Complete Appendices Supplement to
Bayesian Alphas and Mutual Fund Persistence
by Jeffrey A. Busse and Paul J. Irvine

Appendix A

PS (2002a) show that

$$\left(\tilde{\alpha}_N, \tilde{\beta}_N\right)' = \left(I \otimes \left(D + Z'Z\right)^{-1} Z'Z\right) vec\left(\hat{G}\right), \tag{A1}$$

where

$$D = \begin{bmatrix} \dfrac{s^2}{\sigma^2_{\alpha N}} & 0 \\ 0 & 0 \end{bmatrix}, \tag{A2}$$

$$Z = \left(i \quad r_B\right), \tag{A3}$$

and

$$\hat{G} = \left(Z'Z\right)^{-1} Z'r_N . \tag{A4}$$

They also show that

$$\left(\delta_A, c'_A\right) = \left(\Lambda_0 + Z'_A Z_A\right)^{-1}\left(\Lambda_0 \phi_0 + Z'_A r_A\right), \tag{A5}$$

where

$$\Lambda_0 = k\begin{bmatrix} \sigma^2_\delta & 0 \\ 0 & \Phi_c \end{bmatrix}, \tag{A6}$$

$$Z_A = \left(i \quad r_N \quad r_B\right), \tag{A7}$$

$$\phi_0 = \left(\delta_0, c'_0\right)', \tag{A8}$$

and k is an estimate of $E\left(\sigma^2_u\right)$, described in the text.

PS (2002a) show that, conditional on the data,

$$\text{var}\left(\tilde{\alpha}_A\right) = tr\left(V_{\phi_A} V_d\right) + \tilde{d}\,'V_{\phi_A}\tilde{d} + \tilde{\phi}'_A V_d \tilde{\phi}_A, \tag{A9}$$

where

$$V_d = \begin{bmatrix} 0 & 0 & 0 \\ 0 & V_{\alpha_N} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{A10}$$

$$\tilde{d} = \begin{bmatrix} 1 \\ \tilde{\alpha}_N \\ 0 \end{bmatrix}, \tag{A11}$$

$$V_{\phi_A} = \tilde{\sigma}_u^2 \left(\Lambda_0 + Z_A' Z_A\right)^{-1}, \tag{A12}$$

V_{α_N} denotes the covariance matrix of α_N, and $\tilde{\sigma}_u^2$ denotes the posterior variance of the residuals in equation (3). See PS (2002a) for further details.

Appendix B

B.1. Frequentist estimation

Let fund returns at time *t* be represented by $y_t = [r_{A,t}]$, and *X* be a *k* by *T* matrix of ones and nonbenchmark and benchmark asset returns so that each element of *X*, x_t, is defined: $x_t = [1, r_{N,t}, r_{B,t}]$. At each time *t*, β_t is a *k*-vector of estimated coefficients. The model consists of the estimation equation,

$$y_t = x_t' \beta_t + u_t, \quad u_t \sim iid\ N(0, R), \tag{B1}$$

and the transition equation,

$$\beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim iid\ N(0, Q). \tag{B2}$$

Estimates of each β_t conditional on information up to time *t* can be estimated by a sequence of GLS regressions on the transformed estimation equation:

$$Y_t = X_t \beta_t + e_t, \tag{B3}$$

where Y_t is a *t*-vector of dependent variable observations from time *1* through time *t*, β_t is the *k* component vector of time *t* estimates and X_t is a $T \times k$ matrix of observations. e_t is normally distributed with mean zero and covariance matrix: $E(ee') = A_t\ (I_{t-1} \otimes Q) A_t' + RI = \Omega_t$, where A_t is a *t* by *k(t-1)* upper triangular matrix of observations in which row *i* contains independent variable observation *i* and row *t* is a row of zeros. Matrix *R* is initially estimated from the OLS residuals from estimating Equation (3) and the *k* by *k* matrix *Q* is initially estimated from a series of independent quarterly OLS estimates of Equation (3) using daily data.[1] With the initial estimates for R and Q, β_t for each time t is estimated from the GLS equation:

$$\beta_t = \left(X_t'\Omega_t X_t\right)^{-1} X_t'\Omega_t y_t . \tag{B4}$$

The variance of the β_t estimates from initial estimation of Equation (B4) are then used to update the Q matrix and form a new Ω_t matrix, which is used in Equation (B4) to update β_t estimates.[2]

B.2. Bayesian estimation

To accommodate flexible priors similar to PS (2002a,b), we use the Gibbs sampler to implement Bayesian estimation of the time-varying parameter model. The advantage of a Gibbs sampler is that conditional posterior densities can be sampled independently. Thus, a specific posterior density can be sampled from for each parameter or hyperparameter in the model.

