B1. Roadmap

This Internet Appendix contains material of technical nature that is relevant to the published article. This material 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 $\mu_t$ 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.

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

B2. Predictive System: General framework

We begin working with multiple assets, so that $r_t$ and $\mu_t$ 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 $\mu_t$, with coefficients restricted so that $\mu_t$ is the conditional mean of $r_{t+1}$. We also assume that $x_t$ and
$\mu_t$ are stationary with means $E_x$ and $E_r$. The first-order VAR, for example, is

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

(IA.1)

The predictive system in (IA.1) 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 $\pi_t$, where $\pi_t$ has the same dimensions as $r_t$ and contains additional unobserved predictors:

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

(IA.2)

When $B_{13}$ is nonsingular, (IA.1) and (IA.2) are equivalent. It is immediate that (IA.1) implies (IA.2). To see the converse, define

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

which implies

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

(IA.4)

Pre-multiplying both sides of (IA.2) by $[B_{11} B_{12} B_{13}]$, using (IA.3) and (IA.4), gives

$$
\begin{align*}
\mu_{t+1} - E_r &= [C_{11} C_{12} C_{13}]
\begin{bmatrix}
  r_t - E_r \\
  x_t - E_x \\
  -B_{13}^{-1}B_{11}(r_t - E_r) - B_{13}^{-1}B_{12}(x_t - E_x) + B_{13}^{-1}(\mu_t - E_r)
\end{bmatrix} \\
&+ [B_{11} B_{12} B_{13}]
\begin{bmatrix}
  u_{t+1} \\
  v_{t+1} \\
  w_{t+1}
\end{bmatrix} \\
&= A_{31}(r_t - E_r) + A_{32}(x_t - E_x) + A_{33}(\mu_t - E_r) + w_{t+1}.
\end{align*}
$$

(IA.5)

where

$$
\begin{align*}
[C_{11} C_{12} C_{13}] &= [B_{11} B_{12} B_{13}]
[B_{11} B_{12} B_{13}]
[B_{21} B_{22} B_{23}]
[B_{31} B_{32} B_{33}],
\end{align*}

$$

$$
\begin{align*}
w_{t+1} &= [B_{11} B_{12} B_{13}]
\begin{bmatrix}
  u_{t+1} \\
  v_{t+1} \\
  w_{t+1}
\end{bmatrix}
\end{align*}

$$

$$
\begin{align*}
A_{31} &= C_{11} - C_{13}B_{13}^{-1}B_{11} \\
A_{32} &= C_{12} - C_{13}B_{13}^{-1}B_{12} \\
A_{33} &= C_{13}B_{13}^{-1}.
\end{align*}
$$
Combining (IA.2), (IA.3), and (IA.4) gives

\[
\begin{bmatrix}
  r_{t+1} - E_r \\
  x_{t+1} - E_x
\end{bmatrix} =
\begin{bmatrix}
  B_{11} & B_{12} & B_{13} \\
  B_{21} & B_{22} & B_{23}
\end{bmatrix}
\begin{bmatrix}
  r_t - E_r \\
  x_t - E_x \\
  \mu_t - E_r
\end{bmatrix}
+ \begin{bmatrix}
  u_{t+1} \\
  v_{t+1}
\end{bmatrix}
\]

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

(IA.6)

where

\[
A_{21} = B_{21} - B_{23}B_{13}^{-1}B_{11} \\
A_{22} = B_{22} - B_{23}B_{13}^{-1}B_{12} \\
A_{23} = B_{23}B_{13}^{-1}.
\]

(IA.7)

Combining (IA.5) and (IA.6) gives (IA.1).

We assume the disturbances in (IA.1) are distributed identically and independently across \( t \) as

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

(IA.8)

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 \( \tilde{A} \) denote the entire coefficient matrix in (IA.1), and let \( \Sigma \) denote the entire covariance matrix in (IA.8). Then

\[
V_{\zeta\zeta} = \begin{bmatrix}
  V_{rr} & V_{rx} & V_{r\mu} \\
  V_{xr} & V_{xx} & V_{x\mu} \\
  V_{r\mu} & V_{x\mu} & V_{\mu\mu}
\end{bmatrix} = \tilde{A}V_{\zeta\zeta}\tilde{A}' + \Sigma,
\]

