Technical Appendix to

Predictive Systems: Living with Imperfect Predictors

by

Ľuboš Pástor

and

Robert F. Stambaugh

For Supplements and Dataset Section of www.afajof.org

B1. Roadmap

This Technical Appendix, containing material of a technical nature relevant to the published article, is organized as follows. Section B2 discusses the general definition of the predictive system and establishes some notation. Section B3 details the filtering-and-sampling procedure for drawing the time series of the unobservable conditional expected return μ_t conditional on the parameter values. Section B4 characterizes the dependence of estimated expected returns on the full history of returns and predictor realizations. Section B5 describes the prior and posterior distributions of the parameters in the predictive system. Section B6 presents the procedure for maximum likelihood estimation of the predictive system. Section B7 analyzes the R^2 ratio from equation (29) in the paper. Finally, Section B8 provides details regarding the variance decomposition whose results are reported in Table IV in the paper.

The Bayesian analysis of the predictive system proceeds as follows. Let D_T denote the data available to the investor, let θ denote the set of parameters in the predictive system, and let μ denote the full time series of μ_t, $t = 1, \ldots, T$. To obtain the joint posterior distribution of θ and μ, denoted by $p(\theta, \mu | D_T)$, we use an MCMC procedure in which we alternate between drawing μ from the conditional posterior $p(\mu|\theta, D_T)$ and drawing θ from the conditional posterior $p(\theta|\mu, D_T)$. The procedure for drawing μ from $p(\mu|\theta, D_T)$ is described in Section B3. The procedure for drawing θ from $p(\theta|\mu, D_T) \propto p(\theta)p(D_T, \mu|\theta)$ is described in Section B5.

B2. Predictive system: General framework

We begin working with multiple assets, so that r_t and μ_t are vectors (recall x_t can be a vector in any case). We define the predictive system in its most general form as 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 also assume that x_t and μ_t are stationary with means E_x and E_r. The first-order VAR, for example, is

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

The predictive system in (B1.) can be viewed alternatively as simply an unrestricted VAR for returns and predictors when some predictors are unobserved. Specifically, consider an unrestricted VAR for r_t, x_t, and π_t, where π_t has the same dimensions as r_t and contains additional unobserved predictors:

$$\begin{bmatrix} r_{t+1}-E_r \\ x_{t+1}-E_x \\ \pi_{t+1}-E_\pi \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \begin{bmatrix} r_t-E_r \\ x_t-E_x \\ \pi_t-E_\pi \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ \upsilon_{t+1} \end{bmatrix}. \quad \text{(B2.)}$$

When B_{13} is nonsingular, (B1.) and (B2.) are equivalent. It is immediate that (B1.) implies (B2.). To see the converse, define

$$\mu_t = E_r + B_{11}(r_t - E_r) + B_{12}(x_t - E_x) + B_{13}(\pi_t - E_\pi), \quad \text{(B3.)}$$

which implies

$$\pi_t - E_\pi = -B_{13}^{-1}B_{11}(r_t - E_r) - B_{13}^{-1}B_{12}(x_t - E_x) + B_{13}^{-1}(\mu_t - E_r). \quad \text{(B4.)}$$

Pre-multiplying both sides of (B2.) by $[B_{11}\ \ B_{12}\ \ B_{13}]$, using (B3.) and (B4.), gives

$$\begin{aligned} \mu_{t+1} - E_r &= \begin{bmatrix} C_{11} & C_{12} & C_{13} \end{bmatrix} \begin{bmatrix} r_t - E_r \\ x_t - E_x \\ -B_{13}^{-1}B_{11}(r_t - E_r) - B_{13}^{-1}B_{12}(x_t - E_x) + B_{13}^{-1}(\mu_t - E_r) \end{bmatrix} \\ &\quad + \begin{bmatrix} B_{11} & B_{12} & B_{13} \end{bmatrix} \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ \upsilon_{t+1} \end{bmatrix} \\ &= A_{31}(r_t - E_r) + A_{32}(x_t - E_x) + A_{33}(\mu_t - E_r) + w_{t+1}, \quad \text{(B5.)} \end{aligned}$$

where

$$\begin{aligned} \begin{bmatrix} C_{11} & C_{12} & C_{13} \end{bmatrix} &= \begin{bmatrix} B_{11} & B_{12} & B_{13} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \\ w_{t+1} &= \begin{bmatrix} B_{11} & B_{12} & B_{13} \end{bmatrix} \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ \upsilon_{t+1} \end{bmatrix} \\ A_{31} &= C_{11} - C_{13}B_{13}^{-1}B_{11} \\ A_{32} &= C_{12} - C_{13}B_{13}^{-1}B_{12} \\ A_{33} &= C_{13}B_{13}^{-1}. \end{aligned}$$

Combining (B2.), (B3.), and (B4.) gives

$$\begin{bmatrix} r_{t+1}-E_r \\ x_{t+1}-E_x \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \end{bmatrix} \begin{bmatrix} r_t - E_r \\ x_t - E_x \\ -B_{13}^{-1}B_{11}(r_t - E_r) - B_{13}^{-1}B_{12}(x_t - E_x) + B_{13}^{-1}(\mu_t - E_r) \end{bmatrix}$$

$$+\begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & I \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{bmatrix} r_t - E_r \\ x_t - E_x \\ \mu_t - E_r \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix}, \tag{B6.}$$

where

$$\begin{aligned} A_{21} &= B_{21} - B_{23} B_{13}^{-1} B_{11} \\ A_{22} &= B_{22} - B_{23} B_{13}^{-1} B_{12} \\ A_{23} &= B_{23} B_{13}^{-1}. \end{aligned} \tag{B7.}$$

Combining (B5.) and (B6.) gives (B1.).

