Online Appendix for: Macroeconomic Factors Strike Back: A Bayesian Change-Point Model of Time-Varying Risk Exposures and Premia in the U.S. Cross-Section

Daniele Bianchi*, Massimo Guidolin† and Francesco Ravazzolo‡

*Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK. Daniele.Bianchi@wbs.ac.uk
†Department of Finance and IGIER, Bocconi University, Milan, Italy. massimo.guidolin@unibocconi.it.
‡Norges Bank and BI Norwegian Business School, Oslo, Norway. Francesco.Ravazzolo@Norges-Bank.no.

Outline

This Online Appendix provides additional details regarding our methodology, on data used in the article and some additional results. In Section A we describe in detail the Gibbs sampler used for estimating our Multi-Factor Asset Pricing Model (MFAPM). Convergence properties for our MCMC approach can be found in Section B, whereas Section C reports on a prior sensitivity analysis for our framework that uses a simulated data set. In Section D we investigate the impact of different specification of stochastic volatility on in-sample posterior inference. Section E 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 The Gibbs Sampling Algorithm

In this section we derive the full conditional posterior distributions of the latent variables and the model parameters discussed in Section 2 of the main text. Before we describe in detail the different steps of sampler, we need to define the densities that make up the joint density of the data and the latent variables (12). By considering the no-arbitrage restriction such that $\beta_{i0,t} \simeq \lambda_{0,t} + \sum_{j=1}^{K} \lambda_{j,t}\beta_{ij,t-1}$, the likelihood given states and parameters can be written from (4)-(5) as

$$p(r_{it}|F_t, \beta_{it}, \sigma_{it}^2) = \frac{1}{\sqrt{2\pi\sigma_{i,t}^2}} \exp\left(-\frac{\left(r_{it} - \beta_{i0,t} - \sum_{j=1}^{K} \beta_{ij,t}F_{j,t}\right)^2}{2\sigma_{i,t}^2}\right)$$

from (6)-(8) the densities of the latent states can be written as

$$\begin{aligned}
p(\beta_{i,t}|\beta_{i,t-1}, \kappa_{i,t}, q_i^2) &= \prod_{j=0}^{K} \left(\frac{1}{\sqrt{2\pi q_{ij}^2}} \exp\left(-\frac{(\beta_{ij,t} - \beta_{ij,t-1})^2}{2q_{ij}^2}\right)\right)^{\kappa_{ij,t}} (\beta_{ij,t} - \beta_{ij,t-1})^{1-\kappa_{ij,t}} \\
p(\ln \sigma_{it}^2 | \ln \sigma_{it-1}^2, \kappa_{iv,t}, q_{iv}^2) &= \left(\frac{1}{\sqrt{2\pi q_{iv}^2}} \exp\left(-\frac{\left(\ln \sigma_{i,t}^2 - \ln \sigma_{i,t-1}^2\right)^2}{2q_{iv}^2}\right)\right)^{\kappa_{iv,t}} \left(\ln \sigma_{i,t}^2 - \ln \sigma_{i,t-1}^2\right)^{1-\kappa_{iv,t}}
\end{aligned} \tag{A.1}$$

The densities for $\beta_{i,t}$ and $\ln \sigma_{it}^2$ each consist of two parts. First one where breaks occurs and these are drawn from their corresponding distributions. The second component is the case of no break which results in a degenerate distribution of either the $\beta_{ij,t}$ or $\ln \sigma_{it}^2$. Note the latter case my be also represented as a Dirac delta function. For the ease of exposition we summarize the Gibbs sampler for the i_{th} asset.

A.1 *Step 1. Sampling* K_β.

The structural breaks in the conditional dynamics of the factor loadings B, measured by the latent binary state κ_{jt}, are drawn using the algorithm of Gerlach et al. (2000). This algorithm increases the efficiency of the sampling procedure since allows to generate κ_{jt}, without conditioning on the relative regression parameters β_{jt}. The conditional posterior density for κ_{jt}, $t = 1, ..., T, j = 0, ..., K$, is defined as

$$\begin{aligned} p\left(\kappa_{0t,...,}\kappa_{Kt}|\mathcal{K}_{\beta[-t]}, \mathcal{K}_\sigma, \Sigma, \theta, R, F\right) &\propto p\left(R|\mathcal{K}_{\beta t}, \mathcal{K}_\sigma, \Sigma, \theta, F\right) p\left(\kappa_{0t,...,}\kappa_{Kt}|\mathcal{K}_{\beta[-t]}, \mathcal{K}_\sigma, \Sigma, \theta, F\right) \\ &\propto p\left(r_{t+1}, ..., r_T|r_1, ..., r_t, \mathcal{K}_{\beta t}, \mathcal{K}_\sigma, \Sigma, \theta, F\right) p\left(r_t|r_1, ..., r_{t-1}, \kappa_{0t,...,}\kappa_{Kt}, \mathcal{K}_\sigma, \Sigma, \theta, F\right) \\ &\quad p\left(\kappa_{0t,...,}\kappa_{Kt}|\mathcal{K}_{\beta[-t]}, \mathcal{K}_\sigma, \Sigma, \theta, F\right) \end{aligned} \tag{A.2}$$

where $\mathcal{K}_{\beta[-t]} = \left\{\{\kappa_{js}\}_{j=0}^K\right\}_{s=1,s\neq t}^T$. We assume that each of the κ_{js} breaks are independent from each other such that the joint density is defined is defined as $\prod_{j=0}^K \pi_{ij}^{\kappa_{jt}} \left(1 - \pi_{ij}\right)^{1-\kappa_{jt}}$.

The remaining densities $p\left(r_{t+1}, ..., r_T|r_1, ..., r_t, \mathcal{K}_{\beta t}, \mathcal{K}_\sigma, \Sigma, \theta, F\right)$ and $p\left(r_t|r_1, ..., r_{t-1}, \kappa_{0t,...,}\kappa_{Kt}, \mathcal{K}_\sigma, \Sigma, \theta, F\right)$ are evaluated as in Gerlach et al. (2000). Notice that, since κ_{jt} is a binary state the integrating constant is easily evaluated.

A.2 *Step 2. Sampling the Factor Loadings* B.

The full conditional posterior density for the time-varying factor loadings is computed using a standard forward filtering backward sampling as in Carter and Kohn (1994). For each of the $i = 1, ..., N$ assets, the prior distribution of the $\beta_{i0}, ..., \beta_{iK}$ loadings is a multivariate normal with the location parameters corresponding to the OLS parameter estimates and a covariance structure which is diagonal and defined by the variances of the OLS estimates. The initial prior are sequentially updated via the Kalman Filtering recursion, then the parameters are

drawn from the posterior distribution which is generated by a standard backward recursion (see Frühwirth-Schnatter 1994, Carter and Kohn 1994, and West and Harrison 1997).

A.3 *Step 3 and 4. Sampling the Breaks and the Values of the Idiosyncratic Volatility.*

In order to draw the structural breaks $\mathcal{K}_\sigma$ and the idiosyncratic volatilities S we follow a similar approach as above. The stochastic breaks $\mathcal{K}_\sigma$ are drawn by using the Gerlach et al. (2000) algorithm. The conditional variances $\ln \sigma_{it}^2$, does not show a linear structure even though still preserving the standard properties of state space models. The model is rewritten as

$$\ln \left(r_{i,t} - \beta_{0t} - \sum_{j=1}^{K} \beta_{ijt} F_{jt} \right)^2 = \ln \sigma_{it}^2 + u_t$$
$$\ln \sigma_{it}^2 = \ln \sigma_{it-1}^2 + \kappa_{\nu it} \nu_{it} \tag{A.3}$$

where $u_t = \ln \varepsilon_t^2$ has a $\ln \chi^2(1)$. Here we follow Omori et al. (2007) and 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

$$p(u_t) = \sum_{l=1}^{10} \varphi_l \frac{1}{\sqrt{\varpi_l^2 2\pi}} \exp \left(-\frac{(u_t - \mu_l)^2}{2\varpi_l} \right) \tag{A.4}$$

with $\sum_{l=1}^{10} \varphi_l = 1$. The appropriate values for μ_l, φ_l and ϖ_l^2 can be found in Omori et al. (2007). Mechanically in each step of the Gibbs Samplers we simulate at each time t a component of the mixture. Now, given the mixture component we can apply the standard Kalman filter method, such that $\mathcal{K}_\sigma$ and Σ can be sampled in a similar way as $\mathcal{K}_\beta$ and B in the first and second step. The initial prior of the log idiosyncratic volatility $\ln \sigma_0^2$ is normal with mean -1 and conditional variance equal to 0.1.