(IA.9)

which can be solved as

\[
\text{vec} (V_{\zeta\zeta}) = [I - (\tilde{A} \otimes \tilde{A})]^{-1}\text{vec} (\Sigma),
\]

(IA.10)

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

$$E_z = \begin{bmatrix} E_r \\ E_x \end{bmatrix}, \quad V_{zz} = \begin{bmatrix} V_{rr} & V_{rx} \\ V_{xr} & V_{xx} \end{bmatrix}, \quad V_{z\mu} = \begin{bmatrix} V_{r\mu} \\ V_{x\mu} \end{bmatrix}. \quad (IA.11)$$

**B3. Drawing the Time Series of $\mu_t$**

To draw the time series of the unobservable values of $\mu_t$ 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

$$a_t = E(\mu_t | D_{t-1}) \quad b_t = E(\mu_t | D_t) \quad e_t = E(z_t | \mu_t, D_{t-1}) \quad (IA.12)$$
$$f_t = E(z_t | D_{t-1}) \quad P_t = \text{Var}(\mu_t | D_{t-1}) \quad Q_t = \text{Var}(\mu_t | D_t) \quad (IA.13)$$
$$R_t = \text{Var}(z_t | \mu_t, D_{t-1}) \quad S_t = \text{Var}(z_t | D_{t-1}) \quad G_t = \text{Cov}(z_t, \mu'_t | D_{t-1}) \quad (IA.14)$$

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

$$\mu_0 | D_0 \sim N(b_0, Q_0), \quad (IA.15)$$

where $D_0$ denotes the null information set, so that the unconditional moments of $\mu_0$ are given by $b_0 = E_r$ and $Q_0 = V_{\mu\mu}$. Also,

$$\mu_1 | D_0 \sim N(a_1, P_1), \quad (IA.16)$$

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

$$z_1 | D_0 \sim N(f_1, S_1), \quad (IA.17)$$

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

$$G_1 = V_{z\mu} \quad (IA.18)$$

and that

$$z_1 | \mu_1, D_0 \sim N(e_1, R_1), \quad (IA.19)$$

where

$$e_1 = f_1 + G_1 P_1^{-1}(\mu_1 - a_1) \quad (IA.20)$$
$$R_1 = S_1 - G_1 P_1^{-1} G_1' \quad (IA.21)$$


Combining this density with equation (IA.16) using Bayes rule gives

\[ \mu_1|D_1 \sim N(b_1, Q_1), \]  

where

\begin{align*}
    b_1 &= a_1 + P_t(P_1 + G'_tR_t^{-1}G_t)^{-1}G'_tR_t^{-1}(z_t - f_t) \quad \text{(IA.23)} \\
    Q_1 &= P_1(P_1 + G'_tR_t^{-1}G_t)^{-1}P_1. \quad \text{(IA.24)}
\end{align*}

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

\begin{align*}
    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{(IA.25)} \\
    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{(IA.26)} \\
    S_t &= \begin{bmatrix} Q_{t-1} & A'_{t-1}A_{33} \\ A_{23}Q_{t-1} & A_{23}Q_{t-1}A'_{t-1}A_{33} \\ A_{33}Q_{t-1} & A_{33}Q_{t-1}A'_{t-1}A_{33} \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix} \quad \text{(IA.27)} \\
    G_t &= \begin{bmatrix} Q_{t-1}A'_{33} \\ A_{23}Q_{t-1}A'_{33} \end{bmatrix} + \begin{bmatrix} \Sigma_{uw} \end{bmatrix} \quad \text{(IA.28)} \\
    P_t &= A_{33}Q_{t-1}A'_{33} + \Sigma_{ww} \quad \text{(IA.29)} \\
    e_t &= f_t + G_tP_t^{-1}(\mu_t - a_t) \quad \text{(IA.30)} \\
    R_t &= S_t - G_tP_t^{-1}G'_t \quad \text{(IA.31)} \\
    b_t &= a_t + P_t(P_t + G'_tR_t^{-1}G_t)^{-1}G'_tR_t^{-1}(z_t - f_t) \quad \text{(IA.32)} \\
    &= a_t + G'_tS_t^{-1}(z_t - f_t) \quad \text{(IA.33)} \\
    Q_t &= P_t(P_t + G'_tR_t^{-1}G_t)^{-1}P_t. \quad \text{(IA.34)}
\end{align*}

