Appendix for:
A Dynamic Test of Conditional Asset Pricing Models

Outline

This Appendix provides additional details regarding methodology and some additional results. In Section A I compare the modeling framework for the dynamics of alphas and betas with respect to a standard rolling windows OLS estimation scheme. Section B describes in details the Gibbs sampler used for estimating the linear asset pricing models. A prior sensitivity analysis can be found in Section C, whereas Section D reports the convergence properties for the MCMC approach on the basis of a simulated data set. Section E outlines the dynamic hypothesis testing methodology. Section F details the variance decomposition test used to assess the economic performances of the model. Note that all notations and model definitions are similar to those in the main article.

A A Comparison with Rolling Window Time-Series Regressions

In this section, I compare a standard rolling window OLS estimation strategy to capture the dynamics of time-series betas as specified by a standard AR(1) process. I first simulate the following data generating process [DGP]

$$y_t = \beta_{0,t} + \beta_{1,t}x_{1,t} + \beta_{2,t}x_{2,t} + \beta_{3,t}x_{3,t} + \sigma_t\epsilon_t, \qquad \text{for} \qquad t = 1, ..., 200$$

with $\epsilon_t \sim NID(0, 1)$ and $x_{j,t} \sim NID(0, 1)$ for $j = 1, ..., 3$. I simulate betas and (log of) idiosyncratic risks with different levels of persistence. The intercept is set to be a highly stationary process across the simulation sample (i.e. $\delta_0 = 0.5$), zero unconditional mean (i.e. $\overline{\beta_0} = 0$), and high volatility (i.e. $\tau_0 = 1$). This means that I simulate a factor model with a noisy zero unconditional mean pricing error in time series. For the first regressor, I assume a slightly volatile ($\tau_1 = 0.05$) random walk dynamics, such that $\delta_1 = 1$ and undefined unconditional mean. For the second regressor, I assume a moderately persistent process with $\delta_1 = 0.8$, a highly positive unconditional mean $\overline{\beta_2} = 1.5$, and medium volatility (i.e. $\tau_2 = 0.2$). Furthermore, $\beta_{3,t}$ is assumed to be zero and constant across the sample. For the (log of) idiosyncratic volatility, I assume a rather persistent, albeit stationary, dynamics with $\delta_\sigma = 0.95$, with a moderate volatility of volatility (i.e. $\tau_\sigma = 0.1$). I apply the Bayesian estimation framework outlined in Section 2 with $M = 10000$ posterior draws (burn-in of 2000 draws and thin of 2), and different prior settings to investigate the sensitivity of posterior results. As a base case I assume the hyper-parameters outlined in the main text. I set $\underline{m}_\beta^0 = [0 \quad 0.9]$, $\underline{B}_\beta^0 = I_2 \cdot 1e4$, $\underline{\nu}_\beta^0 = 1$ and $\underline{\nu}_\beta^0 \underline{s}_\beta^0 = 10$ for the time-series pricing error, which implies an uninformative prior for the expected beta and a rather relevant impact of the exogenous shock $\eta_{0,t}$. For the other regressors I assume $\underline{m}_\beta^j = [0 \quad 0.9]$, $\underline{B}_\beta^j = I_2 \cdot 1e4$, $\underline{\nu}_\beta^j = .1$ and $\underline{\nu}_\beta^j \underline{s}_\beta^0 = 10$, which instead implies a relatively lower conditional volatility for $j = 1, 2, 3$. As far as the (log of) idiosyncratic risk is concerned, I assume for the base case $\underline{m}_\sigma = [-2 \quad 0.95]$, $\underline{B}_\sigma = I_2 \cdot 1e4$, $\underline{\nu}_\sigma = 0.5$ and $\underline{\nu}_\sigma \underline{s}_\sigma = 10$, which implies highly persistent shocks of moderate size.

[Insert Figure A.1 about here]

Figure A.1 shows the posterior estimates of $\beta_{i,t}$ for $i = 0, 1, 2, 3$ together with the corresponding *true* parameters. The results from this figure show that the estimates are quite accurate. Top panel of Figure A.2 shows the estimates of the time-varying betas based on a standard rolling window time-series regression with a window size equal to $n = 60$ observations. The average estimates of the pricing error are completely off in two periods at the beginning and about a half of the sample. This implies that a time-series test of the simulated factor pricing model would reject the null-hypothesis of no-pricing error although $\beta_{0,t}$ in the DGP. The path of $\hat{\beta}_{2,t}$ in the bottom-left panel also point out that, in the presence of highly unstable risk exposures, changes of the true betas are smoothed out and missed by average estimates.

[Insert Figure A.2 about here]

Bottom panel shows the rolling window OLS estimates with $n = 120$. Again, abrupt changes are massively missed by conditional estimates. Also, the null-hypothesis of a null pricing error can be rejected at the 5% statistical significant level for some observation at the beginning of the testing sample, although $\beta_{0,t}$ in the DGP. As a whole, this simulation exercise shows that rolling window time-series regressions can be potentially misleading in capturing the dynamics of time-varying risks exposures, especially when the true risks exposures are highly unstable through time. As conditional betas are treated as regressors in the second step of standard two-pass methodologies, such inefficiency can heavily affect any hypothesis testing on equilibrium restrictions such as equation (4).

B Gibbs Sampler

In this section I derive the full conditional posterior distributions of the latent variables and the model parameters discussed in Section 2 of the main text. Before describe in detail the different steps of the Gibbs sampler, we need to define the densities that make up the joint density of the data and the latent variables. By considering the no-arbitrage restriction such that $\alpha_{i,t} \simeq \gamma_{0,t} + \gamma_t' \beta_{i,t-1}$, the probability density functions for each portfolios, alphas, betas and log variances can be written from (3)-(4) as

$$p(y_{it}|F_t, \beta_{it}, \beta_{it-1}, \sigma_{it}^2, \gamma) = \frac{1}{\sqrt{2\pi\sigma_{i,t}^2}} \exp\left(-\frac{\left(y_{it} - \gamma_{0,t-1} - \gamma_{t-1}' \beta_{i,t-1} - \sum_{j=1}^K \beta_{ij,t} F_{j,t} \right)^2}{2\sigma_{i,t}^2} \right)$$

from (5)-(7) the densities of the latent states for each portfolio $i = 1, ..., N$ and risk factor $j = 1, ..., K$ can be written as

$$p\left(\alpha_{i,t}|\alpha_{i,t-1}, \delta_{i\alpha}, E_{i\alpha}, \tau_{i\alpha}^2\right) = \frac{1}{\sqrt{2\pi\tau_{i\alpha}^2}} \exp\left(-\frac{\left(\alpha_{i,t} - (1-\delta_{i\alpha}) E_{i\alpha} - \delta_{i\alpha}\alpha_{i,t-1}\right)^2}{2\tau_{i\alpha}^2} \right)$$

$$p\left(\beta_{ij,t}|\beta_{ij,t-1}, \delta_{ij}, E_{\beta_{ij}}, \tau_{ij}^2\right) = \frac{1}{\sqrt{2\pi\tau_{ij}^2}} \exp\left(-\frac{\left(\beta_{ij,t} - (1-\delta_{ij}) E_{\beta_{ij}} - \delta_{ij}\beta_{ij,t-1}\right)^2}{2\tau_{ij}^2} \right)$$

$$p\left(\ln\sigma_{i,t}^2 | \ln\sigma_{i,t-1}^2, \delta_{i\sigma}, E_{\ln\sigma_i^2}, \tau_{i\sigma}^2\right) = \frac{1}{\sqrt{2\pi\tau_{i\sigma}^2}} \exp\left(-\frac{\left(\ln\sigma_{i,t}^2 - (1-\delta_{i\sigma}) E_{\ln\sigma_i^2} - \delta_{i\sigma}\ln\sigma_{i,t-1}^2\right)^2}{2\tau_{i\sigma}^2} \right)$$

Given conditional independence of the latent states, the joint density can be approximated as the product of firm-factor specific densities. Let define $\xi_t' = \left[\alpha_t', \beta_{1,t}', ..., \beta_{K,t}'\right]$ the alphas and betas stacked on each other, with $\alpha'_t = (\alpha_{1,t}, ..., \alpha_{N,t})$ and $\beta'_{j,t} = (\beta_{1j,t}, ..., \beta_{Nj,t})$. The latent variables $\xi = \{\xi_t\}_{t=1}^{T}$ and $\Sigma = \left\{\sigma^2_{i,t}\right\}_{i=1,t=1}^{N \quad T}$, are simulated alongside the model parameters $\theta = \{\theta_i\}_{i=1}^{N}$ and the risk premia $\gamma = \{\gamma_{0t}, \gamma'_t\}_{t=1}^{T}$. At each iteration, the sampler sequentially cycles through the following steps:

1. Draw ξ conditional on Σ, θ, R and F.
2. Draw θ_β conditional on ξ, R and F.
3. Draw Σ conditional on ξ, θ, R and F.
4. Draw θ_σ conditional on ξ, R and F.
5. Draw γ conditional on ξ and θ.