A.4 *Step 5a. Sampling the Risk Premia at Time* t.

The equilibrium restriction in (2) simplify at each time t to a multi-variate linear regression of the N excess returns $r = (r_{1,t}, r_{2,t}, ..., r_{N,t})'$, onto a constant term and past betas $X = (\iota_N, \beta_{1,t-1}, \beta_{2,t-1}, ..., \beta_{K,t-1})$

$$r = X\lambda + e \qquad \text{with} \qquad e \sim N(0, \tau^2 I_N) \tag{A.5}$$

where $\beta_{i,t-1} = (\beta_{1i,t}, \beta_{2i,t}, ..., \beta_{Ni,t})'$. Note here we avoid the time t dependence of regressors for the ease of exposition. We consider independent conjugate priors

$$\lambda \sim MN(\underline{\lambda}, \underline{V}) \qquad \tau^2 \sim IG - 2(\overline{\psi}_0, \Psi_0) \tag{A.6}$$

Posterior updating in the Gibbs sampler evolves as

$$\lambda | X, r \sim MN(\overline{\lambda}, \overline{V}) \qquad \tau^2 | X, r \sim IG - 2(\overline{\psi}, \Psi) \tag{A.7}$$

with

$$\overline{\lambda} = \overline{V}\left(\underline{V}^{-1}\underline{\lambda} + \tau^2 X' r\right) \qquad \text{and} \qquad \overline{V} = \left(\underline{V}^{-1} + \tau^{-2} X' X\right) \tag{A.8}$$

while the posterior hyper-parameters for conditional volatility are defined as

$$\overline{\psi} = \overline{\psi}_0 + N \qquad \text{and} \qquad \Psi = \Psi_0 + ee' \tag{A.9}$$

A.5 *Step 5b. Sampling the Stochastic Breaks Probabilities.*

The full conditional posterior densities for the breaks probabilities $\pi = (\pi_{i1,...,}\pi_{iK})$ is given by

$$p\left(\pi | q^2, B, \mathbf{\Sigma}, \mathcal{K}_\beta, R, F\right) \propto \prod_{j=0}^{K} \pi_{ij}^{a_{ij}-1} (1-\pi_{ij})^{b_{ij}-1} \prod_{t=1}^{T} \pi_{ij}^{\kappa_{ijt}} (1-\pi_{ij})^{1-\kappa_{ijt}} \tag{A.10}$$

and hence the individual π_{ij} parameter can be sampled from a Beta distribution with shape parameters $a_{ij} + \sum_{t=1}^{T} \kappa_{ijt}$ and $b_{ij} + \sum_{t=1}^{T}(1-\kappa_{ijt})$ for $j = 0, ..., K$. Likewise the full conditional

posterior distribution for the breaks probabilities in the idiosyncratic volatilities π_ν is given by

$$p\left(\pi_\nu|q^2,B,\mathbf{\Sigma},\mathcal{K}_\sigma,R,F\right)\propto\pi_{i\nu}^{a_{i\nu}-1}\left(1-\pi_{i\nu}\right)^{b_{i\nu}-1}\prod_{t=1}^{T}\pi_{i\nu}^{\kappa_{i\nu t}}(1-\pi_{i\nu})^{1-\kappa_{i\nu t}}$$

such that the individual $\pi_{i\nu}$ can be sampled from a Beta distribution with shape parameters $a_{i\nu}+\sum_{t=1}^{T}\kappa_{i\nu t}$ and $b_{i\nu}+\sum_{t=1}^{T}(1-\kappa_{i\nu t})$ for $i=1,...,N$.

A.6 *Step 5c. Sampling the Conditional Variance of the States.*

The prior distributions for the conditional volatilities of the factor loadings β_{ijt} for $j=0,...,K$ are inverse-gamma

$$p\left(q_{ij}^2|\pi,B,\mathbf{\Sigma},\mathcal{K}_\beta,\mathcal{K}_\sigma,R,F\right)\propto q_{ij}^{-\nu_{ij}}\exp\left(-\frac{\delta_{ij}}{2q_{ij}^2}\right)\prod_{t=1}^{T}\left(\frac{1}{q_{ij}}\exp\left(\frac{-\left(\beta_{ijt-}\beta_{ijt-1}\right)^2}{2q_{ij}^2}\right)\right)^{\kappa_{ijt}} \tag{A.11}$$

hence q_{ij}^2 is sampled from an inverse-gamma distribution with scale parameter $\nu_{ij}+\sum_{t=1}^{T}\kappa_{ijt}\left(\beta_{ijt-}\beta_{ijt-1}\right)^2$ and degrees of freedom equal to $\nu_{ij}+\sum_{t=1}^{T}\kappa_{ijt}$. Likewise the full conditional of the variance for the idiosyncratic log volatility $q_{i\nu}^2$ is defined as

$$p\left(q_{i\nu}^2|\pi,B,\mathbf{\Sigma},\mathcal{K}_\beta,\mathcal{K}_\sigma,R,F\right)\propto q_{i\nu}^{-\nu_{i\nu}}\exp\left(-\frac{\delta_{i\nu}}{2q_{i\nu}^2}\right)\prod_{t=1}^{T}\left(\frac{1}{q_{i\nu}}\exp\left(\frac{-\left(\ln\sigma_{it}^2-\ln\sigma_{it-1}^2\right)^2}{2q_{i\nu}^2}\right)\right)^{\kappa_{i\nu t}} \tag{A.12}$$

such that $q_{i\nu}^2$ is sampled from an inverted Gamma distribution with scale parameter $\nu_{i\nu}+\sum_{t=1}^{T}\kappa_{i\nu t}\left(\ln\sigma_{it}^2-\ln\sigma_{it-1}^2\right)^2$ and degrees of freedom equal to $\nu_{i\nu}+\sum_{t=1}^{T}\kappa_{i\nu t}$.

B MCMC Convergence Analysis

We report the results of a convergence analysis of the MCMC sampler for the B-TVB-SV model outlined in Section 3 and Appendix A. 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 ques-

tion "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, we 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 B-TVB-SV specification. Based on this comparison we felt most comfortable that with the number of posterior draws set equal to 10000 with a burn-in period of 2000 draws and thinning value of 2, yielding 10000 retained posterior draws, our MCMC sampler would perform satisfactorily.

Tables B.1 provide a summary of the results showing that, for most parameters and latent variables, our MCMC sampler is very efficient and that it requires far less than 5000 retained posterior draws to be able to do a reasonably accurate inferential analysis. In case of the time-invariant parameters $\mathbf{Q}$ and π, with likely values in the 2.3-4.2 range, our sampler is less efficient. Nonetheless, the corresponding inefficiency factors suggest on average a minimum number of draws of less than 4000 to achieve an accurate analysis of these parameters.

We 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 we 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 B-TVB-SV specification estimated over the sample period 1972:01 - 2011:12. Table B.2 shows the results. The convergence diagnostic tests in Table

Table B.1: Summary of Inefficiency Factors

The table summarizes the inefficiency factors, for the posterior values of the model parameters, estimated over the sample period 1972:01 - 2011:12. The estimated inefficiency factors are based on the Bartlett kernel as in Newey and West (1987) with a bandwidth equal to 4% of the 10000 retained draws.