The values of \( \{a_t, b_t, Q_t, S_t, G_t, P_t\} \) for \( t = 1, \ldots, T \) are retained for the next stage. Equations (IA.27) through (IA.29) are derived as

\[
\begin{bmatrix} S_t & G_t \\ G'_t & P_t \end{bmatrix} = \text{Var}(\zeta_t|D_{t-1})
\]

\[
= \tilde{A}\text{Var}(\zeta_{t-1}|D_{t-1})\tilde{A}' + \Sigma
\]

\[
= \tilde{A}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_{t-1} \end{bmatrix} \tilde{A}' + \Sigma
\]

\[
= \begin{bmatrix} Q_{t-1} & Q_{t-1}A'_{23} & Q_{t-1}A'_{33} \\ A_{23}Q_{t-1} & A_{23}Q_{t-1}A'_{23} & A_{23}Q_{t-1}A'_{33} \\ A_{33}Q_{t-1} & A_{33}Q_{t-1}A'_{23} & A_{33}Q_{t-1}A'_{33} \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uw} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vw} \\ \Sigma_{wu} & \Sigma_{ww} & \Sigma_{ww} \end{bmatrix}.
\]
B3.2. Sampling

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

\[
p(\zeta_0, \zeta_1, \ldots, \zeta_T|D_T) = p(\zeta_T|D_T)p(\zeta_{T-1}|\zeta_T, D_{T-1}) \cdots p(\zeta_1|\zeta_2, D_1)p(\zeta_0|\zeta_1, D_0). \tag{IA.35}
\]

We first sample \(\mu_T\) 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 \(\mu_t\) from the conditional density \(p(\zeta_t|\zeta_{t+1}, D_t)\). (Note that the first two subvectors of \(\zeta_t\) are already observed and thus need not be sampled.) To obtain that conditional density, first note that

\[
\zeta_{t+1}|D_t \sim N\left( \begin{bmatrix} f_{t+1} \\ a_{t+1} \end{bmatrix}, \begin{bmatrix} S_{t+1} & G_{t+1} \\ G_{t+1}^T & P_{t+1} \end{bmatrix} \right), \tag{IA.36}
\]

and

\[
\zeta_t|D_t \sim N\left( \begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & Q_t & 0 \end{bmatrix} \right). \tag{IA.37}
\]

and

\[
\text{Cov}(\zeta_t, \zeta_{t+1}|D_t) = \text{Var}(\zeta_t|D_t) \tilde{A}'
\]

\[
= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & Q_t & 0 \end{bmatrix}
\begin{bmatrix} 0 & A_{21}' & A_{31}' \\ 0 & A_{22}' & A_{32}' \\ I & A_{23}' & A_{33}' \end{bmatrix}
\]

\[
= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & Q_tA_{23}' & Q_tA_{33}' \end{bmatrix}. \tag{IA.38}
\]

Therefore,

\[
\zeta_t|\zeta_{t+1}, D_t \sim N(h_t, H_t), \tag{IA.39}
\]

where

\[
h_t = E(\zeta_t|D_t) + \text{Cov}(\zeta_t, \zeta_{t+1}|D_t) \left[ \text{Var}(\zeta_{t+1}|D_t) \right]^{-1} \left[ \zeta_{t+1} - E(\zeta_{t+1}|D_t) \right]
\]

\[
= \begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & Q_tA_{23}' & Q_tA_{33}' \end{bmatrix}
\begin{bmatrix} S_{t+1} & G_{t+1} \\ G_{t+1}^T & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} z_{t+1} - f_{t+1} \\ \mu_{t+1} - a_{t+1} \end{bmatrix}
\]

and

\[
H_t = \text{Var}(\zeta_t|D_t) - \text{Cov}(\zeta_t, \zeta_{t+1}|D_t) \left[ \text{Var}(\zeta_{t+1}|D_t) \right]^{-1} \left[ \text{Cov}(\zeta_t, \zeta_{t+1}|D_t) \right]'
\]

