Market Timing in Bayesian Portfolio Optimisation

Simona Abis
Columbia University

September 13, 2017

Abstract

I present a portfolio allocation model that combines a data-based approach with macroeconomic considerations of the business cycle. The model accounts for the two key features of business cycles, namely co-movement among macroeconomic variables and asymmetric development of the cycles. The joint treatment of these characteristics improves the ability of the model to time market turns. The ensuing regime-dependent probability distributions of returns account for more extreme behaviors in bear markets, and hence more accurately describe non-linear downside and skewness risks. The model has a numerical solution which can be applied recursively in order to optimize the portfolio selection, accounting for market turns. This quantitative method leads to enhanced portfolio gains.

1 Introduction

Macro hedge funds are a type of sophisticated investor who tries to generate positive alpha on the bases of predictability induced by macroeconomic indicators. They express their views by entering various long and short positions in different asset classes, including equities. The main type of skill required to be successful within this asset class is market timing, hence it is crucial for them to be able to time market turns on the bases of macroeconomic analysis. Looking at the HFRI Macro index, which summarizes the performance of macro hedge funds, it can be shown that in the last two big recessions this type of fund outperformed other hedge fund types both in terms of higher excess return and lower volatility. Additionally, if computing the correlation between the returns on the macro hedge fund index and the returns on different stock indexes, it can be shown that in recessions this correlation is negative. These results are even more pronounced when looking at quantitative macro hedge funds. This evidence suggests that it should be possible, through the analysis of macroeconomic indicators, to outperform the market in recessions by obtaining early on a more accurate estimate of market turns.

The work presented in this Chapter is motivated by the behavior of quantitative macro hedge funds and tries to exploit the characteristics of the business cycle within a factor model in order to optimize asset allocation and improve portfolio performance. The model has a numerical solution which can be applied recursively in order to optimize the portfolio selection.

More specifically, I solve the portfolio allocation problem of a quantitative investor who assumes returns to follow a multi-factor model and utilizes information on how these factors jointly behave across financial cycles in order to improve his timing ability. Here the growth rate of financial market is assumed to be a latent variable jointly determined by the co-movement among factors within phases of the cycle and their markov-switching behavior. Hence, the procedure proposed includes both characteristics of business cycles as defined by Burns and Mitchell 1946: co-movement among macroeconomic variables within the cycles and asymmetric development of the cycles themselves. As shown by Kim and Nelson 1998, including both characteristics within the same model improves the timing of market turns. Additionally the quantitative estimation procedure allows to consider a large number of assets and account simultaneously for estimation uncertainty, mispricing uncertainty and structural instability.

The proposed model is able to provide a significantly more accurate estimation of bull and bear markets (from 74% to 94% accuracy as compared to a Markov-switching model without co-movement). This allows generating regime-dependent probability distributions of returns that account for more extreme behaviors in bear markets, hence better accounting for non-linear downside and skewness risks. Ultimately this leads to higher portfolio gains both in-sample and out-of-sample and finally it allows to better time market turns. Failing to include co-movement determines a less precise estimation of regimes and has a significant economic impact on the portfolio decision.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 illustrates the methodology used to solve and estimate the model. Section 4 describes the data to which the model is applied. Section 5 illustrates the results obtained. Section 6 concludes. Appendix **??** features details regarding the model's derivation and estimation.

2 Related Literature

2.1 Classical vs Bayesian portfolio selection

Within the classical portfolio optimization framework, the parameters that dominate the probability distribution of returns are considered as given and a point estimate of such parameters is used as a proxy for the true ones. This type of portfolio optimization ignores parameter uncertainty and its impact on the chosen optimal weights. A way to account for parameter uncertainty in the portfolio optimization problem is by using a Bayesian framework where portfolio weights are jointly estimated with the parameters of the return distribution. Bayesian portfolio optimization consists in solving the following problem:

$$max_{\omega} \int U(W_{t+1}) p(r_{t+1}|\Phi) dr_{t+1} \tag{2.1}$$

Where $p(r_{t+1}|\Phi) \propto \int p(r_{t+1}|\theta, \Phi) p(\theta) L(\theta, \Phi) d\theta$ is the predictive density, θ represents the parameters of the probability distribution of returns ,Φ indicates all available information , $U(.)$ is the chosen utility function, ω are the portfolio weights and W_{t+1} represents wealth at time $(t+1)$

The Bayesian portfolio optimization problem as defined above was first introduced and studied by Bawa and Klein 1976. The use of the posterior distribution of θ accounts for parameter uncertainty and allows incorporating prior information regarding the probability distribution of returns, which can come from fundamental information, asset pricing models, equilibrium relationships, forecasting models or other subjective views. Among the others, this approach is used by Black and Litterman 1992, Pastor 2000 and Pastor and Stambaugh 2001. The latter combine asset pricing theory with a data-based approach by centering the portfolio optimization problem on known asset pricing models and developing a prior for the model parameters that reflected the degree of uncertainty in the model's pricing abilities.

The model proposed in this Chapter is developed into a Bayesian setting since this provides an immediate way to account for parameter uncertainty and estimation error. Additionally the approach developed by Pastor 2000 is used in order to incorporate into the decision process the uncertainty relative to the pricing ability of the chosen model.

2.2 Evidence of regime-switches in macroeconomic variables and its impact on the probability distribution of returns

There exists substantial evidence that many macroeconomic variables and subsequently asset returns follow a regime-switching process which has usually been associated with business cycle dynamics. Ang and Bekaert 2002a, Ang and Bekaert 2002b , Ang and Chen 2002, Guidolin and Timmermann 2006 and Guidolin and Timmermann 2007, among others, study the impact of regime-switching on stock returns and portfolio selection, finding significant evidence in favor of regime-switches. These studies suggests that the dynamic character of the environment should be taken into account as a source of uncertainty within portfolio optimization problems.

Guidolin and Timmermann 2007 apply Kim's approximation of the Hamilton filter to study the effect of regime switches in stock returns on the optimal asset allocation. They find that the optimal allocation with and without regime switching is significantly different in statistical and economical terms. Under the regime switching scenario investors vary their portfolio allocation considerably as they update their estimate of the state probabilities. The chosen estimation method, though, only accounts for one of the sources of uncertainty, structural uncertainty, whereas parameter uncertainty and mispricing uncertainty are ignored. Additionally they do not take into consideration co-movement among factors within the different regimes.

Tu 2010 develops a model that accounts simultaneously for parameter, mispricing and structural uncertainty and applies it to portfolio selection. The treatment of mispricing uncertainty and the portfolio selection problem are in line with Pastor 2000. Tu 2010 solves a Markov-switching factor model through Gibbs sampling which allows to solve the model for a large number of assets and readily obtain the posterior and predictive distributions of returns. He does not account for the co-movement among factors within cycles either, instead he assumes uninformative priors for the probability distribution of the factors and a normally distributed likelihood.

The model proposed adds to the literature by including a more informative structure for the probability distribution of the factors, which stresses the importance of co-movements among variables within regimes. The state of the economy is driven by a latent regime-switching variable and is determined by the co-movements within factors as opposed to being estimated within the asset pricing equation. This is obtained by nesting the returns equation within a state-space structure, whose latent variable is the state-switching growth of the financial market. The two dynamic processes feed into each and are jointly estimated through Gibbs sampling. This procedure provide better estimates of market turns and consequently has a positive impact on portfolio performance, especially during recessions.

3 Methodology

3.1 Portfolio optimization problem

Let's consider a risk averse investor with a one period investment horizon who has to choose a portfolio allocation in order to maximize the expected utility of next period's wealth. He can invest in $(n+k)$ risky assets, n non-benchmark assets and the k benchmark assets (factors), and on the risk-less rate (r_f). Define W_t as the current period's wealth, π as the percentage of wealth invested in the risk-less asset and ω as the vector of weights invested in the risky assets (in excess of the risk-less rate).

Then the wealth at time $(t+1)$ will be given by:

$$W_{t+1} = W_t(1 + r_f + (1-\pi)(\omega' r_{t+1}) \tag{3.1}$$

In line with equation (1), the investor will chose ω to maximize the expected utility of next period's wealth by solving the following optimization problem:

$$\begin{aligned} & max_\omega \int U(W_{t+1}) p(R_{t+1}|\Phi) dR_{t+1} = \\ = \quad & max_\omega \int\int U(W_{t+1}) p(R_{t+1}|\theta,\Phi) p(\theta) L(\theta,\Phi) d\theta dR_{t+1} \\ = \quad & max_\omega \int U(W_{t+1}) \left[\int p(R_{t+1}|\theta,\Phi) p(\theta|\Phi) d\theta\right] dR_{t+1} \end{aligned} \tag{3.2}$$

Where: $p(R_{t+1}|\Phi) = \int p(R_{t+1}|\theta,\Phi)p(\theta|\Phi)d\theta$ is the predictive distribution of returns

The methodology developed below allows to obtain the whole predictive distribution of returns $p(R_{t+1}|\Phi)$, hence to solve the problem for different types of utility functions. In line with Pastor 2000 and Tu 2010, let's first consider a mean-variance investor, whose optimal allocation only depends on the first two moments of the predictive distribution of returns. The optimal mean-variance allocation involves investing a percentage of wealth on the risk-less asset and a percentage on the tangent portfolio, the exact allocation depends on the degree of risk aversion of the investor.

The mean variance investor has to solve the following optimization problem:

$$max_\omega \left(\omega' E[R_{t+1}|\Phi] - \frac{\gamma}{2}\omega' V[R_{t+1}|\Phi]\omega\right) \tag{3.3}$$

Where γ is the coefficient of risk aversion, $E[R_{t+1}|\Phi]$ is the expected value of the predictive distribution and $V[R_{t+1}|\Phi]$ is the variance of the predictive distribution.

The weights invested in the risky assets are independent of the level of risk aversion and are given by:

$$\omega^* = \frac{1}{\gamma}\left\{V[R_{T+1}|\Phi_T]\right\}^{-1} E[R_{T+1}|\Phi_T] \tag{3.4}$$

3.2 State-Switching Factor Model

This session defines the distribution of returns within the context of a Markov-switching factor model and illustrates how to solve for the first two moments of the predictive distribution of returns, which are necessary to solve the portfolio optimization problem just illustrated.

Define F as a matrix of factors of dimensions $(T \times k)$ where k is the number of factors taken into consideration and T is the total number of periods. Now define R as a matrix of asset returns in excess of the risk-free rate of dimensions $(T \times n)$, where n is the total number of available risky assets.

Returns are assumed to follow the factor structure below:

$$r_{i,t} = \alpha_i^s + \beta_i^s f_{i,t} + u_{i,t}^s \tag{3.5}$$

$$u_{i,t}^s \sim iidN(0, \sigma_{i,s}^2) \tag{3.6}$$

For $i = 1...n$ and $s = 1...S$. The above factor model can also be expressed in matrix notation as: $R = XA^s + U^s$, where R is a $(T \times n)$ matrix containing the returns on the n assets, $X = [\iota_T F]$ is a $(T \times (k+1))$ matrix, where ι_T is a T-dimensional vector of ones. $A^s = [\alpha^s B^s]'$ is a matrix of dimensions $((k+1) \times n)$ containing all regression coefficients, where B^s is a matrix of dimensions $(k \times n)$ containing the betas relative to the n regressions and α^s is a vector of dimensions $(1 \times n)$ containing the alphas of the n regressions. Finally U^s is a $(T \times n)$ matrix containing the error terms relative to the n regressions, such that: $U \sim iidN(0, \Sigma^s \otimes I_T)$ and Σ^s is a $(n \times n)$ diagonal matrix.

The model above is state switching, hence all main parameters will be different conditional on the state of the economy, making it a non-linear specification. It is useful to think about this problem in the context of model uncertainty. In fact, conditional on the state, the above specification can be rewritten as a linear Gaussian regression model; we have as many of these models as the number of states of the world (S). Thinking about the problem in this manner greatly simplifies estimation, as most calculations can be done within the context of a linear Gaussian model.

What differentiates this specification from the more classical regime-switching models is that the regime is implicitly derived from the factor structure, in order to account for the co-movement among variables within business cycles and the asymmetric evolution of the cycles themselves. The state-space Markov-switching part below is similar to that of Kim and Nelson 1999, Chapter 10, but it is adapted in order to allow for both the mean and the volatility of the latent regime-switching process to be state dependent.

$$f_{i,t} = \rho_i + \zeta_i(L)\Delta C_t + e_{i,t} \tag{3.7}$$

$$\eta_i(L)e_{it} = \epsilon_{it} \tag{3.8}$$

$$\epsilon_{it} \sim iidN(0, v_i^2) \tag{3.9}$$

$$\lambda(L)(\Delta C_t - \mu_t^s - \delta) = w_t \tag{3.10}$$

$$w_t \sim iidN(0, \sigma_{c,s_t}^2) \tag{3.11}$$

$$\mu_t^s = \mu_0 + \mu_1^s s_t \tag{3.12}$$

$$\sigma_{c,s_t}^2 = \sigma_{c,0}^2(1 - s_t) + \sigma_{c,1}^2 s_t = \sigma_{c,0}^2(1 + h_1 s_t) \tag{3.13}$$

$$P[s_t = 1 | s_{t-1} = 1] = p \tag{3.14}$$

$$P[s_t = 0 | s_{t-1} = 0] = q \tag{3.15}$$

Where $f_{i,t}$ represents the ith indicator, ΔC_t is the growth rate of the financial market, $\mu_1 > 0$, $h_1 > -1$ and $s = 1, ..., S$ and the state probabilities follow a Markov chain.

Assuming that there are only two states of the world and limiting the auto-regressive processes to two lags the model can be simplified. After a few transformations, it can be shown that the model above can be represented with the following matrix notation, which translates into a state-space model (see Appendix .2 for details).

$$F_t^{**} = \Pi\xi_t + v_t \tag{3.16}$$

$$\xi_t = M_{s_t} + \Lambda\xi_{t-1} + \kappa_t \tag{3.17}$$

Equation (3.16) is the measurement equation of the Markov-switching model, while equation (3.17) is the transition equation. The above system of equations would generally be solved with a Kalman filter. In this case, though, this simple solution is not possible because both the state of the economy and the latent factor are unknown and need to be estimated jointly with the rest of the parameters of the model.

Equations (3.5) - (3.6) together with equations (3.7) - (3.15) constitute the full model that needs to be solved. Due to the estimation issues described above Gibbs sampling is used. This allows to derive the full empirical joint distribution of a set of variables given their conditional distributions.

In order to solve the model we need the joint posterior distribution of the latent regime-switching variable, the state of the economy, the factors, the asset returns and all parameters of the model. This information provides us with the predictive distribution of returns whose first two moments, $E[R_{t+1}|\Phi_T]$ and $V[R_{t+1}|\Phi_T]$, are necessary to solve the portfolio allocation problem. The reminder of this Section derives the conditional posterior distributions of all elements of the model and Section 3.3 describes how to use the Gibbs sampler to obtain the predictive distribution of returns, while detailed derivations can be found in Appendix **??**.

3.2.1 *The business cycle latent variable*

From the transformed model, we can derive the business cycle latent variable $\Delta\tilde{c_T} = [\Delta c_1 ... \Delta c_T]$ conditional on all available data ($\tilde{f}_T = [f_1^* ... f_T^*]$), all parameters of the model and the state of the economy. Note that the history of the asset returns doesn't add any information for the derivation of the business cycle variable, once the history of the factors is known. Hence it won't be needed to condition on the distribution of returns in this case. The business cycle latent variable can be obtained through a Kalman filter procedure. It is possible to use a Kalman filter because, conditional on the state of the economy and all parameters, the model is linear Gaussian.

The matrix to be estimated is $\tilde{\xi_T} = [\xi_1 ... \xi_T]'$; all elements can be estimated simultaneously from the following joint distribution: $p(\tilde{\xi}_T|\tilde{f}_T)$. Given that the conditional state-space model is Gaussian also the distributions of ξ_T and ξ_t will be Gaussian, such that:

$$\xi_T|\tilde{f}_T \sim N(E_T[\xi_T], E_T[P_T]) \tag{3.18}$$

$$\xi_t|\xi_{t+1}, \tilde{f}_t \sim N(E_t[\xi_t|\xi_{t+1}], E_t[P_t|\xi_{t+1}]) \tag{3.19}$$

Where P_t indicates the covariance of ξ_t for $t = 1, ..., T$ and $E_t[.] = E[.|\tilde{f}_t]$. The first element of each vector ξ_t, $\xi_t(1)$, gives us an estimate for Δc_t, for $t = 1, ..., T$.