	Inefficiency Factor						
	Parameters	Mean	Median	Min	Max	5%	95%
$\mathbf{B}$	82800	2.9081	2.9222	2.6001	3.5031	2.6842	3.2091
$\mathcal{K}_\beta, \mathcal{K}_\sigma$	91080	2.7886	2.8157	2.0096	4.0311	2.3567	3.4231
$\mathbf{\Sigma}$	8280	2.8121	2.8321	2.0897	3.9421	2.4016	3.4072
$\mathbf{Q}$	253	2.9318	2.9118	2.3314	3.8921	2.3414	3.8532
π	253	3.3478	3.3405	2.6652	4.2307	2.6668	4.2209

B.2 confirm the efficiency of the MCMC sampler we propose. For example, in the case of the $\mathbf{B}$ 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 our factor model appears to be reasonably accurate when we base 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 B.1 and Table B.2.

C Prior Sensitivity Analysis

We investigate in this section the influence of different prior specifications on posterior results. In particular we discuss prior sensitivity for both the expected occurrence probability and expected size of a break for betas and idiosyncratic risks. First, we run a simulation example and directly test how posterior estimates reacts to different prior specifications. Second, we estimate the B-TVB-SV model on the original dataset by using different priors specifications. The goal of both exercises is to assess how the model instability/dynamics implied by posterior estimates is driven by priors on break sizes and probabilities.

Table B.2: Summary of Convergence Diagnostics

The table summarizes the convergence results, for the posterior values of the model parameters, estimated over the sample period 1972:01 - 2011:12. 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 10000 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
$\mathbf{B}$	82800	0.0102	0.0347
$\mathcal{K}_\beta, \mathcal{K}_\sigma$	91080	0.0133	0.0400
$\mathbf{\Sigma}$	8280	0.0108	0.0317
$\mathbf{Q}$	253	0.0000	0.0000
π	253	0.0000	0.0000

C.1 Simulation Example

The first step of the prior sensitivity analysis is based on a simulation example. We base our results on the following data generating process [DGP]

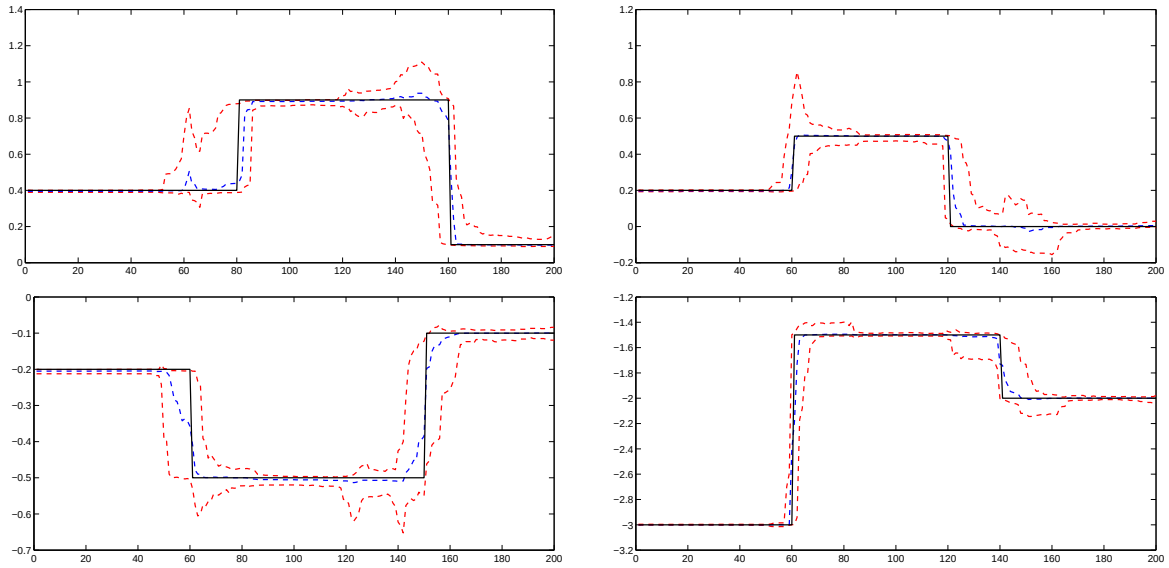
$$y_t = \beta_{0,t} + \beta_{1,t}x_{1,t} + \beta_{2,t}x_{2,t} + \beta_{3,t}x_{3,t} + \sigma_t\epsilon_t, \qquad \text{for} \qquad t = 1, ..., 200$$

with $\epsilon_t \sim NID(0,1)$ and $x_{j,t} \sim NID(0,1)$ for $j = 1, ..., 3$. We simulate discrete breaks both in the betas and idiosyncratic risks. The intercept is set to $\beta_{0,t} = 0$ for $t = 1, ..., 200$, meaning we simulate a factor model where there is no pricing error in the DGP. For the first regressor we take as parameters $\beta_{1,t} = 0.4$ for $t = 1, ..., 80$, $\beta_{1,t} = 0.9$ for $t = 81, ..., 160$, $\beta_{1,t} = 0.1$ and $t = 161, ..., 200$. For the second regressor we have $\beta_{2,t} = 0.2$ for $t = 1, ..., 60$, $\beta_{2,t} = 0.5$ for $t = 61, ..., 120$, and $\beta_{2,t} = 0$ for $t = 121, ..., 200$. Furthermore, $\beta_{3,t} = -0.2$ for $t = 1, ..., 60$, $\beta_{3,t} = -0.5$ for $t = 61, ..., 150$, and $\beta_{3,t} = -0.1$ for $t = 151, ..., 200$. For the (log of) idiosyncratic volatility we assume that $\ln \sigma_t^2 = -3$ for $t = 1, ..., 60$, $\ln \sigma_t^2 = -1.5$ for $t = 61, ..., 140$, and $\ln \sigma_t^2 = -2$ for $t = 141, ..., 200$. Hence we allow for breaks in the parameters at different points in time but we also include breaks which occur at the same time.

We apply our Bayesian estimation framework with structural breaks outlined in Section 3 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 we assume the hyper-

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

This figure plots the posterior distributions of the parameters $\beta_{i,t}$ for $i = 1, 2, 3$ and $\ln \sigma_t^2$ together with the corresponding values implied by the DGP. The blue dashed line reports the median estimates of the parameters. The red dashed lines denote the 20th and 80th percentiles of the posterior distribution. The black solid line displays the values of the data generating process.



parameters outlined in the main text. We set $a_j = 3.2$, $b_j = 60$ and $\gamma_j = 0.5$, $\delta_j = 100$ for $j = 1, 2, 3$ which implies *a priori* a relatively low break probability of having a break with a moderate expected size. As far as the (log of) idiosyncratic risk is concerned, we assume for the base case a low probability of having a break with $a_\nu = 1$, $b_\nu = 99$ while the size of breaks implies $\gamma_j = 0.2$,$\delta_j = 50$. We report in Figure C.1 the posterior estimates of $\beta_{i,t}$ for $i = 1, 2, 3$ and $\ln \sigma_t^2$ together with the corresponding *true* parameters. The results from this figure show that our approach is quite accurate in estimating both the timing and the size of the breaks, where the estimates of $\beta_{i,t}$ for $i = 1, 2, 3$ are more volatile due to our prior choice for the hyper-parameters of the inverse-gamma distributed size of the breaks. As we would expected, conditional volatility of the estimates sensibly increases around the occurrence of breaks in the DGP.

In our prior sensitivity assessment, we consider several alternative prior specifications where increase the prior probability of a break and decrease or increase the expected size of the breaks both across betas and idiosyncratic volatility. In total we consider 12 different prior specifications. A moderately larger probability of a break than in the base case means that we

divide b_j by 5. A more extreme break probability is obtained by dividing b_j by 10. Yet, a higher (lower) expected prior break size is obtained by multiplying (dividing) γ_j and ν_j by 5. As far as the conditional volatility is concerned, we increase the probability of a break by dividing b_ν by 10 and 20, respectively. Table C.3 summarizes the different prior settings. To summarize, by considering $b_j = 6, 12$ we increase the prior expected probability of observing a break in the betas (idiosyncratic risk) to 20% and 35% (10% and 17%) respectively.

Table C.3: Summary of Prior Settings for Different Cases

The table summarizes the different prior settings we used to run the prior sensitivity analysis.