In what follows I provide details of each step of the Gibbs sampler.

B.1 Step 1. Sampling the Conditional Alphas and Betas

The full conditional posterior density for the time-varying factor loadings is computed using a Forward Filtering Backward Sampling (FFBS) approach as in Carter and Kohn (1994). The initial prior are sequentially updated via the Kalman filtering recursion. Let $Z^t = (X^t, Y^t)$ and $X^t = (X_1, ..., X_t)$, $Y^t = (Y_1, ..., Y_t)$ the sample information on risk factors and portfolios returns, respectively. The state space model in (3)-(6) can be written by stacking single equations as

$$y_t = X_t \xi_t + \epsilon_t \qquad \epsilon_t \sim N(0, \Sigma_t) \tag{B.1}$$

$$\xi_t = (1-\delta)E_\xi + \delta \xi_{t-1} + \eta_t \qquad \eta_t \sim N(0, \tau^2) \tag{B.2}$$

with $X_t = [1 \ F'_t] \otimes I_N$. Given the independence of risk exposures across factors and portfolios, the intercept $(1-\delta)E_\xi$ is an $N(K+1)$-dimensional vector with the ij element equal to $(1-\delta_{ij})E_{\beta_{ij}}$. Likewise, $\delta = diag\left(\delta_{10}, ..., \delta_{N0}, \delta_{11}, ..., \delta_{N1}, ..., \delta_{1K}, ..., \delta_{NK}\right)$, and $\tau^2 = diag\left(\tau^2_{1\alpha}, ..., \tau^2_{1\alpha}, \tau^2_{11}, ..., \tau^2_{N1}, \tau^2_{12}, ..., \tau^2_{N2}, ..., \tau^2_{1K}, ..., \tau^2_{NK}\right)$. Conditionally on idiosyncratic risk Σ_t and θ, and assuming an initial distribution $\xi_0 | y_0 \sim N\left(m_0, C_0\right)$, it is straightforward to show that the (see West and Harrison 1997 for more details)

$$\xi_t|Y^{t-1}, X^{t-1}, \theta \sim N(a_t, R_t) \qquad \textit{Propagation Density}$$

$$Y_t|Y^{t-1}, X^{t-1}, \theta \sim N(f_t, Q_t) \qquad \textit{Predictive Density}$$

$$\xi_t|Y^t, X^t, \theta \sim N(m_t, C_t) \qquad \textit{Filtering Density}$$

with

$$\begin{aligned} a_t &= (1-\delta)E_\xi + \delta m_{t-1} & R_t &= \delta C_{t-1}\delta' + \tau \\ f_t &= a_t' X_t & Q_t &= X_t' R_t X_t + \Sigma_t \\ m_t &= a_t + K_t e_t & C_t &= R_t - K_t Q_t K_t' \end{aligned} \tag{B.3}$$

and $K_t = R_t X_t Q_t^{-1}$ and $e_t = y_t - f_t$, the so-called *Kalman gain* and the investors' forecasting error, respectively. Conditional alphas and betas are drawn from the posterior distribution which is generated by backward recursion (see Frühwirth-Schnatter 1994, Carter and Kohn 1994, and West and Harrison 1997).

B.2 Step 2. Sampling the Parameters $\theta_{ij,[\beta]} = \left(\delta_{ij}, E_{\beta_{ij}}, \tau^2_{ij}\right)$ and $\theta_{i,[\alpha]} = \left(\delta_{i\alpha}, E_{\alpha,i}, \tau^2_{i\alpha}\right)$

For the posterior estimates of the parameters for the betas dynamics I consider a Normal-Inverse-Gamma prior structure (see West and Harrison 1997).

$$\left(\delta_{ij}, E_{\beta_{ij}}, \tau^2_{ij}\right) \sim NIG\left(\underline{m}^{ij}_\beta, \underline{B}^{ij}_\beta, \underline{\nu}^{ij}_\beta/2, \underline{\nu}^{ij}_\beta \underline{s}^{ij}_\beta/2\right) \qquad i = 1, ..., N \tag{B.4}$$

where $m^{ij}_\beta, B^{ij}_\beta$ the location and scale hyper-parameters of the normal distribution, $\underline{\nu}^{ij}_\beta$ the initial degrees of freedom and $\underline{\nu}^{ij}_\beta \underline{s}^{ij}_\beta$ the scale parameter of an inverse-gamma distribution. Posterior estimates are obtained once the factor loadings $\beta_{ij,t}$ are sampled for each $t =$

1, ..., T. Given the conjugate prior structure the updating scheme is easily derived as

$$\left(\delta_{ij}, E_{\beta_{ij}} | \tau_{ij}^2, \beta_{ij}^{1:T}\right) \sim N\left(\overline{m}_\beta^{ij}, \tau_{ij}^2 \overline{B}_\beta^{ij}\right) \tag{B.5}$$

$$\left(\tau_{ij}^2 | \beta_{ij}^{1:T}\right) \sim IG\left(\overline{\nu}_\beta^{ij}/2, \overline{\nu}_\beta^{ij} \overline{s}_\beta^{ij}/2\right) \tag{B.6}$$

for $i = 1, ..., N$ and $j = 1, ..., K$, with

$$\begin{aligned}
\overline{B}_\beta^{ij} &= \left(\left(\underline{B}_\beta^{ij}\right)^{-1} + X_\beta' X_\beta\right)^{-1} \\
\overline{m}_\beta^{ij} &= \overline{B}_\beta^{ij}\left(\left(\underline{B}_\beta^{ij}\right)^{-1} \underline{m}_\beta^{ij} + X_\beta' Y_\beta\right) \\
\overline{\nu}_\beta^{ij} &= \underline{\nu}_\beta^{ij} + T \\
\overline{\nu}_\beta^{ij} \overline{s}_\beta^{ij} &= \underline{\nu}_\beta^{ij} \underline{s}_\beta^{ij} + \left(\underline{m}_\beta^{ij}\right)' \left(\underline{B}_\beta^{ij}\right)^{-1} \underline{m}_\beta^{ij} + Y_\beta' Y_\beta - \left(\overline{m}_\beta^{ij}\right)' \left(\overline{B}_\beta^{ij}\right)^{-1} \overline{m}_\beta^{i}
\end{aligned}$$

where $\beta_{ij}^{1:T} = (\beta_{ij,1}, ..., \beta_{ij,T})$, $X_\beta = \left[\iota, \ \beta_{ij}^{1:T-1}\right]$, and $Y_\beta = \beta_{ij}^{2:T}$. Similarly, for the conditional alphas I still consider the same Normal-Inverse-Gamma prior structure.