3.2.2 *The parameters of the k factor equations*

Next we need to derive the posterior distribution of the parameters relative to the k factor equations described in equation (.19), (ζ_i, η_i, v_i^2), conditional on all available data, the business cycle latent variable just estimated and the state of the economy. Also in this case the history of asset returns doesn't add any additional information, hence we won't be conditioning on it. This assumes conditional independence such that:

$$\begin{aligned} p(R, F|\theta, \tilde{s}_T, \Delta\tilde{c}_T, \zeta, \tilde{\eta})) &= \\ = p(R, F|X, \Sigma, E[F], V[F], \tilde{s}_T, \Delta\tilde{c}_T, \zeta, \tilde{\eta}) & \\ = p(R|X, \Sigma, E[F], V[F], \tilde{s}_T) p(F|\Delta\tilde{c}_T, \tilde{s}_T, \zeta, \tilde{\eta}) p(\tilde{s}_T|\Delta\tilde{c}_T) & \end{aligned} \tag{3.20}$$

Also in this case the conditional distributions of the parameters can be easily obtained through Bayesian updating of Gaussian distributions. In fact, conditional on $\Delta\tilde{c_T}$ and s_t this represents a simple system of k equations with uncorrelated disturbances, as illustrated in the equations below.

$$f^*_{i,t} = \zeta_i \Delta c_t + e_{i,t} \tag{3.21}$$

$$\eta_i(L)e_{it} = \epsilon_{it} \tag{3.22}$$

$$\epsilon_{it} \sim iidN(0, v_i^2) \tag{3.23}$$

By assuming a Normal prior for ζ_i and η_i and an Inverted Gamma prior for v_i^2, the posterior conditional distribution of the parameters is:

$$\zeta_i|\tilde{\eta}_i, v_i, \Delta\tilde{c_T}, \tilde{f}_T \sim N\left(\bar{\gamma}_i, \bar{\Gamma}_i\right) \tag{3.24}$$

$$\tilde{\eta}_i|\zeta, v_i, \Delta\tilde{c_T}, \tilde{f}_T \sim N\left(\bar{\gamma}^*_i, \bar{\Gamma}^*_i\right) \tag{3.25}$$

$$v_i^2|\zeta_i, \tilde{\eta}_i, \Delta\tilde{c_T}, \tilde{f}_T \sim IG\left(\frac{j_i + (T-2)}{2}, \frac{z_i + (\tilde{e}_{iT} - E_i\tilde{\eta}_i)'\,(\tilde{e}_{iT} - E_i\tilde{\eta}_i)}{2}\right) \tag{3.26}$$

3.2.3 *The state of the economy and related parameters*

We also need to derive the joint posterior distribution for the state of the economy and all its related parameters $(s_t, \lambda_1, \lambda_2, \mu_0, \mu_1, \sigma^2_{c,0}, \sigma^2_{c,1}, p, q)$, conditional on all available information, $(R, F, , \Delta\tilde{c_T})$, and all other model parameters just estimated. The history of the factors , the asset returns and the other parameters of the system don't add any information beyond that contained in $\Delta\tilde{c}_T$ so conditioning on them is not necessary. Hence the model to be considered in this session is simply an auto-regressive model with Markov switching mean and variance. This is represented by the equations below:

$$\Delta c_t = \lambda_1 \Delta c_{t-1} + \lambda_2 \Delta c_{t-2} + \mu_{s,t} - \lambda_1 \mu_{s,t-1} - \lambda_2 \mu_{s,t-2} + w_t \tag{3.27}$$

$$w_t \sim iidN(0, \sigma^2_{c,s_t}) \tag{3.28}$$

$$\mu^s_t = \mu_0 + \mu^s_1 s_t \tag{3.29}$$

$$\sigma^2_{c,s} = \sigma^2_{c,0}(1 - s_t) + \sigma^2_{c,1} s_t = \sigma^2_0(1 + h_1 s_t) \tag{3.30}$$

$$P[s_t = 1|s_{t-1} = 1] = p \tag{3.31}$$

$$P[s_t = 0|s_{t-1} = 0] = q \tag{3.32}$$

The joint posterior distribution to be estimated is the following:

$$p(\tilde{s}_T, \tilde{\lambda}, \mu_0, \mu_1, \sigma^2_{c,0}, \sigma^2_{c,1}, p, q) = p(\tilde{\lambda}, \mu_0, \mu_1, \sigma^2_{c,0}, \sigma^2_{c,1}|\Delta\tilde{c}_T)p(p, q|\tilde{s}_T)p(\tilde{s}_T|\Delta\tilde{c}_T) \tag{3.33}$$

The posterior distribution is obtained in three steps: first the state is computed from $p(\tilde{s}_T|\Delta\tilde{c}_T)$, then the transition probabilities are drawn from $p(p, q|\tilde{s}_T)$ and finally the other parameters are extracted from $p(\tilde{\lambda}, \mu_0, \mu_1, \sigma^2_{c,0}, \sigma^2_{c,1}|\Delta\tilde{c}_T)$. The procedure for obtaining $p(\tilde{s}_T|\Delta\tilde{c}_T)$ is very similar to that already detailed for the estimation of $\Delta\tilde{c}_T$. Normal prior distributions are assumed for $\tilde{\lambda}, \mu_0, \mu_1$, while Inverted Gamma priors are assumed for $\sigma^2_{c,0}, \sigma^2_{c,1}$. The Bayesian updating yields the following posterior distributions:

$$p(\tilde{s}_T|\Delta\tilde{c}_T) = p(s_T|\Delta\tilde{c}_T) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, \Delta\tilde{c}_t) \tag{3.34}$$

$$p|\tilde{s}_T \sim beta(o_{1,1} + n_{1,1}, o_{1,0} + n_{1,0}) \tag{3.35}$$

$$q|\tilde{s}_T \sim beta(o_{0,0} + n_{0,0}, o_{0,1} + n_{0,1}) \tag{3.36}$$

$$\tilde{\mu}|\tilde{\lambda}, \sigma^2_{c,0}, \sigma^2_{c,1}, \tilde{s}_T, \Delta\tilde{c}_T \sim N(b_1, B_1) \tag{3.37}$$

$$\sigma^2_{c,0}|h_1, \tilde{\mu}, \tilde{\lambda}, \tilde{s}_T, \Delta\tilde{c}_T \sim IG\left(\frac{v^1}{2}, \frac{\delta^1}{2}\right) \tag{3.38}$$

$$\bar{h}_1 = (1 + h_1)|\sigma^2_{c,0}, \tilde{\mu}, \tilde{\lambda}, \tilde{s}_T, \Delta\tilde{c}_T \sim IG\left(\frac{v^{11}}{2}, \frac{\delta^{11}}{2}\right) \tag{3.39}$$

$$\sigma^2_{c,1} = \sigma^2_{c,0}\bar{h}_1 \tag{3.40}$$

3.2.4 *Parameters of the return factor model*

Prior:

Conditional on the state of the economy, the history of the factors, $\Delta\tilde{c}_T$ and all model parameters, in order to estimate the parameters of the factor model it is required to define the prior distribution on the following set of parameters: $\theta = \{\alpha^s, B^s, \Sigma^s\}$. Note that once the probability distribution of the factors and the state of the economy are known, $\Delta\tilde{c}_T$ doesn't add any information, hence it won't be necessary to condition on it.

The prior on α is of particular relevance as it allows to account for mispricing uncertainty, as defined by Pastor and Stambaugh 1999. If having a dogmatic belief on the pricing ability of the chosen model, on average we would expect α to be zero, most tests of asset pricing models use this identity for hypothesis testing. We are aware, though, that, according to hypothesis testing, most models should be rejected, including the most commonly used asset pricing models such as the CAPM. The alternative to using asset pricing models is to use a data-based approach by limiting estimations to historical averages but this approach highly reduces the information set. So Pastor and Stambaugh 1999 proposed an alternative treatment of the problem by which, through the prior on α we can introduce a subjective degree of uncertainty on the level of mispricing of the model, with the consequence that posterior estimates of the parameter will are weighted average of the parameters implied by the asset pricing model and those implied by historical means, using as weighs their relative precision. Such mispricing is introduced through the prior volatility of α. Following Pastor and Stambaugh 1999 we then define the prior distribution of α as:

$$p(\alpha^s|\Sigma^s) \sim N\left(0, \left(\sigma_\alpha^2 \frac{1}{(s^s)^2} \Sigma^s\right)\right) \tag{3.41}$$

Where σ_α^2 identifies the degree of mispricing uncertainty that can be let vary from 0, representing a dogmatic belief in the model, to ∞ representing the case of complete disbelief. Note that the prior belief could also be made state dependent by allowing for a different level of mispricing uncertainty conditional on the state of the economy, by introducing $(\sigma_\alpha^s)^2$. $E[\Sigma^s] = (s^s)^2 \iota_N$ is the prior expectation of Σ^s, $(s^s)^2 = tr\left[(R^s - X^s\hat{A}^s)'(R^s - X^s\hat{A}^s)/T^s\right]/(n+k)$. As pointed out by Pastor and Stambaugh 1999 the prior variance of α is proportional to the variance of the regression residuals as otherwise, through portfolio optimization, it would be possible to obtain extremely high Sharpe Ratios. The reasons behind this phenomenon are exposed in detail in MacKinlay 1995.

From which the prior distribution of all regression parameters is defined as:

$$A^s|\Sigma^s \sim N(\bar{A}^s, \Psi(\Sigma^s)) \tag{3.42}$$

$$\Psi(\Sigma^s) = \begin{bmatrix} \sigma_\alpha^2 \frac{1}{s^2}\Sigma^s & 0 \\ 0 & \Omega^s \end{bmatrix} \tag{3.43}$$

$$(\Sigma^s)^{-1} \sim W\left((H^s)^{-1}, \nu\right) \tag{3.44}$$

Where Ω^s is a diagonal matrix with very large diagonal elements, in order to assure a diffuse prior for betas. It is also independent from Σ, $\bar{A}^s$ represents the prior mean of the regression parameters. The prior mean of the α coefficients is set to zero, whereas the prior mean of the β coefficients is set to their historical average. $W\left((H^s)^{-1}, \nu\right)$ is a Wishart distribution with H parameter matrix and ν degrees of freedom, while $(H^s)^{-1} = (s^s)^2 (\nu - n + k - 1)\iota_{n+k}$.

The joint prior on the regression parameters, conditional on the state of the economy, can be written as:

$$p(A^s, \Sigma^s) = p(A^s|\Sigma^s)p(\Sigma^s) = p(\alpha^s|\Sigma^s)p(\Sigma^s)p(B^s) \tag{3.45}$$

The equality $p(A^s|\Sigma^s)p(\Sigma^s) = p(\alpha^s|\Sigma^s)p(\Sigma^s)p(B^s)$ is only possible because of how the matrix $\Psi(\Sigma^s)$ is defined - α^s and B^s are independent and the variance of B^s is independent from Σ^s.

The joint prior for the regression parameters can be rewritten as a function of only (a^s, Σ^s), where $a^s = vec(A^s)$.

The joint prior on all parameters, conditional on the state, can be defined as:

$$p'(\theta^s) = p(\alpha^s|\Sigma^s)p(\Sigma^s)p(B^s) \tag{3.46}$$

Likelihood:

Conditional on the state of the economy and the factors, the joint likelihood of the asset returns can be defined as follows:

$$p(R^s|A^s, \Sigma^s) \sim MVN(X^s A^s, \Sigma^s) \propto |\Sigma^s|^{-\frac{T^s}{2}} exp\left\{-\frac{1}{2} tr\left[(R^s - X^s A^s)'(R^s - X^s A^s)\left(\Sigma^s\right)^{-1}\right]\right\} \tag{3.47}$$

$$\begin{aligned} &\propto |\Sigma^s|^{-\frac{T^s}{2}} exp\{-\tfrac{1}{2} tr\left[(R^s - X^s \hat{A}^s)'(R^s - X^s \hat{A}^s)\left(\Sigma^s\right)^{-1}\right] + \\ &-\tfrac{1}{2} tr\left[\left(A^s - \hat{A^s}\right)'(X^s)'(X^s)(A^s - \hat{A}^s)\left(\Sigma^s\right)^{-1}\right]\} \\ &\propto |\Sigma^s|^{-\frac{T^s}{2}} exp\left\{-\tfrac{T^s}{2} tr\left[\hat{\Sigma^s}\left(\Sigma^s\right)^{-1}\right] - \tfrac{1}{2} tr\left[\left(a^s - \hat{a^s}\right)'\left(\left(\Sigma^s\right)^{-1} \otimes (X^s)'(X^s)\right)\left(a^s - \hat{a^s}\right)\right]\right\} \end{aligned} \tag{3.48}$$

Where MVN stands for Multivariate Normal distribution, $tr(.)$ computes the trace of a matrix, $\hat{\Sigma^s} = \frac{1}{T}(R^s - X^s\hat{A}^s)(R^s - X^s\hat{A}^s)'$, $\hat{A}^s = (X^{s'}Xv)^{-1}X^{s'}R^s$, $a^s = vec(A^s)$.

Posterior:

The posterior on the regression parameters, conditional on the state of the economy, is obtained as the product of the prior and the likelihood on the regression parameters presented in equations (.116) and (3.48), such that:

$$p(\theta^s|R^s) = p(R^s|\theta^s)p'(\theta^s) \tag{3.49}$$

The solution yields the following probability distributions for the regression parameters, as following Tu 2010:

$$\left((\Sigma^s)^{-1}|R^s, F^s\right) \sim W\left((T^s + \nu - k), (H^s + T\hat{\Sigma^s} + \hat{A^s}' J^s \hat{A^s})^{-1}\right) \tag{3.50}$$

$$\left(a^s|\,(\Sigma^s)^{-1}, R^s, F^s\right) \sim N\left(\tilde{a^s}, \left(\Sigma^s \otimes (G^s)^{-1}\right)\right) \tag{3.51}$$

3.3 Joint estimation through Gibbs sampling

The Gibbs sampler allows to draw from a joint distribution by iteratively drawing samples form the conditional distributions of the parameters involved. In this case the steps to be followed are:

1. Draw $\Delta c_t$from the following distribution

$$\Delta c_t|\Delta c_{t+1}, \tilde{f}_t \sim N(E_t[\xi_t|\Delta c_{t+1}](1), E_t[P_t|\Delta c_{t+1}](1,1))$$

2. Draw the parameters relative to the factors equations from their respective distributions:

$$\zeta_i|\tilde{\eta}_i, v_i, \Delta\tilde{c_T}, \tilde{f}_T \sim N\left(\bar{\tau}_i, \bar{\Gamma}_i\right)$$

$$\tilde{\eta}_i|\zeta, v_i, \Delta\tilde{c_T}, \tilde{f}_T \sim N\left(\bar{\tau}_i^*, \bar{\Gamma}_i^*\right)$$

$$v_i^2|\zeta_i, \tilde{\eta}_i, \Delta\tilde{c_T}, \tilde{f}_T \sim IG\left(\frac{j_{i,s} + (T-2)}{2}, \frac{z_{i,s} + (\tilde{e}_{iT} - E_i\tilde{\eta}_i)'(\tilde{e}_{iT} - E_i\tilde{\eta}_i)}{2}\right)$$

3. Draw s_t from the following distribution: $p(s_t|s_{t+1}, \Delta\tilde{c}_t) \propto p(s_{t+1}|s_t)p(s_t|\Delta\tilde{c}_t)$

4. Draw the parameters relative to the regime-switching latent variable:

$$p|\tilde{s}_T \sim beta(o_{1,1} + n_{1,1}, o_{1,0} + n_{1,0})$$

$$q|\tilde{s}_T \sim beta(o_{0,0} + n_{0,0}, o_{0,1} + n_{0,1})$$

$$\tilde{\mu}|\tilde{\lambda}, \sigma_{c,0}^2, \sigma_{c,1}^2, \tilde{s}_T, \Delta\tilde{c_T} \sim N(b_1, B_1)$$

$$\sigma_{c,0}^2|h_1, \tilde{\mu}, \tilde{\lambda}, \tilde{s}_T, \Delta\tilde{c}_T \sim IG\left(\frac{v^1}{2}, \frac{\delta^1}{2}\right)$$

$$\bar{h}_1|\sigma_{c,0}^2, \tilde{\mu}, \tilde{\lambda}, \tilde{s}_T, \Delta\tilde{c}_T \sim IG\left(\frac{v^{11}}{2}, \frac{\delta^{11}}{2}\right)$$

$$\tilde{\lambda}|\tilde{\mu}, \sigma_{c,0}^2, \sigma_{c,1}^2, \tilde{s}_T, \Delta\tilde{c}_T \sim N(\bar{b}, \bar{B})$$

5. Given s_T and the parameters drawn above, draw s_{T+1} and compute $E[F_{T+1}|\Phi_T]$ and $V[F_{T+1}|\Phi_T]$

6. Separate the returns and the factor data into two groups according to the state of the economy.

7. Draw the parameters relative to the posterior distribution of returns for $s = 1, 2$:

$$((\Sigma^s)^{-1} | R^s, F^s) \sim W\left((T^s + \nu - k), (H^s + T\hat{\Sigma}^s + \hat{A}^{s\prime} J^s \hat{A}^s)^{-1}\right)$$

$$(a^s | (\Sigma^s)^{-1}, R^s, F^s) \sim N\left(\tilde{a^s}, \left(\Sigma^s \otimes (G^s)^{-1}\right)\right)$$

8. Given $s_{T+1} = i$ drawn above, draw $R_{T+1} | a^s, \Sigma^s, F^s$ from the multivariate normal distribution implied by $R = XA^s + U^s$, $U \sim iidN(0, \Sigma^s \otimes I_T)$.

