), with $\epsilon_t \sim N_n(0, \Omega)$. Stacking the VAR system as $y = [y_1 \cdots y_T]$, $X = [X_1 \cdots X_T]$, and $\epsilon = [\epsilon_1 \cdots \epsilon_T]$, the log-likelihood is given by

$$\log p(y|X_1; \Phi, \Omega) = -\frac{nT}{2} \log(2\pi) - \frac{T}{2} \log |\Omega| - \frac{1}{2} \text{tr}[\Omega^{-1} \epsilon \epsilon'], \quad (\text{B.1})$$

where, for convenience, we use the same notation for the random variables as their realizations.

The normal-inverted Wishart prior for (Φ, Ω) is given by

$$\text{vec}(\Phi) | \Omega \sim N_{n(np+1)}(\text{vec}(\Phi_\mu), [\Omega_\Phi \otimes \Omega]), \quad (\text{B.2})$$

$$\Omega \sim IW_n(A, v). \quad (\text{B.3})$$

This means that the sum of the log-likelihood and the log prior is given by

$$\begin{aligned} \log p(y, \Phi, \Omega | X_1) &= -\frac{n(T + np + 1)}{2} \log(2\pi) - \frac{nv}{2} \log(2) - \frac{n(n-1)}{4} \log(\pi) \\ &\quad - \log \Gamma_n(v) - \frac{n}{2} \log |\Omega_\Phi| + \frac{v}{2} \log |A| \\ &\quad - \frac{T + n(p+1) + v + 2}{2} \log |\Omega| \\ &\quad - \frac{1}{2} \text{tr} \left[\Omega^{-1} \left(\epsilon \epsilon' + A + (\Phi - \Phi_\mu) \Omega_\Phi^{-1} (\Phi - \Phi_\mu)' \right) \right]. \end{aligned} \quad (\text{B.4})$$

Using standard ‘‘Zellner’’ algebra, it is straightforward to show that

$$\epsilon \epsilon' + A + (\Phi - \Phi_\mu) \Omega_\Phi^{-1} (\Phi - \Phi_\mu)' = (\Phi - \bar{\Phi}) (XX' + \Omega_\Phi^{-1}) (\Phi - \bar{\Phi})' + S, \quad (\text{B.5})$$

where

$$\begin{aligned} \bar{\Phi} &= (yX' + \Phi_\mu \Omega_\Phi^{-1}) (XX' + \Omega_\Phi^{-1})^{-1}, \\ S &= yy' + A + \Phi_\mu \Omega_\Phi^{-1} \Phi_\mu' - \bar{\Phi} (XX' + \Omega_\Phi^{-1}) \bar{\Phi}'. \end{aligned}$$

Substituting for (B.5) in (B.4), we find that the conjugate normal-inverted Wishart prior gives us a normal posterior for $\Phi | \Omega$ and an inverted Wishart marginal posterior of Ω . Specifically,

$$\text{vec}(\Phi) | \Omega, y, X_1 \sim N_{n(np+1)}(\text{vec}(\bar{\Phi}), [(XX' + \Omega_\Phi^{-1})^{-1} \otimes \Omega]), \quad (\text{B.6})$$

$$\Omega | y, X_1 \sim IW_n(S, T + v). \quad (\text{B.7})$$

Combining these posterior results with equations (B.4) and (B.5) it follows that the log marginal likelihood is given by

$$\begin{aligned} \log p(y | X_1) &= -\frac{nT}{2} \log(\pi) + \log \Gamma_n(T + v) - \log \Gamma_n(v) - \frac{n}{2} \log |\Omega_\Phi| \\ &\quad + \frac{v}{2} \log |A| - \frac{n}{2} \log |XX' + \Omega_\Phi^{-1}| - \frac{T + v}{2} \log |S|. \end{aligned} \quad (\text{B.8})$$

The prior in (B.2) and (B.3) can be implemented through $T_d = n(p + 2) + 1$ dummy observations by prepending the y ($n \times T$) and X ($(np + 1) \times T$) matrices with the following:

$$\begin{aligned} y_{(d)} &= \begin{bmatrix} \lambda_o^{-1} \text{diag}[\delta \odot \omega] & 0_{n \times n(p-1)} & \text{diag}[\omega] & 0_{n \times 1} & \tau^{-1} \text{diag}[\delta \odot \mu] \end{bmatrix} \\ X_{(d)} &= \begin{bmatrix} 0_{1 \times np} & 0_{1 \times n} & \gamma^{-1} & 0_{1 \times n} \\ \lambda_o^{-1} (j_p \otimes \text{diag}[\omega]) & 0_{np \times n} & 0_{np \times 1} & \tau^{-1} (\iota_p \otimes \text{diag}[\mu]) \end{bmatrix}. \end{aligned} \quad (\text{B.9})$$