Break	Exp	Prior Betas				Prior Variance			
Probability	Size	a_j	b_j	γ_j	δ_j	a_ν	b_ν	γ_ν	δ_ν
Base		3.2	60	0.5	100	1	99	0.2	50
Large	Small	3.2	12	0.1	20	1	10	0.04	10
Large	Large	3.2	12	2.5	500	1	10	1	250
Higher	Small	3.2	6	0.1	20	1	5	0.04	10
Higher	Large	3.2	6	2.5	500	1	5	1	250

Figure C.2 reports the posterior estimates of $\beta_{i,t}$ for $i = 1, 2, 3$ and $\ln \sigma_t^2$ by increasing the prior probability of having a break ($a_j = 3.2$, $b_j = 12$, $a_\nu = 1$ and $b_\nu = 10$), as well as rising the prior average size of the breaks ($\gamma_j = 2.5$, $\delta_j = 500$, $\gamma_\nu = 1$ and $\delta_\nu = 250$). The figure makes it clear that posterior medians of the parameter are now quite off in terms of the timing of the breaks, with a large uncertainty for the posterior estimates of $\beta_{1:3,t}$ as well as $\ln \sigma_t^2$. Interestingly, higher uncertainty is more evident for the betas than for the log of conditional variance. Indeed, although posterior median estimates of $\ln \sigma_t^2$ are fairly off from capturing the second break point, the corresponding confidence intervals are still relatively tight. As we would expect, by imposing *a priori* a higher instability in the parameters of the model, the corresponding credibility intervals tend to increase. Figure C.3 shows the posterior estimates of the model parameters by assuming an even larger prior break probability ($a_j = 3.2$, $b_j = 6$, $a_\nu = 1$ and $b_\nu = 5$), while still keeping the ex-ante average break size as before ($\gamma_j = 2.5$, $\delta_j = 500$, $\gamma_\nu = 1$ and $\delta_\nu = 250$). The Figure shows that a higher expected probability and size of a break may lead to much more uncertain posterior estimates. From Figure C.3-C.2, however, is hard to say if less precise estimates comes from a higher expected probability, rather than a higher expected size of a break.

Figure C.2: Posterior Estimates of the Time-Varying Parameters: Large prior-break probabilities and large prior break size

This figure plots the posterior distributions of of the parameters $\beta_{i,t}$ for $i = 1, 2, 3$ and $\ln \sigma_t^2$ together with the corresponding values implied by the DGP. The blue dashed line reports the median estimates of the parameters. The red dashed lines denote the 20th and 80th percentiles of the posterior distributions. The black solid line displays the values used to define the data generating process. Posterior results are now based on a larger prior-break probabilities for both betas and the (log of) idiosyncratic volatility ($a_j = 3.2$, $b_j = 12$, $a_\nu = 1$ and $b_\nu = 10$), also assuming a large expected priors break size ($\gamma_j = 2.5$, $\delta_j = 500$, $\gamma_\nu = 1$ and $\delta_\nu = 250$).

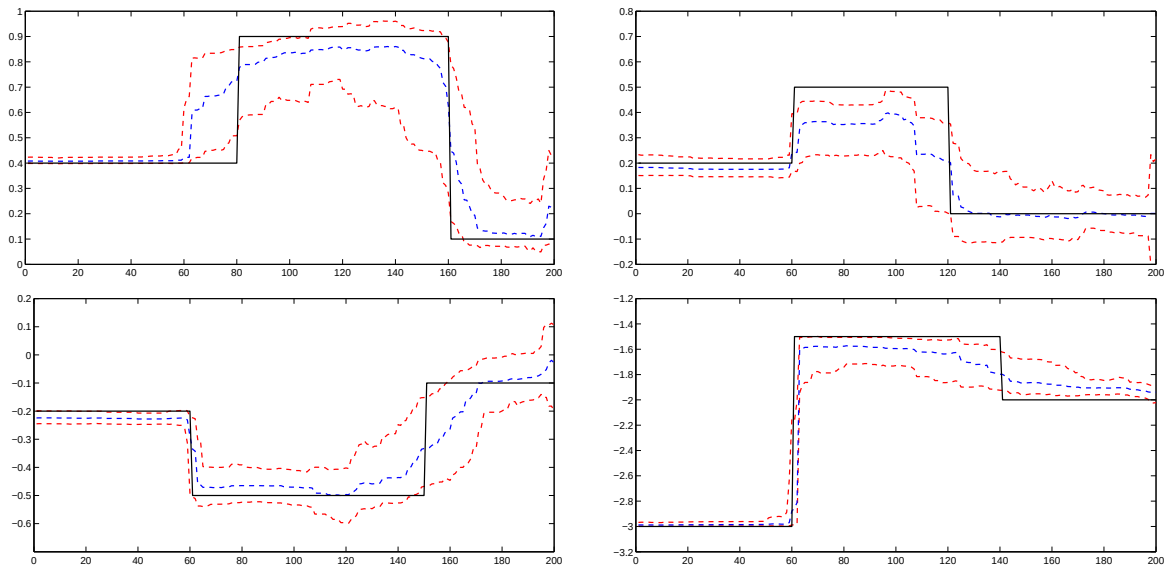


Figure C.3: Posterior Estimates of the Time-Varying Parameters: Higher prior-break probabilities and large prior break size

This figure plots the posterior distributions of the parameters $\beta_{i,t}$ for $i = 1, 2, 3$ and $\ln \sigma_t^2$ together with the corresponding values implied by the DGP. The blue dashed line reports the median estimates of the parameters. The red dashed lines denote the 20th and 80th percentiles of the posterior distributions. The black solid line displays the values used to define the data generating process. Posterior results are now based on a larger prior-break probabilities for both betas and the (log of) idiosyncratic volatility ($a_j = 3.2$, $b_j = 6$, $a_\nu = 1$ and $b_\nu = 5$), also assuming a large expected priors break size ($\gamma_j = 2.5$, $\delta_j = 500$, $\gamma_\nu = 1$ and $\delta_\nu = 250$).

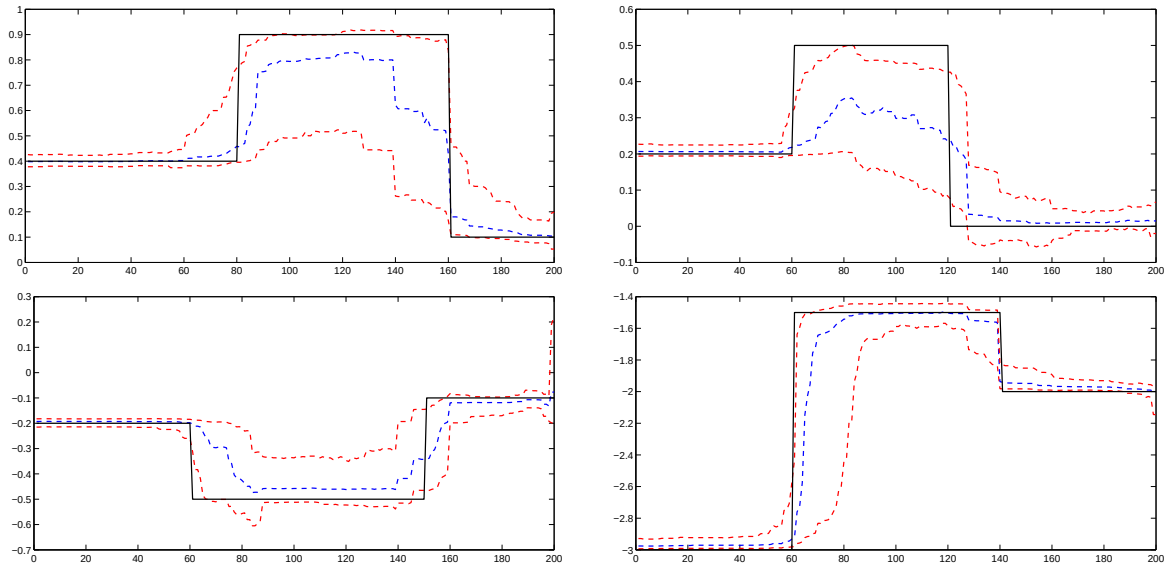
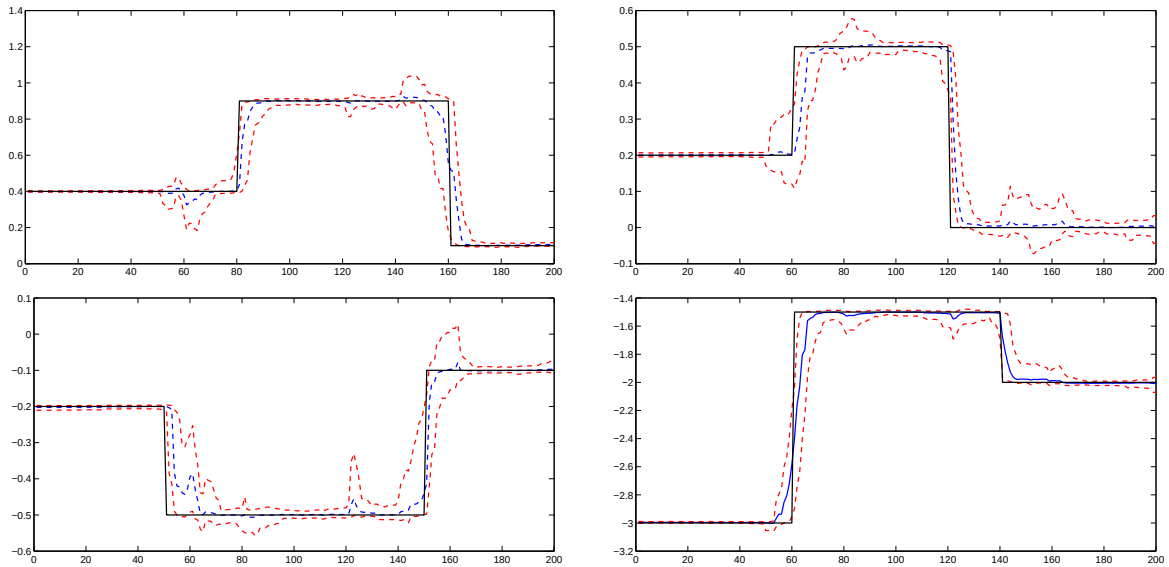