B.2.1 Bayesian time-varying parameter model

Min (1998) constructs a Gibbs sampler for Bayesian estimation of a time-varying parameter model, which we adapt to the PS (2002a,b) framework. The model consists of the estimation equation,

$$y_t = x_t'\beta_t + u_t, \quad u_t \sim iid\ N\left(0, \sigma^2\right), \tag{B5}$$

and the transition equation,

$$\beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim iid\ N\left(0, \delta^2 I\right). \tag{B6}$$

The transition equation requires estimation of the hyperparameter, δ, in addition to the parameters of interest. The marginal posterior density for β_T is obtained by integrating out σ and δ in the joint posterior density for $\left(\beta_T, \sigma, \delta\right)$,

$$\begin{aligned} p\left(\beta_T \mid D_T\right) &= \int p\left(\beta_T, \sigma, \delta \mid D_T\right) d\delta \\ &= \int p\left(\beta_T \mid \sigma, \delta, D_T\right) \cdot p\left(\sigma \mid \delta, D_T\right) \cdot p\left(\delta \mid D_T\right) d\delta, \end{aligned} \tag{B7}$$

[1] See footnote 3 below.
[2] See Kim and Nelson (1999) for further details.

where D_T denotes the data up to time T. The conditional density for β_T, $p(\beta_T \mid \sigma, \delta, D_T)$, is a multivariate normal density. The conditional density, $p(\delta \mid D_T)$, is complex, making analytical integration difficult. However, the conditional densities in the time-varying parameter problem are well defined, so that we can draw samples from them. For example, even though the marginal density $p(\delta \mid D_T)$ is complicated, the conditional density for δ, $p(\delta \mid \beta_T, ..\beta_0, D_T)$, is an inverted gamma.

B.2.2 Full conditional posterior distributions

Using the dimensions of our problem, write equation (B6) as

$$\beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim iid\ N(0, \Delta), \tag{B8}$$

where Δ is a diagonal matrix with elements $(\delta_1^2, \delta_2^2, ..., \delta_{m+n+1}^2)$, and $m+n$ is the total number of benchmark (*m*) and non-benchmark (*n*) factors in the model. If the elements of Δ are equal, then $\Delta = \delta^2 I$, and (B6) and (B8) are identical. Equation (B5) can be stacked and written in matrix notation as a single equation system that can be estimated simultaneously:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_T \end{bmatrix} = \begin{bmatrix} x_1' & 0 & 0 & 0 & 0 \\ 0 & x_2' & 0 & 0 & 0 \\ 0 & 0 & x_3' & 0 & 0 \\ 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & x_T' \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_T \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_T \end{bmatrix}. \tag{B9}$$

If each element of β_t is equal to $\{Vec\,[\alpha_t, \beta_t^{(m)}, \beta_t^{(n)}]\}$, and each diagonal element of X_t is represented as $x_t' = [1 \quad r_1^{nb} \quad r_1^{b}]$, then the beta matrix is $kT \times 1$, where $k = m+n+1$. The transition model in equation (B8) can be similarly written as

$$\begin{bmatrix} \mathrm{I} & 0 & 0 & 0 & 0 \\ -\mathrm{I} & \mathrm{I} & 0 & 0 & 0 \\ 0 & -\mathrm{I} & \mathrm{I} & 0 & 0 \\ 0 & 0 & ... & ... & 0 \\ 0 & 0 & 0 & -\mathrm{I} & \mathrm{I} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_T \end{bmatrix} = \begin{bmatrix} \mathrm{I} \\ 0 \\ . \\ . \\ 0 \end{bmatrix} \beta_0 + \begin{bmatrix} \eta_1 \\ \eta_2 \\ . \\ . \\ \eta_T \end{bmatrix}. \tag{B10}$$

Simplifying the notation, (B9) is written

$$y = X\gamma + u, \qquad u \sim N(0, \sigma^2 I_T), \tag{B11}$$

and (B10) is written

$$A\gamma = J\beta_0 + \eta, \qquad \eta \sim N(0, I_T \otimes \Delta). \tag{B12}$$