Repeat steps 1-8 above until convergence of the Gibbs sampler. Once convergence is reached, successive draws will come from the joint posterior distribution of all parameters involved. The starting values for the initial iteration are chosen arbitrarily, while the starting values for the following ones are the parameters drawn in the previous iteration. The latest value drawn is always used as conditioning variable.

The illustrated iterative process needs to be repeated $M = (I + L)$ times, each iteration provides a vector of draws of the type:

$$(R_{T+1}, \Delta\tilde{c}_T, \tilde{s}_T, \zeta_i, \tilde{\eta_i}, v_i^2, p, q, \tilde{\mu}, \sigma_{c,0}^2, \sigma_{c,1}^2, \tilde{\lambda}, (\Sigma^s)^{-1}, a^s)_m \tag{3.52}$$

For $m = 1, ..., M$. The first I draws are discarded, where I is the number of draws necessary for the convergence of the Gibbs sampler. The remaining $L = M - I$ draws are kept, and represent draws from the joint posterior distribution of returns. These draws are indicated as:

$$(R_{T+1}, \Delta\tilde{c}_T, \tilde{s}_T, \zeta_i, \tilde{\eta_i}, v_i^2, p, q, \tilde{\mu}, \sigma_{c,0}^2, \sigma_{c,1}^2, \tilde{\lambda}, (\Sigma^s)^{-1}, a^s)_l \tag{3.53}$$

And the joint posterior distribution from which they are drawn as:

$$p(R_{T+1}, \Delta\tilde{c}_T, \tilde{s}_T, \zeta_i, \tilde{\eta_i}, v_i^2, p, q, \tilde{\mu}, \sigma_{c,0}^2, \sigma_{c,1}^2, \tilde{\lambda}, (\Sigma^s)^{-1}, a^s) \tag{3.54}$$

Given the L samples from the joint posterior distribution, the moments of the joint distribution and all marginal distributions can be easily computed.

3.4 Portfolio Optimization

Following the principle just described, the first two moments of the predictive distribution can be obtained as:

$$E[R_{T+1}|\Phi_T]^L = \frac{1}{L} \sum_{l=1}^{L} (R_{T+1})_l \tag{3.55}$$

$$V[R_{T+1}|\Phi_T]^L = \frac{1}{L-1} \left[\left((R_{T+1})_l - E[R_{T+1}|\Phi_T]^L\right)' \left((R_{T+1})_l - E[R_{T+1}|\Phi_T]^L\right)\right] \tag{3.56}$$

These can be used to solve the one period ahead portfolio optimization problem for a mean-variance investor, so obtaining the portfolio weights given by:

$$\omega^* = \frac{1}{\gamma} \{V[R_{T+1}|\Phi_T]\}^{-1} E[R_{T+1}|\Phi_T] \tag{3.57}$$

Where the expected value and covariance terms used are the first two moments of the predictive distribution of returns as obtained in equations (3.55) and (3.56) respectively.

The optimal portfolio weights are computed for three different specifications: the one state model (*1S*) the two-state Markov switching model (*MS*) and a two-state Markov switching model where the state are obtained by taking into account the co-movement between variables (*MS_C*).In order to analyze the economic significance of the difference in the optimal allocation, it is not sufficient to compare portfolio returns because even portfolios with very different weights might have a similar portfolio return due to the co-movement between assets. So, as proposed by Pastor and Stambaugh 2000, the certainty equivalent approach is used. This approach consists in computing the difference in the certainty equivalent returns obtained with the two specifications, given that only one is the data generating process the investor believes to be correct.

Given the choice of a mean-variance investor, the certain equivalent of returns and Sharpe Ratio for model i is given by:

$$CER_i = \omega_i' E[R_{T+1}|\Phi_T]^L - \frac{\gamma}{2}\omega_i' V[R_{T+1}|\Phi_T]^L \omega_i \tag{3.58}$$

$$SR_i = \frac{\omega_i' E[R_{T+1}|\Phi_T]^L}{\omega_i V[R_{T+1}|\Phi_T]^L \omega_i} \tag{3.59}$$

The weights are obtained with the two different model specifications, whereas the expected return and variance employed are those of the probability distribution the investor believes to be true. For instance, if assuming the data is Markov switching with co-movement, the Certainty Equivalent difference between the *MS_C* and *MS* models is computed using the moments of the predictive distribution of returns in the Markov-switching case with co-movement. The difference in CER and SR measures the loss in utility and Sharpe Ratio of an investor who, believing in the Markov switching model with co-movement, is forced to invest on the bases of a normal Markov-switching model or a one-state model.

4 Data

The model is applied to the Fama French 10 Industry portfolios. The explanatory variables used in the estimation are the Fama French size and book-to-market factors, the excess return on the market and two additional macroeconomic variables: the growth rate of the Consumer Price Index (CPI) and yield spreads between the Moody's Baa yields and the long term government bond (30 years). The data-set described is specific to the US and the date range considered goes from the first of January 1960 to the first of December 2012. The chosen period includes various financial crises among which the .com bubble and the sub-prime crisis.

As shown by Kim and Nelson 1998, accounting for co-movement among macroeconomic variables significantly improves the estimations of regime switches as compared to the ex-post NBER business cycle indicator. But financial crises don't necessarily coincide with business cycles and, even when they do, we don't have the certainty that co-movement plays a comparatively strong role in this case. For this reason I have created a financial crisis ex-post indicator, similar to that created by Mishkin and White 2002. The 1987 fall is taken as a point of reference in establishing what defines a financial crisis. Both the velocity at which the fall occurs and the duration of the downturn play an important role in defining a financial crisis. For this reason the US financial market is defined to be in a crisis when one of the following conditions is verified: the markets drops of at least 15% in one month, of at least 20% in three months or of at least 40% in six months for ether the NASDAQ, the S&P 500 or the Dow Jones indices. Figure .1 illustrates how the indicator created compares to the NBER business cycle indicator. The two indicators often coincide but their timing isn't identical, sometimes a financial downturn precedes a recession and other times the recession follows a financial crisis. Additionally there are some crises that are only characteristic of the financial markets and do not impact the economy as a whole.

5 Results

The Fama French 10 industry portfolios have notoriously been more difficult to predict using multifactor models than portfolios constructed along other dimensions such as size or book to market. Due to the lower predictability, when estimating a simple Markov-switching model for the regression parameters it is very difficult to capture the true state switches, such that the impact of including a Markov-switching component in the decision problem is low. The results in this section show that simultaneously accounting for mispricing uncertainty, state switches and co-movement during phases of the economic cycle can significantly enhance predictability, having a significant economic impact on portfolio choice.

5.1 Estimated regime switches

Comparing the goodness of the different models in capturing recession periods, Figure .2 shows the predicted state of the economy and the probability of recession obtained using the *MS* model, while Figure .3 displays the same results as obtained with the *MS_C* model. The model with co-movement almost exactly identifies all financial crises, having an accuracy of 94%. Accuracy is defined as the percentage of times in which the estimated probability of financial crisis agrees with the financial crisis index described in Section 4 on the prevailing regime. At times the co-movement model identifies a crisis when none exists, most of the times this is due to the restrictive definition of financial crisis used, in fact usually the conditions for being in a crisis are almost met but the threshold is not crossed. The model without co-movement, despite identifying most crisis periods, is much less precise in their timing, having an overall precision of 74%. It is important to note that in both cases the regime switch is identified solely through the probability distribution of the factors, which are the same in the two models. The only difference is that in the model without co-movement a non-informative prior is given to the probability distribution of the factor, whose likelihood is assumed to be multivariate-normal; in the co-movement model, instead, the joint behavior of the factors is modeled within the state-space component, so considering a more informative behavior of the factors structure.

Finally Figure .4 shows the estimated latent common factor for the MS_C model (ΔC_t), representing the growth rate of the US financial market. It is clearly visible that every time the estimated index drops this exactly coincides with the financial crisis index highlighted.

5.2 Differences in the probability distributions

This section compares the differences in the probability distribution of the factors in recessions and expansions for the two regime-switching models of interest: *MS* and *MS_C*. Figures .5 to .9 show the conditional distribution of the factors in the two regimes for both the *MS* and *MS_C* models.

Focusing on the excess return on the market, we can see that its conditional distribution is clearly different when only one regime is considered from when regime switching is allowed. With both regime-switching models, with and without co-movement, we observe more negative skewness during crises with substantially lower mean, as expected and pointed out in previous studies. What we can additionally see when improving regime estimation by using co-movement is that the crisis regime is more rare but, at the same time, the distribution of the excess return on the market is more skewed to the left with a significantly lower posterior mean. This indicates that when ignoring co-movement we are erroneously still averaging between crisis and tranquil times, leading to an underestimation of tail risk in downturns. A similar reasoning can be applied to the other factors considered.

In order to better understand these results Tables .1 to .8 show the posterior mean and volatility of the distribution of the factors and of the excess return on the 10 industry portfolios. All posterior values are monthly and their standard deviations are reported in parenthesis.

Focusing again on the expected return on the market we can see that, as expected, the posterior mean of the distribution on the one regime model (*1S*) is an average between the bull and bear regimes, which are significantly different from one another in both models that allow for regime switches (*MS* and *MS_C*). The posterior mean is always negative in bear markets, while it is positive in bull markets. What can additionally be seen by comparing the *MS* and *MS_C* models is that the difference in expected excess return between the two regimes is much larger when considering the co-movement model and this difference comes mostly from the bad states, as can be seen by comparing the *(Bull - Bull)* and the *(Bear - Bear)* columns of Table .2. This points out once again the underestimation of the magnitude and negative skewness of downturns when ignoring the information provided by variables co-movement.

If looking at the posterior mean of the excess return on the 10 Industry portfolios in Table .3 we can see that the analysis done above for the excess return on the market applies in the same way here as well.

Looking at volatilities we can see that the posterior volatility of returns is always higher in the bear regime, when introducing the possibility of regime switching, but this time the the difference in volatility between states is less in the co-movement model. If looking at Figure .5 the reason becomes evident. In fact in the model without co-movement the range and dispersion of observations in the crisis period is much wider, this is because it erroneously considers some tranquil times as being crisis times so increasing volatility disproportionately.

Finally Figure .10 compares the predictive distribution of the excess return on the 10 industry portfolios for the three models of interest. Table .9 additionally reports the posterior mean of the predictive distributions for the three models. As it is clearly evident from both the Figure and the Table the mean and volatility of the predictive densities, which are then used for portfolio selection, are clearly different conditionally on the model being considered.

5.3 Differences in portfolio weights

Given the differences in the predictive distributions estimated with the three models, this section analyzes their impact on optimal portfolio choice by comparing optimal weights, Certainty Equivalent of Returns (CERs) and Sharpe Ratios (SRs). Assuming a coefficient of risk-aversion $\gamma = 10$ and a $\sigma_\alpha = 1\%$, Table .10 shows that the impact is quite substantial both in terms of the differences in the single weights and the differences in overall portfolios as measured by the CERs and the SRs. In fact the optimal weight on some of the industry portfolios even goes from positive to negative and magnitudes vary significantly for all industry portfolios. The difference in CER is economically relevant if considering that results are obtained at a monthly horizon. The difference in the optimal allocation and its economic significance become more relevant the higher the level of mispricing uncertainty σ_α. The CER and SR differences also increases substantially when increasing the investment horizon.

5.4 Out-of-sample

In order to analyze the out-of-sample behavior of the modes I left out T^* observations from the initial estimation, for different levels of T^*. For each of them I then estimated the optimal portfolio weights for all three models for the time periods: $N = [1 : (T - T^*)], [1 : (T - T^* + 1)], ..., [1 : (T - T^* + T^*) = 1 : T]$. Then I computed for each of the N runs the realized portfolio returns by multiplying the computed optimal weights by the actual return in the first period excluded in the calculation. I computed an ex-post SR using the N realized returns and compared it by model specification.

Preliminary results suggest that the out-of-sample Sharpe Ratio obtained using the model with co-movement is generally higher than that obtained using the other two models.

6 Conclusion

By taking advantage of the additional information contained in factors co-movement within regimes, in the US financial market, I am able to provide a significantly more accurate estimation of bull and bear markets. This allows to generate regime-dependent probability distributions of returns that account for more extreme behaviors in bear markets, hence better accounting for non-linear downside and skewness risks. Ultimately this leads to higher portfolio gains both in-sample and out-of-sample and finally it allows to better time market turns. Failing to include co-movement determines a less precise estimation of regimes and has a significant economic impact on the portfolio decision.

The results highlighted in this Chapter show that it is possible to use quantitative methods in order to extract private information from public aggregate signals for better portfolio allocation. More work is needed in order to establish a direct connection between the model described in this Chapter and the behavior of quantitative macro hedge funds.

Future work should also analyze the result with a wider spectrum of assets and factors, both in terms of predictability and mostly in terms of portfolio decision. The optimal portfolio choice should be compared conditional on regimes in order to see if the co-movement effect leads to more or less aggressive behavior. The impact of mispricing uncertainty on the prediction should also be explored further. The out-of-sample analysis should be expanded in order to explore the hypothesis that introducing factors co-movement into the decision problem allows to detect and react to crises earlier, possibly providing a form of portfolio hedge in periods when market returns are low. Finally portfolio optimization should be repeated for a different choice of utility function, particularly in the case of skewness aversion, to see if this exacerbates or reduces the effect of factors co-movement on the optimal portfolio choice.

References

Albert, J.H., and S. Chib. 1993. "Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts". *Journal of Business and Economic Statistics* 11 (1): 1–15.

Ang, A., and G. Bekaert. 2002a. "International asset allocation with regime shifts". *Review of Financial Studies* 15:1137–1187.

— . 2002b. "Regime switches in interest rates". *Journal of Business and Economic Statistics* 20:163–182.

Ang, A., and J. Chen. 2002. "Asymmetric correlations of equity portfolios". *Journal of Financial Economics* 63:443–494.

Bawa, V.S., and R.W. Klein. 1976. "The effect of estimation risk on optimal portfolio choice". *Journal of Financial Economics* 3:215–232.

Black, F., and R. Litterman. 1992. "Global portfolio optimization". *Financial Analyst Journal* 48:28–43.

Burns, A.F., and W.C. Mitchell. 1946. *Measuring Business Cycles.* Nber Books.

Carter, C.K., and R. Kohn. 1994. "On Gibbs sampling for state space models". *Biometrika* 81 (3): 541–553.

Casella, G., and E. George. 1992. "Explaining the Gibbs Sampler". *The American Statistician* 46 (3): 167–174.

Guidolin, M., and A. Timmermann. 2007. "Asset allocation under multivariate regime switching". *Journal of Economic Dynamics & Control* 31:3503–3544.

— . 2006. "International asset allocation under regime switching, skew and kurtosis preferences". *The Review of Financial Studies* 21 (2): 889–935.

Kim, Chang-Jin, and Charles R. Nelson. 1998. "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model With Regime Switching". *The Review of Economics and Statistics* 80:188–201.

— . 1999. *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications.* MIT Press Books.

MacKinlay, A.C. 1995. "Multifactor models do not explain deviations from the CAPM". *Journal of Financial Economics* 38 (1): 3–28.

Mishkin, F.S., and E.N. White. 2002. "U.S. stock market crashes and their aftermath implications for monetary policy". *NBER Working Papers Series* Working Paper 8992.

Pastor, L. 2000. "Portfolio selection and asset pricing models". *Journal of Finance* 55:179–223.

Pastor, L., and R.F. Stambaugh. 2000. "Comparing asset pricing models: An investment perspective". *Journal of Financial Economics* 56:335–381.

— . 1999. "Costs of equity capital and model mispricing". *Journal of Finance* 54:67–121.

— . 2001. "The Equity Premium and Structural Breaks". *The Journal of Finance* 56:1207–1239.

Tu, J. 2010. "Is regime switching in stock returns important in portfolio decisions?" *Management Science* 56 (7): 1198–1215.

APPENDIX

.1 General applications of Bayesian updating

The rules described in this session will be useful in understanding the solutions presented throughout the paper.