$$\left(\delta_{i\alpha}, E_{\alpha_i}, \tau_{i\alpha}^2\right) \sim NIG\left(\underline{m}_\alpha^i, \underline{B}_\alpha^i, \underline{\nu}_\alpha^i/2, \underline{\nu}_\alpha^i \underline{s}_\alpha^i/2\right) \qquad i = 1, ..., N \tag{B.7}$$

where $m_\alpha^{ij}, B_\alpha^{ij}$ the location and scale hyper-parameters of the normal distribution, $\underline{\nu}_\alpha^{ij}$ the initial degrees of freedom and $\underline{\nu}_\alpha^{ij} \underline{s}_\alpha^{ij}$ the scale parameter of an inverse-gamma distribution. Posterior estimates are obtained once the conditional alphas $\alpha_{i,t}$ are sampled for each $t = 1, ..., T$. Given the conjugate prior structure the updating scheme is easily derived as

$$\left(\delta_{i\alpha}, E_{\alpha_i} | \tau_{i\alpha}^2, \alpha_i^{1:T}\right) \sim N\left(\overline{m}_\alpha^i, \tau_{i\alpha}^2 \overline{B}_\alpha^i\right) \tag{B.8}$$

$$\left(\tau_{i\alpha}^2 | \alpha_i^{1:T}\right) \sim IG\left(\overline{\nu}_\alpha^i/2, \overline{\nu}_\alpha^i \overline{s}_\alpha^i/2\right) \tag{B.9}$$

for $i = 1, ..., N$, with

$$\overline{B}^i_\alpha = \left(\left(\underline{B}^i_\alpha \right)^{-1} + X'_\alpha X_\alpha \right)^{-1}$$

$$\overline{m}^i_\alpha = \overline{B}^i_\alpha \left(\left(\underline{B}^i_\alpha \right)^{-1} \underline{m}^i_\alpha + X'_\alpha Y_\alpha \right)$$

$$\overline{\nu}^i_\alpha = \underline{\nu}^i_\alpha + T$$

$$\overline{\nu}^i_\alpha \overline{s}^i_\alpha = \underline{\nu}^i_\alpha \underline{s}^i_\alpha + \left(\underline{m}^i_\alpha \right)' \left(\underline{B}^i_\alpha \right)^{-1} \underline{m}^i_\alpha + Y'_\alpha Y_\alpha - \left(\overline{m}^i_\alpha \right)' \left(\overline{B}^i_\alpha \right)^{-1} \overline{m}^i_\alpha$$

where $\alpha_i^{1:T} = (\alpha_{i,1}, ..., \alpha_{i,T})$, $X_\alpha = \left[\iota, \ \alpha_i^{1:T-1} \right]$, and $Y_\alpha = \alpha_i^{2:T}$.

B.3 Step 3 and 4. Sampling the Idiosyncratic Risk and the Corresponding Structural Parameters.

The conditional variances $\ln \sigma^2_{it}$ preserves the standard properties of state space models. From (3) the log of squared residuals for the i_{th} asset can be defined as

$$\ln \left(y_{i,t} - \alpha_{i,t} - F'_t \beta_{i,t} \right)^2 = \ln \sigma^2_{i,t} + u_t \tag{B.10}$$

where $u_t = \ln \varepsilon^2_t$ has a $\ln \chi^2(1)$. As in Omori et al. (2007), I approximate the $\ln \chi^2(1)$ distribution with a finite mixture of ten normal distributions, such that the density of u_t is given by

$$p(u_t) = \sum_{l=1}^{10} \varphi_l \frac{1}{\sqrt{\varpi^2_l 2\pi}} \exp \left(- \frac{(u_t - \mu_l)^2}{2\varpi_l} \right) \tag{B.11}$$

with $\sum_{l=1}^{10} \varphi_l = 1$. The appropriate values for μ_l, φ_l and ϖ^2_l can be found in Omori et al. (2007). At each step of the algorithm I simulate a component of the mixture at each time t. Given the mixture component I apply a Kalman filter method, such that idiosyncratic risk can be sampled as the betas. The initial prior of the log idiosyncratic volatility $\ln \sigma^2_{i,0}$ is normal with mean -1 and conditional variance equal to 10. To sample $\theta_{i,[\sigma]} = \left(\delta_{i\sigma}, \ln \sigma^2_i, \tau^2_{i\sigma} \right)$ from its joint posterior $p \left(\delta_{i\sigma}, \ln \sigma^2_i, \tau^2_{i\sigma} | \ln \Sigma_i \right)$ I proceed as in Step 2, with $\ln \Sigma_i = \left\{ \ln \sigma^2_{it} \right\}_{t=1}^T$. The same updating scheme is adopted by defining $X_\sigma = \left[\iota, \ \ln \left(\sigma^2_{1:T-1} \right) \right]$ and $Y_\sigma = \ln \left(\sigma^2_{2:T} \right)$.

B.4 Step 5. Sampling the Factors Risk Premia

Conditonal on the risk exposures, the estimate of the risk premia coincide with a multivariate linear model with uncorrelated errors. I assume a conjugate prior structure of the form

$$(\gamma, \Omega) \sim NIW\left(\underline{\gamma}, \underline{\Gamma}, \underline{\Omega}, \underline{\omega}\right) \tag{B.12}$$

with $\underline{\gamma}$, $\underline{\Gamma}$, the prior location and scale hyper-parameters of the multivariate normal, and $\underline{\omega}$, $\underline{\Omega}$ the prior degrees of freedom and location hyper-parameters of the inverse-wishart distribution (see West and Harrison 1997). Posterior estimates are then obtained by updating the prior structure as

$$\left(\gamma|\Omega, Z_{t-1}^{\beta}\right) \sim N\left(\overline{\gamma}, \overline{\Gamma}\right) \tag{B.13}$$

$$\left(\Omega|\gamma, Z_{t-1}^{\beta}\right) \sim IW\left(\overline{\Omega}, \overline{\omega}\right) \tag{B.14}$$

for $t = 1, ..., T-1$, with

$$\begin{aligned}
\overline{\Gamma} &= \left(\underline{\Gamma}^{-1} + \left(Z_{t-1}^{\beta}\right)' \Omega^{-1} Z_{t-1}^{\beta}\right)^{-1} \\
\overline{\gamma} &= \overline{\Gamma}\left(\underline{\Gamma}^{-1}\underline{\gamma} + \left(Z_{t-1}^{\beta}\right)' \Omega^{-1} Y_t^{\gamma}\right) \\
\overline{\omega} &= \underline{\omega} + N \\
\overline{\Omega} &= \underline{\Omega} + \hat{e}_t \hat{e}_t'
\end{aligned}$$

where β_{t-1} represents conditional estimates of the factors risks exposures at time $t-1$ for each of the N portfolios and K risk factors, $Z_{t-1}^{\beta} = [\iota,\ \beta_{t-1}]$, $Y_{\gamma} = y_t$ and $\hat{e}_t$ the estimated residuals from equation (4).

C Prior Sensitivity

In this section, I investigate in this section the influence of different prior specifications on posterior results. In particular, I discuss prior sensitivity for the posterior inference on betas and the (log of) idiosyncratic risk. I run a simulation example and directly test how posterior estimates reacts to different prior specifications. For the sake of comparison with the main text I use a pre-sample of 20 observations to "train" the hyper-parameters. The goal is to assess how the model dynamics implied by posterior estimates is driven by initial priors on betas and volatility. The data generating process is the same as in Appendix D. Now, I consider several alternative prior specifications for the conditional volatility of the betas and the (log of) idiosyncratic risks. The reason is twofold. First, conditional volatility could affect the signal-to-noise ratio, provided posterior estimates do not adjust consistently. Second, given the conjugate Normal-Inverse Gamma prior specifications, different priors on conditional volatility similarly affects posterior estimates of each betas. The priors specification for the latter are kept constant and rather uninformative. A moderately higher (lower) conditional variance is obtained by multiplying (dividing) $\underline{\nu}_{\beta}^{j}$ and $\underline{\nu}_{\beta}^{j}\underline{s}_{\beta}^{j}$ by 5, for $j = 0, 1, 2, 3, \sigma$. Table C.1 summarizes the different prior settings.

[Insert Table C.1 about here]

Figure C.2 reports the posterior estimates of $\beta_{i,t}$ for $i = 0, 1, 2, 3$ by decreasing the prior expected conditional variance τ_j^2, as well as rising the prior expected conditional variance of the (log of) idiosyncratic risk. The figure makes it clear that posterior medians are fairly comparable with the base case. Indeed, although posterior median estimates of $\beta_{0,t}$ and $\beta_{3,t}$ are fairly off from capturing the dynamics of the intercept, the corresponding true pricing error is still within the credibility region. Figure C.1 shows the results with the base case priors.

[Insert Figure C.1 about here]

Posterior estimates with such prior hyper-parameters are quite accurate. Figure C.3 shows the posterior estimates of the model parameters by assuming both a lower prior expected conditional variance τ_j^2, as well as decreasing the prior expected conditional variance of the (log of) idiosyncratic risk τ_σ^2. The Figure shows that, again, the true values of the parameters largely fall within the 95% credibility region.