We assume the disturbances in (B1.) are distributed identically and independently across t as

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wu} & \Sigma_{wv} & \Sigma_{ww} \end{bmatrix} \right). \tag{B8.}$$

Define the vector

$$\zeta_t = \begin{bmatrix} r_t \\ x_t \\ \mu_t \end{bmatrix},$$

and let $V_{\zeta\zeta}$ denote its unconditional covariance matrix. Also let $\bar{A}$ denote the entire coefficient matrix in (B1.), and let Σ denote the entire covariance matrix in (B8.). Then

$$V_{\zeta\zeta} = \begin{bmatrix} V_{rr} & V_{rx} & V_{r\mu} \\ V_{xr} & V_{xx} & V_{x\mu} \\ V_{\mu r} & V_{\mu x} & V_{\mu\mu} \end{bmatrix} = \bar{A} V_{\zeta\zeta} \bar{A}' + \Sigma, \tag{B9.}$$

which can be solved as

$$\mathrm{vec}\,(V_{\zeta\zeta}) = [I - (\bar{A} \otimes \bar{A})]^{-1} \mathrm{vec}\,(\Sigma), \tag{B10.}$$

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

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 $D_t = (z_1, \ldots, z_t)$, and note that our complete data consist of D_T. Also define

$$E_z = \begin{bmatrix} E_r \\ E_x \end{bmatrix}, \; V_{zz} = \begin{bmatrix} V_{rr} & V_{rx} \\ V_{xr} & V_{xx} \end{bmatrix}, \; V_{z\mu} = \begin{bmatrix} V_{r\mu} \\ V_{x\mu} \end{bmatrix}. \tag{B11.}$$

B3. 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).

B3.1. Filtering

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

$$a_t = \mathrm{E}(\mu_t|D_{t-1}) \qquad b_t = \mathrm{E}(\mu_t|D_t) \qquad e_t = \mathrm{E}(z_t|\mu_t, D_{t-1}) \tag{B12.}$$

$$f_t = \mathrm{E}(z_t|D_{t-1}) \qquad P_t = \mathrm{Var}(\mu_t|D_{t-1}) \qquad Q_t = \mathrm{Var}(\mu_t|D_t) \tag{B13.}$$

$$R_t = \mathrm{Var}(z_t|\mu_t, D_{t-1}) \qquad S_t = \mathrm{Var}(z_t|D_{t-1}) \qquad G_t = \mathrm{Cov}(z_t, \mu_t'|D_{t-1}) \tag{B14.}$$

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

$$\mu_0|D_0 \sim N(b_0, Q_0), \tag{B15.}$$

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,

$$\mu_1|D_0 \sim N(a_1, P_1), \tag{B16.}$$

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

$$z_1|D_0 \sim N(f_1, S_1), \tag{B17.}$$

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

$$G_1 = V_{z\mu} \tag{B18.}$$

and that

$$z_1|\mu_1, D_0 \sim N(e_1, R_1), \tag{B19.}$$

where

$$e_1 = f_1 + G_1 P_1^{-1}(\mu_1 - a_1) \tag{B20.}$$

$$R_1 = S_1 - G_1 P_1^{-1} G_1'. \tag{B21.}$$

Combining this density with equation (B16.) using Bayes rule gives

$$\mu_1|D_1 \sim N(b_1, Q_1), \tag{B22.}$$

where

$$b_1 = a_1 + P_1(P_1 + G_1' R_1^{-1} G_1)^{-1} G_1' R_1^{-1}(z_1 - f_1) \quad \text{(B23.)}$$

$$Q_1 = P_1(P_1 + G_1' R_1^{-1} G_1)^{-1} P_1. \quad \text{(B24.)}$$

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

$$a_t = (I - A_{31} - A_{33})E_r - A_{32}E_x + A_{31}r_{t-1} + A_{32}x_{t-1} + A_{33}b_{t-1} \quad \text{(B25.)}$$

$$f_t = \begin{bmatrix} b_{t-1} \\ (I - A_{22})E_x - (A_{21} + A_{23})E_r + A_{21}r_{t-1} + A_{22}x_{t-1} + A_{23}b_{t-1} \end{bmatrix} \quad \text{(B26.)}$$

$$S_t = \begin{bmatrix} Q_{t-1} & Q_{t-1}A_{23}' \\ A_{23}Q_{t-1} & A_{23}Q_{t-1}A_{23}' \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix} \quad \text{(B27.)}$$