.1.1 Properties of the multivariate normal distribution

Consider two variables (z_1, z_2) which are jointly distributed as follows:

$$z_1, z_2 | \Phi_{t-1} \sim MVN\left(\left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right), \left[\begin{array}{cc} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{array}\right]\right) \quad (.1)$$

Then the conditional distribution of z_1 on z_2 will be:

$$z_1 | z_2, \Phi_{t-1} \sim N(\mu^*, \Sigma^*) \quad (.2)$$

Where:

$$\mu^* = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(z_2 - \mu_2) \quad (.3)$$

$$\Sigma^* = \Sigma_{1,1} + \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} \quad (.4)$$

.1.2 Bayesian updating for the Multivariate Normal distribution

Consider the following likelihood function, which is distributed as a multivariate normal:

$$\begin{aligned} p(y|\theta) = MVN(P\theta, \Sigma) = & (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}(y - P\theta)'\Sigma^{-1}(y - P\theta)\right\} \quad (.5) \\ = & (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}\theta'(P'\Sigma^{-1}P)\theta + \theta'P'\Sigma^{-1}y - \tfrac{1}{2}y'\Sigma^{-1}y\right\} \\ \propto & exp\left\{-\tfrac{1}{2}\theta'(P'\Sigma^{-1}P)\theta + \theta'P'\Sigma^{-1}y\right\} \end{aligned}$$

The conjugate prior for a multivariate normal likelihood is also multivariate Normal as follows:

$$\begin{aligned} p(\theta) = MVN(m, \Omega) = & (2\pi)^{-\frac{n}{2}}|\Omega|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}(\theta - m)'\Omega^{-1}(\theta - m)\right\} \quad (.6) \\ = & (2\pi)^{-\frac{n}{2}}|\Omega|^{-\frac{1}{2}} exp\left\{-\tfrac{1}{2}\theta'\Omega^{-1}\theta + \theta'\Omega^{-1}m - \tfrac{1}{2}m'\Omega^{-1}m\right\} \\ \propto & exp\left\{-\tfrac{1}{2}\theta'\Omega^{-1}\theta + \theta'\Omega^{-1}m\right\} \end{aligned}$$

Then the posterior will be also multivariate normal and it is obtained as the product of the prior and the likelihood above such that:

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &\propto exp\left\{-\tfrac{1}{2}\theta'(P'\Sigma^{-1}P)\theta + \theta'P'\Sigma^{-1}y\right\} exp\left\{-\tfrac{1}{2}\theta'\Omega^{-1}\theta + \theta'\Omega^{-1}m\right\} \\ &\propto exp\left\{-\tfrac{1}{2}\theta'(\Omega + P'\Sigma^{-1}P)\theta + \theta'(\Omega^{-1}m + P'\Sigma^{-1}y)\right\} \end{aligned} \tag{.7}$$

Hence the posterior follows the distribution below:

$$p(\theta|y) = MVN(\theta^*, \Omega^*) \tag{.8}$$

Where:

$$\theta^* = (\Omega + P'\Sigma^{-1}P)^{-1}(\Omega^{-1}m + P'\Sigma^{-1}y) \tag{.9}$$

$$\Omega^* = (\Omega + P'\Sigma^{-1}P)^{-1} \tag{.10}$$

.2 State-space representation

Given the model defined in equations (3.7)-(3.15) we assume that there are two states of the economy and restrict the auto-regressive processes to two lags such that: $\eta_i(L) = 1 - \eta_{i,1}L - \eta_{i,2}L^2$ and $\lambda(L) = 1 - \lambda_1 L - \lambda_2 L^2$ and additionally define $\zeta_i(L) = \zeta_i$. If we do not operate any transformations, as pointed out by Kim and Nelson 1999 the model is over-identified. Hence it will be required to estimate a model in deviation of the means. In order to do so define: $f^*_{i,t} = f_{it} - \bar{f}_i$ and $\Delta c_t = \Delta C_t - \delta$. Finally by multiplying both sides of equation (3.7) by $\eta_i(L)$ we can solve the following simplified model.

$$\left(f^*_{i,t} - \zeta_i \Delta c_t\right) = -\eta_{i,1}\left(f^*_{i,t-1} - \zeta_i \Delta c_{t-1}\right) + \eta_{i,2}\left(f^*_{i,t-2} - \zeta_i \Delta c_{t-2}\right) + \epsilon_{i,t} \tag{.11}$$

$$\epsilon_{it} \sim iidN(0, v_i^2) \tag{.12}$$

$$(\Delta c_t - \mu_{s,t}) = \lambda_1 (\Delta c_{t-1} - \mu_{s,t-1}) + \lambda_2 (\Delta c_{t-2} - \mu_{s,t-2}) + w_t \tag{.13}$$

$$w_t \sim iidN(0, \sigma^2_{c,s_t}) \tag{.14}$$

$$\mu^s_t = \mu_0 + \mu_1 s_t \tag{.15}$$

$$\sigma^2_{c,s_t} = \sigma^2_{c,0}(1 - s_t) + \sigma^2_{c,1} s_t = \sigma^2_{c,0}(1 + h_1 s_t) \tag{.16}$$

$$P[s_t = 1 | s_{t-1} = 1] = p \tag{.17}$$

$$P[s_t = 0 | s_{t-1} = 0] = q \tag{.18}$$

In order to cast the model into state-space form we need to add another transformation by defining: $f^{**}_{it} = \eta_i(L) f^*_{i,t} = f^*_{i,t} - \eta_{i1} f^*_{i,t-1} - \eta_{i2} f^*_{i,t-2}$, such that:

$$f^{**}_{i,t} = \zeta_i \Delta c_t - \eta_{i,1} \zeta_i \Delta c_{t-1} - \eta_{i,2} \zeta_i \Delta c_{t-2} + \epsilon_{i,t} \tag{.19}$$

$$\epsilon_{it} \sim iidN(0, v_i^2) \tag{.20}$$

$$\Delta c_t = \lambda_1 \Delta c_{t-1} + \lambda_2 \Delta c_{t-2} + \mu_{s,t} - \lambda_1 \mu_{s,t-1} - \lambda_2 \mu_{s,t-2} + w_t \tag{.21}$$

$$w_t \sim iidN(0, \sigma^2_{c,s_t}) \tag{.22}$$

$$\mu^s_t = \mu_0 + \mu^s_1 s_t \tag{.23}$$

$$\sigma^2_{c,s_t} = \sigma^2_{c,0}(1 - s_t) + \sigma^2_{c,1} s_t = \sigma^2_{c,0}(1 + h_1 s_t) \tag{.24}$$

$$P[s_t = 1 | s_{t-1} = 1] = p \tag{.25}$$

$$P[s_t = 0 | s_{t-1} = 0] = q \tag{.26}$$

Writing down the model in matrix notation we have:

$$F^{**}_t = \Pi \xi_t + v_t \tag{.27}$$

$$\xi_t = M_{s_t} + \Lambda \xi_{t-1} + \kappa_t \tag{.28}$$

Or equivalently:

$$\begin{bmatrix} f^{**}_{1,t} \\ \vdots \\ f^{**}_{k,t} \end{bmatrix} = \begin{bmatrix} \zeta_1 & -\zeta_1 \eta_{1,1} & -\zeta_1 \eta_{1,2} \\ \vdots & \vdots & \vdots \\ \zeta_k & -\zeta_k \eta_{k,1} & -\zeta_k \eta_{k,2} \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{k,t} \end{bmatrix} \tag{.29}$$

$$E(v_t v_t') = \Upsilon = \begin{bmatrix} v_1^2 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & v_k^2 \end{bmatrix} \quad (.30)$$

$$\begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \end{bmatrix} = \begin{bmatrix} \lambda(L)\mu_{s_t} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \lambda_1 & \lambda_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \\ 0 \end{bmatrix} \quad (.31)$$

$$E[\kappa_t \kappa_t'] = \Xi_{s_t} = \begin{bmatrix} \sigma^2_{c,s_t} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (.32)$$

.3 Solution of the state-space component

The solution detailed below is similar to that of Kim and Nelson 1999, Chapter 10, with the difference that in their model only the mean of the business cycle variable is Markov switching, whereas in the derivation below both the mean and the variance of the latent variable are Markov switching.

.3.1 Generating the business cycle latent variable

As pointed out in Section 3.2.1, the matrix to be estimated is $\tilde{\xi_T} = [\xi_1 ... \xi_T]'$; all elements can be estimated simultaneously from the following joint distribution: $p(\tilde{\xi}_T|\tilde{f}_T)$. The distribution, thanks to the Markov properties of the system, can be re-written as follows:

$$\begin{aligned} p(\tilde{\xi}_T|\tilde{f}_T) = & \quad p(\xi_T|\tilde{f}_T)p(\xi_{T-1}|\xi_T,\tilde{f}_{T-1})p(\tilde{\xi}_{T-2}|\xi_{T-1},\tilde{f}_{T-2}) \quad (.33) \\ = & \quad p(\xi_T|\tilde{f}_T)p(\xi_{T-1}|\xi_T,\tilde{f}_{T-1})p(\xi_{T-2}|\xi_{T-1},\tilde{f}_{T-2})...p(\xi_1|\xi_2,\tilde{f}_1) \quad = \\ = & \quad p(\xi_T|\tilde{f}_T)\prod_{t=1}^{T-1} p(\xi_t|\xi_{t+1},\tilde{f}_t) \end{aligned}$$

The estimation is obtained through Multimove-Gibbs sampling, as proposed by Carter and Kohn 1994). Given the above simplification the vector ξ_t can be obtained recursively for each time period by first generating ξ_T from $p(\xi_T|\tilde{f}_T)$ and then for $t = (T-1), (T-2), ..., 1$ generating ξ_t from $p(\xi_t|\xi_{t+1}, \tilde{f}_t)$.

before proceeding to show how the estimation can be obtained step by step, note that the matrix Ξ is non-singular, only the element $(1, 1)$ of the matrix is relevant the rest are all zeroes. This is due to the fact that only the first line of the transition equation is relevant, while the other lines are identities. Hence when computing $\xi_t|\xi_{t+1}, \tilde{f}_t$ we won't use the whole vector ξ_{t+1} as a conditioning variable but only its first element:

Δc_{t+1}. This considerations allows to simplify the distributions as follows:

$$p(\tilde{\xi}_T|\tilde{f}_T) = p(\tilde{\Delta c}_T|\tilde{f}_T) = p(\Delta c_T|\tilde{f}_T)\prod_{t=1}^{T-1} p(\Delta c_t|\Delta c_{t+1}, \tilde{f}_t) \quad (.34)$$

$$\Delta c_T|\tilde{f}_T \sim N(E_T[\xi_T](1), E_T[P_T](1,1)) \quad (.35)$$

$$\Delta c_t|\Delta c_{t+1}, \tilde{f}_t \sim N(E_t[\xi_t|\Delta c_{t+1}](1), E_t[P_t|\Delta c_{t+1}](1,1)) \quad (.36)$$

Where $E_T[\xi_T](1)$ is the first element of the vector $E_T[\xi_T]$, $E_T[P_T](1,1)$ is the $(1,1)$ element of the matrix $E_T[P_T]$, $E_t[\xi_t|\Delta c_{t+1}](1)$ is the first element of the vector $E_t[\xi_t|\Delta c_{t+1}]$ and $E_t[P_t|\Delta c_{t+1}](1,1)$ is the $(1,1)$ element of the matrix $E_t[P_t|\Delta c_{t+1}]$.

The estimation can be obtained through the following steps:

First run the Kalman filter to obtain $E_t[\xi_t]$ and $E_t[P_t]$ for $t = 1, ..., T$ and save them. The last iteration provides $E_T[\xi_T]$ and $E_T[P_T]$ which allow to draw Δc_T from $\Delta c_T|\tilde{f}_T \sim N(E_T[\xi_T](1), E_T[P_T](1,1))$.

More in detail, the predictive equations of the Kalman filter, obtained from the transition equation of the model are:

$$E_{t-1}[\xi_t] = M_{s_t} + \Lambda E_{t-1}[\xi_{t-1}] \quad (.37)$$

$$E_{t-1}[P_t] = \Lambda E_{t-1}[P_{t-1}]\Lambda' + \Xi_{s_t} \quad (.38)$$

$$E_{t-1}[\psi_t] = f_t^{**} - E_{t-1}[f_t^{**}] = f_t^{**} - \Pi E_{t-1}[\xi_t] \quad (.39)$$

$$E_{t-1}[g_t] = \Pi E_{t-1}[P_t]\Pi' + \Upsilon \quad (.40)$$

Where ψ_t is the prediction error and g_t is its conditional variance.

The updating equations instead are obtained by applying the simple rule of multivariate normal distributions described in Appendix .1. If ξ_t and $E_{t-1}[\psi_t]$ are assumed to be jointly distributed as a multivariate normal we can easily obtain the distribution of $\xi_t|E_{t-1}[\psi_t], \Phi_{t-1}$as follows:

$$\xi_t, E_{t-1}[\psi_t]|\Phi_{t-1} \sim MVN\left(\left(\begin{array}{c} E_{t-1}[\xi_t] \\ 0 \end{array}\right), \left[\begin{array}{cc} E_{t-1}[P_t] & E_{t-1}[P_t]\Pi' \\ \Pi E_{t-1}[P_t] & E_{t-1}[g_t] \end{array}\right]\right) \quad (.41)$$

Then the updating equations of the Kalman filter will be given by:

$$\xi_t|E_{t-1}[\psi_t], \Phi_{t-1} \sim N(E_t[\xi_t], E_t[P_t]) \quad (.42)$$

$$E_t[\xi_t] = E_{t-1}[\xi_t] + E_{t-1}[P_t]\Pi' E_{t-1}[g_t]^{-1} E_{t-1}[\psi_t] \quad (.43)$$

$$E_t[P_t] = E_{t-1}[P_t] - E_{t-1}[P_t]\Pi' E_{t-1}[g_t]^{-1} \Pi E_{t-1}[P_t] \quad (.44)$$

By defining the Kalman gain as: $K = E_{t-1}[P_t]\Pi' E_{t-1}[g_t]^{-1}$then the updating equations can be re-written as:

$$E_t[\xi_t] = E_{t-1}[\xi_t] + K E_{t-1}[\psi_t] \quad (.45)$$

$$E_t[P_t] = E_{t-1}[P_t] - K\Pi E_{t-1}[P_t] \quad (.46)$$

Once we have drawn Δc_T from $\Delta c_T|\tilde{f}_T \sim N(E_T[\xi_T](1), E_T[P_T](1,1))$ if we consider the generated Δc_{t+1} as additional information to the system, then the distribution of $\Delta c_t|\Delta c_{t+1}, \tilde{f}_t$ can be obtained by applying the updating equations of the Kalman filter recursively for $t = (T-1), (T-2), ..., 1$.

Considering that $\Delta c_{t+1} = \lambda(L)\mu_{s_{t+1}} + \Lambda(1)\xi_t + \kappa_{t+1}(1)$, where $\Lambda(1)$ is the first row of Λ and $\kappa_{t+1}(1)$ is the first element of κ_{t+1}, we have:

$$E_t[\xi_t|\Delta c_{t+1}] = E_t[\xi_t] + E_t[P_t]\Lambda(1)' E_{t+1}[g_t^*]^{-1} E_{t+1}[\psi_t^*] \quad (.47)$$

$$E_t[P_t|\Delta c_{t+1}] = E_t[P_t] - E_t[P_t]\Lambda(1)' E_{t+1}[g_t^*]^{-1} \Lambda(1) E_t[P_t] \quad (.48)$$

$$g_{t+1}^* = \Delta c_{t+1} - \lambda(L)\mu_{s_{t+1}} - \Lambda(1)E_t[\xi_t] \quad (.49)$$

$$\psi_{t+1}^* = \Lambda(1)E_t[P_t]\Lambda(1)' + \sigma^2_{c,s_{t+1}} \quad (.50)$$

Finally Δc_t can be drawn from the distribution:

$$\Delta c_t|\Delta c_{t+1}, \tilde{f}_t \sim N(E_t[\xi_t|\Delta c_{t+1}](1), E_t[P_t|\Delta c_{t+1}](1,1)) \quad (.51)$$

.3.2 Generating the parameters relative to the factor equations

Conditional on $\Delta \tilde{c_T}$ and $\tilde{f}_T$ the system collapses to K equations with auto-correlated disturbances, hence the only relevant equations are:

$$f_{i,t}^* = \zeta_i \Delta c_t + e_{i,t} \quad (.52)$$

$$\eta_i(L)e_{it} = \epsilon_{it} \quad (.53)$$

$$\epsilon_{it} \sim iidN(0, v_i^2) \tag{.54}$$

Generate ζ_i:

First we will generate ζ_i conditional on $\tilde{\eta}_i = [\eta_1 \eta_2]'$, v^2_{i,s_t}, $\Delta\tilde{c_T}$ and $\tilde{f}_T$.