[Insert Figures C.2-C.4 about here]

Figure C.4 shows the posterior estimates of the model parameters by assuming both a higher prior expected conditional variance of betas τ_j^2, and a higher prior expected conditional variance of the (log of) idiosyncratic risk τ_σ^2. The figures shows that by sensibly increasing the size of prior conditional variance might generate a deterioration of posterior median estimates. Indeed, compared to what we see in the previous figures, the credibility region for the β_2 estimates are rather far away from the median values. Yet, the posterior median estimates of the pricing error β_0 are relatively off from the true value, albeit the latter is still within the 95% credibility region.

D MCMC Convergence Analysis

I report the results of a convergence analysis of the MCMC sampler for the model outlined in Section 2 and Appendix A, and with respect to the Fama-French three-factors specification. The convergence analysis involves computing a set of inefficiency factors and t-tests for equality of the means across subsamples of the MCMC chain. (see Geweke 1992, Primiceri 2005 Justiniano and Primiceri 2008, Clark and Davig 2011 and Groen et al. 2013).

For each individual parameter and latent variable, the inefficiency factor answer the question "How much information do we actually have about parameters?", and is measured as $(1 + 2\sum_{f=1}^{\infty} \rho_f)$, where ρ_f is the f_{th} order auto-correlation of the chain of draws. This inefficiency factor equals the variance of the mean of the posterior draws from the MCMC sampler, divided by the variance of the mean assuming independent draws. Then, if we require that the variance of the mean of the MCMC posterior draws should be limited to be at most 1% of the variation due to the data (measured by the posterior variance), the inefficiency factor provides an indication of the minimum number of MCMC draws to achieve this, see Kim et al. (1998). If there are some correlation between successive samples, then we might expect that our sample has not revealed as much information of the posterior distribution of our parameter as we could have gotten if the samples draws were independent. When estimating these inefficiency factors, I use the Bartlett kernel as in Newey and West (1987), with a bandwidth set to 4% of the sample of draws. The inefficiency factor is computed for all the model parameters and applied on a range of choices for the total number of posterior draws as well as burn-in period lengths and thinning for the main model specification. Based on this comparison, I felt most comfortable that with the number of posterior draws set equal to 10000 and thinning value of 2, yielding 5000 retained posterior draws, our MCMC sampler would perform satisfactorily.

Tables D.1 provide a summary of the results showing that, for most parameters and latent variables, the MCMC sampler is very efficient and that it requires less than 5000 retained posterior draws to be able to do a reasonably accurate inferential analysis. In case of the time-invariant parameters $\boldsymbol{\tau}$ and $\boldsymbol{\delta}$, with likely values in the 4.49-16.66 range, the sampler is less efficient. Nonetheless, the corresponding inefficiency factors suggest on average a minimum number of draws of less than 5000 to achieve an accurate analysis of these parameters.

[Insert Table D.1 about here]

I also compute the p-value of the Geweke (1992) t-test for the null hypothesis of equality of the means computed with the first 20 percent and last 40 percent of the sample of retained draws. For this particular convergence diagnostic test, I compute the variances of the respective means using the Newey and West (1987) heteroskedasticity and autocorrelation robust variance estimator with a bandwidth set to 4% of the utilized sample sizes. Such convergence statistics is still computed for the complete model specification depicted in Section 2, estimated over the sample period 1963:07:01 - 2013:12. Table B.2 shows the results.

[Insert Table D.2 about here]

The convergence diagnostic tests in Table D.2 confirm the efficiency of the MCMC sampler. For example, in the case of the **B** parameters the null hypothesis of equal means across sub-samples of the retained draws is hardly ever rejected at the 5% confidence interval. Thus, inference in the factor model appears to be reasonably accurate when based posterior inference on 10000 draws with a burn-in of 2000 and thin value of 2. Such a choice of the number of draws keeps the computational burden relatively low, at the benefit of inference precision as shown in Table D.1 and Table D.2.

E Dynamic Testing Methodology

Testing the conditional CAPM boils down to test a joint restriction of the form

$$\mathcal{H}_0 : H\xi_t = q$$

$$\mathcal{H}_1 : H\xi_t \neq q$$

where H is a $R \times N(K+1)$ selection matrix, q an R-dimensional vector of zeros, R the number of restrictions. Hypothesis testing can be handled by computing standard Bayes factors. A convenient way to calculate Bayes factors comparing a restricted, $\mathcal{H}_0$, to an unrestricted model, $\mathcal{H}_1$, is the so-called Savage-Dickey density (SDD) ratio (see Verdinelli and Wasserman 1995). The main advantage of the SDD ratio is that involves only manipulation of priors and posteriors for the conditional factor pricing model. These are readily available from the estimation output. If the prior structure is common for both the model under the null $\mathcal{H}_0$, and the alternative $\mathcal{H}_1$, the following relationship is satisfied;

$$p\left(\xi_t | H\xi_t = q, \mathcal{H}_1\right) = p\left(\xi_t | \mathcal{H}_0\right)$$

also the (marginal) likelihoods of the observed data are observationally equivalent (see Verdinelli and Wasserman 1995 for a more detailed discussion). Nested models allows to express the Bayes factor corresponding to test the restriction as a ratio of ordinates,

$$\mathcal{BF}_{0,1}^{t} = \frac{p\left(H\xi_t = q|Z^t, \mathcal{H}_1\right)}{p\left(H\xi_t = q|\mathcal{H}_1\right)} \tag{E.15}$$

where $Z^t = (X^t, Y^t)$ and $X^t = (X_1, ..., X_t)$, $Y^t = (Y_1, ..., Y_t)$ the sample information on risk factors and portfolios returns, respectively. To see this note that the Bayes rule implies that

$$\begin{aligned}
\frac{p\left(H\xi_t = q|Z^t, \mathcal{H}_1\right)}{p\left(H\xi_t = q|\mathcal{H}_1\right)} &= \frac{p\left(Z^t|H\xi_t = q, \mathcal{H}_1\right)}{p\left(Z^t|\mathcal{H}_1\right)}, \\
&= \frac{\int p\left(Z^t|\xi_t, H\xi_t = q, \mathcal{H}_1\right) p\left(\xi_t|H\xi_t = q, \mathcal{H}_1\right) d\xi_t}{p\left(Z^t|\mathcal{H}_1\right)} \\
&= \frac{\int p\left(Z^t|\xi_t, \mathcal{H}_0\right) p\left(\xi_t|\mathcal{H}_0\right) d\xi_t}{p\left(Z^t|\mathcal{H}_1\right)} \\
&= \frac{p\left(Z^t|\mathcal{H}_0\right)}{p\left(Z^t|\mathcal{H}_1\right)} = \mathcal{BF}_{0,1}^{t},
\end{aligned}$$

Following Koop et al. (2010) the prior and posterior in (E.15) can be sampled from recursive filtering estimates of conditional moments $E\left(\xi_t|Z^t, \theta\right)$ and $Var\left(\xi_t|Z^t, \theta\right)$ as derived from (B.3) and the output of the MCMC estimation scheme. From (B.3) we can see that the betas of the unrestricted model $\mathcal{H}_1$ are distributed as a multivariate Normal distribution

$$p\left(\xi_t|Z^t, \mathcal{H}_1, \theta\right) = (2\pi)^{-N} \left|C_t\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\xi_t - m_t\right)' C_t^{-1} \left(\xi_t - m_t\right)\right\}$$

Using standard results on multivariate Gaussian distributions

$$p\left(H\xi_t|Z^t, \mathcal{H}_1, \theta\right) = (2\pi)^{-N} \left|HC_tH'\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(H\xi_t - Hm_{i,t}\right)' \left(HC_tH'\right)^{-1} \left(H\xi_t - Hm_t\right)\right\}$$

therefore, the model restriction $\mathcal{H}_0 : H\beta_t = 0$ is distributed as

$$p\left(H\xi_t = q|Z^t, \mathcal{H}_1, \theta\right) = (2\pi)^{-N} \left|HC_tH'\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(q - Hm_{i,t}\right)' \left(HC_tH'\right)^{-1} \left(q - Hm_t\right)\right\}$$

