

Technical Appendix to
Are Stocks Really Less Volatile in the Long Run?

by

Ľuboš Pástor

and

Robert F. Stambaugh

February 14, 2009

This Technical Appendix describes how the predictive variance of multiperiod returns is computed in two alternative frameworks that are different from the basic framework used in the paper. Both alternative frameworks are analyzed for the purpose of assessing the robustness of the paper's results, and both feature noninformative prior beliefs. The first framework is the general VAR form of the predictive system, whose results are reported in Section 4.3 in the paper. This framework is discussed in Section B1 below. The second framework is the perfect-predictor model, whose results are reported in Section 5 in the paper. That framework is discussed in Section B2 below.

B1. Predictive system: General VAR

We begin working with multiple assets, so that not only x_t but also r_t and μ_t are vectors. The predictive system in its most general form is a VAR for r_t , x_t , and μ_t , with coefficients restricted so that μ_t is the conditional mean of r_{t+1} . We assume that x_t and μ_t are stationary with means E_x and E_r . We work with the first-order VAR:^{1 2}

$$\begin{bmatrix} r_{t+1} - E_r \\ x_{t+1} - E_x \\ \mu_{t+1} - E_r \end{bmatrix} = \begin{bmatrix} 0 & 0 & I \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} r_t - E_r \\ x_t - E_x \\ \mu_t - E_r \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ w_{t+1} \end{bmatrix}. \quad (\text{B1})$$

We assume the errors in (B1) are i.i.d. across $t = 1, \dots, T$:

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wu} & \Sigma_{wv} & \Sigma_{ww} \end{bmatrix} \right). \quad (\text{B2})$$

Let \bar{A} denote the entire coefficient matrix in (B1), and let Σ denote the entire covariance matrix in (B2). Define the vector

$$\zeta_t = \begin{bmatrix} r_t \\ x_t \\ \mu_t \end{bmatrix}, \quad (\text{B3})$$

and let $V_{\zeta\zeta}$ denote its unconditional covariance matrix. Then

$$V_{\zeta\zeta} = \begin{bmatrix} V_{rr} & V_{rx} & V_{r\mu} \\ V_{xr} & V_{xx} & V_{x\mu} \\ V_{\mu r} & V_{\mu x} & V_{\mu\mu} \end{bmatrix} = \bar{A} V_{\xi\xi} \bar{A}' + \Sigma, \quad (\text{B4})$$

¹The basic framework used in the paper is a special case of (B1) in which the coefficient matrix is restricted as

$$\begin{bmatrix} 0 & 0 & I \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & I \\ 0 & A & 0 \\ 0 & 0 & B \end{bmatrix}.$$

²The predictive system in (B1) can also be viewed alternatively as an unrestricted VAR for returns and predictors when some predictors are unobserved. See the Technical Appendix to Pástor and Stambaugh (2009).

which can be solved as

$$\text{vec}(V_{\xi\xi}) = [I - (\bar{A} \otimes \bar{A})]^{-1} \text{vec}(\Sigma), \quad (\text{B5})$$

using the well known identity $\text{vec}(DFG) = (G' \otimes D)\text{vec}(F)$.

Let z_t denote the vector of the observed data at time t ,

$$z_t = \begin{bmatrix} r_t \\ x_t \end{bmatrix}.$$

Denote the data we observe through time t as $D_t = (z_1, \dots, z_t)$, and note that our complete data consist of D_T . Also define

$$E_z = \begin{bmatrix} E_r \\ E_x \end{bmatrix}, \quad V_{zz} = \begin{bmatrix} V_{rr} & V_{rx} \\ V_{xr} & V_{xx} \end{bmatrix}, \quad V_{z\mu} = \begin{bmatrix} V_{r\mu} \\ V_{x\mu} \end{bmatrix}. \quad (\text{B6})$$