$$G_t = \begin{bmatrix} Q_{t-1}A_{33}' \\ A_{23}Q_{t-1}A_{33}' \end{bmatrix} + \begin{bmatrix} \Sigma_{uw} \\ \Sigma_{vw} \end{bmatrix} \quad \text{(B28.)}$$

$$P_t = A_{33}Q_{t-1}A_{33}' + \Sigma_{ww} \quad \text{(B29.)}$$

$$e_t = f_t + G_t P_t^{-1}(\mu_t - a_t) \quad \text{(B30.)}$$

$$R_t = S_t - G_t P_t^{-1} G_t' \quad \text{(B31.)}$$

$$b_t = a_t + P_t(P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1}(z_t - f_t) \quad \text{(B32.)}$$

$$= a_t + G_t' S_t^{-1}(z_t - f_t) \quad \text{(B33.)}$$

$$Q_t = P_t(P_t + G_t' R_t^{-1} G_t)^{-1} P_t. \quad \text{(B34.)}$$

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 (B27.) through (B29.) are derived as

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

B3.2. 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,

$$p(\zeta_0, \zeta_1, \ldots, \zeta_T | D_T) = p(\zeta_T | D_T) p(\zeta_{T-1} | \zeta_T, D_{T-1}) \cdots p(\zeta_1 | \zeta_2, D_1) p(\zeta_0 | \zeta_1, D_0). \qquad \text{(B35.)}$$

We first sample μ_T from $p(\mu_T | D_T)$, the normal density obtained in the last step of the filtering. Then, for $t = T-1, T-2, \ldots, 1, 0$, 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

$$\zeta_{t+1} | D_t \sim N\left(\begin{bmatrix} f_{t+1} \\ a_{t+1} \end{bmatrix}, \begin{bmatrix} S_{t+1} & G_{t+1} \\ G'_{t+1} & P_{t+1} \end{bmatrix} \right), \qquad \text{(B36.)}$$

$$\zeta_t | D_t \sim N\left(\begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} \right), \qquad \text{(B37.)}$$

and

$$\begin{aligned} \mathrm{Cov}(\zeta_t, \zeta'_{t+1} | D_t) &= \mathrm{Var}(\zeta_t | D_t) \bar{A}' \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} \begin{bmatrix} 0 & A'_{21} & A'_{31} \\ 0 & A'_{22} & A'_{32} \\ I & A'_{23} & A'_{33} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & Q_t A'_{23} & Q_t A'_{33} \end{bmatrix}. \end{aligned} \qquad \text{(B38.)}$$

Therefore,

$$\zeta_t | \zeta_{t+1}, D_t \sim N(h_t, H_t), \qquad \text{(B39.)}$$

where

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

and

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

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

For the remainder of the Appendix, we deal with the special case in which the coefficient matrix in (B1.) is restricted as

$$\begin{bmatrix} 0 & 0 & I \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & I \\ 0 & A & 0 \\ 0 & 0 & B \end{bmatrix}. \tag{B40.}$$

B4. Expected returns and past values

This section derives equations (8), (11), and (26). We continue to treat the multiple-asset case, in which r_t is a vector of returns. Denoting matrices by uppercase letters, we replace m by M, n by N, λ by Λ, ϕ by Φ, δ by Δ, ω by Ω, and κ by $\mathcal{K}$.

Below, we express the vector of conditional expected returns, $b_t = \mathrm{E}(r_{t+1}|D_t)$, as a function of past returns and predictors. Denote

$$[M_t \; N_t] \equiv P_t(P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1} = G_t' S_t^{-1}, \tag{B41.}$$

so that, from equation (B32.), for $t > 1$,

$$\begin{aligned} b_t &= a_t + [M_t \; N_t](z_t - f_t) \\ &= (I - B)E_r + Bb_{t-1} + [M_t \; N_t] \begin{bmatrix} r_t - b_{t-1} \\ x_t - (I - A)E_x - Ax_{t-1} \end{bmatrix} \\ &= (I - B)E_r + (B - M_t)b_{t-1} + M_t r_t + N_t v_t, \end{aligned} \tag{B42.}$$

or

$$b_t - E_r = B(b_{t-1} - E_r) + M_t(r_t - b_{t-1}) + N_t v_t. \tag{B43.}$$

For $t = 1$, we obtain

$$b_1 - E_r \;\; = \;\; M_1(r_1 - b_0) + N_1 v_1,$$

where v_1 denotes $x_1 - E_x$. Repeated substitution for the lagged values of $(b_t - E_r)$ gives

$$b_t = E_r + \sum_{s=1}^{t} \Lambda_s (r_s - b_{s-1}) + \sum_{s=1}^{t} \Phi_s v_s, \tag{B44.}$$

where

$$\Lambda_s \;\; = \;\; B^{t-s} M_s \tag{B45.}$$

$$\Phi_s \;\; = \;\; B^{t-s} N_s. \tag{B46.}$$

That is, the expected return conditional on data observed through period t can be written as the unconditional mean E_r plus a linear combination of past return forecast errors, $\epsilon_s = r_s - b_{s-1}$, plus a linear combination of past innovations in the predictors. This is equation (8) in the text.