As shown in Appendix .3, by multiplying both sides of equation (.52) by $\eta_i(L) = 1 - \eta_1 L - \eta_2 L^2$ the above system of equations can be re-written as:

$$f^{**}_{i,t} = \zeta_i \Delta c^*_t + \epsilon_{i,t} \tag{.55}$$

$$\epsilon_{it} \sim iidN(0, v_i^2) \tag{.56}$$

Where $f^{**}_{it} = \eta_i(L) f^*_{i,t} = f^*_{i,t} - \eta_{i1} f^*_{i,t-1} - \eta_{i2} f^*_{i,t-2}$ and $\Delta c^*_t = \Delta c_t - \eta_{i1}\Delta c_{t-1} - \eta_{i2}\Delta c_{t-2}$. Or, in matrix form:

$$\tilde{f}^{**}_{i,T} = \zeta_i \Delta\tilde{c}^*_T + \epsilon_i \tag{.57}$$

$$\epsilon_i \sim iidN(0, v_i^2 \iota_{T-2}) \tag{.58}$$

Where $\tilde{f}^{**}_{i,T} = [f^{**}_{i,1} \cdots f^{**}_{i,T}]'$, $\Delta\tilde{c}^*_T = [\Delta c^*_1 \cdots \Delta c^*_T]'$ and $\epsilon_i = [\epsilon_1 \cdots \epsilon_T]'$. Considering a normal prior for ζ_i such that $\zeta_i|\tilde{\eta}_i, v_i \sim N\left(\underline{\tau_i}, \underline{\Gamma_i}\right)$and combining it with its normal likelihood, we obtain the following posterior(see Appendix .1 for more details).

$$\zeta_i|\tilde{\eta}_i, v_i, \Delta\tilde{c_T}, \tilde{f}_T \sim N\left(\bar{\tau}_i, \bar{\Gamma}_i\right) \tag{.59}$$

$$\bar{\tau}_i = \left(\underline{\Gamma_i^{-1}} + v_i^{-2}\Delta\tilde{c}^{*\prime}_T \Delta\tilde{c}^*_T\right)^{-1}\left(\underline{\Gamma_i^{-1}\tau_i} + v_i^{-2}\Delta\tilde{c}^{*\prime}_T \Delta\tilde{f}^{**}_{i,T}\right) \tag{.60}$$

$$\bar{\Gamma}_i = \left(\underline{\Gamma_i^{-1}} + v_i^{-2}\Delta\tilde{c}^{*\prime}_T \Delta\tilde{c}^*_T\right)^{-1} \tag{.61}$$

Generate $\tilde{\eta}_i$:

Next we can generate $\tilde{\eta}_i = [\eta_1 \eta_2]'$ conditional on ζ_i, v_i^2, $\Delta\tilde{c_T}$ and $\tilde{f}_T$. The relevant equations in this case are:

$$e_{it} = \eta_{i1} e_{t-1} + \eta_{i2} e_{t-2} + \epsilon_{it} \tag{.62}$$

$$\epsilon_i \sim iidN(0, v_i^2 \iota_{T-2}) \tag{.63}$$

Or in matrix form:

$$\tilde{e}_{iT} = \tilde{\eta}_i' E_i + \epsilon_i \tag{.64}$$

Where $\tilde{e}_{iT} = [e_{i1} \cdots e_{iT}]' = f^*_{i,t} - \zeta_i \Delta c_t$ and E_i is a $(2 \times (T-2))$ matrix such that $E_i = [\tilde{e}_{iT-1} \tilde{e}_{iT-2}]'$. With a similar procedure as before, assigning a normal prior to $\tilde{\eta}_i$ such that $\tilde{\eta}_i | \zeta_i, v_{is} \sim N\left(\underline{\tau_i}^*, \underline{\Gamma_i}^*\right)$ and multiplying it by its likelihood we obtain the following normal posterior distribution for $\tilde{\eta}_i$:

$$\tilde{\eta}_i | \zeta, v_i, \Delta \tilde{c_T}, \tilde{f}_T \sim N\left(\bar{\tau_i^*}, \bar{\Gamma_i^*}\right) \tag{.65}$$

$$\bar{\tau_i^*} = (\underline{\Gamma_i}^{*-1} + v_i^{-2} E_i E_i')^{-1} (\underline{\Gamma_i}^{*-1} \underline{\tau_i}^* + v_i^{-2} E_i \tilde{e}_{iT}')^{-1} \tag{.66}$$

$$\bar{\Gamma_i^*} = (\underline{\Gamma_i}^{*-1} + v_i^{-2} E_i E_i')^{-1} \tag{.67}$$

Generate v_i^2:

Finally we generate v_i^2conditional on $\tilde{\eta}_i$, ζ_i, $\Delta \tilde{c_T}$, $\tilde{f}_T$ and $\tilde{s}_T$. To do so we assume an inverted gamma prior on v_i^2such that $v_i^2 | \zeta_i, \tilde{\eta}_i \sim IG\left(\frac{j_i}{2}, \frac{z_i}{2}\right)$which, combined with its likelihood yields the following posterior distribution:

$$v_i^2 | \zeta_i, \tilde{\eta}_i, \Delta \tilde{c_T}, \tilde{f}_T \sim IG\left(\frac{j_i + (T-2)}{2}, \frac{z_i + \left(\tilde{e}_{iT} - \tilde{\eta}_i' E_i\right)' \left(\tilde{e}_{iT} - \tilde{\eta}_i' E_i\right)}{2}\right) \tag{.68}$$

.3.3 Generating the parameters relative to the state of the economy

Finally we need to generate the state of the economy and all parameters related to it: $\tilde{s}_T, \lambda_1, \lambda_2, \mu_0, \mu_1, \sigma^2_{c,0}, \sigma^2_{c,1}, p, q$ conditional on $\Delta \tilde{c}_T$. The procedure is very similar to that adopted in Appendix .3.1 and .3.2.

Generating $\tilde{s}_T$:

The probability distribution $p(\tilde{s}_T | \Delta \tilde{c}_T)$ can be transformed in the same manner as equation (**??**) due to the Markov property of the system, such that:

$$p(\tilde{s}_T | \Delta \tilde{c}_T) = p(s_T | \Delta \tilde{c}_T) \prod_{t=1}^{T-1} p(s_t | s_{t+1}, \Delta \tilde{c}_t) \tag{.69}$$

Similarly to what was done in Appendix .3.1 this distribution can be obtained in two steps:

1. First $p(s_t | \Delta \tilde{c}_t)$ is obtained by running a Hamilton filter (the discrete version of a Kalman filter) for $t = 1, ..., T$ and all values are stored. From the last iteration we obtain $p(s_T | \Delta \tilde{c}_T)$ from which s_T can be drawn.
2. Then $p(s_t | s_{t+1}, \Delta \tilde{c}_t)$ can be obtained by iterating backwards starting from s_T obtained above for $t = (T-1), ..., 1$ by drawing s_t from the following distribution:

$$p(s_t | s_{t+1}, \Delta \tilde{c}_t) = \frac{p(s_{t+1} | s_t) p(s_t | \Delta \tilde{c}_t)}{p(s_{t+1} | \Delta \tilde{c}_t)} \propto p(s_{t+1} | s_t) p(s_t | \Delta \tilde{c}_t) \tag{.70}$$

Where $p(s_t|\Delta\tilde{c}_t)$ has been obtained in step 1 and $p(s_{t+1}|s_t)$ are the transition probabilities.

s_t is obtained from the distribution above by first computing:

$$Pr[s_t = 1|s_{t+1} = 1, \Delta\tilde{c}_t] = \frac{p(s_{t+1}|s_t = 1)p(s_t = 1|\Delta\tilde{c}_t)}{\sum_{j=0}^{1} p(s_{t+1}|s_t = j)p(s_t = j|\Delta\tilde{c}_t)} \quad (.71)$$

Then drawing a random number from a normal distribution bounded between $[0, 1]$ and assigning $s_t = 0$ if the random number generated is less than or equal to the value of $Pr[s_t = 1|s_{t+1} = 1, \Delta\tilde{c}_t]$; otherwise $s_t = 1$.

Generating p, q:

The second step consists in generating p, q conditional on $\tilde{s}_T = [s_1, ..., s_T]$. Assuming beta priors for p, q such that $p \sim beta(o_{11}, o_{1,0})$ and $q \sim beta(o_{0,0}, o_{0,1})$ and the joint prior is given by;

$$p(p, q) \propto p^{o_{1,1}-1}(1-p)^{o_{1,0}-1}q^{o_{0,0}-1}(1-q)^{o_{0,1}-1} \quad (.72)$$

These can be combined with the following likelihood:

$$L(p, q|\tilde{s}_t) = p^{n_{1,1}}(1-p)^{n_{1,0}}q^{n_{0,0}}(1-q)^{n_{0,1}} \quad (.73)$$

Where $n_{i,j}$ indicates how many times the system switched from state i to state j which can be easily counted from $\tilde{s}_T$, previously generated.

By combining the prior with the likelihood we obtain the following posterior distribution:

$$p(p, q|\tilde{s}_t) \propto p^{o_{1,1}+n_{1,1}-1}(1-p)^{o_{1,0}+n_{1,0}-1}q^{o_{0,0}+n_{0,0}-1}(1-q)^{o_{0,1}+n_{0,1}-1} \quad (.74)$$

Hence p, q can be drawn from the following posterior distributions:

$$p|\tilde{s}_T \sim beta(o_{1,1} + n_{1,1}, o_{1,0} + n_{1,0}) \quad (.75)$$

$$q|\tilde{s}_T \sim beta(o_{0,0} + n_{0,0}, o_{0,1} + n_{0,1}) \quad (.76)$$

Generating $\tilde{\mu}$:

Then $\tilde{\mu} = [\mu_0 \mu_1]'$ can be generated conditional on $\tilde{s}_T, \Delta\tilde{c}_T, \tilde{\lambda}, \sigma^2_{c,0}, \sigma^2_{c,1}$. In order to do so consider the following part of the system:

$$\Delta c_t = \lambda_1 \Delta c_{t-1} + \lambda_2 \Delta c_{t-2} + \mu_{s,t} - \lambda_1 \mu_{s,t-1} - \lambda_2 \mu_{s,t-2} + w_t \quad (.77)$$

$$w_t \sim iidN(0, \sigma^2_{c,s_t}) \quad (.78)$$

$$\mu_t^s = \mu_0 + \mu_1^s s_t \tag{.79}$$

Defining $\Delta c_t^{**} = L(\lambda)\Delta c_t = \Delta c_t - \lambda_1 \Delta c_{t-2} - \lambda_2 \Delta c_{t-2}$ and substituting equation (.79) into equation (.77) we have:

$$\Delta c_t^{**} = \mu_0(1 - \lambda_1 - \lambda_2) + \mu_1(s_t - \lambda_1 s_{t-1} - \lambda_2 s_{t-2}) + w_t \tag{.80}$$

Dividing both sides of equation (.80) by σ_{c,s_t}^2 and define $\Delta c_t^x = \frac{\Delta c_t^{**}}{\sigma_{c,s_t}}$, $x = \frac{(1-\lambda_1-\lambda_2)}{\sigma_{c,s_t}}$ and $x_{1,t} = \frac{(s_t - \lambda_1 s_{t-1} - \lambda_2 s_{t-2})}{\sigma_{c,s_t}}$ we obtain:

$$\Delta c_t^x = \mu_0 x_{0,t} + \mu_1 x_{1,t} + w_t^* \tag{.81}$$

$$w_t^* = \frac{w_t}{\sigma_{c,s_t}} \sim iidN(0,1) \tag{.82}$$

Which in matrix notation yields:

$$\Delta c^x = X\tilde{\mu} + W \tag{.83}$$

$$W \sim N(0, \iota_T) \tag{.84}$$

Now we can attribute a Normal prior to $\tilde{\mu}$ and proceed to obtaining the posterior as illustrated in Appendix .1.

Let the prior distribution of $\tilde{\mu}$ be such that $\tilde{\mu}|\tilde{\lambda}, \sigma_{c,0}^2, \sigma_{c,1}^2 \sim N(b_0, B_0)$ then combining it with its likelihood we obtain the following posterior distribution for $\tilde{\mu}$:

$$\tilde{\mu}|\tilde{\lambda}, \sigma_{c,0}^2, \sigma_{c,1}^2, \tilde{s}_T, \Delta\tilde{c_T} \sim N(b_1, B_1) \tag{.85}$$

$$b_1 = (B_0^{-1} + X'X)^{-1}(B_0^{-1} b_0 + X'\Delta c^x) \tag{.86}$$

$$B_1 = (B_0^{-1} + X'X)^{-1} \tag{.87}$$

Draws of $\tilde{\mu}$ from the posterior distribution above will only be kept if $\mu_1 > 0$.

Generating $\sigma_{c,0}^2$:

The equation that illustrates the behavior of σ_{c,s_t}^2 is the following:

$$\sigma_{c,s}^2 = \sigma_{c,0}^2(1 - s_t) + \sigma_{c,1}^2 s_t = \sigma_0^2(1 + h_1 s_t) \tag{.88}$$

In order to generate $\sigma_{c,0}^2$, similarly to what was done above, consider equation (.80) and transform it by dividing both sides by by $\sqrt{1 + h_1 s_t}$ such that:

$$\Delta c_t^{xx} = \mu_0 x_{0,t}^x + \mu_1 x_{1t}^x + w_t^x \tag{.89}$$

$$w_t^x = \frac{w_t}{\sqrt{1+h_1 s_t}} \sim N(0, \sigma_{c,0}^2) \tag{.90}$$

Where:

$$\Delta c_t^{xx} = \frac{\Delta c_t^{**}}{\sqrt{1+h_1 s_t}} \tag{.91}$$

$$x_{0,t}^x = \frac{(1-\lambda_1-\lambda_2)}{\sqrt{1+h_1 s_t}} \tag{.92}$$

$$x_{1t}^x = \frac{(s_t - \lambda_1 s_{t-1} - \lambda_2 s_{t-2})}{\sqrt{1+h_1 s_t}} \tag{.93}$$

Then defining an inverted gamma prior for $\sigma_{c,0}^2$ such that:

$$\sigma_{c,0}^2 | h_1, \tilde{\mu}, \tilde{\lambda} \sim IG\left(\frac{v^0}{2}, \frac{\delta^0}{2}\right) \tag{.94}$$

Then combining it with its posterior as shown in Appendix .1 yields:

$$\sigma_{c,0}^2 | h_1, \tilde{\mu}, \tilde{\lambda}, \tilde{s}_T, \Delta\tilde{c}_T \sim IG\left(\frac{v^1}{2}, \frac{\delta^1}{2}\right) \tag{.95}$$

Where:

$$v^1 = v^0 + T \tag{.96}$$

$$\delta^1 = \delta^0 + \sum_{t=1}^{T} \left(\Delta c_t^{xx} - \mu_0 x_{0,t}^x + \mu_1 x_{1t}^x\right)^2 \tag{.97}$$

Generating $\sigma_{c,1}^2$:

In order to generate $\sigma_{c,1}^2$we first generate $\bar{h}_1 = (1+h_1)$ then $\sigma_{c,1}^2$ is obtained through the relationship $\sigma_{c,1}^2 = \sigma_{c,0}^2 \bar{h}_1$. In order to generate $\bar{h}_1$ we transform equation (.80) as follows:

$$\Delta c_t^{xxx} = \mu_0 x_{0,t}^{xx} + \mu_1 x_{1t}^{xx} + w_t^{xx} \tag{.98}$$

Where:

$$\Delta c_t^{xxx} = \frac{\Delta c_t^{**}}{\sigma_{c,0}} \tag{.99}$$

$$x_{0,t}^x = \frac{(1-\lambda_1-\lambda_2)}{\sigma_{c,0}} \tag{.100}$$

$$x_{1t}^{x} = \frac{(s_t - \lambda_1 s_{t-1} - \lambda_2 s_{t-2})}{\sigma_{c,0}} \tag{.101}$$

$$w_t^{x} = \frac{w_t}{\sigma_{c,0}} \sim N(0, 1 + h_1 s_t) \tag{.102}$$

Then if we define the prior for $\bar{h}_1$as $\bar{h}_1|\sigma_{c,0}^2, \tilde{\mu}, \tilde{\lambda} \sim IG\left(\frac{v^{00}}{2}, \frac{\delta^{00}}{2}\right)$and we combine it with its likelihood, we obtain the following posterior distribution:

$$\bar{h}_1|\sigma_{c,0}^2, \tilde{\mu}, \tilde{\lambda}, \tilde{s}_T, \Delta\tilde{c}_T \sim IG\left(\frac{v^{11}}{2}, \frac{\delta^{11}}{2}\right) \tag{.103}$$

Where:

$$v^{11} = v^{00} + T_1 \tag{.104}$$

$$\delta^{11} = \delta^{00} + \sum^{N1} \left(\Delta c_t^{xxx} - \mu_0 x_{0,t}^{xx} + \mu_1 x_{1t}^{xx}\right)^2 \tag{.105}$$

Where T_1 is the number of times the system is in state 1, which can be counted from $\tilde{s}_T$ and the symbol $\sum^{N1}$indicates that the sum is computed only on the data that corresponds to the time periods when $s_t = 1$.