In equation (B11), y is a $T \times 1$ vector of fund returns, X is $T \times kT$, γ is $kT \times 1$, and u is $T \times 1$. In equation (B12), A is the $kT \times kT$ matrix described in (B10), J is $kT \times k$, and β_0 is a $k \times 1$ vector of initial parameter estimates. The disturbance, η, is $kT \times 1$, and the transition equation covariance matrix, Δ, is $k \times k$. The full conditional posteriors of the parameters of the model, $(\beta_0, \gamma, \sigma, \Delta)$, can be expressed as $p(\beta_0, \gamma \mid \sigma, \Delta, D_T)$ and $p(\sigma, \Delta \mid \beta_0, \gamma, D_T)$. Then, random draws of a full conditional distribution can be implemented using the following relations:

$$p(\beta_0, \gamma \mid \sigma, \delta, D_T) = p(\beta_0 \mid \sigma, \Delta, D_T) \cdot p(\gamma \mid B_0, \sigma, \Delta, D_T), \tag{B13}$$

$$p(\sigma, \Delta \mid \beta_0, \gamma, D_T) = p(\sigma \mid \beta_0, \gamma, D_T) \cdot p(\Delta \mid \sigma, \beta_0, \gamma, D_T). \tag{B14}$$

Thus, we can obtain a random sample from $p(\beta_0, \gamma \mid \sigma, \Delta, D_T)$ by first drawing β_0^* from $p(\beta_0 \mid \sigma, \Delta, D_T)$ and then drawing γ^* from $p(\gamma \mid \beta_0, \sigma, \Delta, D_T)$. Similarly, we first draw σ^* from $p(\sigma \mid \beta_0, \gamma, D_T)$ and then draw Δ^* from $p(\Delta \mid \sigma, \beta_0, \gamma, D_T)$.

B.2.3 Gibbs sampler

Min (1998) shows that diffuse priors generate the conditional posterior densities:

$$p(\beta_0 \mid \sigma, \Delta, D_T) \sim N(\hat{\beta}_0, V) \tag{B15a}$$

$$p(\gamma \mid \sigma, \Delta, D_T) \sim N(\hat{\gamma}, W) \tag{B15b}$$

$$p(\sigma \mid \gamma, D_T) \propto \frac{1}{\sigma^{T+1}} exp\left\{-(y - X\gamma)'(y - X\gamma)/2\sigma^2\right\} \tag{B15c}$$

$$p(\delta_i \mid B_0, \gamma) \propto \frac{1}{\delta_i^{T+1}} exp\left\{-\sum_{t=1}^{T}(\beta_{it} - \beta_{it-1})^2 / 2\delta_i^2\right\} \tag{B15d}$$

With the conditional densities described above, our Gibbs sampler is designed as follows:

i. Initialize σ and Δ.[3]

ii. Repeat (a) - (d) for $i = 1, \ldots, \tau + N$:

[3] We initially estimate the elements of Δ using the time-series variation in OLS estimation parameters, so that $\Delta_{ii} = \frac{1}{3}(\beta_{it}^{OLS} - \beta_{it-1}^{OLS})^2 / T_{quarters}$. We initialize σ^2 using the fund investment objective mean cross-sectional variance.

(a) Sample $B_0^{(i)}$ from $p\left(\beta_0 \mid \sigma^{(i-1)}, \Delta^{(i-1)}, D_T\right)$, which is multivariate normal, as above

(b) Sample $\gamma^{(i)}$ from $p\left(\gamma \mid \beta_0^{(i)}, \sigma^{(i-1)}, \Delta^{(i-1)}, D_T\right)$, which is multivariate normal, as above

(c) Sample $\sigma^{(i)}$ from $p\left(\sigma \mid \gamma^{(i)}, D_T\right)$, which is an inverted gamma density, as above

(d) Sample $\Delta^{(i)}$ from $p\left(\Delta \mid \beta_0^{(i)}, \gamma^{(i)}\right)$, which is an inverted gamma density, as above.