The conditional expected return b_t can be rewritten so that past forecast errors are replaced by returns in excess of the unconditional mean E_r. To do so, modify equation (B42.) as

$$b_t - E_r = (B - M_t)(b_{t-1} - E_r) + M_t(r_t - E_r) + N_t v_t \tag{B47.}$$

so that repeated substitution for the lagged values of $(b_t - E_r)$ then yields

$$b_t = E_r + \sum_{s=1}^{t} \Omega_s (r_s - E_r) + \sum_{s=1}^{t} \Delta_s v_s \tag{B48.}$$

where

$$\Omega_s = \begin{cases} (B - M_t)(B - M_{t-1}) \cdots (B - M_{s+1}) M_s & \text{for } s < t \\ M_s & \text{for } s = t \end{cases} \tag{B49.}$$

$$\Delta_s = \begin{cases} (B - M_t)(B - M_{t-1}) \cdots (B - M_{s+1}) N_s & \text{for } s < t \\ N_s & \text{for } s = t \end{cases} \tag{B50.}$$

That is, b_t is then equal to the unconditional mean return E_r plus linear combinations of past returns in excess of E_r and past innovations in the predictors. This is equation (11) in the text.

If E_r is replaced by the sample mean in equation (11), then the estimate of b_t becomes

$$\hat{b}_t = \sum_{s=1}^{t} \mathcal{K}_s r_s + \sum_{s=1}^{t} \Delta_s v_s, \tag{B51.}$$

where

$$\mathcal{K}_s = \frac{1}{t} \left(I - \sum_{l=1}^{t} \Omega_l \right) + \Omega_s, \tag{B52.}$$

and $\sum_{s=1}^{t} \mathcal{K}_s = I$. This is a generalized version of equation (26) in the text.

In the rest of the Appendix, we discuss the special case (implemented in the paper) in which r_t is a scalar. This simplification turns μ_t, E_r, and B into scalars as well. Therefore, we now turn back to the notation from the text in which B is replaced by β and the relevant Σ's by σ's.

B5. 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.

B5.1. Prior distributions

First, we discuss the prior on (E_x, A, E_r, β). We require both x_t and μ_t to be stationary, so that all eigenvalues of A must lie inside the unit circle and $\beta \in (-1, 1)$. Apart from this restriction, our prior is noninformative about A but informative about β, $\beta \sim N(0.99, 0.15^2)$ (see Figure 5). We put a mildly informative prior on E_r, $E_r \sim N(\bar{\mu}, \sigma^2_{E_r})$, centered at the sample mean return with a large prior standard deviation of 1% per quarter. We use a noninformative prior for E_x, $E_x \sim N(0, \sigma^2_{E_x} I_K)$ with a large σ_{E_x}. All four parameters, A, β, E_μ, and E_x, are independent a priori.

The prior on Σ is more complicated. Recall that, with r_t being a scalar, Σ is defined as

$$\Sigma \equiv \begin{bmatrix} \sigma^2_u & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \Sigma_{vv} & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma^2_w \end{bmatrix}.$$

We divide the elements of Σ into two subsets: first, the 2×2 submatrix Σ_{11}, where

$$\Sigma_{11} \equiv \begin{bmatrix} \sigma^2_u & \sigma_{uw} \\ \sigma_{wu} & \sigma^2_w \end{bmatrix},$$

and second, the elements of Σ that involve v: $\Sigma_{(v)} \equiv (\Sigma_{vv}, \sigma_{vu}, \sigma_{vw})$. We choose a prior that is informative about Σ_{11} but noninformative about $\Sigma_{(v)}$. Such a prior is obtained as a posterior of Σ when a noninformative prior is updated with a hypothetical sample in which there are T_0 observations of (u, w) but only $S_0 \ll T_0$ observations of v (see Stambaugh, 1997). We choose T_0 equal to one fifth of the sample size, which makes the prior on Σ_{11} informative (five times less informative than the actual sample). We choose $S_0 = K + 3$, where K is the number of predictors, which makes the prior on $\Sigma_{(v)}$ virtually noninformative (as informative as a sample of only $K + 3$ observations, where $K = 1$ or 3).

The prior on Σ_{11} is inverted Wishart, $\Sigma_{11} \sim IW(T_0 \hat{\Sigma}_{11,0}, T_0 - K)$, so the prior mean is $\text{E}(\Sigma_{11}) = \hat{\Sigma}_{11,0}\,(T_0/(T_0 - K - 3))$. Denote the (i, j) element of $\hat{\Sigma}_{11,0}$ by M_{ij}, for $i = 1, 2$ and $j = 1, 2$. The value of M_{11} is chosen such that the prior mean of σ^2_u is equal to 95% of the sample variance of market returns. The value of M_{22} is chosen to deliver the prior mean of σ^2_w which, combined with β of 0.97, sets the variance of μ_t equal to 5% of the sample variance of market returns. These values of M_{11} and M_{22} lead to a prior for the R^2 from the regression of r_{t+1} on μ_t that we find plausible (see Figure 5). To be able to put different priors on ρ_{uw} while keeping the same prior on σ^2_u and σ^2_w, we adopt a hyperparameter approach. We assume that M_{12} is an unknown hyperparameter with a uniform prior distribution on the interval $(-\underline{c}\sqrt{M_{11}M_{22}}, \overline{c}\sqrt{M_{11}M_{22}})$. Since the prior mean of ρ_{uw} is approximately equal to $M_{12}/\sqrt{M_{11}M_{22}}$, this prior mean is approximately uniformly distributed as $U(-\underline{c}, \overline{c})$. For all three priors on ρ_{uw}, we specify $\underline{c} = -0.90$ and we vary $\overline{c}$ as follows: 0.9 for the noninformative prior, -0.35 for the less informative prior, and -0.87 for the more informative prior. These choices produce the priors on ρ_{uw} plotted in Figure 5.