Generating $\tilde{\lambda}$:

To generate $\tilde{\lambda}$ conditional on $\tilde{s}_T, \Delta\tilde{c_T}, \tilde{\mu}, \sigma_{c,0}^2, \sigma_{c,1}^2$consider the following transformation of equation (.77):

$$y_t^* = \lambda_1 y_{t-1}^* + \lambda_2 y_{t-2}^* + w_t^* \tag{.106}$$

Where:

$$y_t^* = \frac{(\Delta c_t - \mu_{s_t})}{\sigma_{c,s_t}} \tag{.107}$$

$$w_t^* = \frac{w_t}{\sigma_{c,s_t}} \sim iidN(0,1) \tag{.108}$$

Then in matrix form we have:

$$Y^* = Y^L \tilde{\lambda} + W \tag{.109}$$

If defining the prior on $\tilde{\lambda}$ as $\tilde{\lambda}|\tilde{\mu}, \sigma_{c,0}^2, \sigma_{c,1}^2 \sim N(\underline{b}, \underline{B})$ and combining it with its likelihood we obtain the following posterior distribution from which $\tilde{\lambda}$ can be drawn:

$$\tilde{\lambda}|\tilde{\mu}, \sigma_{c,0}^2, \sigma_{c,1}^2, \tilde{s}_T, \Delta\tilde{c}_T \sim N(\bar{b}, \bar{B}) \tag{.110}$$

Where:

$$\bar{b} = \left(\underline{B}^{-1} + Y^{L'}Y^{L}\right)^{-1}\left(\underline{B}^{-1}\underline{b} + Y^{L'}Y^{*}\right) \quad (.111)$$

$$\bar{B} = \left(\underline{B}^{-1} + Y^{L'}Y^{L}\right)^{-1} \quad (.112)$$

.4 Details on the derivation of the posterior of the returns factor model

.4.1 Priors

The joint prior on the regression parameters, conditional on the state of the economy, can be written as:

$$p(A^s, \Sigma^s) = p(A^s|\Sigma^s)p(\Sigma^s) = p(\alpha^s|\Sigma^s)p(\Sigma^s)p(B^s) \quad (.113)$$

$$\propto |\Psi(\Sigma^s)|^{\frac{1}{2}} exp\left\{-\tfrac{1}{2}(A^s - \bar{A}^s)'\Psi(\Sigma^s)(A^s - \bar{A}^s)\right\} |\Sigma^s|^{\frac{-(v+n+1)}{2}} exp\left\{-\tfrac{1}{2}tr\left((\Sigma^s)^{-1}H^s\right)\right\}$$

$$\propto |\Sigma^s|^{\frac{-(v+n+2)}{2}} exp\left\{(B^s - B^s)'(\Omega^s)^{-1}(B^s - \bar{B^s})\right\} \quad (.114)$$

$$exp\left\{-\tfrac{1}{2}(\alpha^s)'\left(\frac{\sigma_\alpha^2}{(s^s)^2}\Sigma^s\right)^{-1}(\alpha^s) - \tfrac{1}{2}tr\left((\Sigma^s)^{-1}H^s\right)\right\}$$

The equality $p(A^s|\Sigma^s)p(\Sigma^s) = p(\alpha^s|\Sigma^s)p(\Sigma^s)p(B^s)$ is only possible because of how the matrix $\Psi(\Sigma^s)$ is defined - α^s and B^s are independent and the variance of B^s is independent from Σ^s.

Additionally the term $(B^s - B^s)'(\Omega^s)^{-1}(B^s - \bar{B^s})$ goes to zero as Ω^s is chosen to be a diagonal matrix with very large diagonal elements in order to provide a non-informative prior for B^s.

Hence the joint prior on the regression parameters can be considered as:

$$p(A^s, \Sigma^s) \propto |\Sigma^s|^{\frac{-(v+n+2)}{2}} exp\left\{-\frac{1}{2}(\alpha^s)'\left(\frac{\sigma_\alpha^2}{(s^s)^2}\Sigma^s\right)^{-1}(\alpha^s) - \frac{1}{2}tr\left((\Sigma^s)^{-1}H^s\right)\right\} \quad (.115)$$

If further considering the transformation proposed by Tu 2010 such that: $\left[(\alpha^s)'\left(\frac{\sigma_\alpha^2}{(s^s)^2}\Sigma^s\right)^{-1}(\alpha^s)\right] = \left[(a^s)'\left(\Sigma^{-1} \otimes D\right)(a^s)\right]$,

where $a = vec(a)$ and D is a matrix of zeroes, whose $(1,1)$ element equals $\frac{s^2}{\sigma_\alpha^2}$. Then the joint prior for the regression parameters can be rewritten as a function of only (a, Σ) as follows:

$$p(A^s,\Sigma^s) \propto |\Sigma^s|^{\frac{-(\nu+n+2)}{2}} exp\left\{-\frac{1}{2}(a)'\left((\Sigma^s)^{-1}\otimes D^s\right)^{-1}(a^s) - \frac{1}{2}tr\left((\Sigma^s)^{-1}H^s\right)\right\} \tag{.116}$$

.4.2 Posterior

The derivation of the posterior distributions below follows Tu 2010.

$$\begin{aligned} & p(a^s,\Sigma^s|R^s,F^s) \propto \qquad (.117) \\ \propto |\Sigma^s|^{-\frac{k+1}{2}} exp & \left\{-\tfrac{1}{2}(a^s)'((\Sigma^s)^{-1}\otimes D^s)(a^s) - \tfrac{1}{2}tr\left[(a^s-\hat{a^s})'((\Sigma^s)^{-1}\otimes X^{s'}X^s)(a^s-\hat{a^s})\right]\right\} \\ & \times|\Sigma^s|^{-\frac{(T+\nu+n-k+1)}{2}} exp\left\{-\tfrac{1}{2}tr\left(H^s+T^s\hat{\Sigma^s}\right)(\Sigma^s)^{-1}\right\} \end{aligned}$$

Defining: $G = D^s + (X^{s'}X^s)$ and $J^s = X^{s'}X^s - X^{s'}X^s\,(G^s)^{-1}\,X^{s'}X^s$ and completing the square on c we obtain:

$$\begin{aligned} p(a^s,\Sigma^s|R^s,F^s) \propto |\Sigma^s|^{-\frac{k+1}{2}} exp\left\{-\tfrac{1}{2}(a^s-\tilde{a^s})'((\Sigma^s)^{-1}\otimes G^s)(a^s-\tilde{a^s})\right\} \\ \times|\Sigma^s|^{-\frac{(T+\nu+n-k+1)}{2}} exp\left\{-\tfrac{1}{2}tr\left(H^s+T^s\hat{\Sigma^s}+\hat{A^s}'J^s\hat{A^s}\right)(\Sigma^s)^{-1}\right\} \end{aligned} \tag{.118}$$

From which we can derive the following posterior distributions for a^s and Σ^s, conditional on the state of the economy:

$$((\Sigma^s)^{-1}|R^s,F^s) \sim W\left((T^s+\nu-k),(H^s+T^s\hat{\Sigma}^s+\hat{A^s}'J^s\hat{A^s})^{-1}\right) \tag{.119}$$

$$(a^s|\,(\Sigma^s)^{-1},R^s,F^s) \sim N\left(\tilde{a^s},\left(\Sigma^s\otimes(G^s)^{-1}\right)\right) \tag{.120}$$

Where $\tilde{a^s} = \left(I_N\otimes\left((G^s)^{-1}X^{s'}X^s\right)\right)\hat{a^s}$.

.5 Gibbs Sampler

The Gibbs sampling procedure allows to estimate empirically the joint and marginal distribution of a number of variables by only using their conditional distributions.

This can be achieved as follows:

Consider N variables y_i for $i = 1,...N$ whose joint distribution is given by:

$$p(y_1,y_2,...,y_N) \tag{.121}$$

Further define their marginal distributions as:

$$p(y_i) = \int \cdots \int p(y_1, y_2, ..., y_N) dy_1 ... dy_N \tag{.122}$$

While the conditional distributions can be defined as:

$$p(y_i | y_1, y_2, ..., y_N) \tag{.123}$$

If the conditional distributions are known, the empirical marginal and joint distributions can be obtained with the following iterative process:

0. Choose arbitrary initial values

1. Draw y_1 from its conditional distribution $p(y_1 | y_2, ..., y_N)$

2. Draw y_2 from its conditional distribution $p(y_2 | y_1, y_3, ..., y_N)$, note the value of y_1 drawn in the step above is used here

...

N. Draw y_N from its conditional distribution $p(y_{2N} | y_1, y_3, ..., y_{N-1})$, note the values of $(y_1, ..., y_{N-1})$ drawn in the steps above are used here

Repeat steps 1-N above for $M = (I + L)$ times, where I is the number of iterations needed for convergence of the Gibbs sampler. Note that the initial values at the beginning of each iteration are the values drawn in the previous iteration, excluding the first iteration, where initial values are chosen arbitrarily. Each time the loop above is repeated we have a sample of the type $(y_1, y_2, ..., y_N)_m$ for $m = 1, ...M$. The first I samples can be discarded while the L remaining samples are saved and used for the computation of the moments of the joint and marginal distributions. In fact, after the sampler has converged, further draws are taken from the joint distribution of the variables. So this procedure allows to have L samples $(y_1, y_2, ..., y_N)_l$ for $l = 1, ..., L$, drawn from the joint distribution $p(y_1, y_2, ..., y_N)$. At this point empirical moments of the joint distribution and all marginal distributions can be easily computed.

For instance the mean of the marginal distribution for y_i can be obtained as:

$$\mu_i = \frac{1}{K} \sum_{m=I+1}^{M} y_i \tag{.124}$$

More details on Gibbs sampling can be found in Casella and George 1992 and Albert and Albert and Chib 1993.

Tables and Figures

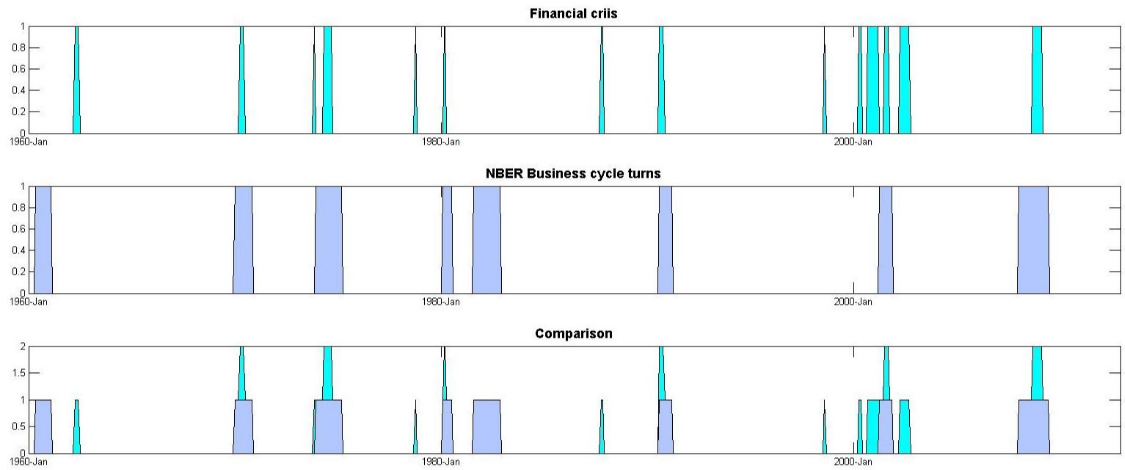

Figure .1: **Comparison of financial crises and recessions:** comparison of the NBER recession dates (panel 2) with an indicator of financial crises constructed as follows (panel 1). The US financial market is defined to be in a crisis when: the markets drops of at least 15% in one month OR of at least 20% in three months OR of at least 40% in six months for ether the NASDAQ, the S&P 500 or the Dow Jones indices.

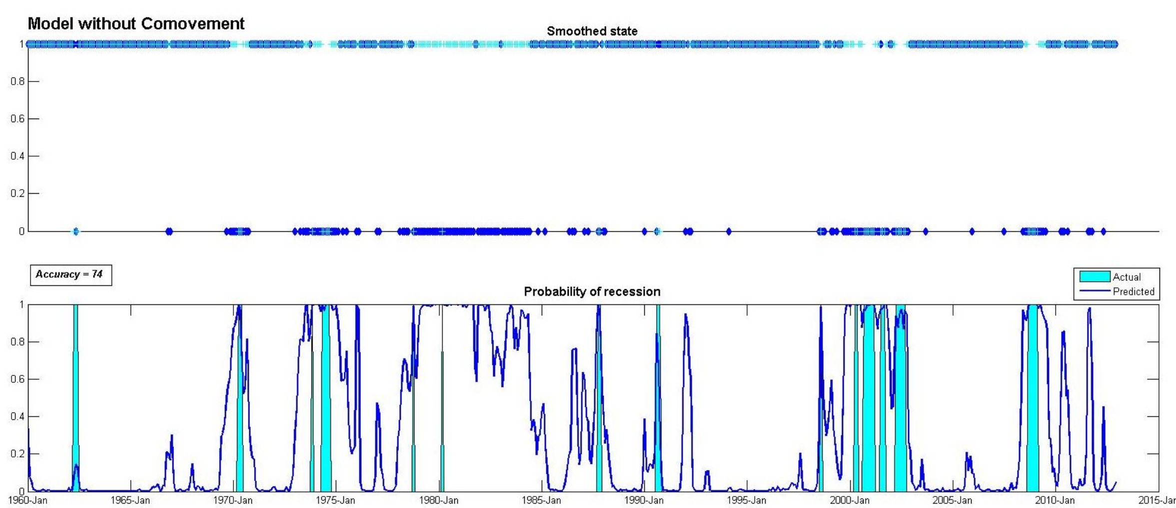

Figure .2: **Regime and probability of financial crisis using the Markov-Switching model without Co-movement:** comparison of the probability of crisis estimated using the *MS* model with the finacial crisis indicator described in Figure .1

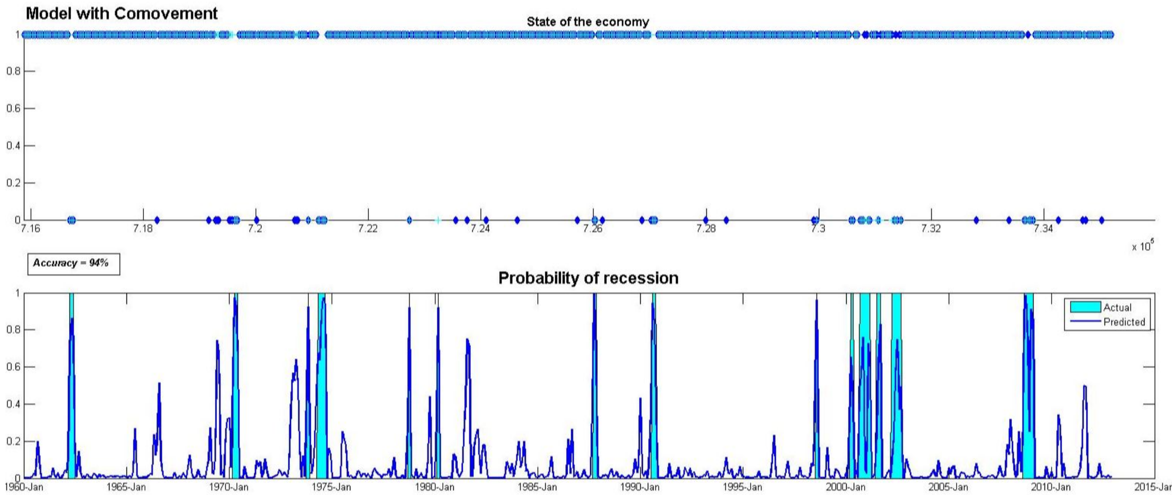

Figure .3: **Regime and probability of financial crisis using the Markov-Switching model with Co-movement:** comparison of the probability of crisis estimated using the *MS_C* model with the finacial crisis indicator described in Figure .1