Let ϕ denote the full set of parameters in the model, $\phi \equiv (\bar{A}, \Sigma, E_z)$, and let μ denote the full time series of μ_t , $t = 1, \dots, T$. To obtain the joint posterior distribution of ϕ and μ , denoted by $p(\phi, \mu | D_T)$, we use an MCMC procedure in which we alternate between drawing μ from the conditional posterior $p(\mu | \phi, D_T)$ and drawing ϕ from the conditional posterior $p(\phi | \mu, D_T)$. The procedure for drawing μ from $p(\mu | \phi, D_T)$ is described in Section B1.1. The procedure for drawing ϕ from $p(\phi | \mu, D_T) \propto p(\phi)p(D_T, \mu | \phi)$ is described in Section B1.2.

B1.1. Drawing the time series of μ_t

To draw the time series of the unobservable values of μ_t conditional on the current parameter draws, we apply the *forward filtering, backward sampling* (FFBS) approach developed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994). See also West and Harrison (1997, chapter 15).

B1.1.1. Filtering

The first stage follows the standard methodology of Kalman filtering. Define

$$a_t = E(\mu_t | D_{t-1}) \quad b_t = E(\mu_t | D_t) \quad e_t = E(z_t | \mu_t, D_{t-1}) \quad (\text{B7})$$

$$f_t = E(z_t | D_{t-1}) \quad P_t = \text{Var}(\mu_t | D_{t-1}) \quad Q_t = \text{Var}(\mu_t | D_t) \quad (\text{B8})$$

$$R_t = \text{Var}(z_t | \mu_t, D_{t-1}) \quad S_t = \text{Var}(z_t | D_{t-1}) \quad G_t = \text{Cov}(z_t, \mu'_t | D_{t-1}) \quad (\text{B9})$$

Conditioning on the (unknown) parameters of the model is assumed throughout but suppressed in the notation for convenience. First observe that

$$\mu_0 | D_0 \sim N(b_0, Q_0), \quad (\text{B10})$$

where D_0 denotes the null information set, so that the unconditional moments of μ_0 are given by $b_0 = E_r$ and $Q_0 = V_{\mu\mu}$. Also,

$$\mu_1|D_0 \sim N(a_1, P_1), \quad (\text{B11})$$

where $a_1 = E_r$ and $P_1 = V_{\mu\mu}$, and

$$z_1|D_0 \sim N(f_1, S_1), \quad (\text{B12})$$

where $f_1 = E_z$ and $S_1 = V_{zz}$. Note that

$$G_1 = V_{z\mu} \quad (\text{B13})$$

and that

$$z_1|\mu_1, D_0 \sim N(e_1, R_1), \quad (\text{B14})$$

where

$$e_1 = f_1 + G_1 P_1^{-1} (\mu_1 - a_1) \quad (\text{B15})$$

$$R_1 = S_1 - G_1 P_1^{-1} G_1'. \quad (\text{B16})$$

Combining this density with equation (B11) using Bayes rule gives

$$\mu_1|D_1 \sim N(b_1, Q_1), \quad (\text{B17})$$

where

$$b_1 = a_1 + P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} G_1' R_1^{-1} (z_1 - f_1) \quad (\text{B18})$$

$$Q_1 = P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} P_1. \quad (\text{B19})$$

Continuing in this fashion, we find that all conditional densities are normally distributed, and we obtain all the required moments for $t = 2, \dots, T$:

$$a_t = (I - A_{31} - A_{33})E_r - A_{32}E_x + A_{31}r_{t-1} + A_{32}x_{t-1} + A_{33}b_{t-1} \quad (\text{B20})$$

$$f_t = \begin{bmatrix} b_{t-1} \\ (I - A_{22})E_x - (A_{21} + A_{23})E_r + A_{21}r_{t-1} + A_{22}x_{t-1} + A_{23}b_{t-1} \end{bmatrix} \quad (\text{B21})$$

$$S_t = \begin{bmatrix} Q_{t-1} & Q_{t-1}A'_{23} \\ A_{23}Q_{t-1} & A_{23}Q_{t-1}A'_{23} \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix} \quad (\text{B22})$$