The prior on $\Sigma_{(v)}$ is obtained by changing variables from $(\Sigma_{vv}, \sigma_{vu}, \sigma_{vw})$ to the slope C $(K \times 2)$ and the residual covariance matrix Ω $(K \times K)$ from the regression of v_t on (u_t, w_t), with zero intercept. That is, $C = [\sigma_{vu}\ \sigma_{vw}]\Sigma_{11}^{-1}$, and $\Omega = \Sigma_{vv} - C\Sigma_{11}C'$. We then put a normal-inverted-Wishart prior on C and Ω: $\Omega \sim IW(S_0\hat{\Omega}_0, S_0)$ and vec $(C)|\Omega \sim N(\hat{c}_0, \Omega \otimes (X_0'X_0)^{-1})$, where $\hat{\Omega}_0$, $\hat{c}_0$, and $X_0'X_0$ represent estimates from the S_0 periods in the hypothetical sample in which both v_t and (u_t, w_t) are available. The choices of $\hat{\Omega}_0$ and $\hat{c}_0$ are inconsequential because they represent means of distributions with large variances. We choose a very small value for S_0, as explained above, so the prior variance of Ω is large . We then choose the 2×2 matrix $X_0'X_0$ equal to a small positive number times the identity matrix, so $(X_0'X_0)^{-1}$, and thus the prior variance of C, is large. As a result, the priors on C and Ω are noninformative.

As mentioned above, these priors on Σ_{11}, C, and Ω can be thought of as posteriors. After changing variables from Σ to (Σ_{11}, C, Ω), the diffuse prior on Σ, $p(\Sigma) \propto |\Sigma|^{-(K+3)/2}$, translates into $p(\Sigma_{11}, C, \Omega) \propto |\Sigma_{11}|^{(K-3)/2}|\Omega|^{-(K+3)/2}$. When this noninformative prior is updated with the hypothetical sample of T_0 observations of (u, w) and S_0 observations of v, the posteriors of Σ_{11}, C, and Ω are exactly the same as the priors described above. See Stambaugh (1997), with the additional restriction that the population means of u_t and w_t are zero.

B5.2. Posterior distributions

B5.2.1. Drawing (E_x, A, E_r, β) given Σ

Equations (4) and (5) can be written as

$$\underbrace{\begin{pmatrix} x_{t+1} \\ \mu_{t+1} \end{pmatrix}}_{q_{t+1}} - \underbrace{\begin{pmatrix} A & 0 \\ 0 & \beta \end{pmatrix}}_{L_1} \underbrace{\begin{pmatrix} x_t \\ \mu_t \end{pmatrix}}_{q_t} - \underbrace{\begin{pmatrix} I_K - A & 0 \\ 0 & 1-\beta \end{pmatrix}}_{L_2} \underbrace{\begin{pmatrix} E_x \\ E_\mu \end{pmatrix}}_{E_{x\mu}} = \begin{pmatrix} v_{t+1} \\ w_{t+1} \end{pmatrix},$$

where the covariance matrix of the residuals is

$$\Sigma_{(vw)} \equiv \begin{bmatrix} \Sigma_{vv} & \sigma_{vw} \\ \sigma_{wv} & \sigma_w^2 \end{bmatrix}.$$

The prior for $E_{x\mu}$ is

$$E_{x\mu} \sim N\left(E_{x\mu_0}, V_{x\mu_0}\right),$$

where

$$E_{x\mu_0} \equiv \begin{pmatrix} 0 \\ \bar{\mu} \end{pmatrix}$$
$$V_{x\mu_0} \equiv \begin{pmatrix} \sigma^2_{E_x} I_K & 0 \\ 0 & \sigma^2_{E_\mu} \end{pmatrix}.$$

Since both the prior and the likelihood are normally distributed, the full conditional posterior distribution of $E_{x\mu}$ is also normal,

$$E_{x\mu}|\cdot \sim N\left(\tilde{E}_{x\mu}, \tilde{V}_{x\mu}\right), \tag{B53.}$$

where $\tilde{V}_{x\mu} = (V^{-1}_{x\mu_0} + T L_2' \Sigma^{-1}_{(vw)} L_2)^{-1}$ and $\tilde{E}_{x\mu} = \tilde{V}_{x\mu}\left[V^{-1}_{x\mu_0} E_{x\mu_0} + L_2' \Sigma^{-1}_{(vw)} \sum_{t=1}^{T} (q_{t+1} - L_1 q_t)\right]$.

