Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Servaas van Bilsen

Measuring Systemic Risk

A Markov Switching Regression Approach

RM Thesis 2011-051

Measuring Systemic Risk: A Markov Switching Regression Approach

by

S. van Bilsen

24.26.44

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy in Economics

Tilburg University

August, 2011

Supervised by:

dr. R.J.A. Laeven

prof. dr. T. Nijman

Abstract

This thesis aims to contribute to the debate on systemic risk. The methodology developed in this thesis is applied to a small set of US bank holding companies using publicly available data. Specifically, we investigate the performance of the financial system conditional on the bank holding company being under financial distress. To this end, we propose a Markov switching regression model to infer the regime in which the bank holding company was at any historical date. Conditional on the bank holding company being in a particular regime, we model the returns of a global Bank Index as an affine function of macro-economic state variables and the bank holding company's stock price returns. Systemic risk contribution is assessed using correlations, changes in parameter estimates and a measure of regime co-movement. The results indicate that correlations of returns increase during times of turmoil. The results are, however, mixed regarding the change in the fraction of shocks that are on average transmitted from the bank holding company's stock price to the global Bank Index. We also briefly consider a Markov switching regression model with time-varying transition probabilities.

Keywords: Systemic Risk, Markov switching regression model, Correlations, Bank Holding Companies, Financial System

Contents

List of Figures

1 Introduction

The global financial system is a large, complex and dynamic network consisting of commercial banks, investment banks, brokers, hedge funds, insurance companies, money market funds, private equity funds and many other financial players. The recent financial crisis has revealed the fragility and the interconnectedness of the financial system, and has ultimately caused an unprecedented level of failures of large financial institutions. Because this crisis had its origin in the run-up of real estate prices, the initially affected companies were especially involved in mortgage lending activities. However, as the recent crisis began to unfold, it became apparent that the failure of small mortgage lenders rapidly spilled over to several major financial institutions, including Lehman Brothers, Freddie Mac, Fannie Mae, Merrill Lynch and AIG. This "propagation mechanism" is closely related to the notion of systemic risk, which is the risk of the collapse of the system exacerbated by idiosyncratic shocks. Systemic risk can be studied along multiple dimensions. The traditional literature has focused on systemic risk in bank run models and bank contagion models (see De Bandt and Hartmann [10] for an extensive survey). We examine systemic risk in the context of financial markets and investigate whether a shock originating from a US bank holding company (BHC, for short) affects the stability of the financial system. By way of motivation, Figure 1 depicts Bank of America's stock price and a global Bank Index. Idiosyncratic shocks (indicated by black arrows) tend to lead in a sequential fashion to system-wide shocks (indicated by blue ellipses). This observation suggests directionality of correlation and the existence of an underlying economic mechanism through which shocks are propagated from Bank of America's stock price to the global Bank Index. Billio et al. [7] has conducted Granger-causality tests to determine the directionality of correlation among four indexes (i.e. Banks, Brokers, Insurers and Hedge funds). They found that during the period 2001-2008, the four indexes became highly interlinked and all (Granger) caused the S&P 500 returns.

The models, which are presented in the subsequent sections, explore the extent to which index i's log return behavior contributes to index j's log return behavior conditional on index i being under financial distress. Economic agents, such as supervisors, are often interested in the financial institution's risk contribution to the overall risk in the system. We call this the institution's contribution to systemic risk.

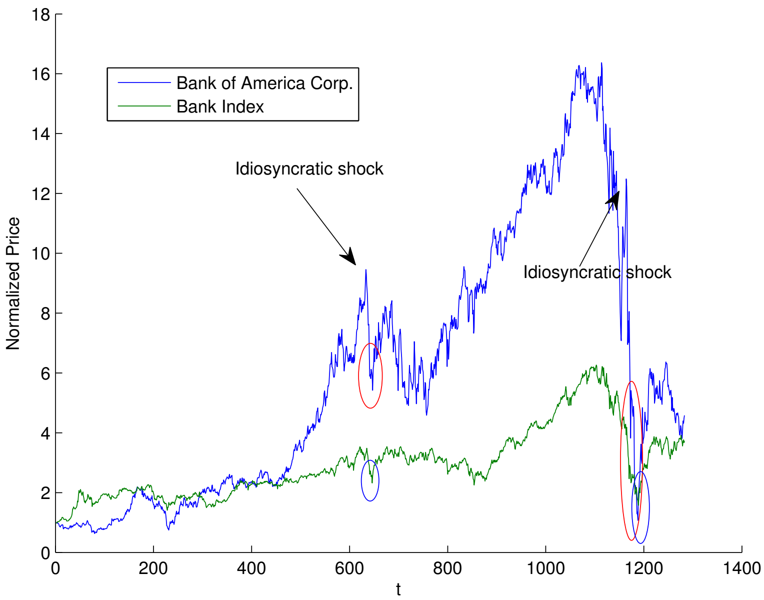

Figure 1: **Bank of America's stock price and a global Bank Index**
This figure depicts the evolution of Bank of America's stock price and a global Bank Index over time (May 26, 1986 - December 31, 2010).

The papers of e.g. Acharya et al. [1] and Huang et al. [27] consider the systemic risk exposure, or more precisely the institution's performance conditional on the performance of the system. Both aspects of systemic risk are analyzed in this thesis. Our setup shares some similarity with the proposed setup for measuring CoVaR, which is a systemic risk measure introduced by Adrian and Brunnermeier [2]. CoVaR measures the marginal effect of the institution's tail risk on the tail risk of the system. Adrian and Brunnermeier [2] assess the tail risk using a Value-at-Risk (VaR) criterion and employ the technique of quantile regressions to obtain CoVaR estimates. This thesis differs in at least two respects from the paper of Adrian and Brunnermeier [2].