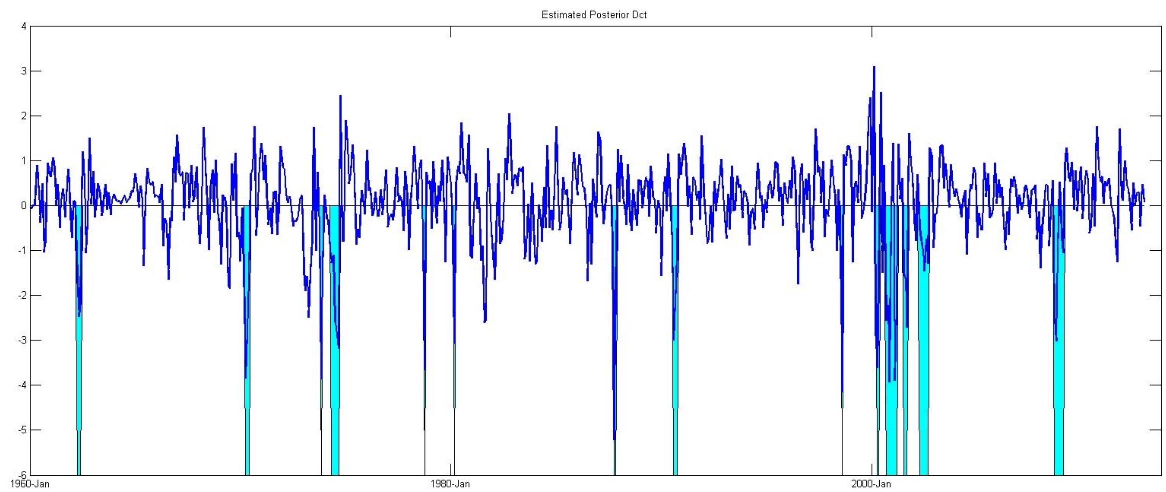

Figure .4: **Estimated common latent variable:** this figure displays the common latent variable reppresenting the growth rate of the financial market estimated using the *MS_C* model. The highlighted area reppresents the finacnial crisis periods computed as described in Figure .1

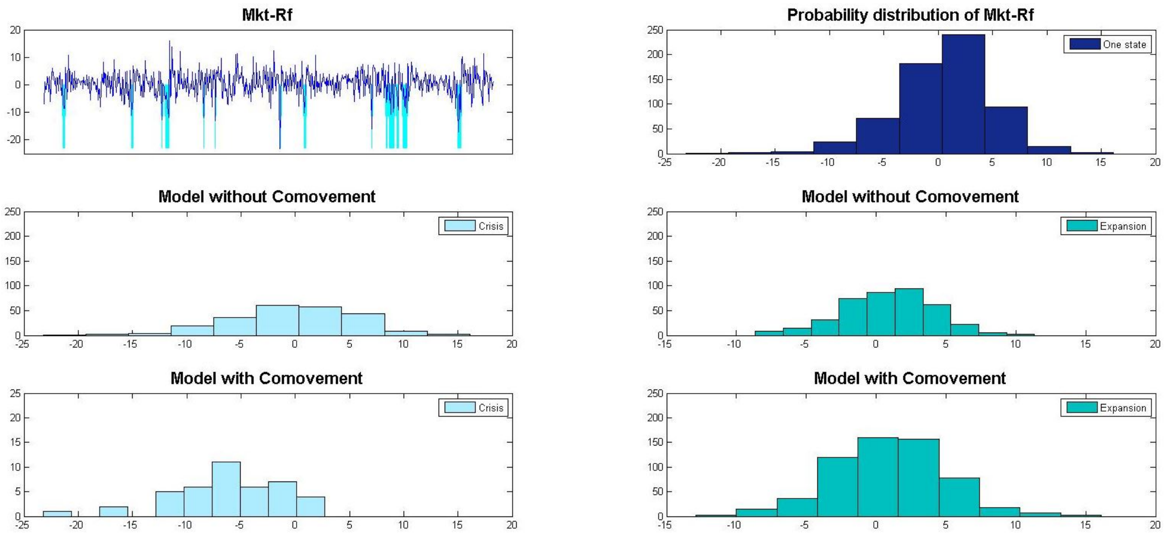

Figure .5: **Probability distribution of the excess return on the market:** panel 1 shows the excess return on the market, financial crisis periods are highlighted; panel 2 displays the probability distribution of the excess return in the one-state model *1s*; panels 3 and 4 display the probability distribution of the excess return in tranquil times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively; panels 5 and 6 display the probability distribution of the excess return in crises times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively

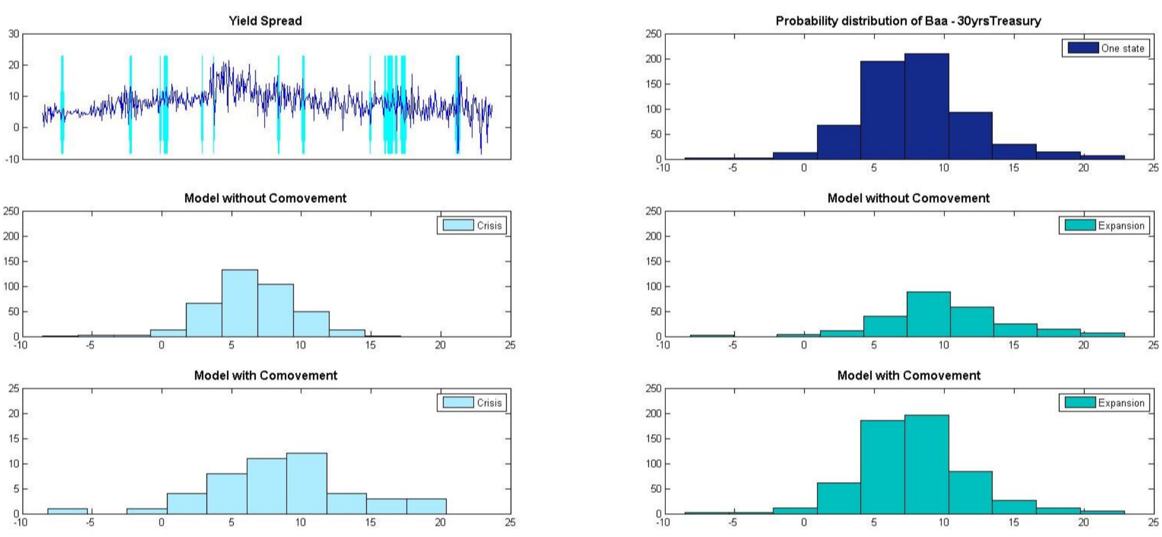

Figure .6: **Probability distribution of the yield spread:** panel 1 shows the yield spread, financial crisis periods are highlighted; panel 2 displays the probability distribution of the yield spread in the one-state model *1s*; panels 3 and 4 display the probability distribution of the yield spread in tranquil times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively; panels 5 and 6 display the probability distribution of the yield spread in crises times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively

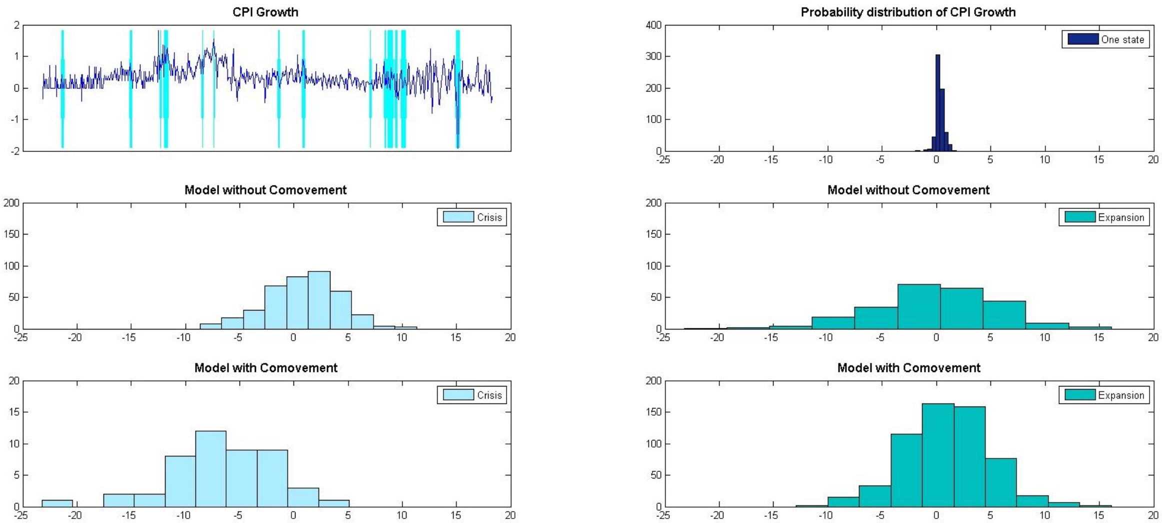

Figure .7: **Probability distribution of CPI growth:** panel 1 shows CPI growth, financial crisis periods are highlighted; panel 2 displays the probability distribution of CPI growth in the one-state model *1s*; panels 3 and 4 display the probability distribution of CPI growth in tranquil times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively; panels 5 and 6 display the probability distribution of CPI growth in crises times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively

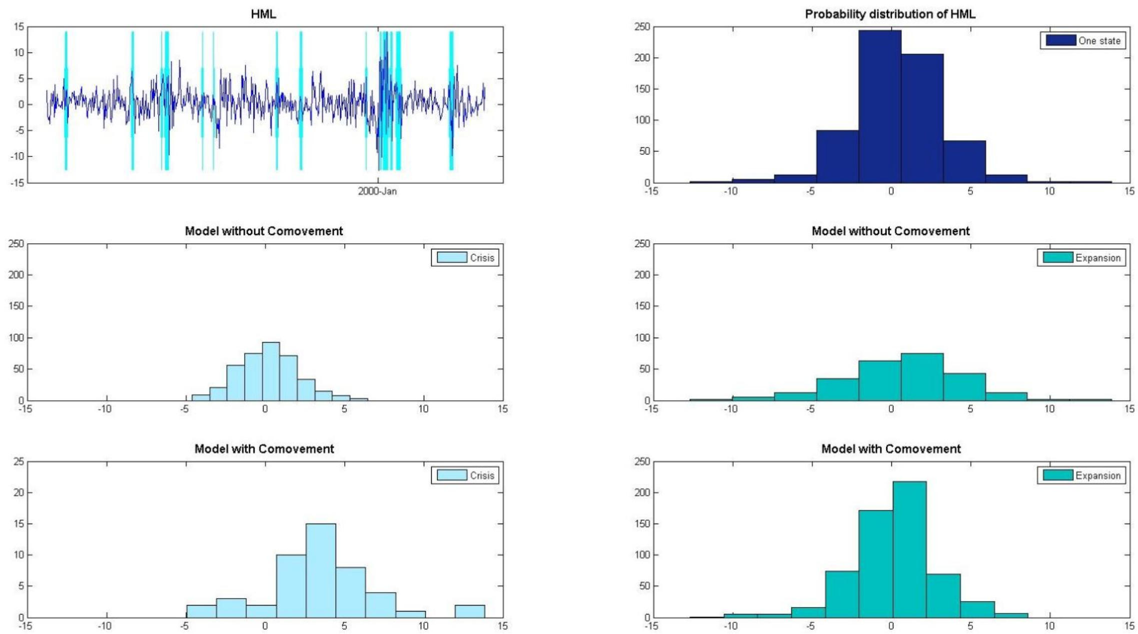

Figure .8: **Probability distribution of HML:** panel 1 shows HML, financial crisis periods are highlighted; panel 2 displays the probability distribution of HML in the one-state model *1s*; panels 3 and 4 display the probability distribution of HML in tranquil times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively; panels 5 and 6 display the probability distribution of HML in crises times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively

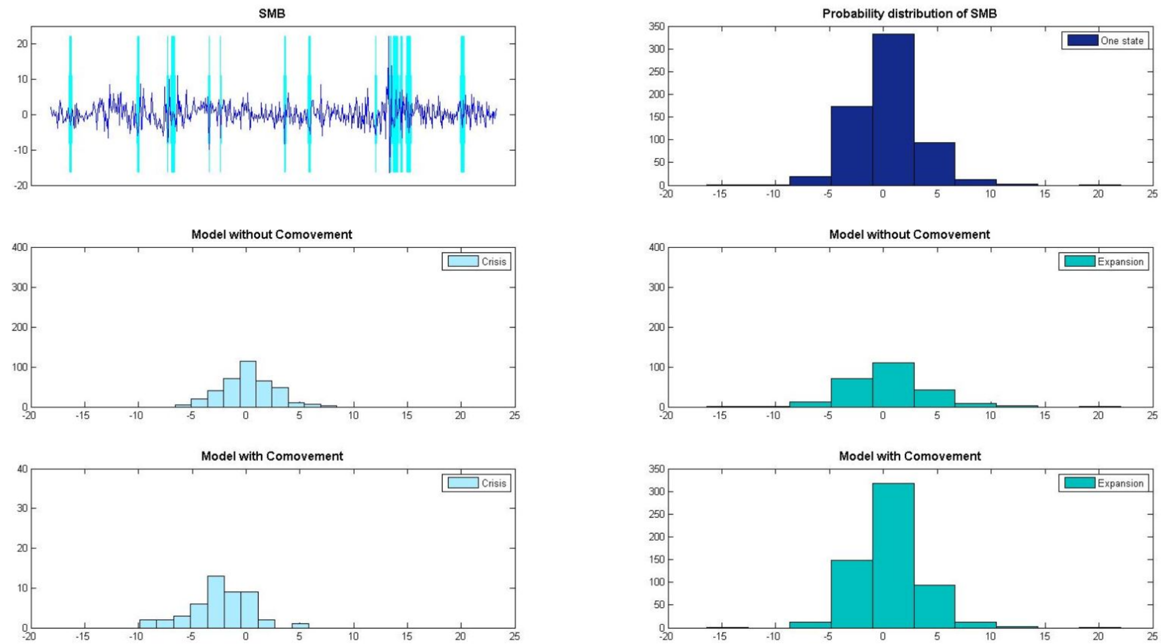

Figure .9: **Probability distribution of SMB:** panel 1 shows SMB, financial crisis periods are highlighted; panel 2 displays the probability distribution of SMB in the one-state model *1s*; panels 3 and 4 display the probability distribution of SMB in tranquil times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively; panels 5 and 6 display the probability distribution of SMB in crises times for the model without co-movement *MS* and the model with co-movement *MS_C* respectively

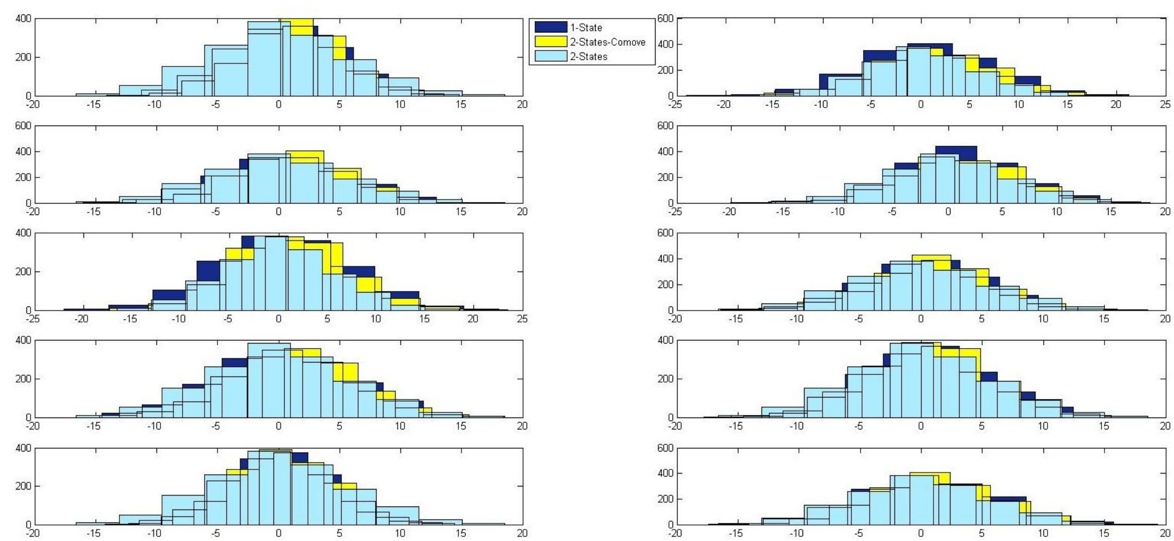

Figure .10: **Predictive distribution of return:** predictive distribution of returns for time $(t+1)$ for the 10 Fama–French industry portfolios, obtained using the three models of interest (*1s*, *MS* and *MS_C*)

Table .1: **Posterior Mean of the Factors**
This table shows the posterior mean of the distribution of the factors. The factors considered are: the yield spreads between the Moody's Baa yields and the long term (30 yrs) government bond ($Yield\ Spread$), the growth rate of the Consumer Price Index (CPI), the excess return on the market ($Mkt - Rf$) and the Fama–French size (SMB) and book-to-market (HML) factors. Model (1) shows the results for the one-state model ($1S$). Models (2-4) display results for the 2-state Markov switching model without co-movement, models (2) and (3) show the posterior mean of the distribution for the Bull and the Bear states respectively, model (4) displays the difference in posterior mean between the two states ($Bull - Bear$). Models (5-7) display results for the 2-state Markov switching model with co-movement, models (5) and (6) show the posterior mean of the distribution for the Bull and the Bear states respectively, model (7) displays the difference in posterior mean between the two states ($Bull - Bear$). All posterior values are monthly