Now given the output from the MCMC algorithm, the numerator in (E.15) can be numerically approximated as

$$\hat{p}\left(H\xi_t = q|Z^t, \mathcal{H}_1\right) = \frac{1}{G}\sum_{g=1}^{G} p\left(H\xi_t = q|Z^t, \mathcal{H}_1, \theta^{(g)}\right)$$

with g the number of draws from the MCMC algorithm.[1] The same strategy can be applied to evaluate the prior at the denominator in (E.15). The hierarchical prior can be derived from the initial distribution $p\left(\xi_0|Y_0, Z_0\right) = N\left(m_0, C_0\right)$ and the recursion of the state equation ξ_t in (B.2). As such,

$$\xi_t = E_\xi + \delta^t \xi_0 + \sum_{j=0}^{t-1} \delta^j \tau^{1/2} \eta_{t-j}$$

$$\xi_t = E_\xi + \delta^t m_0 + \delta^t C_0^{1/2} z + \sum_{j=0}^{t-1} \delta^j \tau^{1/2} \eta_{t-j} \tag{E.16}$$

with $z \sim N\left(0, I_p\right)$ with $p = N(K+1)$. As such, the recursive prior has a location parameter equal to

$$\underline{\xi_t} = E\left(\xi_t|\mathcal{H}_1, \theta\right) = E_\xi + \delta^t m_0$$

and prior variance

$$\underline{V_t} = Var\left(\xi_t|\mathcal{H}_1, \theta\right) = \delta^t C_0 \left(\delta^t\right)' + \sum_{j=0}^{t-1} \delta^j \tau \left(\delta^j \tau\right)'$$

Therefore

$$p\left(\xi_t|\mathcal{H}_1, \theta\right) = (2\pi)^{-N} \left|\underline{V_t}\right|^{-1/2} \exp\left\{ -\frac{1}{2}\left(\xi_t - \underline{\xi_t}\right)' \underline{V_t}^{-1} \left(\xi_t - \underline{\xi_t}\right) \right\}$$

such that

$$p\left(H\xi_t = q|\mathcal{H}_1, \theta\right) = (2\pi)^{-N} \left|H\underline{V_t}H'\right|^{-1/2} \exp\left\{ -\frac{1}{2}\left(q - \underline{\xi_t}\right)' \left(H\underline{V_t}H'\right)^{-1} \left(q - \underline{\xi_t}\right) \right\}$$

[1] As usual the number of draws can be chosen to optimize the accuracy of the approximation, and standard diagnostics can be used to analyze convergence properties.

Again, given the output of the MCMC algorithm, I can approximate the marginal prior as

$$\hat{p}\left(H\xi_t = q|\mathcal{H}_1\right) = \frac{1}{G}\sum_{g=1}^{G} p\left(H\xi_t = q|\mathcal{H}_1, \theta^{(g)}\right)$$

Under the assumption of correct specification of conditional factor models, the cross-sectional intercept $\gamma_{0,t}$ should not be statistically different from zero. Therefore, a cross-sectional test involves to investigate the null $\mathcal{H}_0 : H\tilde{\gamma}_t = q$, against the alternative $\mathcal{H}_1 : H\tilde{\gamma}_t \neq q$, where H is a $R \times (K+1)$ selection matrix, q an R-dimensional vector of zeros, R the number of restrictions (i.e. $R = 1$ in this case), and $\tilde{\gamma}_t = (\gamma_{0,t}, \gamma_t')$. As above, hypothesis testing can be handled by computing a Bayes factor comparing the unrestricted, $\mathcal{H}_1$, to the restricted $\mathcal{H}_0$ cross-sectional regression at time t;

$$\mathcal{BF}_{0,1}^{t} = \frac{p\left(H\tilde{\gamma}_t = q|Z_{t-1}^{\beta}, \mathcal{H}_1\right)}{p\left(H\tilde{\gamma}_t = q|\mathcal{H}_1\right)} \tag{E.17}$$

with $Z_{t-1}^{\beta} = \left[\iota, \beta_{t-1}, Y^t\right]$. Again, to see this note that the Bayes rule implies that

$$\begin{aligned}
\frac{p\left(H\tilde{\gamma}_t = q|Z_{t-1}^{\beta}, \mathcal{H}_1\right)}{p\left(H\tilde{\gamma}_t = q|\mathcal{H}_1\right)} &= \frac{p\left(Z_{t-1}^{\beta}|H\tilde{\gamma}_t = q, \mathcal{H}_1\right)}{p\left(Z_{t-1}^{\beta}|\mathcal{H}_1\right)}, \\
&= \frac{\int p\left(Z_{t-1}^{\beta}|\tilde{\gamma}_t, H\tilde{\gamma}_t = q, \mathcal{H}_1\right) p\left(\tilde{\gamma}_t|H\tilde{\gamma}_t = q, \mathcal{H}_1\right) d\tilde{\gamma}_t}{p\left(Z_{t-1}^{\beta}|\mathcal{H}_1\right)} \\
&= \frac{\int p\left(Z_{t-1}^{\beta}|\tilde{\gamma}_t, \mathcal{H}_0\right) p\left(\tilde{\gamma}_t|\mathcal{H}_0\right) d\tilde{\gamma}_t}{p\left(Z_{t-1}^{\beta}|\mathcal{H}_1\right)} \\
&= \frac{p\left(Z_{t-1}^{\beta}|\mathcal{H}_0\right)}{p\left(Z_{t-1}^{\beta}|\mathcal{H}_1\right)} = \mathcal{BF}_{0,1}^{t},
\end{aligned}$$

Given the independence of $\tilde{\gamma}_t$ estimates across time, both the numerator and the denominator can be conveniently sampled from the output of the MCMC scheme. Conditional on the parameters $\theta = \Omega$ and the prior hyperparameters $\underline{\Gamma}$ and $\underline{\gamma}$, the posterior probability of the factors risk premia is equal to

$$p\left(\tilde{\gamma}_t|\beta_{t-1},\mathcal{H}_1,\theta\right)=(2\pi)^{-(K+1)}\left|\overline{\Gamma}\right|^{-1/2}\exp\Big\{-\frac{1}{2}\left(\tilde{\gamma}_t-\overline{\gamma}\right)'\overline{\Gamma}^{-1}\left(\tilde{\gamma}_t-\overline{\gamma}\right)\Big\}$$

where the hyperparameters $\overline{\Gamma}$ and $\overline{\gamma}$ are defined as

$$\overline{\Gamma}=\left(\underline{\Gamma}^{-1}+\left(Z_{t-1}^{\beta}\right)'\Omega^{-1}Z_{t-1}^{\beta}\right)^{-1}$$
$$\overline{\gamma}=\overline{\Gamma}\left(\underline{\Gamma}^{-1}\underline{\gamma}+\left(Z_{t-1}^{\beta}\right)'\Omega^{-1}Y_t^{\gamma}\right)$$

with $y_t=(y_{1,t},...,y_{N,t})$, and the posterior covariance structure Ω distributed as an Inverse-Wishart $IW\left(\overline{\Omega},\overline{\omega}\right)$, with

$$\overline{\omega}=\underline{\omega}+N,\qquad \text{and} \qquad \overline{\Omega}=\underline{\Omega}+\hat{e}_t\hat{e}_t'$$

where $\hat{e}_t$ the estimated residuals from equation (4), and N the number of portfolios. From the properties of the multivariate normal distribution it can be shown that

$$p\left(H\tilde{\gamma}_t|Z_{t-1}^{\beta},\mathcal{H}_1,\theta\right)=(2\pi)^{-(K+1)}\left|H\overline{\Gamma}H\prime\right|^{-1/2}\exp\Big\{-\frac{1}{2}\left(H\tilde{\gamma}_t-H\overline{\gamma}\right)'\left(H\overline{\Gamma}H'\right)^{-1}\left(H\tilde{\gamma}_t-H\overline{\gamma}\right)\Big\}$$

such that the pricing restriction is distributed as

$$p\left(H\tilde{\gamma}_t=q|Z_{t-1}^{\beta},\mathcal{H}_1,\theta\right)=(2\pi)^{-(K+1)}\left|H\overline{\Gamma}H\prime\right|^{-1/2}\exp\Big\{-\frac{1}{2}\left(q-H\overline{\gamma}\right)'\left(H\overline{\Gamma}H'\right)^{-1}\left(q-H\overline{\gamma}\right)\Big\}$$

Given the output of the MCMC algorithm, I can approximate the marginal prior as

$$\hat{p}\left(H\tilde{\gamma}_t=q|Z_{t-1}^{\beta},\mathcal{H}_1\right)=\frac{1}{G}\sum_{g=1}^{G}p\left(H\tilde{\gamma}_t=q|\beta_{t-1},\mathcal{H}_1,\theta^{(g)}\right)$$