Figure C.4 shows the posterior estimates of $\beta_{i,t}$ for $i = 1, 2, 3$ and $\ln \sigma_t^2$ by assuming an higher prior probability of having a break and a smaller prior expected size of the breaks. The figures makes clear that much of the posterior medians deterioration showed in figure C.3 and C.2 would likely come from a higher expected size of the breaks. In particular assuming *a priori* small sized breaks will result in a less disperse posterior estimate and that the timing of breaks are mostly precisely estimated. A general pattern we observe is that when the prior settings correspond to a higher probability of smaller breaks than in the base case, the posterior estimates of $\beta_{1:3,t}$ and $\ln \sigma_t^2$ are consistent with those implied by the data generating process. However, by imposing ex-ante a larger size of breaks, the precision of posterior median estimates deteriorates. As a whole, the posterior estimates seems to be more sensible to prior hyper-parameters on the breaks size rather than to prior probabilities of breaks on itself.

Figure C.4: Posterior Estimates of the Time-Varying Parameters: Higher prior-break probabilities and small prior break size

This figure plots the posterior distributions of the parameters $\beta_{i,t}$ for $i = 1, 2, 3$ and $\ln \sigma_t^2$ together with the corresponding values implied by the DGP. The blue dashed line reports the median estimates of the parameters. The red dashed lines denote the 20th and 80th percentiles of the posterior distributions. The black solid line displays the values used to define the data generating process. Posterior results are now based on a higher prior-break probabilities for both betas and the (log of) idiosyncratic volatility, while assuming a small expected priors break size.



C.2 Empirical Example

The second step of the prior sensitivity analysis is based on an empirical exercise. We base our results on the full B-TVB-SV model estimated on the original dataset of 23 stock and bond portfolios sorted on size, industry and maturity, and 9 macroeconomic risk factors (see Table 1 in the main text). The different prior specifications are those reported in Table C.3. Figure C.5 shows the distribution of posterior break probabilities for each of the macroeconomic risk factors, and averaged across the 23 stock and bond portfolios. The red dashed line corresponds to the posterior under the base case prior. The black line corresponds to the posterior under the "Large" case ($a_j = 3.2$, $b_j = 12$, $a_\nu = 1$ and $b_\nu = 10$), while the blue dot-dashed line represents the posterior distribution under the "Higher" case ($a_j = 3.2$, $b_j = 6$, $a_\nu = 1$ and $b_\nu = 5$). Figure C.5 makes clear that by assuming, *a priori* a higher probability of having a break does not lead to sensibly different posterior estimates of the instability of the betas in the full B-TVB-SV model. In fact, posterior estimates tend to largely overlap across explanatory macroeconomic factors. The same applies by looking at the posterior distribution of break probabilities for idiosyncratic variances. Figure C.6 reports the average posterior probabilities of having a break in $\ln \sigma_{i,t}^2$ under different priors. Again, posterior estimates tend to largely overlap under different priors.

As a further assessment we investigate the role of priors on break probabilities for $\ln \sigma_{i,t}^2$ in isolation. Figure C.7 shows the distribution of posterior break probabilities for each of the macroeconomic risk factors, and averaged across the 23 stock and bond portfolios. The prior structure for the betas is kept constant to the base case ($a_j = 3.2$, $b_j = 60$, and $\gamma_j = 0.5$, $\delta_j = 100$ for $j = 1, .., 10$). The red dashed line corresponds to the posterior under the base case prior ($a_\nu = 1$ and $b_\nu = 99$). The black line corresponds to the posterior under the "Large" case ($a_\nu = 1$ and $b_\nu = 10$), while the blue dot-dashed line represents the posterior distribution under the "Higher" case ($a_\nu = 1$ and $b_\nu = 5$). Figure C.7 makes clear that posterior estimates are rather robust with respect to different priors specifications for break probabilities on $\ln \sigma_{i,t}^2$. Interestingly, the data seem to be rather informative in defining the amount instability required by the dynamics of betas and idiosyncratic volatility. In fact, different priors do not lead to dramatically different results.

Figure C.5: Posterior Distributions of Break Probabilities of the Betas for Fixed Break Sizes

This figure plots the posterior distributions of the break probabilities for the betas averaged across portfolios and for each of the 9 macroeconomic factors depicted in Table B.1 in the main text. The red dashed line corresponds to the posterior under the base case prior. The black line corresponds to the posterior under the "Large" case ($a_j = 3.2$, $b_j = 12$, $a_\nu = 1$ and $b_\nu = 10$), while the blue dot-dashed line represents the posterior distribution under the "Higher" case ($a_j = 3.2$, $b_j = 6$, $a_\nu = 1$ and $b_\nu = 5$).

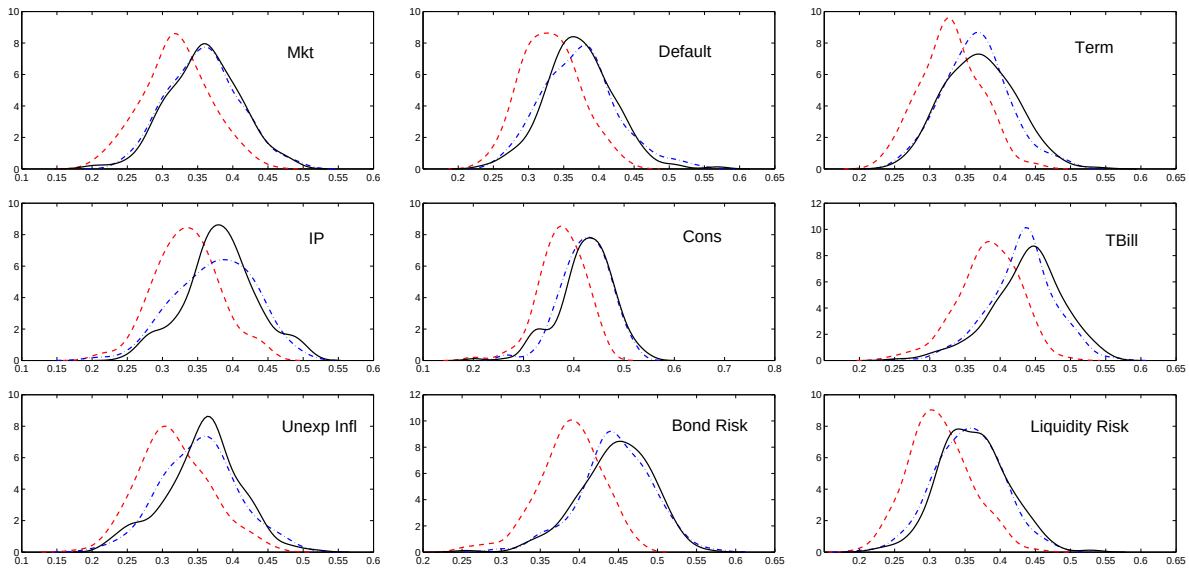


Figure C.6: Posterior Distributions of Break Probabilities of Idiosyncratic Volatility for Fixed Break Sizes