First, instead of using a VaR criterion, we employ a Markov switching regression model to draw inference regarding index i's different states[1]. Specifically, we allow the data to determine the probabilities of being in some state at time t (i.e. so-called "smoothed state probabilities"). These probabilities are calculated using the entire sample period. Consequently, inference is not only based on the cross-sectional distribution of equity returns, but also on the (transitional) dynamics of equity returns. Hamilton [22] [23] was the first to study Markov switching autoregression models. He used Markov switching regression (as introduced by Quandt [39] and Goldfeld and Quandt [20]) to characterize changes in the parameters of the model.

[1] We use the words "state" and "regime" interchangeably.

After the seminal work of Hamilton, many authors have employed Markov switching models to describe regime switches in interest rates (Ang and Bekaert [4], Cai [8] and Gray [21]), business cycles (Hamilton [22], Filardo [17] and Owyang et al. [37]), foreign exchange rates (Engel and Hamilton [15] and Bekaert and Holdrick [6]) and stock returns (Ang and Bekaert [3]).

Secondly, to relate index i's and index j's log return behavior, we propose a model for index j's conditional mean. Expressed differently, conditional on index i being under financial distress, we are concerned with the marginal effect of index i's log return behavior on index j's conditional mean (and not on index j's conditional quantiles).

The reason why we restrict ourselves to a small set of BHCs is at least twofold. First, the banking sector performs essential roles in the economy. Schumpeter [43] has already emphasized the importance of the banking sector in promoting economic growth and welfare. Banks can potentially spur economic growth by accepting deposits and channeling those deposits to profitable investments. In the United States, the banking sector provided \$ 38 trillion of credit to various sectors in 2008 (Worldbank, 2008). Because banks are at the heart of the economy, it is of special importance to study systemic risk for banks. Secondly, banks may be a natural source of systemic risk. This view is also supported by Billio et al. [7]. In times of market turmoil, banks may be unable to unwind positions due to the illiquidity of bank assets. According to Diamond and Dybvig [12], the illiquidity of bank assets and the liquidity of bank liabilities provide "the rationale for the existence of banks and for their vulnerability of runs". Once a bank fails, its losses may spread to the rest of the economy through e.g. real and informational channels. Ultimately, bank runs may result in large welfare losses and costs on the economy, as argued by Friedman and Schwartz [19].

Systemic risk can be based upon (daily, weekly) market data or accounting information, the latter often only being available at a quarterly or yearly time horizon. For example, Adrian and Brunnermeier [2] focus on systemic risk related to the growth rate of the market valued total financial assets, which is based upon both daily equity prices and quarterly balance sheet data. We study, however, systemic risk related to the growth rate of equity prices. As noted by Hull and White [28], asset return correlation is closely related to equity return correlation. Under the Black-Scholes assumptions, it can even be shown that the asset return correlation is equal to the equity return correlation (see Appendix A of Huang et al. [27] for a simple proof). The logic behind this is that equity represents an implicit call option on the underlying assets.

Moreover, under efficient markets, equity prices are forward-looking: they appropriately reflect changes in information on present and future performance of the company. This is in sharp contrast to accounting information, which is often reported with a substantial time lag.

Markov switching regression models: potential to reproduce empirical stylized facts.

There is ample evidence of the fact that empirical return distributions are not well described by a Gaussian distribution, which motivates the search for alternative return specifications. In what follows, we consider three empirical stylized facts and motivate that a Markov switching regression model is potentially capable of reproducing each empirical stylized fact.

Skewness

The skewness of empirical return distributions is often found to be significant. A nonzero skewness coefficient is compatible with a Markov switching regression model, provided that the mean switches between different regimes. To motivate this point, consider the following setting. Suppose that nature flips an unfair coin, with the probability of a head showing up equal to 0.8. Conditional on a head (tail) coming up, the equity return is normally distributed with mean 2 (-2) and variance 1 (9). It follows that equity returns possess a mixture distribution with a density as shown in Figure 2.

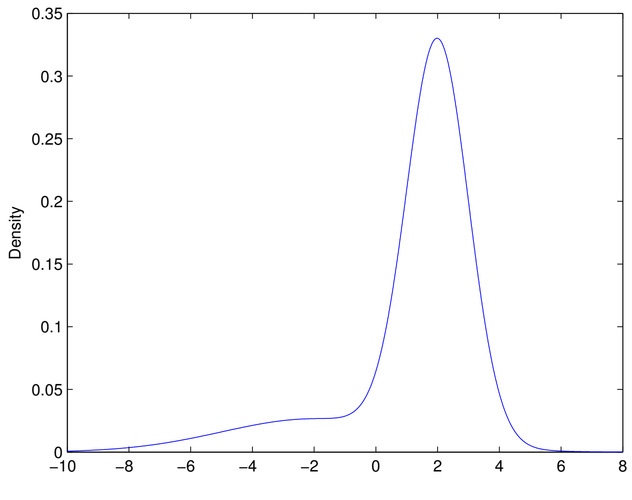

Figure 2: **Plot of a mixture density of two univariate normal components**
This figure plots a mixture density of two univariate normal components in proportions 0.8 and 0.2 with variances $\sigma_1^2 = 1$ and $\sigma_2^2 = 9$ and means $\mu_1 = 2$ and $\mu_2 = -2$.

Consequently, conditional on the outcome of the coin toss, equity returns are normally distributed. Unconditionally, equity returns are, however, negatively skewed and possess a non-normal distribution.