\[
= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & Q_t \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & Q_tA_{23}' & Q_tA_{33}' \end{bmatrix}
\begin{bmatrix} S_{t+1} & G_{t+1} \\ G_{t+1}^T & P_{t+1} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & Q_t \\ 0 & 0 & A_{23}Q_t \\ 0 & 0 & A_{33}Q_t \end{bmatrix}
\]

\[6\]
The mean and covariance matrix of \( \mu_t \) 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 (IA.1) is restricted as

\[
\begin{bmatrix}
0 & 0 & I \\
A_{21} & A_{22} & A_{23} \\
A_{31} & A_{32} & A_{33}
\end{bmatrix} = \begin{bmatrix}
0 & 0 & I \\
0 & A & 0 \\
0 & 0 & B
\end{bmatrix}. \tag{IA.40}
\]

### 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 \( N \), \( \lambda, \phi \) by \( \Delta, \omega \) by \( \Omega \), and \( \kappa \) by \( K \).

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

\[
[M_t, N_t] = P_t(P_t + G_t R_t^{-1} G_t)^{-1} G_t' R_t^{-1} = G_t' S_t^{-1}, \tag{IA.41}
\]

so that, from equation (IA.32), for \( t > 1 \),

\[
b_t = a_t + [M_t, N_t](z_t - f_t) \\
= (I - B) \text{E}_r + B b_{t-1} + [M_t, N_t] \begin{bmatrix}
r_t - b_{t-1} \\
x_t - (I - A) \text{E}_x - A x_{t-1}
\end{bmatrix} \\
= (I - B) \text{E}_r + (B - M_t) b_{t-1} + M_t r_t + N_t v_t. \tag{IA.42}
\]

or

\[
b_t - \text{E}_r = B(b_{t-1} - \text{E}_r) + M_t(r_t - b_{t-1}) + N_t v_t. \tag{IA.43}
\]

For \( t = 1 \), we obtain

\[
b_1 - \text{E}_r = M_1(r_1 - b_0) + N_1 v_1,
\]

where \( v_1 \) denotes \( x_1 - \text{E}_x \). Repeated substitution for the lagged values of \( (b_t - \text{E}_r) \) gives

\[
b_t = \text{E}_r + \sum_{s=1}^{t} \Lambda_s (r_s - b_{s-1}) + \sum_{s=1}^{t} \Phi_s v_s, \tag{IA.44}
\]

where

\[
\Lambda_s = B^{t-s} M_s \tag{IA.45} \\
\Phi_s = B^{t-s} N_s. \tag{IA.46}
\]
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 (IA.42) as

\[
 b_t - E_r = (B - M_t)(b_{t-1} - E_r) + M_t(r_t - E_r) + N_t v_t \tag{IA.47}
\]

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

\[
 b_t = E_r + \sum_{s=1}^{t} \Omega_s (r_s - E_r) + \sum_{s=1}^{t} \Delta_s v_s \tag{IA.48}
\]

where

\[
 \Omega_s = \begin{cases} 
 (B - M_t)(B - M_{t-1}) \cdots (B - M_{s+1}) M_s & \text{for } s < t \\
 M_s & \text{for } s = t 
\end{cases} \tag{IA.49}
\]

\[
 \Delta_s = \begin{cases} 
 (B - M_t)(B - M_{t-1}) \cdots (B - M_{s+1}) N_s & \text{for } s < t \\
 N_s & \text{for } s = t 
\end{cases} \tag{IA.50}
\]

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 (11), then the estimate of \( b_t \) becomes

\[
 \hat{b}_t = \sum_{s=1}^{t} \mathcal{K}_s r_s + \sum_{s=1}^{t} \Delta_s v_s, \tag{IA.51}
\]

where

\[
 \mathcal{K}_s = \frac{1}{t} \left( I - \sum_{l=1}^{t} \Omega_l \right) + \Omega_s, \tag{IA.52}
\]

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 \( \mu_t, E_r, \) and \( B \) into scalars as well. Therefore, we now turn back to the notation from the text in which \( B \) is replaced by \( \beta \) and the relevant \( \Sigma \)'s by \( \sigma \)'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 \( \mu_t \).