	One State	*Two States*			*Two States with Comovement*		
	$1S$	*Bull*	*Bear*	*Diff*	*Bull*	*Bear*	*Diff*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Yield Spread	7.86	6.71	9.85	-3.14	7.79	8.77	-0.98
	(0.25)	(0.30)	(0.46)	(0.49)	(0.26)	(1.10)	(1.12)
CPI Growth	0.32	0.25	0.46	-0.21	0.31	0.45	-0.14
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.11)	(0.11)
$(Mkt - Rf)$	0.47	0.86	-0.26	1.12	1.01	-6.46	7.47
	(0.18)	(0.18)	(0.43)	(0.49)	(0.23)	(1.44)	(1.39)
SMB	0.22	0.19	0.27	-0.08	0.48	-2.97	3.44
	(0.12)	(0.13)	(0.28)	(0.32)	(0.14)	(0.89)	(0.88)
HML	0.38	0.25	0.64	-0.39	0.16	3.26	-3.10
	(0.11)	(0.11)	(0.28)	(0.31)	(0.12)	(0.96)	(0.96)

Standard deviations in parenthesis

Table .2: **Differences in Posterior Mean of the Factors**
This table shows differences in the posterior mean of the distribution of the factors for the two Markov switching models with and without comovement ($MS\_C - MS$). The factors considered are the same as described in Table .1. Models (1) and (2) display the difference in the posterior mean of the distribution for the Bull and the Bear states respectively. Model (3) displays difference in the difference between the posterior distribution of the factors between the Bull and Bear states in the two considered models ($MS\_C - MS$). All posterior values are monthly

	Comparison : $MS\_C - MS$		
	$Bull - Bull$	$Bear - Bear$	$[(Bull - Bear) - (Bull - Bear)]$
	(1)	(2)	(3)
Yield Spread	1.08	-1.08	2.17
	(0.30)	(1.13)	(1.21)
CPI Growth	0.06	0.00	0.07
	(0.02)	(0.11)	(0.11)
$(Mkt - Rf)$	0.15	-6.20	6.39
	(0.29)	(1.50)	(1.49)
SMB	0.28	-3.24	3.47
	(0.19)	(0.93)	(0.96)
HML	-0.09	2.62	-2.69
	(0.16)	(1.00)	(0.99)

Standard deviations in parenthesis

Table .3: **Posterior mean of the excess return on the 10 industry portfolios**
This table shows the posterior mean of the distribution of the 10 Fama–French industry portfolios. The industries considered are: ... Model (1) shows the results for the one-state model ($1S$). Models (2-4) display results for the 2-state Markov switching model without co-movement, models (2) and (3) show the posterior mean of the distribution for the Bull and the Bear states respectively, model (4) displays the difference in posterior mean between the two states ($Bull - Bear$). Models (5-7) display results for the 2-state Markov switching model with co-movement, models (5) and (6) show the posterior mean of the distribution for the Bull and the Bear states respectively, model (7) displays the difference in posterior mean between the two states ($Bull - Bear$). All posterior values are monthly

	One State	*Two States*			*Two States with Comovement*		
	$1S$	*Bull*	*Bear*	*Diff*	*Bull*	*Bear*	*Diff*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
NoDur	0.54	0.87	-0.06	0.93	0.98	-4.85	5.83
	(0.17)	(0.19)	(0.39)	(0.46)	(0.20)	(1.35)	(1.33)
Durbl	0.64	0.90	0.09	0.81	1.22	-6.68	7.91
	(0.24)	(0.26)	(0.56)	(0.64)	(0.28)	(1.85)	(1.83)
Manuf	0.55	0.92	-0.15	1.07	1.11	-6.38	7.49
	(0.20)	(0.21)	(0.47)	(0.54)	(0.24)	(1.60)	(1.57)
Enrgy	0.74	1.11	0.14	0.97	1.14	-4.18	5.32
	(0.21)	(0.23)	(0.47)	(0.55)	(0.23)	(1.54)	(1.56)
HiTec	0.56	0.99	-0.23	1.22	1.32	-8.88	10.20
	(0.26)	(0.27)	(0.61)	(0.70)	(0.31)	(2.05)	(1.98)
Telcm	0.43	0.81	-0.19	1.00	0.82	-4.38	5.20
	(0.18)	(0.20)	(0.42)	(0.49)	(0.21)	(1.31)	(1.30)
Shops	0.54	0.90	-0.16	1.06	1.12	-6.55	7.67
	(0.20)	(0.22)	(0.48)	(0.56)	(0.24)	(1.59)	(1.55)
Hlth	0.45	0.91	-0.32	1.23	0.91	-4.94	5.85
	(0.19)	(0.22)	(0.43)	(0.51)	(0.23)	(1.30)	(1.28)
Utils	0.21	0.57	-0.37	0.94	0.47	-2.57	3.04
	(0.15)	(0.17)	(0.34)	(0.40)	(0.18)	(1.12)	(1.15)
Other	0.49	0.95	-0.31	1.26	1.08	-6.64	7.72
	(0.21)	(0.23)	(0.49)	(0.57)	(0.26)	(1.64)	(1.62)

Standard deviations in parenthesis

Table .4: **Differences in Posterior Mean of the Excess Return on the 10 Industry Portfolios**

This table shows differences in the posterior mean of the distribution of the excess return on the 10 Fama–French industry portfolios for the two Markov switching models with and without comovement ($MS\_C - MS$). The industries considered are the same as described in Table .3. Models (1) and (2) display the difference in the posterior mean of the distribution for the Bull and the Bear states respectively. Model (3) displays difference in the difference between the posterior distribution of the factors between the Bull and Bear states in the two considered models ($MS\_C - MS$). All posterior values are monthly

	$Comparison: \ MS\_C \ - \ MS$		
	$Bull - Bull$	$Bear - Bear$	$[(Bull - Bear) - (Bull - Bear)]$
	(1)	(2)	(3)
$NoDur$	0.12	-4.78	4.90
	(0.28)	(1.40)	(1.41)
$Durbl$	0.32	-6.77	7.10
	(0.39)	(1.95)	(1.96)
$Manuf$	0.19	-6.23	6.42
	(0.32)	(1.67)	(1.67)
$Enrgy$	0.03	-4.32	4.35
	(0.32)	(1.61)	(1.65)
$HiTec$	0.34	-8.65	8.98
	(0.41)	(2.14)	(2.10)
$Telcm$	0.01	-4.19	4.20
	(0.29)	(1.38)	(1.39)
$Shops$	0.22	-6.40	6.62
	(0.33)	(1.66)	(1.65)
$Hlth$	0.00	-4.62	4.62
	(0.31)	(1.37)	(1.38)
$Utils$	-0.10	-2.19	2.10
	(0.24)	(1.17)	(1.21)
$Other$	0.13	-6.33	6.46
	(0.34)	(1.72)	(1.72)

Standard deviations in parenthesis

Table .5: **Posterior Volatility of the Factors**
This table shows the posterior volatility of the distribution of the factors. The factors considered are the same as in Table .1. Model (1) shows the results for the one-state model ($1S$). Models (2-4) display results for the 2-state Markov switching model without co-movement, models (2) and (3) show the posterior mean of the distribution for the Bull and the Bear states respectively, model (4) displays the difference in posterior mean between the two states ($Bull - Bear$). Models (5-7) display results for the 2-state Markov switching model with co-movement, models (5) and (6) show the posterior mean of the distribution for the Bull and the Bear states respectively, model (7) displays the difference in posterior mean between the two states ($Bull - Bear$). All posterior values are monthly

	One State	*Two States*			*Two States with Comovement*		
	$1S$	*Bull*	*Bear*	*Diff*	*Bull*	*Bear*	*Diff*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Yield Spread	4.14	3.12	4.90	-1.78	4.06	5.18	-1.13
	(0.16)	(0.18)	(0.37)	(0.40)	(0.17)	(0.93)	(0.97)
CPI Growth	0.36	0.27	0.45	-0.18	0.35	0.52	-0.17
	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.11)	(0.12)
$(Mkt - Rf)$	4.51	3.41	5.94	-2.53	3.97	5.55	-1.59
	(0.17)	(0.18)	(0.34)	(0.35)	(0.19)	(0.99)	(0.94)
SMB	3.08	2.40	4.03	-1.64	2.91	3.72	-0.82
	(0.12)	(0.11)	(0.25)	(0.24)	(0.12)	(0.72)	(0.73)
HML	2.86	1.97	3.97	-1.99	2.68	3.79	-1.12
	(0.11)	(0.10)	(0.24)	(0.22)	(0.11)	(0.66)	(0.66)

Standard deviations in parenthesis

Table .6: **Differences in Posterior Volatility of the Factors**
This table shows differences in the posterior volatility of the distribution of the factors for the two Markov switching models with and without comovement ($MS\_C - MS$). The factors considered are the same as described in Table .1. Models (1) and (2) display the difference in the posterior volatility of the distribution for the Bull and the Bear states respectively. Model (3) displays difference in the difference between the posterior mean of the factors between the Bull and Bear states in the two considered models ($MS\_C - MS$). All posterior values are monthly

	$Comparison: \ MS\_C \ - \ MS$		
	$Bull - Bull$	$Bear - Bear$	$[(Bull - Bear) - (Bull - Bear)]$
	(1)	(2)	(3)
$Yield\ Spread$	0.94	0.29	0.65
	(0.21)	(0.97)	(1.03)
$CPI\ Growth$	0.07	0.07	0.01
	(0.02)	(0.12)	(0.12)
$(Mkt - Rf)$	0.56	-0.39	0.95
	(0.22)	(1.03)	(1.00)
SMB	0.51	-0.31	0.82
	(0.14)	(0.75)	(0.77)
HML	0.70	-0.18	0.88
	(0.13)	(0.69)	(0.70)

Standard deviations in parenthesis

Table .7: **Posterior Volatility of the Excess Return on the 10 Industry Portfolios** This table shows the posterior volatility of the distribution of the 10 Fama–French industry portfolios. The industries considered are: Consumer NonDurables (*NoDur*), Consumer Durables (*Durbl*), Manufacturing (*Manuf*), Oil, Gas, and Coal Extraction and Products (*Enrgy*), Business Equipment (*HiTec*), Telephone and Television Transmission (*Telcm*), Wholesale, Retail, and Some Services – Laundries, Repair Shops (*Shops*), Healthcare (*Hlth*), Utilities (*Utils*), Other (*Other*). Model (1) shows the results for the one-state model (1*S*). Models (2-4) display results for the 2-state Markov switching model without co-movement, models (2) and (3) show the posterior volatility of the distribution for the Bull and the Bear states respectively, model (4) displays the difference in posterior volatility between the two states (*Bull* − *Bear*). Models (5-7) display results for the 2-state Markov switching model with co-movement, models (5) and (6) show the posterior volatility of the distribution for the Bull and the Bear states respectively, model (7) displays the difference in posterior volatility between the two states (*Bull* − *Bear*). All posterior values are monthly

	One State	*Two States*			*Two States with Comovement*		
	1*S*	*Bull*	*Bear*	*Diff*	*Bull*	*Bear*	*Diff*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
NoDur	4.41	3.64	5.53	-1.89	3.99	6.36	-2.37
	(0.16)	(0.18)	(0.31)	(0.32)	(0.17)	(1.25)	(1.21)
Durbl	6.29	4.95	8.18	-3.23	5.77	8.58	-2.81
	(0.23)	(0.27)	(0.50)	(0.55)	(0.25)	(1.67)	(1.64)
Manuf	5.03	3.92	6.57	-2.65	4.48	6.97	-2.48
	(0.19)	(0.20)	(0.39)	(0.40)	(0.21)	(1.35)	(1.29)
Enrgy	5.37	4.22	6.93	-2.71	5.05	7.22	-2.17
	(0.20)	(0.22)	(0.41)	(0.42)	(0.22)	(1.23)	(1.22)
HiTec	6.60	5.12	8.65	-3.54	5.94	7.97	-2.03
	(0.25)	(0.25)	(0.49)	(0.49)	(0.26)	(1.28)	(1.24)
Telcm	4.72	3.91	5.85	-1.94	4.43	6.07	-1.64
	(0.18)	(0.20)	(0.32)	(0.35)	(0.19)	(0.92)	(0.91)
Shops	5.26	4.13	6.84	-2.71	4.72	7.15	-2.43
	(0.20)	(0.21)	(0.38)	(0.38)	(0.21)	(1.43)	(1.38)
Hlth	5.05	4.29	6.13	-1.84	4.73	6.27	-1.54
	(0.19)	(0.20)	(0.34)	(0.36)	(0.19)	(1.05)	(1.03)
Utils	4.06	3.35	5.08	-1.73	3.87	5.73	-1.86
	(0.15)	(0.16)	(0.29)	(0.29)	(0.16)	(0.97)	(0.96)
Other	5.36	4.16	7.00	-2.84	4.82	7.30	-2.48
	(0.20)	(0.22)	(0.42)	(0.43)	(0.22)	(1.46)	(1.41)

Standard deviations in parenthesis

Table .8: **Differences in Posterior Volatility of the Excess Return on the 10 Industry Portfolios**
This table shows differences in the posterior volatility of the distribution of the excess return on the 10 Fama–French industry portfolios for the two Markov switching models with and without comovement ($MS\_C - MS$). The industries considered are the same as described in Table .3. Models (1) and (2) display the difference in the posterior volatility of the distribution for the Bull and the Bear states respectively. Model (3) displays difference in the difference between the posterior distribution of the factors between the Bull and Bear states in the two considered models ($MS\_C - MS$). All posterior values are monthly

	$Comparison: \; MS\_C \; - \; MS$		
	$Bull - Bull$	$Bear - Bear$	$[(Bull - Bear) - (Bull - Bear)]$
	(1)	(2)	(3)
$NoDur$	0.35	0.83	-0.48
	(0.21)	(1.27)	(1.25)
$Durbl$	0.82	0.40	0.42
	(0.31)	(1.72)	(1.73)
$Manuf$	0.56	0.39	0.17
	(0.24)	(1.37)	(1.34)
$Enrgy$	0.83	0.29	0.54
	(0.26)	(1.26)	(1.27)
$HiTec$	0.83	-0.68	1.51
	(0.30)	(1.35)	(1.35)
$Telcm$	0.52	0.22	0.29
	(0.23)	(0.96)	(0.97)
$Shops$	0.59	0.31	0.28
	(0.25)	(1.46)	(1.43)
$Hlth$	0.43	0.14	0.30
	(0.23)	(1.07)	(1.08)
$Utils$	0.52	0.65	-0.14
	(0.19)	(0.99)	(0.99)
$Other$	0.66	0.29	0.36
	(0.27)	(1.49)	(1.48)

Standard deviations in parenthesis

Table .9: **Posterior Mean of the Predictive Distribution of Return**
This table displays the posterior mean of the predictive distribution of return for the 10 Fama–French industry portfolios. Model (1) displays the results for the one state model($1S$). Models (2) and (3) show the results for the two states models without (MS) and with co-movement respectively (MS_C).

	One State	*Two States*	
	$1S$	MS	MS_C
	(1)	(2)	(3)
NoDur	0.45	-0.01	1.11
Durbl	0.60	0.24	1.25
Manuf	0.51	-0.16	1.18
Enrgy	0.74	0.44	1.01
HiTec	0.54	-0.19	1.25
Telcm	0.38	-0.08	0.83
Shops	0.45	-0.17	1.20
Hlth	0.50	-0.26	0.91
Utils	0.07	-0.18	0.41
Other	0.44	-0.20	1.08

Standard deviations in parenthesis

Table .10: **Optimal Portfolio Weights, CER and SR**

This table compares the one-period-ahead optimal portfolio weights allocated to the 10 Fama French industry portfolios using the three models of interest: one state model (Model 1 – 1S), two state model without co-movement (Model 2 – MS) and two state model with co-movement (Model 3 – MS_C). It additionally displays for each model the certainty equivalent of return (CER) and the Sharpe Ratio (SR) based on a $100 of investment. Models (4) and (5) display the difference in such CER and SR for between the 1S and MS and between the MS_C and MS models respectively.

	One State	*Two States*		*Comparison*	
	$1S$	MS	MS_C	$(MS-1S)$	(MS_C-MS)
	(1)	(2)	(3)	4	5
NoDur	35.25	50.39	83.79		
Durbl	14.37	35.16	3.41		
Manuf	-17.37	-52.26	22.04		
Enrgy	36.52	34.70	27.28		
HiTec	-3.81	3.86	5.96		
Telcm	8.50	-0.24	8.06		
Shops	-3.43	-13.18	5.04		
Hlth	13.25	-10.88	-11.50		
Utils	-41.45	-28.11	-45.83		
Other	-17.69	-21.50	-29.19		
CER	16.78	35.00	54.25	18.22	19.25
SR	25.76	30.31	33.27	4.55	2.96

Values computed per $100