Heavy-taildness

Empirical work has provided convincing evidence that equity price data contain outliers generated by heavy-tailed distributions. There is also some support in favor of using Gaussian mixture models to describe equity price data that contain outliers (see e.g. Richardson and Smith [40] and Roll [41]). To motivate this point, consider the return behavior of the Dow Jones Industrial Average Index, which is depicted in Figure 3A. The return behavior involves shifts over time between periods of relative market stability ("low volatility regimes") and periods of relative market instability ("high volatility regimes"). Figure 3B presents both a histogram that describes the distribution of the returns in the low volatility regime and a smoothed bell-shaped curve that approximates the shape of the distribution. A Gaussian distribution seems capable of capturing relevant empirical characteristics. The same conclusion seems to hold true for the histogram depicted in Figure 3C. On the contrary, Figure 3D shows the histogram that describes the distribution of all returns. We observe that the data sample contains outliers and hence is not well-captured by a Gaussian distribution.

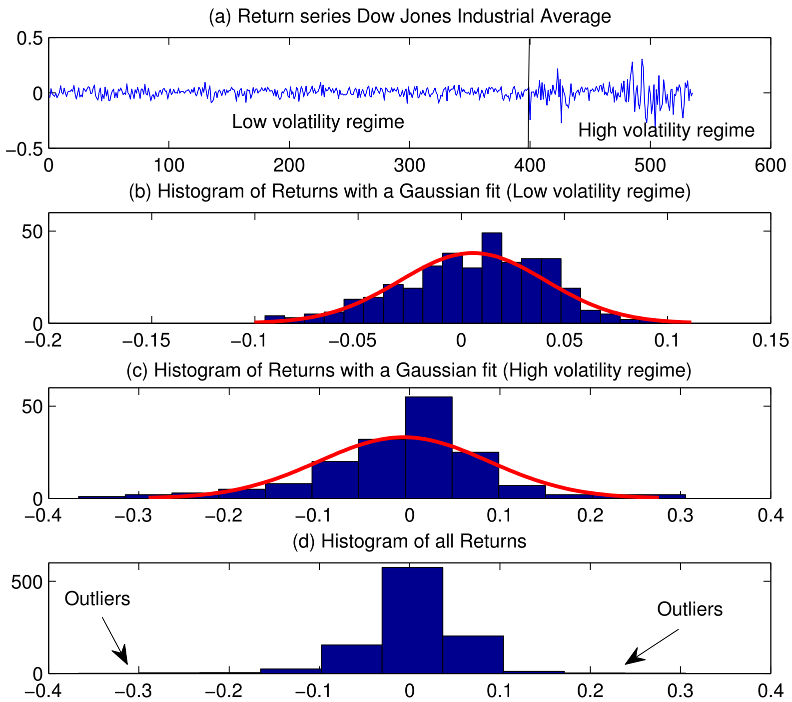

Figure 3: **Histograms Dow Jones Industrial Average**
This figure depicts (a) Return series Dow Jones Industrial Average Index; (b) Histogram of Returns with a Gaussian fit (Low volatility regime); (c) Histogram of Returns with a Gaussian fit (High volatility regime); (d) Histogram of all Returns.

Volatility clustering

Financial time series data often exhibit abrupt changes in their dynamics. For instance, the Dow Jones Industrial Average Index felled 22.6% on October 19, 1987. After a dramatic break, many stock prices, especially at a high frequency, have a tendency to behave quite differently compared to before the break. As shown by Figure 4, the volatility of the Dow Jones Index has changed considerable over time. The literature has widely recognized the stylized fact that stock price volatility is time-varying and, as noted by Mandelbrot [36], "large price changes tend to be followed by large price changes - of either sign - and small price changes tend to be followed by small price changes". There is also evidence that stock price volatility increases around financial crises, as motivated by Schwert [44]. Many authors have employed (G)ARCH-type models (as introduced by Engle [16]) and other stochastic volatility models to account for the fact that financial time series data exhibit volatility clustering. A Markov switching regression model is compatible with volatility clustering, as it allows the volatility parameter to switch between different regimes. Furthermore, in Section 3, we explore a (bivariate) model, which allows both index i's and index j's volatility parameter to depend on index i's regimes.

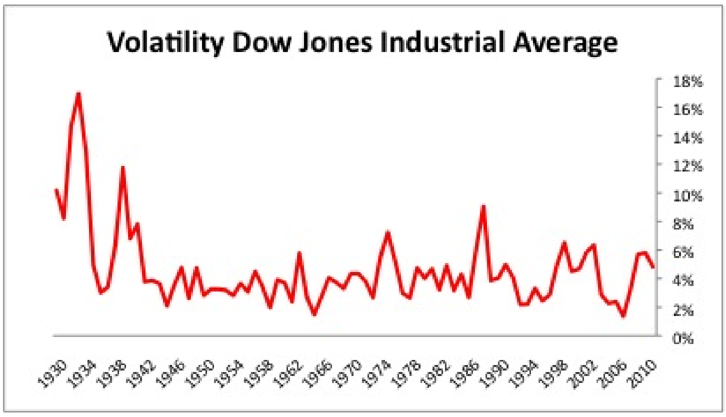

Figure 4: **Volatility of Dow Jones Industrial Average**
For each year, the figure depicts the estimated standard deviation (based on 12 observations) of the monthly log returns.

Systemic Risk Assessments

As motivated by the recent global financial crisis and other financial episodes, financial markets seem to become increasingly intertwined during times of turmoil and, moreover, idiosyncratic shocks that originated in one economic entity (e.g. individual institution, sector, industry, market, economy, country or even a group of countries) tend to propagate to other economic entities.