$$G_t = \begin{bmatrix} Q_{t-1}A'_{33} \\ A_{23}Q_{t-1}A'_{33} \end{bmatrix} + \begin{bmatrix} \Sigma_{uw} \\ \Sigma_{vw} \end{bmatrix} \quad (\text{B23})$$

$$P_t = A_{33}Q_{t-1}A'_{33} + \Sigma_{ww} \quad (\text{B24})$$

$$e_t = f_t + G_t P_t^{-1} (\mu_t - a_t) \quad (\text{B25})$$

$$R_t = S_t - G_t P_t^{-1} G_t' \quad (\text{B26})$$

$$b_t = a_t + P_t (P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1} (z_t - f_t) \quad (\text{B27})$$

$$= a_t + G_t' S_t^{-1} (z_t - f_t) \quad (\text{B28})$$

$$Q_t = P_t (P_t + G_t' R_t^{-1} G_t)^{-1} P_t. \quad (\text{B29})$$

The values of $\{a_t, b_t, Q_t, S_t, G_t, P_t\}$ for $t = 1, \dots, T$ are retained for the next stage. Equations (B22) through (B24) are derived as

$$\begin{aligned} \begin{bmatrix} S_t & G_t \\ G_t' & P_t \end{bmatrix} &= \text{Var}(\zeta_t | D_{t-1}) \\ &= \bar{A} \text{Var}(\zeta_{t-1} | D_{t-1}) \bar{A}' + \Sigma \\ &= \bar{A} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_{t-1} \end{bmatrix} \bar{A}' + \Sigma \\ &= \begin{bmatrix} Q_{t-1} & Q_{t-1} A'_{23} & Q_{t-1} A'_{33} \\ A_{23} Q_{t-1} & A_{23} Q_{t-1} A'_{23} & A_{23} Q_{t-1} A'_{33} \\ A_{33} Q_{t-1} & A_{33} Q_{t-1} A'_{23} & A_{33} Q_{t-1} A'_{33} \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wu} & \Sigma_{wv} & \Sigma_{ww} \end{bmatrix}. \end{aligned}$$

B1.1.2. Sampling

We wish to draw $(\mu_0, \mu_1, \dots, \mu_T)$ conditional on D_T . The backward-sampling approach relies on the Markov property of the evolution of ζ_t and the resulting identity,

$$p(\zeta_0, \zeta_1, \dots, \zeta_T | D_T) = p(\zeta_T | D_T) p(\zeta_{T-1} | \zeta_T, D_{T-1}) \cdots p(\zeta_1 | \zeta_2, D_1) p(\zeta_0 | \zeta_1, D_0). \quad (\text{B30})$$

We first sample μ_T from $p(\mu_T | D_T)$, the normal density obtained in the last step of the filtering. Then, for $t = T - 1, T - 2, \dots, 1, 0$, we sample μ_t from the conditional density $p(\zeta_t | \zeta_{t+1}, D_t)$. (Note that the first two subvectors of ζ_t are already observed and thus need not be sampled.) To obtain that conditional density, first note that

$$\zeta_{t+1} | D_t \sim N \left(\begin{bmatrix} f_{t+1} \\ a_{t+1} \end{bmatrix}, \begin{bmatrix} S_{t+1} & G_{t+1} \\ G_{t+1}' & P_{t+1} \end{bmatrix} \right), \quad (\text{B31})$$

$$\zeta_t | D_t \sim N \left(\begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} \right), \quad (\text{B32})$$

and

$$\text{Cov}(\zeta_t, \zeta'_{t+1} | D_t) = \text{Var}(\zeta_t | D_t) \bar{A}'$$

$$\begin{aligned}
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} \begin{bmatrix} 0 & A'_{21} & A'_{31} \\ 0 & A'_{22} & A'_{32} \\ I & A'_{23} & A'_{33} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & Q_t A'_{23} & Q_t A'_{33} \end{bmatrix}. \tag{B33}
\end{aligned}$$