**B5.1. Prior distributions**

First, we discuss the prior on \( (E_x, A, E_r, \beta) \). We require both \( x \) and \( \mu_t \) 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 $\beta$, $\beta \sim N(0.99, 0.15^2)$ (see Figure 5). We put a mildly informative prior on $E_r$, $E_r \sim N(\bar{\mu}, \sigma_{E_r}^2)$, centered at the sample mean return with a large prior standard deviation of 1% per quarter. We use a noninformative prior for $E_x$, $E_x \sim N(0, \sigma_{E_x}^2 I_K)$ with a large $\sigma_{E_x}$. All four parameters, $A$, $\beta$, $E_\mu$, and $E_x$, are independent a priori.

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

$$
\Sigma = \begin{bmatrix}
\sigma_{uu}^2 & \sigma_{uw} & \sigma_{mw} \\
\sigma_{uw} & \Sigma_{vv} & \sigma_{vw} \\
\sigma_{mw} & \sigma_{vw} & \sigma_{ww}^2
\end{bmatrix}.
$$

We divide the elements of $\Sigma$ into two subsets: first, the $2 \times 2$ submatrix $\Sigma_{11}$, where

$$
\Sigma_{11} = \begin{bmatrix}
\sigma_{uu}^2 & \sigma_{uw} \\
\sigma_{wu} & \sigma_{ww}^2
\end{bmatrix},
$$

and second, the elements of $\Sigma$ that involve $v$: $\Sigma_{(v)} = (\Sigma_{vv}, \sigma_{vu}, \sigma_{vw})$. We choose a prior that is informative about $\Sigma_{11}$ but noninformative about $\Sigma_{(v)}$. Such a prior is obtained as a posterior of $\Sigma$ 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 $\Sigma_{11}$ 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 $\Sigma_{11}$ is inverted Wishart, $\Sigma_{11} \sim IW(T_0 \hat{\Sigma}_{11,0}, T_0 - K)$, so the prior mean is $E(\Sigma_{11}) = \hat{\Sigma}_{11,0} (T_0/ (T_0 - K - 3))$. 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 $\sigma_{uu}^2$ is equal to 95% of the sample variance of market returns. The value of $M_{22}$ is chosen to deliver the prior mean of $\sigma_{ww}^2$ which, combined with $\beta$ of 0.97, sets the variance of $\mu_t$ 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 $\mu_t$ that we find plausible (see Figure 5). To be able to put different priors on $\rho_{uw}$ while keeping the same prior on $\sigma_u^2$ and $\sigma_w^2$, we adopt a hyperparameter approach. We assume that $M_{12}$ is an unknown hyperparameter with a uniform prior distribution on the interval $(-\xi \sqrt{M_{11} M_{22}}, \eta \sqrt{M_{11} M_{22}})$. Since the prior mean of $\rho_{uw}$ is approximately equal to $M_{12}/ \sqrt{M_{11} M_{22}}$, this prior mean is approximately uniformly distributed as $U(-\xi, \eta)$. For all three priors on $\rho_{uw}$, we specify $\xi = -0.90$ and we vary $\eta$ 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 $\rho_{uw}$ 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 $\Omega$ ($K \times K$) 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 $\Omega$: $\Omega \sim IW(S_0 \hat{\Omega}_0, S_0)$ and vec $(C) | \Omega \sim N(\hat{C}_0, \Omega \otimes (X_0' X_0)^{-1})$, where
\( \hat{\Omega}_0, \hat{c}_0, \) and \( \hat{X}'_0X_0 \) 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 \( \Omega \) is large. We then choose the \( 2 \times 2 \) matrix \( X'_0X_0 \) equal to a small positive number times the identity matrix, so \( (X'_0X_0)^{-1} \), and thus the prior variance of \( C \), is large. As a result, the priors on \( C \) and \( \Omega \) are noninformative.

As mentioned above, these priors on \( \delta_{11}, C, \) and \( \Omega \) can be thought of as posteriors. After changing variables from \( \delta \) to \( \delta_{11}, C, \) the diffuse prior on \( \delta \), \( p(\delta) \propto |\Sigma|^{-(K+3)/2} \), 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 \( \delta \), the posteriors of \( \Sigma_{11}, C, \) and \( \Omega \) 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 \( (E_x, A, E_r, \beta) \) given \( \Sigma \)