This figure plots the posterior distributions of the break probabilities for the (log of) idiosyncratic variances averaged across the 23 portfolios reported in Table 1 in the main text. The red dashed line corresponds to the posterior under the base case prior. The black line corresponds to the posterior under the "Large" case ($a_j = 3.2$, $b_j = 12$, $a_\nu = 1$ and $b_\nu = 10$), while the blue dot-dashed line represents the posterior distribution under the "Higher" case ($a_j = 3.2$, $b_j = 6$, $a_\nu = 1$ and $b_\nu = 5$).

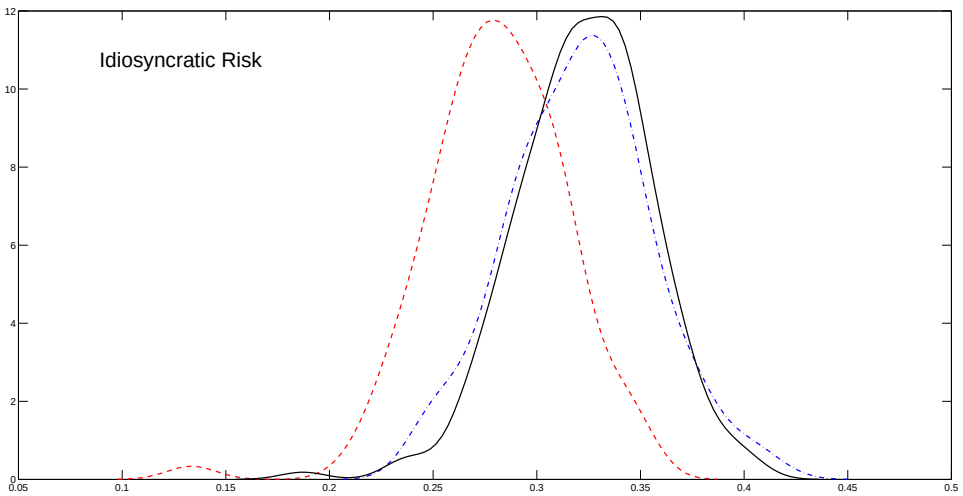
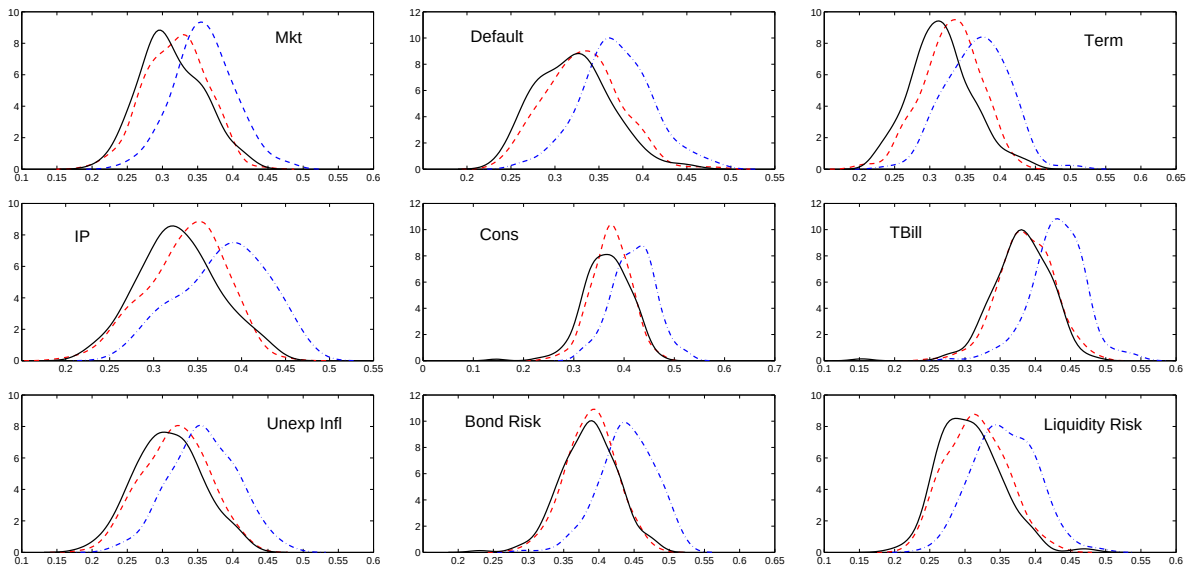


Figure C.7: Posterior Distributions of Break Probabilities of the Betas by Changing Priors on Idiosyncratic Volatility, and Keeping Fixed Break Sizes and Betas Prior Structure

This figure plots the posterior distributions of the break probabilities for the betas averaged across portfolios and for each of the 9 macroeconomic factors depicted in Table 1 in the main text. The prior structure on the betas is kept constant, while priors on $\ln \sigma_{i,t}^2$ change. The red dashed line corresponds to the posterior under the base case prior. The black line corresponds to the posterior under the "Large" case ($a_\nu = 1$ and $b_\nu = 10$), while the blue dot-dashed line represents the posterior distribution under the "Higher" case ($a_\nu = 1$ and $b_\nu = 5$).



D In-Sample Posterior Inference on Stochastic Volatility

A number of studies in the finance literature have compared alternative models of time-varying volatility of asset returns (e.g. Hansen and Lunde 2005, Geweke and Amisano 2010, and Clark and Ravazzolo 2014). More recently Eisenstat and Strachan (2014) discuss estimation of volatility in the context of inflation, in particular whether it should be modelled as a stationary process or as a random walk. In our B-TVB-SV model when a break arrives, log-volatility follows a random walk. While such random walk assumption might be indeed useful for practical reasons, it can be criticized as inappropriate since it implies that the range of possible values of volatility is unbounded in probability in the limit, which is obviously something we do not observe in financial markets. On the other hand, stationary processes, say an AR(1) dynamics for log-volatility, are bounded in the limit, even though are close to be non-stationary at monthly frequencies, and also substantially increase the parameter space have to be estimated.

In this section, we report some in-sample properties of different specifications of the stochastic volatility component of our general model reported in Section 3 in the main text. The purpose is to investigate if the functional for volatility implied by the change-point dynamics sensibly affects the results in comparison of standard stationary and random walk specifications. For the model in-sample estimation we used the original Dataset of 23 portfolios of stocks sorted by size and industry, and bond portfolios sorted by maturity. Data are monthly and cover the sample period 1972:01 - 2011:12. The first ten years of data are used to calibrate the priors for each model.