Therefore,

$$\zeta_t | \zeta_{t+1}, D_t \sim N(h_t, H_t), \tag{B34}$$

where

$$\begin{aligned}
h_t &= E(\zeta_t | D_t) + [\text{Cov}(\zeta_t, \zeta'_{t+1} | D_t)] [\text{Var}(\zeta_{t+1} | D_t)]^{-1} [\zeta_{t+1} - E(\zeta_{t+1} | D_t)] \\
&= \begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & Q_t A'_{23} & Q_t A'_{33} \end{bmatrix} \begin{bmatrix} S_{t+1} & G_{t+1} \\ G'_{t+1} & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} z_{t+1} - f_{t+1} \\ \mu_{t+1} - a_{t+1} \end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
H_t &= \text{Var}(\zeta_t | D_t) - [\text{Cov}(\zeta_t, \zeta'_{t+1} | D_t)] [\text{Var}(\zeta_{t+1} | D_t)]^{-1} [\text{Cov}(\zeta_t, \zeta'_{t+1} | D_t)]' \\
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & Q_t A'_{23} & Q_t A'_{33} \end{bmatrix} \begin{bmatrix} S_{t+1} & G_{t+1} \\ G'_{t+1} & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & Q_t \\ 0 & 0 & A_{23} Q_t \\ 0 & 0 & A_{33} Q_t \end{bmatrix}
\end{aligned}$$

The mean and covariance matrix of μ_t are taken as the relevant elements of h_t and H_t .

In the rest of the Appendix, we discuss the special case (implemented in the paper) in which r_t and μ_t are scalars. The dimensions of \bar{A} and Σ are then $(K + 2) \times (K + 2)$, where K is the number of predictors in the vector x_t .

B1.2. Drawing the parameters

This section describes how we obtain the posterior draws of all parameters conditional on the current draw of the time series of μ_t . The parameters are $(\bar{A}, \Sigma, E_{xr})$, where $E_{xr} \equiv (E'_x E_r)'$.

B1.2.1. Prior distributions

We specify noninformative priors for all parameters. The prior on \bar{A} is flat, except for the stationarity restriction that all eigenvalues of \bar{A} must lie inside the unit circle. The prior on E_{xr} is normal, $E_{xr} \sim N(E_{xr0}, V_{xr0})$. The prior covariance matrix V_{xr0} has very large elements on the main diagonal and zeros off the diagonal, so the choice of E_{xr0} is unimportant. The prior on Σ is inverted Wishart with a small number of degrees of freedom: $\Sigma \sim IW(\Sigma_0, \nu)$. We

choose $\nu = K + 4$, so that $\nu = 7$ in our basic specification with three predictors. The choice of Σ_0 , the prior mean of Σ , is unimportant because the prior standard deviation of Σ is very large, corresponding to the posterior from a hypothetical sample of only seven observations. All parameters are independent a priori.

B1.2.2. Posterior distributions

Drawing Σ given (E_{xr}, \bar{A})

Given the normal likelihood and the inverted Wishart prior for Σ , the posterior of Σ is also inverted Wishart, $\Sigma|\cdot \sim IW(S + \Sigma_0, (T - 1) + \nu)$, where $S = (Y - XB)'(Y - XB)$ and

$$Y = \begin{pmatrix} r_2 - E_r & x'_2 - E_x \iota'_K & \mu_2 - E_r \\ \vdots & \vdots & \vdots \\ r_T - E_r & x'_T - E_x \iota'_K & \mu_T - E_r \end{pmatrix}$$

$$X = \begin{pmatrix} r_1 - E_r & x'_1 - E_x \iota'_K & \mu_1 - E_r \\ \vdots & \vdots & \vdots \\ r_{T-1} - E_r & x'_{T-1} - E_x \iota'_K & \mu_{T-1} - E_r \end{pmatrix}$$

$$B = \bar{A}'.$$

Above, ι_K denotes a $K \times 1$ vector of ones. The dimensions of both X and Y are $(T - 1) \times (K + 2)$.