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

\[
\begin{pmatrix}
  x_{t+1} \\
  \mu_{t+1}
\end{pmatrix}
= \begin{pmatrix}
  A & 0 \\
  0 & \beta
\end{pmatrix}
\begin{pmatrix}
  x_t \\
  \mu_t
\end{pmatrix}
- \begin{pmatrix}
  I_K - A & 0 \\
  0 & 1 - \beta
\end{pmatrix}
\begin{pmatrix}
  E_x \\
  E_\mu
\end{pmatrix}
= \begin{pmatrix}
  v_{t+1} \\
  w_{t+1}
\end{pmatrix},
\]

where the covariance matrix of the residuals is

\[
\Sigma_{(v w)} = \begin{bmatrix}
  \Sigma_{vv} & \sigma_{vw} \\
  \sigma_{vw} & \sigma_w^2
\end{bmatrix}.
\]

The prior for \( E_{x\mu} \) is

\[
E_{x\mu} \sim N(E_{x\mu_0}, V_{x\mu_0}),
\]

where

\[
E_{x\mu_0} = \begin{pmatrix}
  0 \\
  \tilde{\mu}
\end{pmatrix},
\]

\[
V_{x\mu_0} = \begin{pmatrix}
  \sigma_{E_x}^2 I_K & 0 \\
  0 & \sigma_{E_\mu}^2
\end{pmatrix}.
\]

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

\[
E_{x\mu}|. \sim N(\tilde{E}_{x\mu}, \tilde{V}_{x\mu}), \quad (IA.53)
\]

10
where \( \hat{V}_{x\mu} = (V_{x\mu0}^{-1} + (T-1)L_2\Sigma_{(vw)}^{-1}L_2)^{-1} \) and \( \hat{E}_{x\mu} = \hat{V}_{x\mu} \left[ V_{x\mu0}^{-1} E_{x\mu0} + L_2^{-1} \Sigma_{(vw)}^{-1} \sum_{t=1}^{T-1} (q_{t+1} - L_1 q_t) \right] \).

Let \( x^k = (x_1^k, \ldots, x_T^k)' \) 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 \( E_{x\mu} \) be the \( k \)-th element of \( E_x \). Denote

\[
\begin{pmatrix}
    x^1 - t_{T-1} E_{x^1} \\
    \vdots \\
    x^K - t_{T-1} E_{x^K} \\
    \mu - t_{T-1} E_{\mu}
\end{pmatrix},
\end{pmatrix}
\begin{pmatrix}
    x_{(l)} - t_{T-1} E_{x_{(l)}} & 0 & 0 & 0 \\
    0 & \ddots & 0 & 0 \\
    0 & 0 & x_{(l)} - t_{T-1} E_{x_{(l)}}' & 0 \\
    0 & 0 & 0 & \mu_{(l)} - t_{T-1} E_{\mu}'
\end{pmatrix},
\]

where \( t_{T-1} \) 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

\[
b \sim N (b_0, V_{b0}) \times 1_{b \in S},
\]

where \( b_0 \) and \( V_{b0} \) are chosen as explained earlier and \( 1_{b \in S} \) is equal to one when \( x_t \) and \( \mu_t \) are stationary and zero otherwise. Let \( \hat{V}_b = \left[ Z' (\Sigma_{(vw)}^{-1} \otimes I_{T-1}) Z \right]^{-1} \) and \( \hat{b} = \hat{V}_b Z' (\Sigma_{(vw)}^{-1} \otimes I_{T-1}) z \). The full conditional posterior distribution of \( b \) is then given by

\[
|b| \sim N \left( \hat{b}, \hat{V}_b \right) \times 1_{b \in S}, \quad \text{(IA.54)}
\]

where \( \hat{V}_b = (V_{b0}^{-1} + \hat{V}_b^{-1})^{-1} \) and \( \hat{b} = \hat{V}_b \left( V_{b0}^{-1} b_0 + \hat{V}_b^{-1} \hat{b} \right) \). We obtain the posterior draws of \( b \) by making draws from \( N \left( \hat{b}, \hat{V}_b \right) \) and retaining only draws that satisfy \( b \in S \). The posterior draws of \( A \) and \( \beta \) are constructed from the posterior draws of \( b \) from the definition \( b = (\text{vec} (A')' \beta)' \).