Let $x^k \equiv (x^k_2, \ldots, x^k_T)'$ denote the $(T-1)\times 1$ vector of realizations of predictor k in periods $2, \ldots, T$, for $k = 1, \ldots, K$. Also, let $x_{(l)}$ denote the $(T-1)\times K$ vectors of realizations of all K predictors in periods $1, \ldots, T-1$. Similarly, let $\mu \equiv (\mu_2, \ldots, \mu_T)'$ and $\mu_{(l)} \equiv (\mu_1, \ldots, \mu_{T-1})'$, and let E_{x^k} be the k-th element of E_x. Denote

$$z = \begin{pmatrix} x^1 - \iota_{T-1} E_{x^1} \\ \vdots \\ x^K - \iota_{T-1} E_{x^K} \\ \mu - \iota_{T-1} E_\mu \end{pmatrix}, \quad Z = \begin{pmatrix} x_{(l)} - \iota_{T-1} E_x' & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & x_{(l)} - \iota_{T-1} E_x' & 0 \\ 0 & 0 & 0 & \mu_{(l)} - \iota_{T-1} E_\mu \end{pmatrix},$$

where ι_{T-1} is a $(T-1)\times 1$ vector of ones, the dimensions of z are $[(T-1)(K+1)]\times 1$, and the dimensions of Z are $[(T-1)(K+1)]\times(K^2+1)$. Then we can write the equations (4) and (5) as

$$z = Zb + \text{ errors },$$

where $b = (\text{vec}\,(A')'\ \beta)'$ and the covariance matrix of the error terms is $\Sigma_{(vw)} \otimes I_{T-1}$. The prior distribution on b is given by

$$b \sim N\,(b_0, V_{b_0}) \times 1_{b\in S},$$

where b_0 and V_{b_0} are chosen as explained earlier and $1_{b\in S}$ is equal to one when x_t and μ_t are stationary and zero otherwise. Let $\hat{V}_b = \left[Z'(\Sigma^{-1}_{(vw)} \otimes I_{T-1})Z\right]^{-1}$ and $\hat{b} = \hat{V}_b Z'(\Sigma^{-1}_{(vw)} \otimes I_{T-1})z$. The full conditional posterior distribution of b is then given by

$$b|\cdot \sim N\left(\tilde{b}, \tilde{V}_b\right) \times 1_{b\in S}, \tag{B54.}$$

where $\tilde{V}_b = (V^{-1}_{b_0} + \hat{V}^{-1}_b)^{-1}$ and $\tilde{b} = \tilde{V}_b\left(V^{-1}_{b_0} b_0 + \hat{V}^{-1}_b \hat{b}\right)$. We obtain the posterior draws of b by making draws from $N\left(\tilde{b}, \tilde{V}_b\right)$ and retaining only draws that satisfy $b \in S$. The posterior draws of A and β are constructed from the posterior draws of b from the definition $b = (\text{vec}\,(A')'\ \beta)'$.

B5.2.2. Drawing Σ given (E_x, A, E_r, β)

Recall that we change variables from Σ, where

$$\Sigma \equiv \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \Sigma_{vv} & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix},$$

to the set of (Σ_{11}, C, Ω), where

$$\Sigma_{11} \equiv \begin{bmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{wu} & \sigma_w^2 \end{bmatrix},$$

and C and Ω are the slope and the residual covariance matrix from the regression of v on (u, w).

The prior for Σ_{11} is conditional on the hyperparameter M_{12}. This hyperparameter can be drawn from its full conditional posterior density, $p(M_{12}|\cdot, D_t)$, which is given by

$$p(M_{12}|\Sigma_{11}) \propto |\hat{\Sigma}_{11,0}|^{\frac{T_0-K}{2}} \exp\left\{-\frac{T_0}{2}\mathrm{tr}(\Sigma_{11}^{-1}\hat{\Sigma}_{11,0})\right\}, \quad M_{12} \in (-\underline{c}\sqrt{M_{11}M_{22}}, \overline{c}\sqrt{M_{11}M_{22}}), \tag{B55.}$$

where M_{12} is the $(1, 2)$ element of $\hat{\Sigma}_{11,0}$. Although this is not a density of a well known distribution, we can make posterior draws of M_{12} easily. We approximate this density by a piecewise linear function, using a fine (250-point) grid on the interval $(-\underline{c}\sqrt{M_{11}M_{22}}, \overline{c}\sqrt{M_{11}M_{22}})$. For a random draw $z \sim U(0, 1)$, we find the points on the grid whose cumulative probability densities are immediately above and below z, and we compute the value of M_{12} by linear interpolation.