This propagation mechanism is often hard to analyze on the basis of macro-economic fundamentals (i.e. price movements are too large to be explained by fundamentals). Given that index i is inferred to be under financial distress (in Section 2 we propose a classifier that infers index i's different regimes), we investigate the extent to which index i and index j are tightly coupled. Market connectedness can be assessed using a variety of methods. For instance, Kritzman et al. [35] measure the fraction of the total variance of a set of assets that is absorbed by a fixed number of eigenvectors. They introduce the so-called absorption ratio. This ratio is large when sources of risk are more unified. The risk assessments used in this thesis are based on (linear) correlations, parameter estimates, a so-called measure of regime co-movement and transition probabilities[2].

Estimated correlation coefficient

A wide range of literature and research has been conducted on correlation shifts: a significant increase in correlation during times of turmoil (see e.g. Calvo and Reinhardt [9] and King and Wadhwani [32]). In our bivariate setting, we use the estimated correlation coefficient to characterize the degree of association between index i and index j. The correlation coefficient is estimated on the basis of the different model specifications. A high correlation coefficient means that idiosyncratic shocks are more likely to propagate to other markets. We also decompose the correlation coefficient to account for increased volatility. As motivated by Forbes and Rigobon [18], an increase in index i's log return volatility may cause an increase in the correlation coefficient between index i and index j, even though the true propagation mechanism (i.e. the fraction of shocks that are transmitted from one index to another index) remains constant. The first component of our decomposition captures the change in correlation driven by the dynamics of macro-economic fundamentals and index i's idiosyncratic shocks. The remainder term captures the change in correlation caused by a change in the underlying propagation mechanism.

Parameter estimates

We obtain parameter estimates that characterize the fraction of shocks that are (on average) instantaneously transmitted from index i to index j. We conduct a direct test to infer whether the change in the transmission mechanism is significantly different from zero.

[2]The probability of moving from one regime to another regime.

Measure of regime co-movement

We also propose a measure that assesses the extent to which index i's and index j's regimes are positively related. This measure is defined through the probabilities of being in some state at time t. The joint probability of being in the same regime reflects the regime co-movement between index i and index j.

Time-varying transition probabilities

In Section 4, we explore a Markov switching regression model with time-varying transition probabilities. The economic rationale is that transition probabilities may be driven by exogenous (economic) variables and hence need not be constant over time.

This thesis is structured as follows. In Section 2, we describe index i's model specification and propose a classifier to infer index i's states on the basis of smoothed state probabilities. Given inference regarding index i's states, we relate index j's and index i's log return behavior through a linear regression model. Details are given in Section 2. In Section 3, we use a refined estimation technique to relate index i's and index j's log return behavior. Section 4 examines a Markov switching regression model with time-varying transition probabilities. Section 5 conducts an empirical study. Finally, in Section 6, we gather a few concluding remarks.

2 Model Formulation

This section formulates the model we propose to measure the extent to which index i's activities contribute to index j's performance. The estimation procedure consists of two stages: inference regarding index i being in state $S_t^i = s$ at time t (the different states are specified in Section 2.1) and estimating the extent to which index i's log return behavior contributes (in a non-causal sense) to index j's log return behavior conditional on index i being in state $S_t^i = s$ at time t. Index i's states are inferred *ex post* from the data. That is, the inference is based on all the information contained in the sample.

This section is structured as follows. In Section 2.1, we describe the models formally. Finally, in Section 2.2, we introduce measures to assess systemic risk.

2.1 Formal Description

We propose a parsimonious model in which index i's log return behavior is characterized by a relatively stable period and a relatively volatile period[3]. To be more precise, we describe index i's log return behavior by a two-state Markov switching regression model (Hamilton [22] [23]) with discrete hidden (i.e. unobserved) state variable $S_t^i \in \{0,1\}$ and switching parameter $\sigma^i_{S_t^i} = (1 - S_t^i)\sigma_0^i + S_t^i\sigma_1^i$. By using a Markov switching regression model, we allow the data to determine the parameters that characterize each state and the probability of moving from some state to another state. We also include a $(1 \times k)$ vector M_{t-1} of (lagged) macro-economic fundamentals to control for the state-of-the economy. The formal description proceeds as follows.

Denote by T the sample length. Let $(R_t^i : t \in \{1, ..., T\})$ and $(R_t^j : t \in \{1, ..., T\})$ be index i's and index j's log return time series, respectively. Define $\mathcal{R}_t^i = (R_s^i : s \in \{1, ..., t\})$, which is the complete history of index i's log return behavior observed through the date t. The model specification of R_t^i is described as follows

$$R_t^i = \begin{cases} \alpha_0^i + M_{t-1}\gamma^i + \sigma_0^i\epsilon_t^i & \text{if } S_t^i = 0 \\ \alpha_1^i + M_{t-1}\gamma^i + \sigma_1^i\epsilon_t^i & \text{if } S_t^i = 1 \end{cases} \qquad (t = 1, ..., T), \tag{2.1}$$

where $\epsilon_t^i \overset{\text{i.i.d.}}{\sim} N(0,1)$.

The discrete hidden state variable S_t^i is governed by a first-order ergodic irreducible[4] Markov chain with transition probability matrix

$$P^i = \begin{pmatrix} p_{00}^i & p_{01}^i \\ p_{10}^i & p_{11}^i \end{pmatrix},$$

where $p_{xy}^i = \mathbb{P}\{S_t^i = y | S_{t-1}^i = x, S_{t-2}^i, ..., S_1^i\} = \mathbb{P}\{S_t^i = y | S_{t-1}^i = x\}$ denotes the probability of moving from state x to state y. Since the transition probability from state x to some state y must be one, we have the restriction $p_{x0} + p_{x1} = 1$ for each $x \in \{0,1\}$.