After a τ-period burn-in, averaging the Gibbs sampler β_T parameters over N draws produces the unconditional parameters of interest. To adapt this general procedure to the specific prior distributions used by PS (2002a,b), we rewrite transition equation (B12) as

$$\gamma = A^{-1}J\beta_0 + A^{-1}\eta \ , \tag{B16}$$

and substitute it into estimation equation (B14), giving

$$y = XA^{-1}J\beta_0 + XA^{-1}\eta + u = X^*\beta_0 + u^* , \tag{B17}$$

where $X^* = XA^{-1}J$, and $u^* \sim N(0,Q)$. With a diffuse prior for β_0 as above, the conditional posterior is:

$$p(\beta_0 \mid \sigma, \Delta, D_T) \sim N\left(\hat{\beta}_0, V\right), \tag{B18}$$

where $\hat{\beta}_0 = \left(X^{*'}Q^{-1}X^*\right)^{-1}\left(X^{*'}Q^{-1}y\right)$, $V = \left(J'A^{-1'}X'Q^{-1}XA^{-1}J\right)^{-1}$, and $Q = \left(\sigma^2 I + XA^{-1}(I_T \otimes \Delta)A^{-1'}X'\right)$.

B.2.4 Flexible prior beliefs in managerial skill

Instead of using a diffuse prior for β_0, we incorporate flexible prior beliefs in managerial skill into the estimate of the conditional posterior. Define the conjugate priors for the coefficients as $\beta_0 = \{\delta_0, c_{AN}, c_{AB}\}$, where δ_0 is the initial estimate of managerial skill. Thus, the initial $\hat{\beta}_0$ estimates for the fund return equation (3) will be

$$\hat{\beta}_0 = \left(\Sigma_0^{-1} + X^{*'}Q^{-1}X^*\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + X^{*'}Q^{-1}y\right), \tag{B19}$$

which is a standard Bayesian equation given the normal errors in (B17). Specifically, the priors for β_0 are set equal to the priors described in the text under the stationary beta assumption. Likewise, the prior for Σ_0^{-1} is identical to the prior Λ_0 defined in equation

(A6). We use $\hat{\beta}_0$ and V as the moments of the conditional posterior distribution in (B15a) and sample to obtain the posteriors from iteration (1): $\tilde{\beta}_0^{(1)} = \{\tilde{\delta}_0^{(1)}, \tilde{c}_{AN}^{(1)}, \tilde{c}_{AB}^{(1)}\}$.[4] Thus, the prior on managerial skill only directly affects the initial values of the time-varying parameters. Through the conditional posteriors of $\tilde{\beta}_0$, this prior indirectly affects all of the other parameters in the model. We then use the $\tilde{\beta}_0^{(1)}$ posteriors to update the moments of the conditional posterior distribution (B15b), where $\hat{\gamma} = \{X'y/\sigma^2 + A'(I_T \otimes \Delta)^{-1} J\tilde{\beta}_0^{(1)}\}$, and $W = \{X'X/\sigma^2 + A'(I_T \otimes \Delta)^{-1} A\}^{-1}$. The initial sample from this distribution is designated $\tilde{\gamma}^{(1)}$. We use $\tilde{\gamma}^{(1)}$ and its individual elements, β_{it}, to sample from the inverted gamma distributions in (B15c) and (B15d) to obtain $\sigma^{2(1)}$ and each diagonal element of the Δ matrix, $\delta_i^{2(1)}$.[5] Using the updated values of $\sigma^{2(1)}$ and $\Delta_{ii}^{(1)}$ to form a new Q matrix, we then form new parameters for the conditional posterior distribution in (B15a) to resample, as outlined above in Section B.2.3 (ii). We repeat the process $\tau+N$ times. After a burn-in of $\tau = 200$ draws, we calculate the sample averages of the $N = 800$ draws from the Gibbs sampler for different degrees of prior belief in managerial skill, σ_δ^2.

---

[4] Since the prior for β_0 is no longer diffuse, we adjust the variance of the conditional posterior for this draw so that $V = \left(\Sigma_0^{-1} + J'A^{-1'}X'Q^{-1}XA^{-1}J\right)^{-1}$.

[5] We also adjust the posterior inverted gamma distribution for σ^2 for the information in the empirical Bayesian prior, Σ_0^{-1}, as in Gelfand et al. (1990). The conditional posterior distribution for Δ assumes a diffuse prior. However, alternative priors on the drift variance are easy to implement.

Appendix References

Gelfand, A., S. Hills, A. Racine-Poon, and A. Smith, 1990, "Illustration of Bayesian inference in normal data models using Gibbs sampling," *Journal of the American Statistical Association* 85, 972-985.

Kim, C., and C. Nelson, 1999, *State-space models with regime switching*, MIT Press, Cambridge, Massachusetts.

Min, C., 1998, "A Gibbs sampling approach to estimation and prediction of time-varying parameter models," *Computational Statistics & Data Analysis* 27, 171-198.