**B5.2.2. Drawing \( \Sigma \) given \( (E_x, A, E_r, \beta) \)**

Recall that we change variables from \( \Sigma \), 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 \(( \Sigma_{11}, C, \Omega) \), where

\[
\Sigma_{11} \equiv \begin{bmatrix}
    \sigma_u^2 & \sigma_{uw} \\
    \sigma_{wu} & \sigma_w^2
\end{bmatrix}.
\]
and $C$ and $\Omega$ are the slope and the residual covariance matrix from the regression of $v$ on $(u, w)$.

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

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

(IA.55)

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 $(-\sqrt{\frac{1}{M_{11}M_{22}}} \hat{c} \sqrt{M_{11}M_{22}})$. 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 $\Sigma_{11}$. In addition, conditional on $(E_x, A, E_r, \beta)$, we have the sample of the residuals $(u_t, v_t, w_t)$, $t = 2, \ldots, T$. Let $X$ denote the $(T - 1) \times 2$ matrix of $[u_t \ w_t]$, let $Y_{2,T-1}$ denote the $(T - 1) \times K$ matrix of $v_t$. The sample estimates from the regression of $Y_{2,T-1}$ on $X$ are given by $\hat{C} = (X'X)^{-1}X'Y_{2,T-1}$, $\hat{\Omega} = (Y_{2,T-1} - X\hat{C})'(Y_{2,T-1} - X\hat{C})/(T - 1)$, and $\hat{\Sigma}_{11} = X'X/(T - 1)$. The posterior of $\Sigma_{11}$ has an inverted Wishart distribution:

$$|\Sigma_{11}| \sim IW(T_0 \hat{\Sigma}_{11,0} + (T - 1)\hat{\Sigma}_{11}, T - 1 + T_0 - K).$$

(IA.56)

In addition, let $V_C = (X_0'X_0 + X'X)^{-1}$, $\tilde{C} = V_C \left[ (X_0'X_0)\hat{C}_0 + (X'X)\hat{C} \right]$, $\tilde{c} = \text{vec}(\tilde{C})$, 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 $\Omega$ has an inverted Wishart distribution:

$$|\Omega| \sim IW(S_0\hat{\Omega}_0 + (T - 1)\hat{\Omega} + D, T - 1 + S_0).$$

(IA.57)

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

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

(IA.58)

Given the posterior draws of $(\Sigma_{11}, C, \Omega)$, we construct the remaining (non-$\Sigma_{11}$) elements of $\Sigma$ as follows: $[\sigma_{uy} \ \sigma_{uw}] = C \Sigma_{11}$ and $\Sigma_{uv} = \Omega + C \Sigma_{11} C'$.

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 $\rho_{w,u}, \rho_{x_{12}}, R^2(\mu_t \text{ on } x_t), R^2(\mu_{t+1} \text{ on } \mu_t)$, 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

$$\text{Cov} \left[ \begin{array}{c} \varepsilon_t \\ v_t \end{array} \right], \left[ \begin{array}{c} \varepsilon_t \\ v_t \end{array} \right] = \Sigma^* = \left[ \begin{array}{cc} \sigma^2 & \sigma_{v\varepsilon} \\ \sigma_{v\varepsilon} & \Sigma_{vv} \end{array} \right],$$

(IA.59)
Maximum likelihood estimates are computed as the values of $E_z$, $\beta$, $m$, $n$, $A$, $\sigma_{\varepsilon}^2$, $\sigma_{\nu \varepsilon}$, and $\Sigma_{uv}$ that minimize

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

(IA.60)

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

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

$$\sigma_r^2 = (1 - \beta^2)^{-1} \left[ n' \Sigma_{uu} n + (1 - \beta^2 + m^2) \sigma_{\varepsilon}^2 + 2m \sigma_{\varepsilon \nu} n \right],$$

$$\sigma_{xr} = (I - \beta A)^{-1} \left[ A \Sigma_{uv} n + [I - (\beta - m)A] \sigma_{\nu \varepsilon} \right],$$

$$\hat{z}_{t\mid t-1} = E_z + F_{11} (z_{t-1} - E_z) + F_{12} \Sigma^* V_{t-1\mid t-2}^{-1} (z_{t-1} - \hat{z}_{t-1\mid t-2}), \quad t = 2, \ldots, T,$n

$$V_{t\mid t-1} = F_{12} (\Sigma^* - \Sigma^* V_{t-1\mid t-2}^{-1} \Sigma^*) F_{12}' + \Sigma^*, \quad t = 2, \ldots, T,$$

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

and $V_{xx}$ is given by (IA.9), (IA.10), and (IA.40).