[3] We restrict our attention to a *two-state* Markov switching regression model for a couple of reasons. (i) The estimation procedure becomes computationally burdensome as the number of states exceeds two. In particular, the number of parameters to be estimated grows exponentially in the number of states. (ii) A *two-state* Markov switching regression model is intuitively appealing, since it only allows for periods of turmoil and tranquility. (iii) We can achieve a reasonable level of flexibility with a *two-state* Markov switching regression model, while maintaining parsimony.

[4] A Markov chain is said to be ergodic if it is aperiodic (i.e. states occur at irregular times) and positive recurrent (i.e. expected return time is finite and expected number of visits is infinite).

Denote by $\Theta^i = \{(\alpha_0^i, \alpha_1^i, \gamma^{i'}, \sigma_0^i, \sigma_1^i, p_{00}^i, p_{11}^i)' \in \mathbb{R}^{6+k} : (p_{00}^i, p_{11}^i) \in [0,1]^2, (\sigma_0^i, \sigma_1^i) \in \mathbb{R}_+^2\}$ the parameter space. θ^i is a typical element of Θ^i. $\theta_0^i \in \Theta^i$ is understood as the true parameter value and $\hat{\theta}^i \in \Theta^i$ (we do not make the dependence on T visible in the notation) denotes a consistent estimator of θ_0^i.

In (2.1), γ^i is a $(k \times 1)$ vector of non-switching parameters. Although there is some evidence of non-linear and asymmetric responses to macro-economic fundamentals, we restrict our attention to the non-switching case. In Section 5, we test the validity of this assumption and examine the implications for the inference regarding index i's states. In particular, we found that a specification with switching parameters has little consequences for the inference regarding index i's regimes. As motivated by Timmermann [46], a necessary condition for the Markov switching regression model to generate skewness is that the mean of state 0 differs from the mean of state 1. Therefore, we allow the intercept to switch between the two different regimes. The switching parameter $\sigma^i_{S_t^i}$ is especially included to account for empirically observed volatility clustering in financial time series data.

By convention and without loss of generality, we associate $S_t^i = 0$ with a period of tranquility ("low volatility regime") and $S_t^i = 1$ with a period of turmoil ("high volatility regime").

The model parameters [5][6] are estimated by employing maximum likelihood optimization techniques and Hamilton's nonlinear filter for the hidden state variable process. As a by-product, we obtain filtered and smoothed (state) probabilities from the estimation procedure. The smoothed probabilities are deduced from the optimal nonlinear smoother and represent inference regarding index i being in state $S_t^i = s$ at time t conditional upon $\hat{\theta}^i$ and all the information contained in the sample. These probabilities are used to draw statistical inference about index i being in state $S_t^i = s$ at time t. In particular, we propose the following classifier[7]

$$\hat{S}_t^i = \begin{cases} 0 & \text{if } \mathbb{P}\{S_t^i = 0 | \mathcal{R}_T^i; \hat{\theta}^i\} \geq \tau_0 \\ 1 & \text{if } \mathbb{P}\{S_t^i = 1 | \mathcal{R}_T^i; \hat{\theta}^i\} \geq \tau_1 \\ \text{inconclusive} & \text{otherwise} \end{cases} \qquad (t = 1, ..., T), \tag{2.2}$$

[5] We denote by $S_0^i = s$ the initial state. We use that $\mathbb{P}\{S_0^i = s\} = \frac{1}{2}$ for each $s \in \{0, 1\}$.

[6] We partly use the software developed by Perlin [38]. This software has been made available via MATLAB Central.

[7] For notational simplicity, we suppress the dependence on τ_0 and τ_1 and the model parameters.

where $(\tau_0, \tau_1) \in [0,1]^2$ and $\mathbb{P}\{S_t^i = s|\mathcal{R}_T^i; \hat{\theta}^i\}$ represents the estimated smoothed probability of being in state $S_t^i = s$ at time t.

Denote by $\mathcal{T}_0 = (t : t \in \{1, ..., T\}, \hat{S}_t^i = 0)$ and $\mathcal{T}_1 = (t : t \in \{1, ..., T\}, \hat{S}_t^i = 1)$.

Index i is inferred to be in state s at time t if the estimated smoothed probability of being in state s at time t exceeds the threshold τ_s.

We now present the following lemma.

Lemma 2.1. *Let $\{X_t\}_{t=1}^T$ be a stochastic process defined on some underlying filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=1}^T, \mathbb{P}_X)$. The risk function*

$$\mathbb{P}\{S_t \neq \kappa_t(X)\}$$

is minimized over the set of all measurable maps from $\kappa_t : \Omega \mapsto \{0, 1\}$ by

$$\kappa_t(X) = 1_{\mathbb{P}\{S_t=1|X=x\}>\frac{1}{2}}. \tag{2.3}$$

Proof. By definition,

$$\begin{aligned}\mathbb{P}\{S_t \neq \kappa_t(X)\} &= \mathbb{P}\{S_t = 1, \kappa_t(X) = 0\} + \mathbb{P}\{S_t = 0, \kappa_t(X) = 1\} \\ &= \int_\Omega \mathbb{P}\{S_t = 1|X = x\}1_{\kappa_t(x)=0} + \mathbb{P}\{S_t = 0|X = x\}1_{\kappa_t(x)=1}\mathrm{d}\mathbb{P}_X(\omega).\end{aligned}$$

The integral is minimized by minimizing the integrand for each x, i.e.

$$\min_\kappa \mathbb{P}\{S_t = 1|X = x\}1_{\kappa_t(x)=0} + \mathbb{P}\{S_t = 0|X = x\}1_{\kappa_t(x)=1}.$$

$\kappa_t(x)$ is either zero or one. Hence, $\kappa_t(x)$ must be chosen such that

$$\kappa_t(x) = \begin{cases} 1 & \text{if } \mathbb{P}\{S_t = 1|X = x\} > \mathbb{P}\{S_t = 0|X = x\} \\ 0 & \text{if } \mathbb{P}\{S_t = 0|X = x\} > \mathbb{P}\{S_t = 1|X = x\} \end{cases}.$$

Q.E.D.