For both discussion and estimation, we use a common observation equation specification for each of the excess returns $r_{t.t}$ on the 23 portfolios. For the sake of simplicity we assume a constant mean model. In general, we write this as

$$r_{i,t} = \mu + \sigma_{it}\epsilon_{i,t} \qquad i = 1, ..., N,$$

We restrict the discussion to two specifications, namely our change-point dynamics as in Section 3, and a standard AR(1) stationary process. The change-point specification for log-volatility, is

defined as (see the main text for more details)

$$\ln(\sigma_{i,t}^2) = \ln(\sigma_{i,t-1}^2) + \kappa_{i\upsilon,t}\upsilon_{i,t} \quad \text{with} \quad \upsilon_{i,t} \sim N(0, q_i) \quad \text{and} \quad \kappa_{i\upsilon,t} = \begin{cases} 1 & \pi_{i\upsilon} \\ 0 & 1 - \pi_{i\upsilon} \end{cases} \tag{D.13}$$

The alternative stationary specification we consider is a standard AR(1) dynamics

$$\ln(\sigma_{i,t}^2) = (1 - \delta_i)\,\overline{\ln(\sigma_i^2)} + \delta_i \ln(\sigma_{i,t-1}^2) + \upsilon_{i,t} \qquad \text{with} \qquad \upsilon_{i,t} \sim N(0, q_i) \tag{D.14}$$

with $\overline{\ln(\sigma_i^2)}$ the long-run mean, and δ_i the asset specific persistence parameter of log-volatility. Clearly both the change-point model and the AR(1) nests a random walks dynamics when $\pi_{i\upsilon} = 1$ and $\delta_i = 1$, respectively. Figure D.8 shows the median estimates of $\ln(\sigma_{i,t}^2)$ according to (D.13) and (D.14) respectively. We report the results for the size-sorted portfolios for the sake of readability. The blue line corresponds to the stationary AR(1), while the red line is the median estimate under the B-TVB-SV model. This figure makes clear that both of the models specifications helps to capture spikes in conditional volatility, for instance around the period 2000/2002, across different assets. In other words, at least in finite samples, there is not clear benefit in using a stationary dynamics as the AR(1) as opposed to the full B-TVB-SV model. One potential reason is that, indeed, at the monthly frequency, the AR(1) dynamics of log-volatility is close to be non-stationary, i.e. $\delta_i = 1$. Figure D.9 shows the posterior distribution of the persistence parameters δ_i across the same set of size-sorted portfolios reported in Figure D.8. Indeed, shocks to the AR(1) log-volatility turn out to have a largely persistent effect. The average posterior median estimate of δ_i is well above 0.9 on a monthly frequency.

A general pattern we observe is that, at least in finite samples, both a highly persistent AR(1) and a change-point dynamics may help to capture the same dynamic features of log-volatility. The question is now why we argue the latter might be indeed better to fit the data. We compared the log-marginal likelihoods of the specifications (D.13)-(D.14) together with a standard random walk dynamics. Table D.4 shows the results. The full B-TVB-SV model delivers the highest log-marginal likelihood across all the size-sorted portfolios. Interestingly, the data are clearly in favor of the stationary AR(1) dynamics as opposed to the random walk one.

As a whole, posterior estimates of log-volatility would not radically change by using a more

Figure D.8: Posterior Median Estimates of Log-Volatility Under the Change-Point Dynamics vs. Stationary AR(1)

This figure plots the posterior median estimates of the log-volatility for a set of size-sorted portfolios across the period 1972:01 - 2011:01. The blue line corresponds to the stationary AR(1), while the red line is the median estimate under the B-TVB-SV model. Prior hyper-parameters are trained in both cases by using a pre-sample period of ten years.

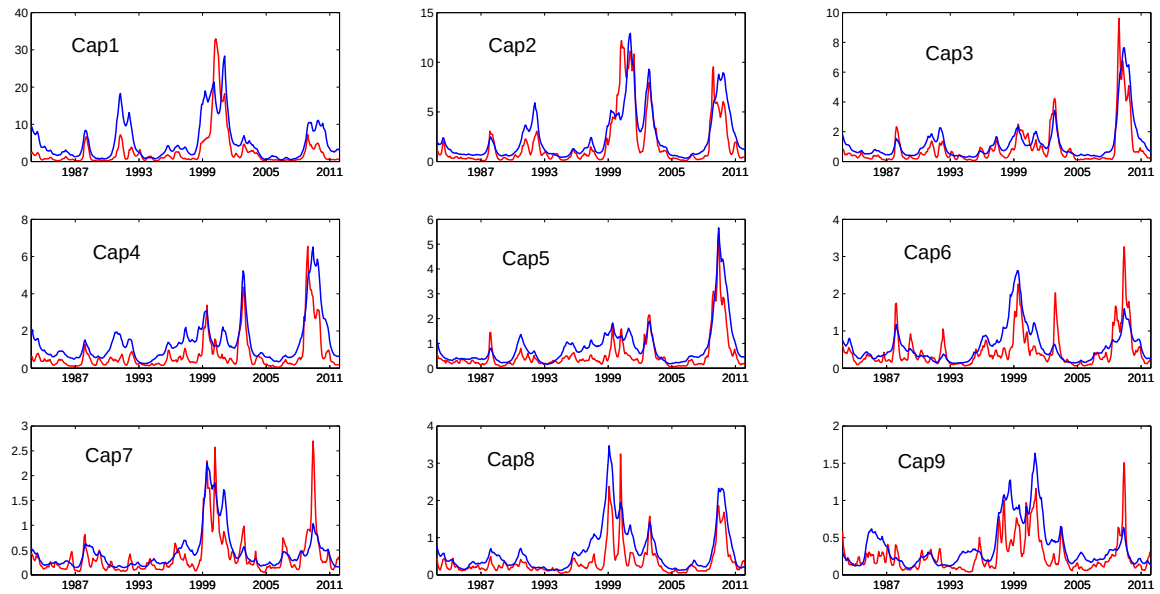
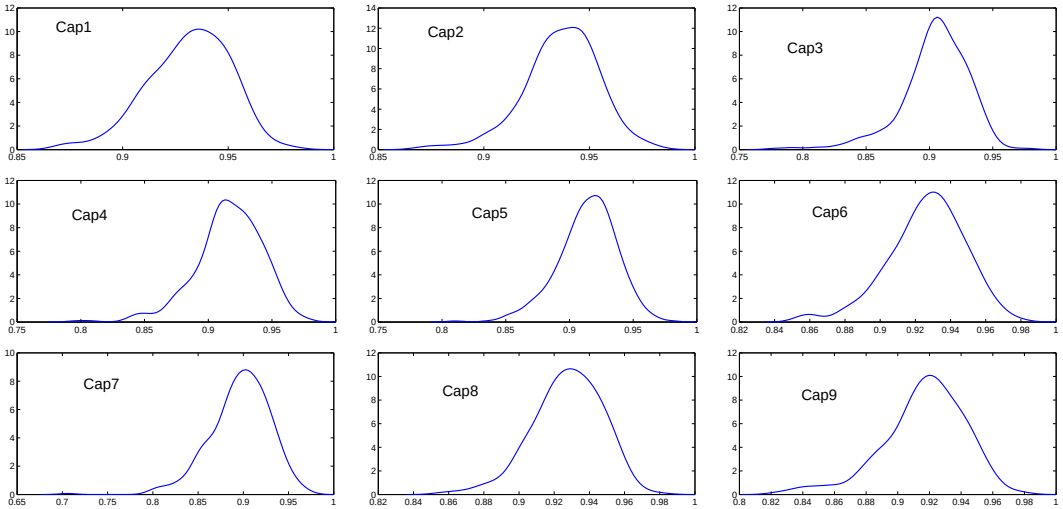


Figure D.9: Posterior Estimates of the Persistence Parameters for the AR(1) Log-Volatility

This figure plots the posterior distribution estimates of the persistence parameters δ_i for the log-volatility for a set of size-sorted portfolios across the period 1972:01 - 2011:01. The blue line corresponds to the stationary AR(1), while the red line is the median estimate under the B-TVB-SV model. Prior hyper-parameters are trained in both cases by using a pre-sample period of ten years.



standard stationary AR(1) dynamics. The latter however, is strongly rejected by the data in typical set of 40 years of post-war data on size-sorted stocks. Interestingly, the full B-TVB-SV closely behaves as a highly persistent, stationary, AR(1) process (of course, this is true in finite samples and not asymptotically).