The vector ι_p is a p -dimension unit vector, while the $p \times p$ matrix $j_p = \text{diag}[1 \cdots p]$. The hyperparameter $\lambda_o > 0$ gives the overall tightness in the Minnesota prior, the cross-equation tightness is set to unity, while the harmonic lag decay hyperparameter is equal to 2. The hyperparameter $\tau > 0$ handles shrinkage for the sum of coefficients prior on $(I_n - \sum_{i=1}^p \Phi_i)$, where $\tau \rightarrow 0$ means that the prior on the sum of the lag coefficients approach the case of exact differences, and where shrinkage decreases as τ becomes larger. The n -dimensional vector δ gives the prior mean of the diagonal of Φ_1 , ω is a vector of scale parameters for the residuals ϵ_{it} , while μ is a vector that reflects the mean of y_{it} . Finally, γ reflects the overall tightness on Φ_0 .

The relationship between the dummy observations and the prior parameters $(\Phi_\mu, \Omega_\Phi, A, v)$ are:

$$\begin{aligned} \Phi_\mu &= y_{(d)} X'_{(d)} \left(X_{(d)} X'_{(d)} \right)^{-1}, & \Omega_\Phi &= \left(X_{(d)} X'_{(d)} \right)^{-1}, \\ A &= (y_{(d)} - \Phi_\mu X_{(d)}) (y_{(d)} - \Phi_\mu X_{(d)})', & v &= T_d - (np + 1) + 2. \end{aligned}$$

This guarantees that the prior mean of Ω exists. Letting $y_\star = [y_{(d)} \ y]$ and $X_\star = [X_{(d)} \ X]$, it follows that the posterior parameters

$$\begin{aligned} \bar{\Phi} &= y_\star X'_\star \left(X_\star X'_\star \right)^{-1}, \\ X X' + \Omega_\Phi^{-1} &= X_\star X'_\star, \\ S &= (y_\star - \bar{\Phi} X_\star) (y_\star - \bar{\Phi} X_\star)'. \end{aligned}$$

In the empirical application, $\tau = 10\lambda_o$, i.e. a relatively loose prior on the sum of the autoregressive matrices. The hyperparameters $\delta_i = 0$ if y_{it} is a first differenced variable and $\delta_i = 1$ when y_{it} is a levels variable. The scale parameters ω_i is given by the within-sample residual standard deviation from an AR(p) model for y_{it} , while μ_i is equal to the within-sample mean of y_{it} . The parameter $\varsigma = \gamma^{-1}$ is set to a very small number, which takes care of having an improper prior on Φ_0 .

The formula suggested by Bańbura et al. (2010) for selecting λ_o can be expressed as

$$\bar{\lambda}_o(\phi) = \arg \min_{\lambda_o} \left| \phi - \frac{1}{q} \sum_{j=1}^q \frac{\sigma_j^2(\lambda_o)}{\sigma_j^2(0)} \right|,$$

where $\phi \in (0, 1)$ is the desired fit, and $\sigma_j^2(\tilde{\lambda}_o)$ is the one-step-ahead mean square forecast error of variable j when $\lambda_o = \tilde{\lambda}_o$. The one-step-ahead within-sample mean square forecast errors used in

the selection scheme are based on the sample 1985Q1–1998Q4. With $\phi = 0.5$, $q = 3$ using real GDP growth, the GDP deflator, and the short-term nominal interest rate, this selection scheme sets $\bar{\lambda}_o = 0.0693$ when $p = 4$.

It should be noted that having an improper prior on Φ_0 technically means that Ω_Φ is singular. This needs to be taken into account when computing, e.g., the log marginal likelihood in (B.8). To deal with this, let

$$X = \begin{bmatrix} \iota'_T \\ Y \end{bmatrix}, \quad X_{(d)} = \begin{bmatrix} 0_{1 \times T_d} \\ Y_{(d)} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \Phi_1 & \dots & \Phi_p \end{bmatrix}, \quad \Omega_\Phi = \begin{bmatrix} \gamma^2 & 0_{1 \times np} \\ 0_{np \times 1} & \Omega_\Gamma \end{bmatrix},$$

where ι_T is a $T \times 1$ unit vector. The prior for the BVAR is now expressed as

$$\text{vec}(\Gamma) | \Omega \sim N_{n^2 p}(\text{vec}(\Gamma_\mu), [\Omega_\Gamma \otimes \Omega]), \quad (\text{B.10})$$

while $p(\Phi_0) = 1$ and the prior of Ω is given by (B.3). Let $Z = y - \Gamma Y$, $\bar{\Phi}_0 = T^{-1} Z \iota'_T$, and let

$$D = I_T - T^{-1} \iota_T \iota'_T,$$

a $T \times T$ symmetric and idempotent matrix. Through the usual Zellner algebra we have that

$$\epsilon \epsilon' = Z D Z' + (\Phi_0 - \bar{\Phi}_0) \iota'_T \iota_T (\Phi_0 - \bar{\Phi}_0)'$$