The minimal risk classifier (2.3) is a function of the true smoothed probabilities, which are, however, unknown. Therefore, we base our classifier on the estimated smoothed probabilities, which are consistent estimates of the true smoothed probabilities.

We examine the finite sample performance of $\hat{S}_t^i$. In particular, we compute the simulated mean squared error (SMSE) and the simulated mean squared prediction error (SMSPE) based on 1,000 replications and $T = 1,000$. The SMSE and SMSPE are defined as follows

$$\begin{aligned} SMSE(S_t^i) &= \tfrac{1}{1{,}000} \textstyle\sum_{m=1}^{1{,}000} (S_{t,m}^i - \hat{S}_{t,m}^i)^2 \\ SMSPE(S_t^i) &= \tfrac{1}{1{,}000} \textstyle\sum_{t=1}^{1{,}000} (S_{t,m}^i - \hat{S}_{t,m}^i)^2 \end{aligned},$$

where $S_{t,m}^i$ denotes the simulated value of S_t^i in simulation run m and $\hat{S}_{t,m}^i$ denotes the predicted value (based on a sample of length $T = 1,000$) of $S_{t,m}^i$ in simulation run m.

Table 1 contains the (fictitious) parameter values.

τ_0	0.5
τ_1	0.5
α_0^i	0
α_1^i	0
γ^i	0
σ_0^i	0.03
σ_1^i	0.06
p_{00}^i	0.95
p_{11}^i	0.85

Table 1: **(Fictitious) parameter values**

Figure 5 depicts the SMSE for each $t \in \{1, ..., T\}$ and the SMSPE for each replication m. We observe that both the SMSE and the SMSPE fluctuate around the level 0.15. The SMSPE is, however, more volatile than the SMSE. Since $\hat{\theta}^i$ is fixed for each m, averaging over time causes the SMSPE to be more volatile than the SMSE.

We are interested in index j's performance conditional upon index i being under financial distress. We propose a linear regression model to relate index i's and index j's log return behavior. Specifically, the conditional mean of R_t^j given the value of the explanatory variables is modeled as an affine function of the explanatory variables.

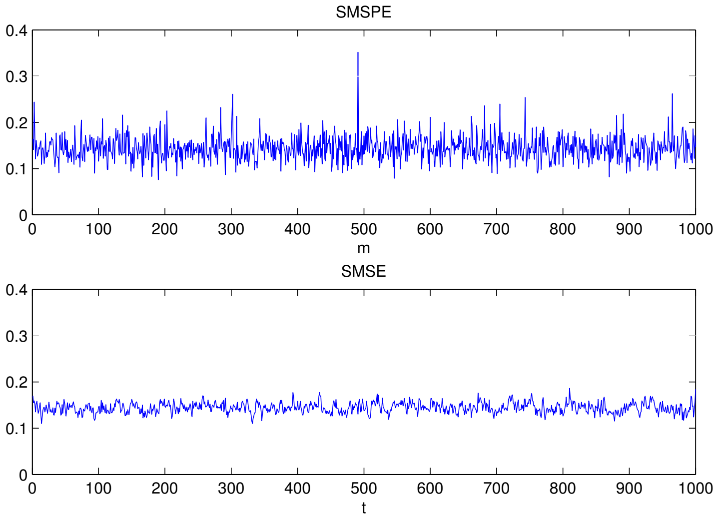

Figure 5: **Simulated MSE and Simulated MSPE**

The model specification of R_t^j is formally written as follows

$$R_t^j = \begin{cases} \alpha_0^{j|i} + M_{t-1}\gamma^{j|i} + \beta_0^{j|i} R_t^i + \sigma^{j|i}\epsilon_t^{j|i} & \text{if } \hat{S}_t^i = 0 \\ \alpha_1^{j|i} + M_{t-1}\gamma^{j|i} + \beta_1^{j|i} R_t^i + \sigma^{j|i}\epsilon_t^{j|i} & \text{if } \hat{S}_t^i = 1 \end{cases} \qquad t \in \mathcal{T}_0 \cup \mathcal{T}_1, \tag{2.4}$$

where $\epsilon_t^{j|i} \overset{\text{i.i.d.}}{\sim} N(0,1)$.

We again include a $(1 \times k)$ vector M_{t-1} of (lagged) macro-economic fundamentals to control for the state-of-the economy.

Denote the parameter space by $\Theta^{j|i} = \{(\alpha_0^{j|i}, \alpha_1^{j|i}, \gamma^{j|i\prime}, \beta_0^{j|i}, \beta_1^{j|i}, \sigma^{j|i})' \in \mathbb{R}^{5+k} : \sigma^{j|i} \in \mathbb{R}_+\}$. $\theta^{j|i}$ is a typical element of $\Theta^{j|i}$. $\theta_0^{j|i}$ is understood as the true parameter value and $\hat{\theta}^{j|i} \in \Theta^{j|i}$ denotes a consistent estimator of $\theta_0^{j|i}$.

The model is estimated by classical ordinary least squares (OLS). OLS estimation yields consistent estimates, even in the presence of heteroscedastic residuals. For instance, if we want to test $\beta_0^{j|i} - \beta_1^{j|i} < 0$, then Student's t-test is still valid, provided that heteroskedasticity consistent standard errors are used.