The conditional prior $p\left(H\tilde{\gamma}_t=q|\mathcal{H}_1,\theta\right)$ can be directly sampled from a multivariate normal distribution with prior hyperparameters $\underline{\Gamma}$ and $\underline{\gamma}$. Then its marginal $\hat{p}\left(H\tilde{\gamma}_t=q|\mathcal{H}_1\right)$ can be approximated as above from the output of the MCMC scheme. Assuming equal prior over the null and the alternative hypothesis, $p\left(\mathcal{H}_0\right)=p\left(\mathcal{H}_1\right)$, we can compute (see Robert 2007, Ch.5);

$$p\left(\mathcal{H}_0|Z^t\right) = \left[1 + \frac{1}{\mathcal{BF}^t_{0,1}}\right]^{-1} = \frac{\mathcal{BF}^t_{0,1}}{1 + \mathcal{BF}^t_{0,1}} \tag{E.18}$$

Note $p\left(\mathcal{H}_0|Z^t\right)$ might be interpreted as a p-value. Unlike standard p-value, however, the posterior probability naturally penalizes for the complexity of the model, being a direct function of the Bayes factor $BF^t_{0,1}$. This address the so-called Lindleys paradox which is the apparent conflict between standard frequentist and Bayesian hypothesis testing. The conflict arises since standard t-statistics and corresponding p-values tend to go in favour of the null as the sample size increases. The posterior probability (E.18) makes clear this is not the case under the methodology I propose.

F Variance Decomposition Tests

We use the posterior densities of the time series of factor loadings and risk premia to perform a number of tests that allow us to assess whether a posited asset pricing framework may explain an adequate percentage of excess asset returns. Equation (4) decomposes excess asset returns in a component related to risk, represented by the term $\gamma_t'\beta_{i,t}$ plus a residual $\gamma_{0,t}+e_{i,t+1}$. In principle, a multi-factor model is as good as the implied percentage of total variation in excess returns explained by its first component, $\gamma_t'\beta_{i,t}$. However, here we should recall that even though (4) refers to excess returns, it remains a statistical implementation of the framework in (3). This implies that in practice it may be naive to expect that $\gamma_t'\beta_{i,t}$ be able to explain much of the variability in excess returns. A more sensible goal seems to be that$\gamma_t'\beta_{i,t}$ ought to at least explain the *predictable* variation in excess returns. We therefore follow earlier literature, such as Karolyi and Sanders (1998), and adopt the following approach. First, the excess return on each asset is regressed onto a set of M instrumental variables that proxy for available information at time $t-1$, $\mathbf{Z}_{t-1}$,

$$y_{i,t} = \lambda_{i0} + \sum_{m=1}^{M} \lambda_{im} Z_{m,t-1} + \xi_{i,t}, \tag{F.19}$$

to compute the sample variance of fitted values,

$$Var[P(y_{i,t}|\mathbf{Z}_{t-1})] \equiv Var\left[\lambda_{i0} + \sum_{m=1}^{M} \lambda_{im} Z_{m,t-1}\right], \tag{F.20}$$

where the notation $P(y_{i,t}|\mathbf{Z}_{t-1})$ means "linear projection" of x_{it} on a set of instruments, $\mathbf{Z}_{t-1}$. Second, for each asset $i = 1, ..., N$, a time series of fitted (posterior) risk compensations, $\gamma_t'\beta_{i,t-1}$, is regressed onto the instrumental variables,

$$\gamma_t'\beta_{i,t} = \lambda_{i0}' + \sum_{m=1}^{M} \lambda_{im}' Z_{m,t-1} + \xi_{i,t}' \tag{F.21}$$

to compute the sample variance of fitted risk compensations:

$$Var\left[P\left(\gamma_t'\beta_{i,t}|\mathbf{Z}_{t-1}\right)\right] \equiv Var\left[\lambda_{i0}' + \sum_{m=1}^{M} \lambda_{im}' Z_{m,t-1}\right]. \tag{F.22}$$

The predictable component of excess returns in (F.19) not captured by the model is then the sample variance of the fitted values from the regression of the residuals $\gamma_{0,t} + e_{i,t+1}$ on the instruments:

$$Var\left[\gamma_{0,t} + e_{i,t+1}\right] = Var\left[P\left(\gamma_{0,t} + e_{i,t+1}|\mathbf{Z}_{t-1}\right)\right]. \tag{F.23}$$

At this point, it is informative to compute and report two variance ratios, commonly called $VR1$ and $VR2$, after Ferson and Harvey (1991):

$$\mathcal{VR}1 \equiv \frac{Var\left[P\left(\gamma_t'\beta_{i,t}|\mathbf{Z}_{t-1}\right)\right]}{Var[P(y_{i,t}|\mathbf{Z}_{t-1})]} > 0 \tag{F.24}$$

VR1 should be equal to 1 if the multi-factor model is correctly specified, which means that all the predictable variation in excess returns is captured by variation in risk compensations; at the same time, VR2 should be equal to zero if the multi-factor model is correctly specified. Importantly, when these decomposition tests are implemented using the estimation outputs obtained from the MCMC scheme, drawing from the joint posterior densities of the factor loadings $\beta_{i,t}$ and the implied risk premia γ_t, $i = 1, ..., N$ and $t = 1, ..., T$, and holding the instruments fixed over time, it is possible to compute $\mathcal{VR}1$ in correspondence to each of such draws and hence obtain their posterior distributions.

References

Carter, C., and R. Kohn. 1994. On Gibbs sampling for state-space models. *Biometrika* pp. 541–553.

Ferson, W., and C. Harvey. 1991. The Variation of Economic Risk Premiums. *Journal of Political Economy* 99:385–415.

Frühwirth-Schnatter, S. 1994. Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis* 15:183–202.

Groen, J., R. Paap, and F. Ravazzolo. 2013. Real-time Inflation Forecasting in a Changing World. *Journal of Business and Economic Statistics* 31:29–44.

Karolyi, G., and A. Sanders. 1998. The Variation of Economic Risk Premiums in Real Estate Returns. *Journal of Real Estate Finance and Economics* 17:245–262.

Koop, G., R. Leon-Gonzalez, and R. Strachan. 2010. Dynamic Probabilities of Restrictions in State Space Models: An Application to the Phillips Curve. *Journal of Business and Economic Statistics* 28:370–379.

Omori, Y., S. Chib, N. Shepard, and J. Nakajima. 2007. Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference. *Journal of Econometrics* 140:425–449.

Primiceri, G. 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies* 72:821–852.

Robert, C. 2007. *The Bayesian Choice, Second Edition*. Springer.

Verdinelli, I., and L. Wasserman. 1995. Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association* pp. 614–618.

West, M., and J. Harrison. 1997. *Bayesian forecasting and dynamics models*. Springer.

Table C.1. Summary of Prior Settings for Different Cases

Prior settings. This table summarizes the different prior settings used for the prior sensitivity analysis. Together with the base case used in the empirical analysis I consider different settings for the conditional variance of risk exposures, alphas and idiosyncratic risks.

		Prior Intercept		Prior Regressors		Prior Variance	
Betas	Variance	$\underline{\nu}_\beta^0$	$\underline{\nu}_\beta^0 \underline{s}_\beta^0$	$\underline{\nu}_\beta^{1:3}$	$\underline{\nu}_\beta^{1:3} \underline{s}_\beta^{1:3}$	$\underline{\nu}_\sigma$	$\underline{\nu}_\sigma \underline{s}_\sigma$
Base	Base	1	10	0.1	10	0.5	10
Low	High	0.2	5	0.02	5	2.5	50
Low	Low	0.2	5	0.02	5	0.1	2
High	Low	5	50	0.5	50	2.5	50
High	High	5	50	0.5	50	0.1	2

Table D.1. Summary of Inefficiency Factors

The table summarizes the inefficiency factors, for the posterior values of the model parameters, estimated over the sample period 1963:07 - 2013:12. The estimated inefficiency factors are based on the Bartlett kernel as in Newey and West (1987) with a bandwidth equal to 4% of the 5000 retained draws.