E 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. (5) decomposes excess asset returns in a component related to risk, represented by the term $\sum_{j=1}^{K} \lambda_{j,t}\beta_{ij,t-1}$ plus a residual $\lambda_{0,t} + e_{i,t}$. In principle, a multi-factor model is as good as the implied percentage of total variation in excess returns explained by its first component, $\sum_{j=1}^{K} \lambda_{j,t}\beta_{ij,t-1}$. However, here we should recall that even though (5) refers to excess returns, it remains a statistical implementation of the framework in (4). This implies that in practice it may be naive to expect that $\sum_{j=1}^{K} \lambda_{j,t}\beta_{ij,t-1}$ be able to explain much of the variability in excess returns. A more sensible goal seems to be that

Table D.4: Log-Marginal Likelihoods Across Alternative Stochastic Volatility Specifications

This table the values of the log-marginal likelihoods for different specifications of stochastic volatility. The values of log marginal likelihoods are reported for ten stocks portfolios sorted on size. *Change-Point* stands for the full model proposed in the main text, while *Stationary* and *Random Walk*, respectively represents a model with a stationary and random walk dynamics for the stochastic volatility process.

	10 Size-Sorted Portfolios, Value Weighted		
	Change-Point	**Stationary**	**Random Walk**
Decile 1	-534.140	-613.760	-805.997
Decile 2	-444.477	-598.132	-776.736
Decile 3	-307.812	-479.321	-734.209
Decile 4	-279.771	-460.944	-733.441
Decile 5	-232.713	-415.034	-717.713
Decile 6	-217.594	-432.015	-719.466
Decile 7	-168.411	-445.438	-705.954
Decile 8	-148.350	-312.585	-704.538
Decile 9	-96.393	-433.774	-691.064
Decile 10	-43.428	-222.075	-683.426

$\sum_{j=1}^{K} \lambda_{j,t}\beta_{ij,t-1}$ 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 $t-1$, $\mathbf{Z}_{t-1}$,

$$x_{i,t} = \theta_{i0} + \sum_{m=1}^{M} \theta_{im} Z_{m,t-1} + \xi_{i,t}, \tag{E.15}$$

to compute the sample variance of fitted values,

$$Var[P(x_{it}|\mathbf{Z}_{t-1})] \equiv Var\left[\hat{\theta}_{i0} + \sum_{m=1}^{M} \hat{\theta}_{im} Z_{m,t-1}\right], \tag{E.16}$$

where the notation $P(x_{it}|\mathbf{Z}_{t-1})$ 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, $\sum_{j=1}^{K} \lambda_{j,t}\beta_{ij,t-1}$, is regressed onto the instrumental variables,

$$\sum_{j=1}^{K} \lambda_{j,t}\beta_{ij,t-1} = \theta'_{i0} + \sum_{m=1}^{M} \theta'_{im} Z_{m,t-1} + \xi'_{i,t} \tag{E.17}$$

to compute the sample variance of fitted risk compensations:

$$Var\left[P\left(\sum_{j=1}^{K}\lambda_{j,t}\beta_{ij,t-1}|\mathbf{Z}_{t-1}\right)\right] \equiv Var\left[\hat{\theta}'_{i0}+\sum_{m=1}^{M}\hat{\theta}'_{im}Z_{m,t-1}\right]. \tag{E.18}$$

The predictable component of excess returns in (E.15) not captured by the model is then the sample variance of the fitted values from the regression of the residuals $\hat{\xi}_{i,t}$ on the instruments:

$$Var\left[\hat{\xi}_{i,t}\right] = Var\left[P\left(\lambda_{0,t}+e_{i,t}|\mathbf{Z}_{t-1}\right)\right]. \tag{E.19}$$

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

$$VR1 \equiv \frac{Var\left[P\left(\sum_{j=1}^{K}\lambda_{j,t}\beta_{ij,t|t-1}|\mathbf{Z}_{t-1}\right)\right]}{Var[P(x_{it}|\mathbf{Z}_{t-1})]} > 0 \tag{E.20}$$

$$VR2 \equiv \frac{Var\left[P\left(\lambda_{0,t}+e_{i,t}|\mathbf{Z}_{t-1}\right)\right]}{Var[P(x_{it}|\mathbf{Z}_{t-1})]} > 0. \tag{E.21}$$

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 our B-TVB-SV framework, drawing from the joint posterior densities of the factor loadings $\beta_{ij,t-1}$ and the implied risk premia $\lambda_{j,t}$, $i = 1, ..., N$, $j = 1, ..., K$, and $t = 1, ..., T$, and holding the instruments fixed over time, it is possible to compute VR1 and VR2 in correspondence to each of such draws and hence obtain their posterior distributions.[1]

Finally, the predictable variation of returns due to the multi-factor model may be further decomposed into the components imputed to each of the individual systematic risk factors, by

[1]Notice that $VR1 = 1$ does not imply that $VR2 = 0$ and viceversa, because

$$Var[P(x_{it}|\mathbf{Z}_{t-1})] \neq Var\left[P\left(\sum_{j=1}^{K}\hat{\lambda}_{j,t}\hat{\beta}_{ij,t-1}|\mathbf{Z}_{t-1}\right)\right] + Var\left[P\left(r_{i,t}-\hat{\theta}_{i0}-\sum_{m=1}^{M}\hat{\theta}_{im}Z_{m,t-1}|\mathbf{Z}_{t-1}\right)\right].$$

computing the factoring of $Var[P(\sum_{j=1}^{K}\lambda_{j,t}\beta_{ij,t-1}|\mathbf{Z}_{t-1})]$ as

$$\sum_{j=1}^{K} Var\left[P\left(\lambda_{j,t}\beta_{ij,t-1}|\mathbf{Z}_{t-1}\right)\right] + \sum_{j=1}^{K}\sum_{k=1}^{K} Cov[P\left(\lambda_{j,t}\beta_{ij,t-1}|\mathbf{Z}_{t-1}\right), P\left(\lambda_{k,t}\beta_{ik,t-1}|\mathbf{Z}_{t-1}\right)] \quad \text{(E.22)}$$

and tabulating $Var\left[P\left(\lambda_{j,t}\beta_{ij,t-1}|\mathbf{Z}_{t-1}\right)\right]$ for $j = 1, ..., K$ as well as the residual factor $\sum_{j=1}^{K}\sum_{k=1}^{K} Cov[$ $P\left(\lambda_{j,t}\beta_{ij,t-1}|\mathbf{Z}_{t-1}\right),\ P\left(\lambda_{k,t}\beta_{ik,t-1}|\mathbf{Z}_{t-1}\right)]$ to pick up any interaction terms. Note that because of the existence of the latter term, the equality

$$\sum_{j=1}^{K} \frac{Var\left[P\left(\lambda_{j,t}\beta_{ij,t-1}|\mathbf{Z}_{t-1}\right)\right]}{Var\left[P\left(\sum_{j=1}^{K}\lambda_{j,t}\beta_{ij,t-1}|\mathbf{Z}_{t-1}\right)\right]} = 1 \quad \text{(E.23)}$$

fails to hold, i.e., the sum of the K risk compensations should not equal the total predictable variation from the asset pricing model because of the covariance among individual risk compensations. This derives from the fact that even though in (1) the risk factors are assumed to be orthogonal, this does not imply that their time-varying total risk compensations ($\lambda_{j,t}\beta_{ij,t-1}$ for $j = 1, ..., K$) should be orthogonal.

References

Carter, C. and Kohn, R. (1994). On gibbs sampling for state-space models. *Biometrika*, (81):541–553.

Clark, T. and Ravazzolo, F. (2014). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, Forthcoming.

Eisenstat, E. and Strachan, R. (2014). Modelling inflation volatility. *CAMA Working Paper 21/2014*.

Ferson, W. and Harvey, C. (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.

Gerlach, R., Carter, C., and Kohn, R. (2000). Efficient bayesian inference for dynamic mixture models. *Journal of the American Statistical Association*, (95):819–828.

Geweke, J. and Amisano, G. (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26:216–230.

Groen, J., Paap, R., and Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business and Economic Statistics*, 31:29–44.

Hansen, P. and Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20:873–889.

Karolyi, G. and Sanders, A. (1998). The variation of economic risk premiums in real estate returns. *Journal of Real Estate Finance and Economics*, 17:245–262.

Omori, Y., Chib, S., Shepard, N., and Nakajima, J. (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.

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