Even though the regime sequence is not directly observable by the researcher, model (2.4) relates index j's and index i's log return behavior by imposing *a priori* index i's regimes.

OLS estimates are derived under the assumption that the explanatory variables are exogenous and uncorrelated with the error term. The parameter estimates are generally inconsistent and biased if the true data generating process is as follows

$$R_t^j = \begin{cases} \alpha_0^{j|i} + M_{t-1}\gamma^{j|i} + \beta_0^{j|i} R_t^i + \sigma^{j|i}\eta_t^{j|i} & \text{if } S_t^i = 0 \\ \alpha_1^{j|i} + M_{t-1}\gamma^{j|i} + \beta_1^{j|i} R_t^i + \sigma^{j|i}\eta_t^{j|i} & \text{if } S_t^i = 1 \end{cases} \qquad t \in \mathcal{T}_0 \cup \mathcal{T}_1, \tag{2.5}$$

where $\eta_t^{j|i} \overset{\text{i.i.d.}}{\sim} N(0,1)$.

Since $\hat{S}_t^i$ is random, we have that $\hat{S}_t^i = S_t^i + \nu_t^i$ (here, ν_t^i represents the prediction error). Substitution into (2.5) yields

$$R_t^j = \alpha_0^{j|i} + (\alpha_1^{j|i} - \alpha_0^{j|i})\hat{S}_t^i + M_{t-1}\gamma^{j|i} + \beta_0^{j|i} R_t^i(1 - \hat{S}_t^i) + \beta_1^{j|i} R_t^i \hat{S}_t^i + \sigma^{j|i}\epsilon_t^{j|i},$$

where $\epsilon_t^{j|i} = \eta_t^{j|i} - (\sigma^{j|i})^{-1}\nu_t^i[R_t^i(\beta_0^{j|i} + \beta_1^{j|i}) + (\alpha_1^{j|i} - \alpha_0^{j|i})]$.

Obviously, the explanatory variables are correlated with the error term unless $\nu_t^i = 0$ for each time instance. In Section 3, we consider model specification (2.5) with a regime dependent volatility parameter and develop an alternative estimation technique, which yields consistent estimates.

The interconnectedness of markets suggests the existence of a mechanism through which shocks are transmitted from index i to index j. This instantaneous transmission mechanism is captured by the $\beta_0^{j|i}$ and $\beta_1^{j|i}$ parameters in (2.4).

2.2 Systemic Risk Assessments

There are different definitions of systemic risk measurement. Hautsch et al. [26] define systemic risk as the marginal effect of index i's tail risk on index j's tail risk. This definition is also in accordance with the notion of CoVaR. We consider, in particular, the marginal effect of index i's log return behavior on index j's conditional mean.

Systemic risk measurement is of particular importance during times of financial turmoil. During such times, an individual failure could cause a cascading failure, which could potentially bring down the entire financial system. In our setup, a period of financial distress is determined by index i's log return behavior. Specifically, the indicator variable $\hat{S}_t^i$, as defined in (2.2), characterizes index i's crisis and non-crisis periods.

The degree of interdependence is related, among many other things, to the $\beta_0^{j|i}$ and $\beta_1^{j|i}$ parameters in (2.4) and the correlation coefficient between index i and index j. In what follows, we consider different assessments of systemic risk.

Estimated correlation coefficient

The first assessment is based on estimating changes in correlation coefficients. In particular, we estimate the increase in the Pearson correlation coefficient between index i and index j during a period of financial distress (where a period of financial distress corresponds with $\hat{S}_t^i = 1$). As suggested by Forbes and Rigobon [18], heteroscedasticity in returns affects estimates of correlation coefficients. In general, it causes the correlation coefficient to be biased upwards. Therefore, we decompose the correlation coefficient to account for increased volatility in R_t^i. To this end, we propose the following decomposition rule for the change in correlation

$$\Delta\rho^{j|i} = \Delta\rho^{j|i}_{\Delta\beta^{j|i}} + \Delta\rho^{j|i}_{\Delta\Sigma^M} + \Delta\rho^{j|i}_{\Delta\sigma^i},$$

where Σ_M denotes the variance-covariance matrix of the (lagged) macro-economic fundamentals.

In Appendix A.1, we provide some details on this decomposition.

The last two terms on the right-hand side, $\Delta\rho^{j|i}_{\Delta\Sigma^M}$ and $\Delta\rho^{j|i}_{\Delta\sigma^i}$, particularly represent the change in correlation due to a change in the volatility of R_t^i. The volatility of R_t^i is driven by the variance-covariance matrix of the macro-economic fundamentals M and the volatility of the Gaussian innovations. $\Delta\rho^{j|i}_{\Delta\Sigma^M}$ captures the change in correlation caused by a change in the variance-covariance matrix of the fundamentals. The other term, $\Delta\rho^{j|i}_{\Delta\sigma^i}$, represents the change in correlation due to a change in the volatility of the Gaussian innovations. This volatility only affects the volatility of R_t^i, while the variance-covariance matrix of M affects both the volatility of R_t^i and R_t^j.

Moreover, an increase in the volatility of R_t^i may cause an increase in the correlation coefficient between R_t^j and R_t^i, even though the transmission mechanism, as captured by the $\beta_0^{j|i}$ and $\beta_1^{j|i}$ parameters in (2.4), remains constant.

The first term on the right-hand side, $\Delta\rho^{j|i}_{\Delta\sigma^i}$, represents the change in correlation due to a change in the transmission channel. That is, it represents the change in correlation after controlling for a change in the variance-covariance matrix of M and a change in the volatility of index i's idiosyncratic shocks.