Drawing E_{xr} given (Σ, \bar{A})

The conditional posterior of E_{xr} is normal, $E_{xr}|\cdot \sim N(\tilde{E}_{xr}, \tilde{V}_{xr})$, where

$$\tilde{V}_{xr} = [V_{xr0}^{-1} + (T - 1)Q' \Sigma^{-1} Q]^{-1} \quad (\text{B35})$$

$$\tilde{E}_{xr} = \tilde{V}_{xr} \left[V_{xr0}^{-1} E_{xr0} + Q' \Sigma^{-1} \sum_{t=1}^{T-1} (\zeta_{t+1} - A \zeta_t) \right] \quad (\text{B36})$$

$$Q \equiv \begin{bmatrix} 0 & 0 \\ I_K - A_{22} & -(A_{21} + A_{23}) \\ -A_{32} & 1 - A_{31} - A_{33} \end{bmatrix}. \quad (\text{B37})$$

Above, Q is $(K + 2) \times (K + 1)$, I_K is a $K \times K$ identity matrix, and ζ_t is defined in equation (B3).

Drawing \bar{A} given (E_{xr}, Σ)

In order to draw \bar{A} , we need to draw the $(K + 2) \times (K + 1)$ matrix

$$A = \begin{bmatrix} A'_{21} & A'_{31} \\ A'_{22} & A'_{32} \\ A'_{23} & A'_{33} \end{bmatrix}. \quad (\text{B38})$$

First, we establish some notation. Denote $Z = I_{K+1} \otimes X$, where I_{K+1} is a $(K+1) \times (K+1)$ identity matrix, so that Z is $[(T-1)(K+1)] \times [(K+2)(K+1)]$. Let

$$\bar{Y} = \begin{pmatrix} x'_2 - E_{x'l'_K} & \mu_2 - E_r \\ \dots & \dots \\ x'_T - E_{x'l'_K} & \mu_T - E_r \end{pmatrix}$$

and let $z = \text{vec}(\bar{Y})$, so that z is $[(T-1)(K+1)] \times 1$. Denote

$$u = \begin{bmatrix} u_2 \\ \vdots \\ u_T \end{bmatrix}, V = \begin{bmatrix} v'_2 \\ \vdots \\ v'_T \end{bmatrix}, w = \begin{bmatrix} w_2 \\ \vdots \\ w_T \end{bmatrix}, r = \begin{bmatrix} r_2 \\ \vdots \\ r_T \end{bmatrix}, \mu_{(l)} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{T-1} \end{bmatrix}, \quad (\text{B39})$$

where u, w, r , and $\mu_{(l)}$ are $(T-1) \times 1$ vectors and V is a $(T-1) \times K$ matrix. Let $v = \text{vec}(V)$. Define Σ_{12} , Σ_{21} , and Σ_{22} so that

$$\Sigma \equiv \begin{bmatrix} \Sigma_{uu} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

That is, $\Sigma_{12} \equiv [\Sigma_{uv} \ \Sigma_{uw}]$ and

$$\Sigma_{22} \equiv \begin{bmatrix} \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wv} & \Sigma_{ww} \end{bmatrix}.$$

With the above notation, we can write the system (B1) without its first equation as

$$z = Zb + \begin{bmatrix} v \\ w \end{bmatrix},$$

where $b = \text{vec}(A)$. Conditional on u , the error terms $[v' \ w']'$ are normally distributed with the mean of $\Sigma_{21} \Sigma_{uu}^{-1} \otimes u$ and the covariance matrix of $(\Sigma_{22} - \Sigma_{21} \Sigma_{uu}^{-1} \Sigma_{12}) \otimes I_{T-1}$.