Conditional on M_{12}, we have the matrix $\hat{\Sigma}_{11,0}$ in the prior distribution for Σ_{11}. In addition, conditional on (E_x, A, E_r, β), we have the sample of the residuals (u_t, v_t, w_t), $t = 1, \dots, T$. Let X denote the $T \times 2$ matrix of $[u_t\ w_t]$, let $Y_{2,T}$ denote the $T \times K$ matrix of v_t. The sample estimates from the regression of $Y_{2,T}$ on X are given by $\hat{C} = (X'X)^{-1}X'Y_{2,T}$, $\hat{\Omega} = (Y_{2,T} - X\hat{C})'(Y_{2,T} - X\hat{C})/T$, and $\hat{\Sigma}_{11} = X'X/T$. The posterior of Σ_{11} has an inverted Wishart distribution:

$$\Sigma_{11}|\cdot \sim IW(T_0\hat{\Sigma}_{11,0} + T\hat{\Sigma}_{11}, T + T_0 - K). \tag{B56.}$$

In addition, let $V_C = (X_0'X_0 + X'X)^{-1}$, $\tilde{C} = V_C\left[(X_0'X_0)\hat{C}_0 + (X'X)\hat{C}\right]$, $\tilde{c} = \mathrm{vec}\,(\tilde{C})$, and $D = \hat{C}_0'X_0'X_0\hat{C}_0 + \hat{C}'X'X\hat{C} - \tilde{C}'V_C^{-1}\tilde{C}$. The posterior of Ω has an inverted Wishart distribution:

$$\Omega|\cdot \sim IW(S_0\hat{\Omega}_0 + T\hat{\Omega} + D, T + S_0), \tag{B57.}$$

and the conditional posterior of $c = \mathrm{vec}\,(C)$ is normal:

$$c|\Omega, \cdot \sim N(\tilde{c}, \Omega \otimes V_C). \tag{B58.}$$

Given the posterior draws of (Σ_{11}, C, Ω), we construct the remaining (non-Σ_{11}) elements of Σ as follows: $[\sigma_{vu}\ \sigma_{vw}] = C\Sigma_{11}$ and $\Sigma_{vv} = \Omega + C\Sigma_{11}C'$.

Our inference is based on 25,000 draws from the posterior distribution. First, we generate a sequence of 76,000 draws. We discard the first 1,000 draws as a "burn-in" and take every third draw from the rest to obtain a series of 25,000 draws that exhibit little serial correlation. The posterior draws of the relevant quantities such as ρ_{uw}, $\rho_{x\mu}$, $R^2(\mu_t \text{ on } x_t)$, $R^2(r_{t+1} \text{ on } \mu_t)$, etc. are constructed easily from the posterior draws of the basic parameters in the model.

B6. Maximum likelihood estimation

Denote the variance-covariance matrix of the disturbances in equations (4) and (28) as

$$\text{Cov}(\begin{bmatrix} \epsilon_t \\ v_t \end{bmatrix}, \begin{bmatrix} \epsilon_t & v_t \end{bmatrix}) = \Sigma^* = \begin{bmatrix} \sigma_\epsilon^2 & \sigma_{v\epsilon}' \\ \sigma_{v\epsilon} & \Sigma_{vv} \end{bmatrix}. \tag{B59.}$$

Maximum likelihood estimates are computed as the values of E_z, β, m, n, A, σ_ϵ^2, $\sigma_{v\epsilon}$, and Σ_{vv} that minimize

$$-2\ln L = \sum_{t=1}^{T} \left[\ln \left| V_{t|t-1} \right| + (z_t - \hat{z}_{t|t-1})' V_{t|t-1}^{-1} (z_t - \hat{z}_{t|t-1}) \right], \tag{B60.}$$

where $\hat{z}_{1|0} = E_z$,

$$V_{1|0} = \begin{bmatrix} \sigma_r^2 & \sigma_{xr}' \\ \sigma_{xr} & V_{xx} \end{bmatrix},$$

$$\begin{aligned}
\sigma_r^2 &= (1-\beta^2)^{-1} \left[n'\Sigma_{vv}n + (1-\beta^2+m^2)\sigma_\epsilon^2 + 2m\sigma_{v\epsilon}'n \right], \\
\sigma_{xr} &= (I-\beta A)^{-1} \left[A\Sigma_{vv}n + [I-(\beta-m)A]\sigma_{v\epsilon} \right], \\
\hat{z}_{t|t-1} &= E_z + F_{11}(z_{t-1}-E_z) + F_{12}\Sigma^* V_{t-1|t-2}^{-1} \left(z_{t-1} - \hat{z}_{t-1|t-2} \right), \quad t = 2, \ldots, T, \\
V_{t|t-1} &= F_{12} \left(\Sigma^* - \Sigma^* V_{t-1|t-2}^{-1} \Sigma^* \right) F_{12}' + \Sigma^*, \quad t = 2, \ldots, T,
\end{aligned}$$

$$F_{11} = \begin{bmatrix} \beta & 0 \\ 0 & A \end{bmatrix}, \quad F_{12} = \begin{bmatrix} -(\beta-m) & n' \\ 0 & 0 \end{bmatrix},$$

and V_{xx} is given by (B9.), (B10.), and (B40.).

B7. The R^2 ratios

The numerator of the R^2 ratio in equation (29) is computed as

$$R^2(\mu_t \text{ on } x_t) = \frac{\text{Var}(\text{E}(\mu_t|x_t))}{\text{Var}(\mu_t)} = \frac{\text{Var}(\text{E}(\mu_t) + V_{\mu x}V_{xx}^{-1}(x_t - \text{E}(x_t)))}{\text{Var}(\mu_t)} = \frac{V_{\mu x}V_{xx}^{-1}V_{\mu x}'}{V_{\mu\mu}}, \tag{B61.}$$

where V_{xx}, $V_{\mu\mu}$, and $V_{x\mu}$ are given by (B9.), (B10.), and (B40.).

