Online Appendix to:
Marginalized Predictive Likelihood Comparisons of Linear Gaussian State-Space Models with Applications to DSGE, DSGE-VAR, and VAR Models

Anders Warne, Günter Coenen and Kai Christoffel

April 2015

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

$$y_t = y_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T, \tag{A.1}$$

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

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

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

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

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

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

where

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

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.

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

Normal Approximation of Marginal Predictive Likelihood

When forecasting with the random walk model it holds that

$$E\bigl[y_{T+h}\big|y, y_0, \Omega\bigr] = y_T, \quad h = 1, \ldots, h^*. \tag{A.6}$$

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

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

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 $E[y_{T+h}\big|y, y_0]$. 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,

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

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 $n \times n_h$ matrix of columns from I_n which selects $y_{s,T+h} = S'y_{T+h}$.[2] Similarly, let $S_\perp$ the the $n \times (n - n_h)$ matrix which selects the remaining variables from the y_{T+h} vector. Define

$$M = \bigl[S \;\; S_\perp\bigr], \tag{A.9}$$

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_h-dimensional inverted Wishart with location matrix $S'\varepsilon\varepsilon'S$ and $T - n + n_h$ degrees of freedom.

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

[2] We here suppress the subscript h from S_h.

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 matrix

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

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

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

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

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

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

Recognizing that the two terms involving Ω_S is the kernel of an n_h-dimensional inverted Wishart distribution with location matrix $S'\varepsilon\varepsilon'S + h^{-1}\varepsilon_{s,T,h}\varepsilon'_{s,T,h}$ and $T-n+n_h+1$ degrees of freedom, it follows that the integral of the density $p\left(y_{s,T+h},\Omega_S\middle|y,y_0\right)$ 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 $IW_{n_h}(S'\varepsilon\varepsilon'S + h^{-1}\varepsilon_{s,T,h}\varepsilon'_{s,T,h}, T-n+n_h+1)$ distribution. We therefore find that

$$p\left(y_{s,T+h}\middle|y,y_0\right) = \frac{\Gamma_{n_h}(T-n+n_h+1)\left|hS'\varepsilon\varepsilon'S\right|^{-1/2}}{\pi^{n_h/2}\Gamma_{n_h}(T-n+n_h)\left|I_{n_h} + \left(hS'\varepsilon\varepsilon'S\right)^{-1}\varepsilon_{s,T,h}\varepsilon'_{s,T,h}\right|^{(T-n+n_h+1)/2}}. \tag{A.13}$$

In other words (and as expected), the density of $y_{s,T+h}|y,y_0$ is an n_h-dimensional t-distribution with mean $y_{s,T}$, covariance matrix given in (A.10), and $T-n+n_h$ degrees of freedom; see, e.g., Bauwens et al. (1999, Appendix A) for details.[3]

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).

[3] Notice also that $\left|I_{n_h} + \left(hS'\varepsilon\varepsilon'S\right)^{-1}\varepsilon_{s,T,h}\varepsilon'_{s,T,h}\right| = 1 + \varepsilon'_{s,T,h}\left(hS'\varepsilon\varepsilon'S\right)^{-1}\varepsilon_{s,T,h}$; 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

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

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

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

$$\Omega \sim IW_n\big(A, v\big). \tag{B.3}$$

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

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

Using standard "Zellner" algebra, it is straightforward to show that

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

where

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

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,

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

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

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

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

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:

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

The vector $\imath_p$ is a p-dimension unit vector, while the $p \times p$ matrix $j_p = \mathrm{diag}[1 \; \cdots \; p]$. The hyperparameter $\lambda_o > 0$ gives the overall tightness in the Minnesota prior, the cross-equation tightness is set to unity, while the harmonc lag decay hyperparameter is equal to 2. The hyperparameter $\tau > 0$ handles shrinkage for the sum of cofficients prior on $(I_n - \sum_{i=1}^{p} \Phi_i)$, where $\tau \to 0$ means that the prior on the sum of the lag cofficients 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:

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

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

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

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

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

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

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

while $p(\Phi_0) = 1$ and the prior of Ω is given by (B.3). Let $Z = y - \Gamma Y$, $\bar{\Phi}_0 = T^{-1} Z \imath_T$, and let

$$D = I_T - T^{-1}\imath_T \imath_T',$$

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

$$\epsilon\epsilon' = ZDZ' + \big(\Phi_0 - \bar{\Phi}_0\big)\imath_T'\imath_T\big(\Phi_0 - \bar{\Phi}_0\big)'.$$

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

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

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

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

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

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

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

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

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

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

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

$$\bar{\Gamma} = \tilde{y}_\star\tilde{Y}_\star'\big(\tilde{Y}_\star\tilde{Y}_\star'\big)^{-1},$$
$$\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1} = \tilde{Y}_\star\tilde{Y}_\star',$$
$$S = \big(\tilde{y}_\star - \bar{\Gamma}\tilde{Y}_\star\big)\big(\tilde{y}_\star - \bar{\Gamma}\tilde{Y}_\star\big)'.$$

Appendix C: The Missing Observation Consistent Kalman Filter

This section presents the necessary equations for computing the marginalized conditional log-likelihood in linear Gaussian state-space models with a missing observations consistent Kalman filter. Let the n-dimensional vector of observable variables y_t be linked to a vector of state variables ξ_t of dimension r through equation

$$y_t = \mu + H'\xi_t + w_t, \quad t = 1, \ldots, T. \tag{C.15}$$

The measurement errors, w_t, are assumed to be i.i.d. $N(0, R)$, with R being an $n \times n$ positive semidefinite matrix, while the state variables are determined from a VAR system:

$$\xi_t = F\xi_{t-1} + B\eta_t, \quad t = 1, \ldots, T. \tag{C.16}$$

The state shocks, η_t, are of dimension q and i.i.d. $N(0, I_q)$, while F is an $r \times r$ matrix, and B is $r \times q$. The parameters of this model, (μ, H, R, F, B), are all uniquely determined by the vector of parameters θ_m from model m. Provided that $H'\xi_t$ is stationary, the vector μ is the population mean of y_t conditional on θ_m.

In order to handle marginalization, let $\mu_{s,i} = S_i'\mu$, $H_{s,i} = HS_i$, and $R_{s,i} = S_i'RS_i$ when $n_i \geq 1$ for $i = 1, \ldots, h$. The selection matrix S_i has dimension $n \times n_i$ with $n_i \in \{0, 1, \ldots, n\}$ for $i = 1, \ldots, h$, has rank equal to n_i and is known. For example, the columns of S_i are taken from I_n such that n_i unique entries of y_{T+i} are selected.

When $n_i \geq 1$ the one-step-ahead forecasts of $y_{s,T+i}$ and its forecast error covariance matrix conditional on information available at time $T + i - 1$ are given by

$$y_{s,T+i|T+i-1} = \mu_{s,i} + H_{s,i}'\xi_{T+i|T+i-1},$$
$$\Sigma_{y_s,T+i|T+i-1} = H_{s,i}'P_{T+i|T+i-1}H_{s,i} + R_{s,i}.$$

If $n_i = 0$ we let these forecasts and forecast error covariances be empty, while the marginalised conditional log-likelihood is equal to zero.