Recall that the prior distribution on b is given by $1_{b \in S}$, which is equal to one when the stationarity restriction on \bar{A} is satisfied and zero otherwise. Given the normal likelihood, the full conditional posterior distribution of b is then given by

$$b|\cdot \sim N(\tilde{b}, \tilde{V}_b) \times 1_{b \in S}, \quad (\text{B40})$$

where

$$\tilde{b} = (Z'Z)^{-1} Z' [z - \Sigma_{21} \Sigma_{uu}^{-1} \otimes (r - \mu_{(l)})] \quad (\text{B41})$$

$$\tilde{V}_b = (\Sigma_{22} - \Sigma_{21} \Sigma_{uu}^{-1} \Sigma_{12}) \otimes (X'X)^{-1}. \quad (\text{B42})$$

We obtain the posterior draws of b by making draws from $N(\tilde{b}, \tilde{V}_b)$ and retaining only draws that satisfy $b \in S$. The posterior draws of A are constructed from the posterior draws of $b = \text{vec}(A)$.

B1.3. Predictive variance for general VAR

In addition to the notation from equations (B1), (B2), and (B3), define also

$$E_\zeta = \begin{bmatrix} E_r \\ E_x \\ E_r \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix}, \quad (\text{B43})$$

and

$$\zeta_t^e = \zeta_t - E_\zeta. \quad (\text{B44})$$

Equation (B1) can then be written as

$$\zeta_{t+1}^e = \bar{A}\zeta_t^e + \epsilon_{t+1}. \quad (\text{B45})$$

For $i > 1$, successive substitution using (B45) gives

$$\zeta_{t+i}^e = \bar{A}^i \zeta_t^e + \bar{A}^{i-1} \epsilon_{t+1} + \bar{A}^{i-2} \epsilon_{t+2} + \cdots + \epsilon_{t+i}. \quad (\text{B46})$$

Define

$$\begin{aligned} \zeta_{T,T+K} &= \sum_{i=1}^K \zeta_{T+i}, \\ \zeta_{T,T+K}^e &= \sum_{i=1}^K \zeta_{T+i}^e = \zeta_{T,T+K} - KE_\zeta. \end{aligned}$$

Summing (B45) over K periods then gives

$$\begin{aligned} \zeta_{T,T+K}^e &= \left(\sum_{i=1}^K \bar{A}^i \right) \zeta_T^e + (I + \bar{A} + \cdots + \bar{A}^{K-1}) \epsilon_{t+1} \\ &\quad + (I + \bar{A} + \cdots + \bar{A}^{K-2}) \epsilon_{t+2} + \cdots + \epsilon_{t+K} \\ &= (\bar{\Lambda}_{K+1} - I) \zeta_T^e + \bar{\Lambda}_K \epsilon_{t+1} + \bar{\Lambda}_{K-1} \epsilon_{t+2} + \cdots + \epsilon_{t+K}, \end{aligned} \quad (\text{B47})$$

where

$$\begin{aligned} \bar{\Lambda}_i &= I + \bar{A} + \cdots + \bar{A}^{i-1} \\ &= (I - \bar{A})^{-1} (I - \bar{A}^i). \end{aligned} \quad (\text{B48})$$

It then follows that

$$\begin{aligned} E(\zeta_{T,T+K}^e | D_T, \phi, \mu_T) &= (\bar{\Lambda}_{K+1} - I) \zeta_T^e, \\ E(\zeta_{T,T+K} | D_T, \phi, \mu_T) &= (\bar{\Lambda}_{K+1} - I) \zeta_T^e + KE_\zeta, \end{aligned} \quad (\text{B49})$$

and

$$\begin{aligned}\text{Var}(\zeta_{T,T+K}|D_T, \phi, \mu_T) &= \text{Var}(\zeta_{T,T+K}^e|D_T, \phi, \mu_T) \\ &= \sum_{i=1}^K \bar{\Lambda}_i \Sigma \bar{\Lambda}_i'.\end{aligned}\quad (\text{B50})$$