Furthermore, with D being symmetric and idempotent we may define $\tilde{Z} = Z D$, such that $\tilde{y} = y D$, $\tilde{Y} = Y D$ and $Z D Z' = \tilde{Z} \tilde{Z}'$. The Zellner algebra now provides us with

$$\tilde{Z} \tilde{Z}' + (\Gamma - \Gamma_\mu) \Omega_\Gamma^{-1} (\Gamma - \Gamma_\mu)' + A = (\Gamma - \bar{\Gamma}) \left(\tilde{Y} \tilde{Y}' + \Omega_\Gamma^{-1} \right) (\Gamma - \bar{\Gamma})' + S,$$

where

$$\begin{aligned} \bar{\Gamma} &= \left(\tilde{y} \tilde{Y}' + \Gamma_\mu \Omega_\Gamma^{-1} \right) \left(\tilde{Y} \tilde{Y}' + \Omega_\Gamma^{-1} \right)^{-1} \\ S &= \tilde{y} \tilde{y}' + A + \Gamma_\mu \Omega_\Gamma^{-1} \Gamma_\mu' - \bar{\Gamma} \left(\tilde{Y} \tilde{Y}' + \Omega_\Gamma^{-1} \right) \bar{\Gamma}'. \end{aligned}$$

It can therefore be shown that the normal-inverted Wishart posterior for the VAR parameters is given by

$$\Phi_0 | \Gamma, \Omega, y, X_1 \sim N_n(\bar{\Phi}_0, T^{-1} \Omega), \quad (\text{B.11})$$

$$\text{vec}(\Gamma) | \Omega, y, X_1 \sim N_{n^2 p}(\text{vec}(\bar{\Gamma}), [(\tilde{Y} \tilde{Y}' + \Omega_\Gamma^{-1})^{-1} \otimes \Omega]) \quad (\text{B.12})$$

$$\Omega | y, X_1 \sim IW_n(S, T + v - 1). \quad (\text{B.13})$$

Hence, the improper prior on Φ_0 results in a loss of degrees of freedom for the posterior of Ω . Furthermore, the log marginal likelihood is

$$\begin{aligned} \log p(y | X_1) &= -\frac{n(T-1)}{2} \log(\pi) + \log \Gamma_n(T+v-1) - \log \Gamma_n(v) - \frac{n}{2} \log |\Omega_\Gamma| \\ &\quad + \frac{v}{2} \log |A| - \frac{n}{2} \log(T) - \frac{n}{2} \log |\tilde{Y} \tilde{Y}' + \Omega_\Gamma^{-1}| - \frac{T+v-1}{2} \log |S|, \end{aligned} \quad (\text{B.14})$$

where the term $\log(T)$ stems from $T = \iota_T' \iota_T$ and is obtained when integrating out Φ_0 from the joint posterior. The relationship between the dummy observations and the prior parameters is

$$\begin{aligned}\Gamma_\mu &= y_{(d)} Y'_{(d)} \left(Y_{(d)} Y'_{(d)} \right)^{-1}, & \Omega_\Gamma &= \left(Y_{(d)} Y'_{(d)} \right)^{-1}, \\ A &= \left(y_{(d)} - \Gamma_\mu Y_{(d)} \right) \left(y_{(d)} - \Gamma_\mu Y_{(d)} \right)', & v &= T_d - (np + 1) + 2.\end{aligned}$$

Letting $\tilde{y}_\star = [y_{(d)} \tilde{y}]$ and $\tilde{Y}_\star = [Y_{(d)} \tilde{Y}]$, it follows that the posterior parameters

$$\begin{aligned}\bar{\Gamma} &= \tilde{y}_\star \tilde{Y}'_\star \left(\tilde{Y}_\star \tilde{Y}'_\star \right)^{-1}, \\ \tilde{Y} \tilde{Y}' + \Omega_\Gamma^{-1} &= \tilde{Y}_\star \tilde{Y}'_\star, \\ S &= \left(\tilde{y}_\star - \bar{\Gamma} \tilde{Y}_\star \right) \left(\tilde{y}_\star - \bar{\Gamma} \tilde{Y}_\star \right)'.\end{aligned}$$

REFERENCES

- Bańbura, M., Giannone, D., and Reichlin, L. (2010), “Large Bayesian Vector Auto Regressions,” *Journal of Applied Econometrics*, 25, 71–92.
- Bauwens, L., Lubrano, M., and Richard, J. F. (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, Oxford.
- Lubik, T. and Schorfheide, F. (2006), “A Bayesian Look at New Open Economy Macroeconomics,” in M. L. Gertler and K. S. Rogoff (Editors), *NBER Macroeconomics Annual 2005*, 313–366, MIT Press.
- Magnus, J. R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, Chichester.