	Inefficiency Factor						
	Parameters	Mean	Median	Min	Max	5%	95%
$\mathbf{B}$	58320	3.1021	3.3654	1.9804	9.9822	2.2152	5.4129
$\boldsymbol{\Sigma}$	14580	3.6471	3.6145	2.8181	4.2604	3.1461	4.0403
$\boldsymbol{\tau}$	150	9.4377	9.3805	4.4951	13.892	5.3414	12.853
$\boldsymbol{\delta}$	150	5.3567	6.6122	3.0722	16.663	3.4644	13.202

Table D.2. Summary of Convergence Diagnostics

The table summarizes the convergence results, for the posterior values of the model parameters, estimated over the sample period 1963:07:01 - 2013:12. For each of these, we compute the p-value of the Geweke (1992) t-test for the null hypothesis of equality of the means computed for the first 20% and the last 40% of the retained 5000 draws. The variances of the means are estimated with the Newey and West (1987) variance estimator using a bandwidth of 4% of the respective sample sizes.

Summary of Convergence Diagnostics			
	Parameters	5% Reject Rate	10% Reject Rate
$\mathbf{B}$	58320	0.0011	0.0411
$\mathbf{\Sigma}$	14580	0.0005	0.0951
$\boldsymbol{\tau}$	150	0.0001	0.0013
$\boldsymbol{\delta}$	150	0.0012	0.0090

Figure A.1. Posterior Estimates of Simulated Time-Varying Parameters

Simulation example estimates. This figure plots the posterior distributions of the parameters $\beta_{i,t}$ for $i = 0, 1, 2, 3$ together with the corresponding values implied by the true data-generating process. The blue solid line reports the median estimates of the parameters. The red dashed line denotes the true value of the parameter. The black dashed lines denote the 5th and 95th percentiles of the posterior distribution.

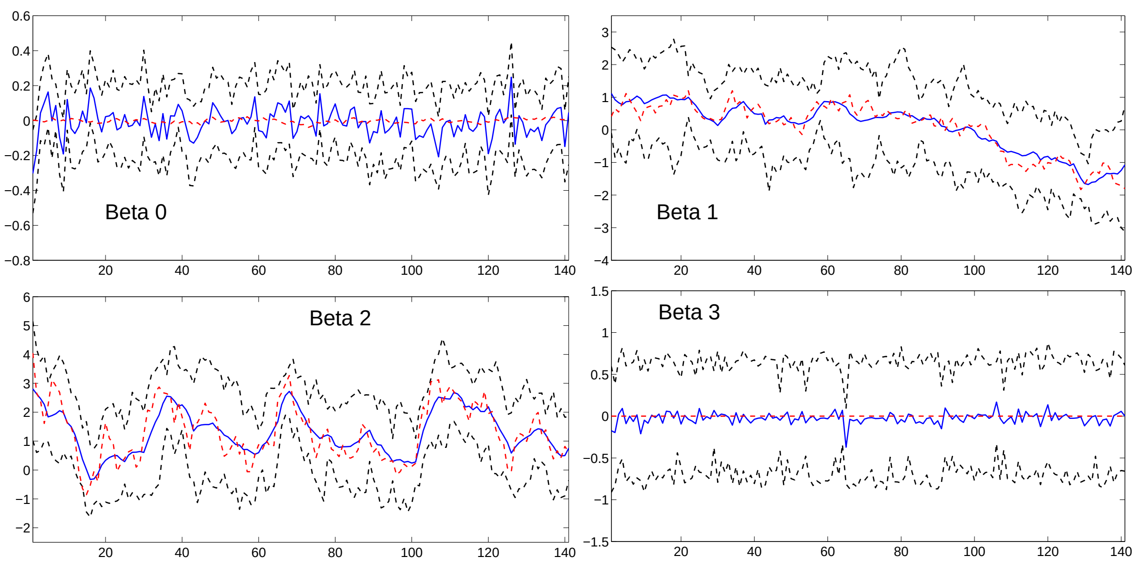

Figure A.2. Posterior Estimates of Time-Varying Parameters Using Rolling Window OLS

Simulation example estimates, OLS. This figure plots the OLS estimates of the conditional betas $\beta_{i,t}$ for $i = 0, 1, 2, 3$ from a rolling window regression. The blue solid line reports the mean estimates of the parameters. The red dashed line denotes the true value of the parameter. The black dashed lines denote the 5th and 95th percentiles of the posterior distribution. Top panel shows the results obtained using a window of length $n = 60$, while bottom panel shows the results obtained with $n = 120$.

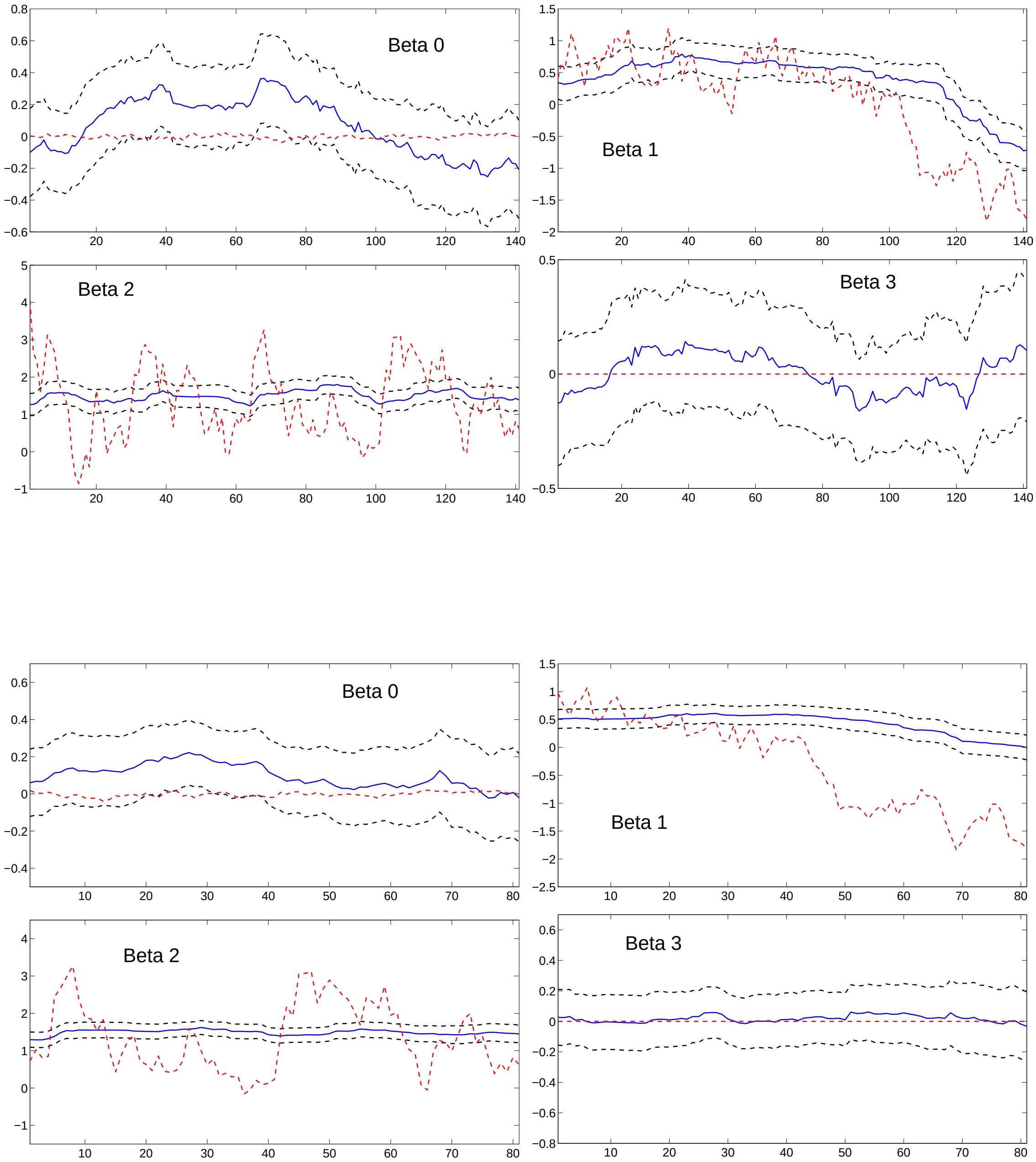

Figure C.1. Posterior Estimates of the Time-Varying Parameters: Base Case Priors

Simulation example estimates. This figure plots the posterior distributions of the parameters $\beta_{i,t}$ for $i = 0, 1, 2, 3$ together with the corresponding values implied by the true data-generating process. The blue solid line reports the median estimates of the parameters. The red dashed line denotes the true value of the parameter. The black dashed lines denote the 5th and 95th percentiles of the posterior distribution. Posterior results are based on the prior hyper-parameters for the base case.