The first and second moments of (B49) given D_T and ϕ are given by

$$\text{E}(\zeta_{T,T+K}|D_T, \phi) = (\bar{\Lambda}_{K+1} - I) \begin{bmatrix} r_T - E_r \\ x_T - E_x \\ b_T - E_r \end{bmatrix} + KE_\xi \quad (\text{B51})$$

and

$$\text{Var}[\text{E}(\zeta_{T,T+K}|D_T, \phi, \mu_T)|D_T, \phi] = (\bar{\Lambda}_{K+1} - I) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_T \end{bmatrix} (\bar{\Lambda}_{K+1} - I)'. \quad (\text{B52})$$

Combining (B50) and (B52) gives

$$\begin{aligned}\text{Var}(\zeta_{T,T+K}|D_T, \phi) &= \text{E}[\text{Var}(\zeta_{T,T+K}|D_T, \phi, \mu_T)|D_T, \phi] + \text{Var}[\text{E}(\zeta_{T,T+K}|D_T, \phi, \mu_T)|D_T, \phi] \\ &= \sum_{i=1}^K \bar{\Lambda}_i \Sigma \bar{\Lambda}_i' + (\bar{\Lambda}_{K+1} - I) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_T \end{bmatrix} (\bar{\Lambda}_{K+1} - I)'.\end{aligned}\quad (\text{B53})$$

By evaluating (B51) and (B53) for repeated draws of ϕ from its posterior, the predictive variance of $\zeta_{T,T+K}$ can be computed using the decomposition,

$$\text{Var}(\zeta_{T,T+K}|D_T) = \text{E}\{\text{Var}(\zeta_{T,T+K}|\phi, D_T)|D_T\} + \text{Var}\{\text{E}(\zeta_{T,T+K}|\phi, D_T)|D_T\}. \quad (\text{B54})$$

Finally, the predictive variance of $r_{T,T+K}$ is the (1,1) element of $\text{Var}(\zeta_{T,T+K}|D_T)$.

B2. Perfect predictors

Note: The notation in this section is distinct from the notation in the previous section.

The paper focuses on the realistic scenario in which the observable predictors x_t are imperfect, in that they do not perfectly capture the conditional expected return μ_t . In this section, we discuss the calculation of predictive variance in an alternative framework with perfect predictors, for which $\mu_t = a + b'x_t$. In that case, the predictive system is replaced by a model consisting of equations

$$r_{t+1} = a + b'x_t + e_{t+1} \quad (\text{B55})$$

$$x_{t+1} = \theta + Ax_t + v_{t+1}, \quad (\text{B56})$$

combined with the following distributional assumption on the residuals:

$$\begin{bmatrix} e_t \\ v_t \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_e^2 & \sigma_{ev} \\ \sigma_{ve} & \Sigma_{vv} \end{bmatrix} \right). \quad (\text{B57})$$

Let δ denote the full set of parameters in equations (B55), (B56), and (B57). Let Ω denote the covariance matrix in (B57), let B denote the matrix of the slope coefficients in (B55) and (B56),

$$B = \begin{bmatrix} a & \theta' \\ b & A' \end{bmatrix},$$

and let $c = \text{vec}(B)$. Note that δ consists of the elements of c and Ω .

B2.1. Posterior distributions under perfect predictors

We specify the prior distribution on δ as $p(\delta) = p(c)p(\Omega)$. The priors on c and Ω are non-informative, except for the restriction that the spectral radius of A , $\rho(A)$, is less than 1. The prior on c is $p(c) \propto I[\rho(A) < 1]$, where $I[\cdot]$ denotes the indicator function. The prior on Ω is $p(\Omega) \propto |\Omega|^{-(m+1)/2}$, where m is the number of rows in Ω (i.e., x_t is $(m-1) \times 1$).