If $n_{i-1} = 0$, the one-step-ahead forecast of the state variables and the corresponding forecast error covariance matrix are

$$\begin{aligned}
\xi_{T+i|T+i-1} &= F\xi_{T+i-1|T+i-2} \\
&= F\xi_{T+i-1|T+i-1}, \\
P_{T+i|T+i-1} &= FP_{T+i-1|T+i-2}F' + BB' \\
&= FP_{T+i-1|T+i-1}F' + BB',
\end{aligned}$$

while if $n_{i-1} \geq 1$ we instead obtain

$$\begin{aligned}
\xi_{T+i|T+i-1} &= F\xi_{T+i-1|T+i-2} + K_{T+i-1}\big(y^o_{s,T+i-1} - y_{s,T+i-1|T+i-2}\big) \\
&= F\xi_{T+i-1|T+i-1}, \\
P_{T+i|T+i-1} &= \big(F - K_{T+i-1}H'_{s,i-1}\big)P_{T+i-1|T+i-2}\big(F - K_{T+i-1}H'_{s,i-1}\big)' \\
&\quad + K_{T+i-1}R_{s,i-1}K'_{T+i-1} + BB' \\
&= FP_{T+i-1|T+i-1}F' + BB', \\
K_{T+i-1} &= FP_{T+i-1|T+i-2}H_{s,i-1}\Sigma^{-1}_{y_s,T+i-1|T+i-2}.
\end{aligned}$$

The filtering equations are initialized by $\xi_{T|T}$, $P_{T|T}$, and K_T, respectively, obtained from the Kalman filter estimates of the state variables using $\mathcal{Y}^o_T$, while $n_0 = n$ if all entries in Y_T are observed.

The log of the marginalized conditional likelihood is for $h \geq 1$ given by

$$\log p\big(\mathcal{Y}^o_{s,T,h}\big|\mathcal{Y}^o_T, \theta_m, m\big) = \sum_{i=1}^{h} \log p\big(y^o_{s,T+i}\big|\mathcal{Y}^o_{s,T,i-1}, \mathcal{Y}^o_T, \theta_m, m\big), \tag{C.17}$$

where $\mathcal{Y}^o_{s,T,0}$ is empty by definition. If $n_i \geq 1$ the marginalized conditional log-likelihood value at $T+i$ is

$$\begin{aligned}
\log p\big(y^o_{s,T+i}\big|\mathcal{Y}^o_{s,T,i-1}, \mathcal{Y}^o_T, \theta_m, m\big) &= -\frac{n_i}{2}\log(2\pi) - \frac{1}{2}\log\left|\Sigma_{y_s,T+i|T+i-1}\right| \\
-\frac{1}{2}\left(y^o_{s,T+i} - y_{s,T+i|T+i-1}\right)' &\Sigma^{-1}_{y_s,T+i|T+i-1}\left(y^o_{s,T+i} - y_{s,T+i|T+i-1}\right),
\end{aligned} \tag{C.18}$$

while it is zero when $n_i = 0$. Notice that the Kalman filter based approach generates a *bottom-up* evaluation of the marginalized conditional likelihood.

An interesting special case of the above filtering equations arises when we are concerned with the marginal h-step-ahead forecast of $y_{s,T+h}$. We here have that $n_i = 0$ for $i = 1, \ldots, h-1$, while

$$\begin{aligned}
y_{s,T+h|T} &= \mu_{s,h} + H'_{s,h}F^h\xi_{T|T}, \\
\Sigma_{y_s,T+h|T} &= H'_{s,h}P_{T+h|T}H_{s,h} + R_{s,h}, \\
P_{T+i|T} &= FP_{T+i-1|T}F' + BB', \quad i = 1, \ldots, h.
\end{aligned}$$

The marginalized conditional log-likelihood value at $T+h$ is:

$$\begin{aligned}\log p\big(y^o_{s,T+h}\big|\mathcal{Y}^o_T,\theta_m,m\big) &= -\frac{n_h}{2}\log(2\pi) - \frac{1}{2}\log\left|\Sigma_{y_s,T+h|T}\right| \\ &-\frac{1}{2}\left(y^o_{s,T+h} - y_{s,T+h|T}\right)'\Sigma^{-1}_{y_s,T+h|T}\left(y^o_{s,T+h} - y_{s,T+h|T}\right).\end{aligned} \tag{C.19}$$

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.