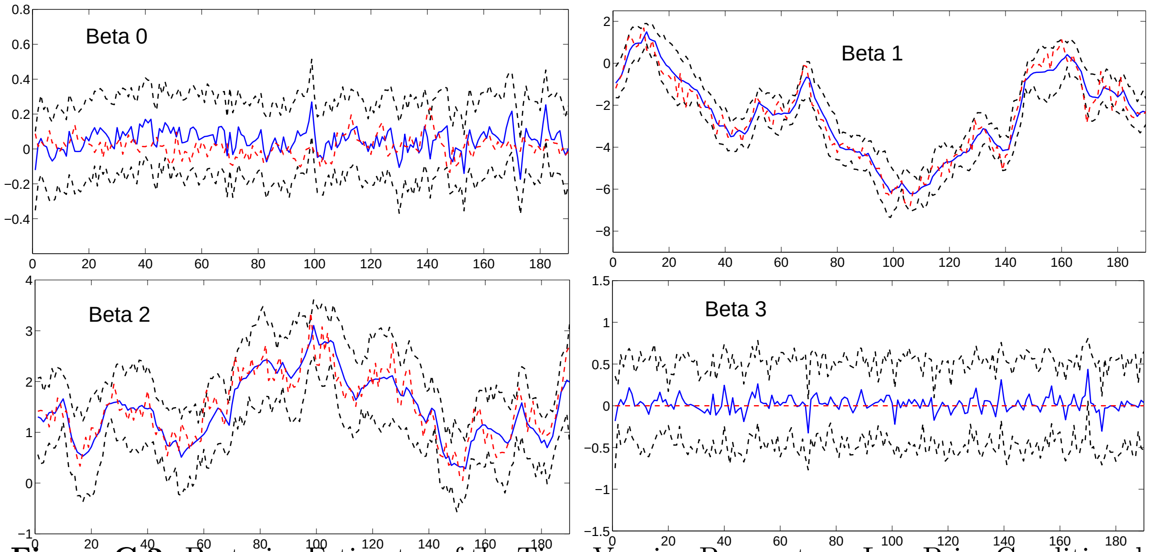

Figure C.2. Posterior Estimates of the Time-Varying Parameters: Low Prior Conditional Variance for Betas and High Prior Conditional Variance for the (log of) Idiosyncratic Risk

Simulation example estimates. This figure plots the posterior distributions of the parameters $\beta_{i,t}$ for $i = 0, 1, 2, 3$ together with the corresponding values implied by the true data-generating process. The blue solid line reports the median estimates of the parameters. The red dashed line denotes the true value of the parameter. The black dashed lines denote the 5th and 95th percentiles of the posterior distribution. Posterior results are now based on a lower prior for the expected conditional variance τ_j^2 and a higher prior for the conditional volatility of the (log of) idiosyncratic risk τ_σ^2

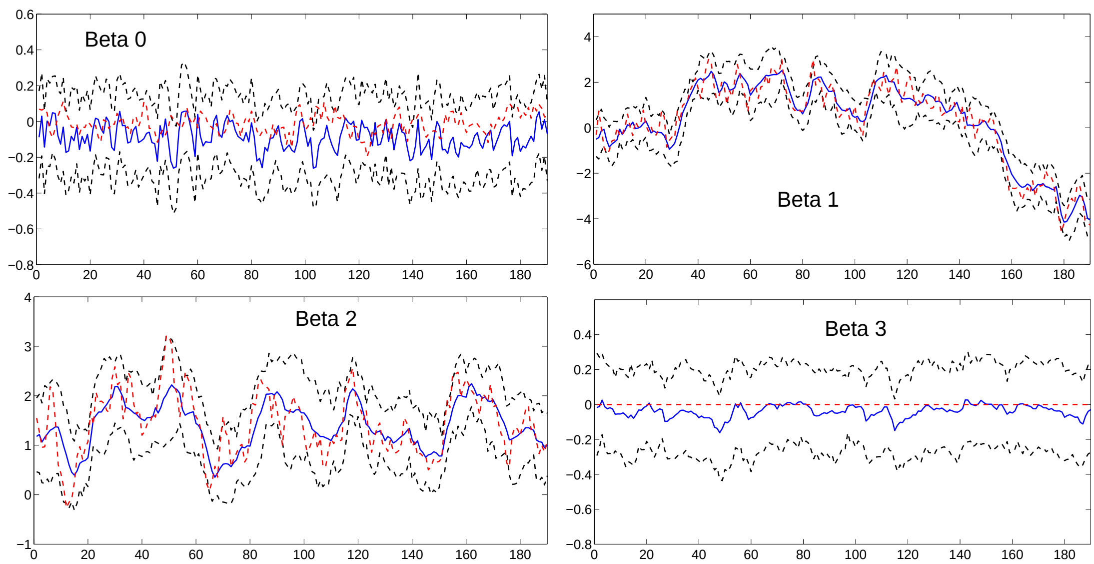

Figure C.3. Posterior Estimates of the Time-Varying Parameters: Low Prior Conditional Variance for Betas and Low Prior Conditional Variance for the (log of) Idiosyncratic Risk

Simulation example estimates. This figure plots the posterior distributions of the parameters $\beta_{i,t}$ for $i = 0, 1, 2, 3$ together with the corresponding values implied by the true data-generating process. The blue solid line reports the median estimates of the parameters. The red dashed line denotes the true value of the parameter. The black dashed lines denote the 5th and 95th percentiles of the posterior distribution. Posterior results are now based on a lower prior for the expected conditional variance τ_j^2 and a low prior for the conditional volatility of the (log of) idiosyncratic risk τ_σ^2

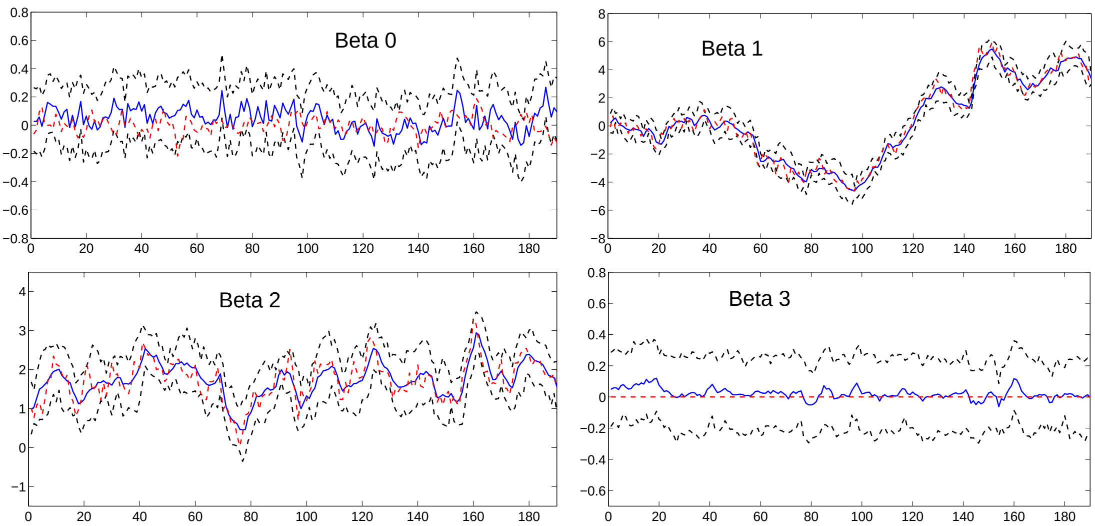

Figure C.4. Posterior Estimates of the Time-Varying Parameters: High Prior Conditional Variance for Betas and High Prior Conditional Variance for the (log of) Idiosyncratic Risk

This figure plots the posterior distributions of the parameters $\beta_{i,t}$ for $i = 0, 1, 2, 3$ together with the corresponding values implied by the true data-generating process. The blue solid line reports the median estimates of the parameters. The red dashed line denotes the true value of the parameter. The black dashed lines denote the 5th and 95th percentiles of the posterior distribution. Posterior results are now based on a higher prior for the expected conditional variance τ_j^2 and a higher prior for the conditional volatility of the (log of) idiosyncratic risk τ_σ^2