To obtain the posterior distribution of δ , the prior $p(\delta)$ is combined with the normal likelihood function $p(D_T|\delta)$ implied by equation (B57). The posterior draws of δ can be obtained by applying standard results from the multivariate regression model (e.g., Zellner, 1971). Define the following notation: $r = [r_1 \ r_2 \ \cdots \ r_T]'$, $Q^+ = [x_1 \ x_2 \ \cdots \ x_T]'$, $Q = [x_0 \ x_1 \ \cdots \ x_{T-1}]'$, $X = [1_T \ Q]$, where 1_T denotes a $T \times 1$ vector of ones, $Y = [r \ Q^+]$, $\hat{B} = (X'X)^{-1}X'Y$, and $S = (Y - X\hat{B})'(Y - X\hat{B})$. We first draw Ω^{-1} from a Wishart distribution with $T-m$ degrees of freedom and parameter matrix S^{-1} . Given that draw of Ω^{-1} , we then draw c from a normal distribution with mean $\hat{c} = \text{vec}(\hat{B})$ and covariance matrix $\Omega \otimes (X'X)^{-1}$. That draw of δ is retained as a draw from $p(\delta|D_T)$ if $\rho(A) < 1$.

B2.2. Predictive variance under perfect predictors

The conditional moments of the k -period return $r_{T,T+k}$ are given by

$$E(r_{T,T+k}|D_T, \delta) = ka + b'\Psi_{k-1}\theta + b'\Lambda_k x_T \quad (\text{B58})$$

$$\text{Var}(r_{T,T+k}|D_T, \delta) = k\sigma_e^2 + 2b'\Psi_{k-1}\sigma_{ve} + b' \left(\sum_{i=1}^{k-1} \Lambda_i \Sigma_{vv} \Lambda_i' \right) b, \quad (\text{B59})$$

where

$$\Lambda_i = I + A + \dots + A^{i-1} = (I - A)^{-1}(I - A^i) \quad (\text{B60})$$

$$\Psi_{k-1} = \Lambda_1 + \Lambda_2 + \dots + \Lambda_{k-1} = (I - A)^{-1}[kI - (I - A)^{-1}(I - A^k)]. \quad (\text{B61})$$

The first term in (B59) reflects i.i.d. uncertainty. The second term reflects correlation between unexpected returns and innovations in future x_{T+i} 's, which deliver innovations in future μ_{T+i} 's. That term can be positive or negative and captures any mean reversion. The third term, always positive, reflects uncertainty about future x_{T+i} 's, and thus uncertainty about future μ_{T+i} 's. This third term, which contains a summation, can also be written without the summation as

$$\begin{aligned} b' \left(\sum_{i=1}^{k-1} \Lambda_i \Sigma_{vv} \Lambda_i' \right) b &= (b' \otimes b') [(I - A)^{-1} \otimes (I - A)^{-1}] \left[kI - \Lambda_k \otimes I - I \otimes \Lambda_k \right. \\ &\quad \left. + (I - A \otimes A)^{-1}(I - (A \otimes A)^k) \right] \text{vec}(\Sigma_{vv}). \end{aligned}$$

Applying the standard variance decomposition

$$\text{Var}(r_{T,T+k} | D_T) = \text{E}\{\text{Var}(r_{T,T+k} | D_T, \delta) | D_T\} + \text{Var}\{\text{E}(r_{T,T+k} | D_T, \delta) | D_T\}, \quad (\text{B62})$$

the predictive variance $\text{Var}(r_{T,T+k} | D_T)$ can be computed as the sum of the posterior mean of the right-hand side of equation (B59) and the posterior variance of the right-hand side of equation (B58). These posterior moments are computed from the posterior draws of δ , which are described in Section B2.1.

References

- Carter, Chris K., and Robert Kohn, 1994, On Gibbs sampling for state space models, *Biometrika* 81, 541–553.
- Frühwirth-Schnatter, Sylvia, 1994, Data augmentation and dynamic linear models, *Journal of Time Series Analysis* 15, 183–202.
- Pástor, Ľuboš, and Robert F. Stambaugh, 2009, Predictive systems: Living with imperfect predictors, *Journal of Finance*, forthcoming.
- West, Mike, and Jeff Harrison, 1997, *Bayesian Forecasting and Dynamic Models* (Springer-Verlag, New York, NY).