**B7. The $R^2$ Ratios**

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

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

(IA.61)

where $V_{xx}$, $V_{\mu x}$, and $V_{x \mu}$ are given by (IA.9), (IA.10), and (IA.40).

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

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

(IA.62)

where $Q_t$ is given in equation (IA.34). We replace $Q_t$ by its steady-state value, $Q$, which can be shown to be equal to a solution of a quadratic equation:

$$Q = \frac{\sqrt{\xi_1^2 - 4 \xi_2} - \xi_1}{2},$$

(IA.63)

$$\xi_1 = (1 - \beta^2)(\sigma_u^2 - \sigma_{uv} \Sigma_{uv}^{-1} \sigma_{vu}) + 2 \beta (\sigma_{uw} - \sigma_{wv} \Sigma_{uv}^{-1} \sigma_{vu}) - (\sigma_w^2 - \sigma_{wv} \Sigma_{uv}^{-1} \sigma_{vw})$$

$$= (1 - \beta^2) \text{Var}(u \mid v) + 2 \beta \text{Cov}(u, w \mid v) - \text{Var}(w \mid v)$$

$$\xi_2 = (\sigma_{uw} - \sigma_{wv} \Sigma_{uv}^{-1} \sigma_{vu})^2 - (\sigma_u^2 - \sigma_{uv} \Sigma_{uv}^{-1} \sigma_{vu})(\sigma_w^2 - \sigma_{wv} \Sigma_{uv}^{-1} \sigma_{vw})$$

$$= \text{Cov}(u, w \mid v)^2 - \text{Var}(u \mid v) \text{Var}(w \mid v) < 0$$

13
The value of $Q$ is also used in computing the steady-state values of $M_t$ and $N_t$ from equation (IA.41), denoted by $m_t$ and $n_t$ in the scalar case:

$$m = (\beta Q + \text{Cov}(u, w|v))(Q + \text{Var}(u|v))^{-1}$$  \hspace{1cm} (IA.64)

$$n = (\sigma_{uw} - m\sigma_{uv})\Sigma_{vv}^{-1}.$$  \hspace{1cm} (IA.65)

**B8. Variance Decomposition of Expected Return**

In equation (34), the conditional expected return $\mu_t$ 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. $C3 = \sum_{i=0}^{\infty} (\beta^i I_K - A^i) v_{t-i}$, an infinite sum of current and lagged predictor innovations,

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

$$R^2(\mu_t \text{ on } C) = \frac{V' \mu C V^{-1} V \mu C}{V \mu \mu}.$$  \hspace{1cm} (IA.66)

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

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

$$\text{Var}(C2) = \sigma_u^2 (1 - \beta^2)^{-1},$$

$$\text{vec} \left( \text{Var}(C3) \right) = \left[ (1 - \beta^2)^{-1} I_{K^2} - (I_K - \beta A)^{-1} \otimes I_K - I_K \otimes (I_K - \beta A)^{-1} + (I_{K^2} - A \otimes A)^{-1} \right] \text{vec} (\Sigma_{vu}),$$

$$\text{Cov}(C1, C2) = (I_K - \beta A)^{-1} \sigma_{uv},$$

$$\text{Cov}(C2, C3) = \left[ (1 - \beta^2)^{-1} I_K - (I_K - \beta A)^{-1} \right] \sigma_{uv},$$

$$\text{vec} \left( \text{Cov}(C1, C3') \right) = \left[ I_K \otimes (I_K - \beta A)^{-1} + (I_{K^2} - A \otimes A)^{-1} \right] \text{vec} (\Sigma_{vu}),$$

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

$$\text{Cov}(\mu_t, C1') = \Psi_v \text{Var}(C1) + \Psi_u \text{Cov}(C1, C2) + \Psi_v \text{Cov}(C1, C3')'$$

$$\text{Cov}(\mu_t, C2) = \Psi_u \text{Var}(C2) + \Psi_v \text{Cov}(C1, C2) + \Psi_v \text{Cov}(C2, C3)$$

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