The denominator of the R^2 ratio in equation (29) is computed as

$$R^2(\mu_t \text{ on } D_t) = \frac{\text{Var}(\text{E}(\mu_t|D_t))}{\text{Var}(\mu_t)} = \frac{\text{Var}(\mu_t) - \text{Var}(\mu_t|D_t)}{\text{Var}(\mu_t)} = 1 - \frac{Q_t}{V_{\mu\mu}}, \tag{B62.}$$

where Q_t is given in equation (B34.). We replace Q_t by its steady-state value, Q, which can be shown to be equal to a solution of a quadratic equation:

$$\begin{aligned} Q &= \frac{\sqrt{\xi_1^2 - 4\xi_2} - \xi_1}{2}, \qquad\qquad \text{(B63.)} \\ \xi_1 &= (1-\beta^2)(\sigma_u^2 - \sigma_{uv}\Sigma_{vv}^{-1}\sigma_{vu}) + 2\beta(\sigma_{uw} - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vu}) - (\sigma_w^2 - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vw}) \\ &= (1-\beta^2)\text{Var}(u|v) + 2\beta\text{Cov}(u, w|v) - \text{Var}(w|v) \\ \xi_2 &= (\sigma_{uw} - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vu})^2 - (\sigma_u^2 - \sigma_{uv}\Sigma_{vv}^{-1}\sigma_{vu})(\sigma_w^2 - \sigma_{wv}\Sigma_{vv}^{-1}\sigma_{vw}) \\ &= \text{Cov}(u, w|v)^2 - \text{Var}(u|v)\text{Var}(w|v) < 0 \end{aligned}$$

The value of Q is also used in computing the steady-state values of M_t and N_t from equation (B41.), denoted by m_t and n_t in the scalar case:

$$m = (\beta Q + \text{Cov}(u, w|v))(Q + \text{Var}(u|v))^{-1} \tag{B64.}$$

$$n = (\sigma_{wv} - m\sigma_{uv})\Sigma_{vv}^{-1}. \tag{B65.}$$

B8. Variance decomposition of expected return

In equation (34), the conditional expected return μ_t depends on three time-varying variables:

1. C1 $= x_t$, the current predictor values
2. C2 $= \sum_{i=0}^{\infty} \beta^i u_{t-i}$, an infinite sum of current and lagged unexpected returns
3. C3 $= \sum_{i=0}^{\infty} \left(\beta^i I_K - A^i\right) v_{t-i}$, an infinite sum of current and lagged predictor innovations ,

plus an error term. In the variance decomposition in Table IV, we consider regressions of μ_t on various subsets of $(C1, C2, C3)$. Let C denote a given subset of $(C1, C2, C3)$. The R^2 from the regression of μ_t on C is equal to

$$R^2(\mu_t \text{ on } C) = \frac{V_{\mu C}' V_C^{-1} V_{\mu C}}{V_{\mu\mu}}. \tag{B66.}$$

The matrix V_C, the covariance matrix of C, is pieced together from

$$\text{Var(C1)} = V_{xx}$$

$$
\begin{aligned}
\text{Var(C2)} &= \sigma_u^2(1-\beta^2)^{-1} \\
\text{vec}\,(\text{Var(C3)}) &= \left[(1-\beta^2)^{-1} I_{K^2} - (I_K-\beta A)^{-1} \otimes I_K - I_K \otimes (I_K-\beta A)^{-1} + \right. \\
&\quad \left. + \, (I_{K^2} - A \otimes A)^{-1}\right] \text{vec}\,(\Sigma_{vv}) \\
\text{Cov(C1, C2)} &= (I_K-\beta A)^{-1}\sigma_{vu} \\
\text{Cov(C2, C3)} &= \left[(1-\beta^2)^{-1} I_K - (I_K-\beta A)^{-1}\right]\sigma_{vu} \\
\text{vec}\,(\text{Cov(C1, C3}')) &= \left[I_K \otimes (I_K-\beta A)^{-1} + (I_{K^2} - A \otimes A)^{-1}\right] \text{vec}\,(\Sigma_{vv}),
\end{aligned}
$$

and $V_{\mu C}$, the vector of covariances between μ_t and C, is built from

$$
\begin{aligned}
\text{Cov}(\mu_t, \text{C1}') &= \Psi_v \text{Var(C1)} + \Psi_u \text{Cov(C1, C2)}' + \Psi_v \text{Cov(C1, C3}')' \\
\text{Cov}(\mu_t, \text{C2}) &= \Psi_u \text{Var(C2)} + \Psi_v \text{Cov(C1, C2)} + \Psi_v \text{Cov(C2, C3)} \\
\text{Cov}(\mu_t, \text{C3}') &= \Psi_v \text{Var(C3)} + \Psi_v \text{Cov(C1, C3}') + \Psi_u \text{Cov(C2, C3)}'.
\end{aligned}
$$