Parameter estimates

The $\beta_0^{j|i}$ and $\beta_1^{j|i}$ parameters capture the mechanism through which shocks are instantaneously transmitted from index i to index j. The next assessment is based on the $\hat{\beta}_0^{j|i}$ and $\hat{\beta}_1^{j|i}$ parameter estimates. We provide Student's t-values, which can be used to perform a test to determine whether $\beta_0^{j|i}$ is significantly smaller than $\beta_1^{j|i}$. The alternative hypothesis is then as follows

$$H_1 : \beta_0^{j|i} - \beta_1^{j|i} < 0.$$

The test statistic T is given by

$$\mathrm{T} = \frac{\hat{\beta}_0^{j|i} - \hat{\beta}_1^{j|i}}{\sqrt{\mathbb{V}(\hat{\beta}_0^{j|i} - \hat{\beta}_1^{j|i})}},$$

where $\mathbb{V}(\cdot)$ denotes the variance.

The variance of $\hat{\beta}_0^{j|i} - \hat{\beta}_1^{j|i}$ can be decomposed as follows

$$\mathbb{V}(\hat{\beta}_0^{j|i} - \hat{\beta}_1^{j|i}) = \mathbb{V}(\hat{\beta}_0^{j|i}) + \mathbb{V}(\hat{\beta}_1^{j|i}) - 2\mathbb{C}(\hat{\beta}_0^{j|i}, \hat{\beta}_1^{j|i}),$$

where $\mathbb{C}(\cdot,\cdot)$ represents the covariance. To control for heteroscedastic residuals, we use *heteroscedasticity-consistent* standard errors (White [47]).

Under the null hypothesis, T follows Student's t-distribution with approximately T degrees of freedom (T large).

Measure of regime co-movement

We also introduce a measure of regime co-movement. Following Barberis et al. [5], co-movement is understood as a pattern of positive correlation.

Regime pairs $(S_t^i, S_t^j) = (1,1)$ and $(S_t^i, S_t^j) = (0,0)$ should therefore be regarded as perfect co-movement. To assess the relative co-movement, we impose the condition that regime pairs $(S_t^i, S_t^j) = (1,1)$ and $(S_t^i, S_t^j) = (0,0)$ lead to a co-movement of one ("perfect positive correlation") and regime pairs $(S_t^i, S_t^j) = (1,0)$ and $(S_t^i, S_t^j) = (0,1)$ lead to a co-movement of minus one ("perfect negative correlation") . We define our measure as the expected regime co-movement. Formally,

$$\Phi^{i,j} = 1 - 2\mathbb{E}[\mathbb{P}\{S_t^i = 0, S_t^j = 1\} + \mathbb{P}\{S_t^i = 1, S_t^j = 0\}], \tag{2.6}$$

where the 2 is a normalizing constant.

To estimate the joint probabilities $\mathbb{P}\{S_t^i = 0, S_t^j = 1\}$ and $\mathbb{P}\{S_t^i = 1, S_t^j = 0\}$, we require four regimes in a multivariate setting which is computationally intensive. Therefore, following Billio et al. [7], we estimate the joint probabilities by assuming conditional independence. Particularly,

$$\begin{aligned}
\mathbb{P}\{S_t^i = 0, S_t^j = 1 | \mathcal{R}_T^i, \mathcal{R}_T^j; \hat{\theta}^i, \hat{\theta}^j\} &= \mathbb{P}\{S_t^i = 0 | \mathcal{R}_T^i; \hat{\theta}^i\}\mathbb{P}\{S_t^j = 1 | \mathcal{R}_T^j; \hat{\theta}^j\}, \\
\mathbb{P}\{S_t^i = 0, S_t^j = 1 | \mathcal{R}_T^i, \mathcal{R}_T^j; \hat{\theta}^i, \hat{\theta}^j\} &= \mathbb{P}\{S_t^i = 1 | \mathcal{R}_T^i; \hat{\theta}^i\}\mathbb{P}\{S_t^j = 0 | \mathcal{R}_T^j; \hat{\theta}^j\}.
\end{aligned}$$

As argued by Billio et al. [7], to the extent that regimes are positively dependent, these estimators are biased downwards (i.e. underestimate the true joint probabilities). An estimator of (2.6) is now given by

$$\hat{\Phi}^{i,j} = 1 - 2\frac{1}{T}\sum_{t=1}^{T}\left[\mathbb{P}\{S_t^i = 0 | \mathcal{R}_T^i; \hat{\theta}^i\}\mathbb{P}\{S_t^j = 1 | \mathcal{R}_T^j; \hat{\theta}^j\} + \mathbb{P}\{S_t^i = 1 | \mathcal{R}_T^i; \hat{\theta}^i\}\mathbb{P}\{S_t^j = 0 | \mathcal{R}_T^j; \hat{\theta}^j\}\right].$$

3 A Refined Estimation Procedure

As shown in the previous section, OLS yields inconsistent and biased estimates if one wishes to estimate model (2.5) instead of model (2.4). We present an alternative consistent estimation technique to estimate model (2.5).

This section is structured as follows. In Section 3.1, we describe two models. The first model can be viewed as (a version of) a finite mixture model with known (time-independent) mixing weights. It is illustrative to consider this model first. The second model is more complicated in the sense that the mixing weights are time-varying.

Section 3.2 outlines the general principles of the expectation-maximization (EM) algorithm, which is an iterative method for obtaining maximum likelihood parameter estimates. In Section 3.3, we consider the particular form of the EM algorithm for both models. Finally, Section 3.4 examines the finite sample behavior of the estimation procedure.

3.1 Theoretical Description

3.1.1 Finite Mixture Model

For clarity of exposition, we present a model in which index j's log return behavior only depends on index i's states. These states are in turn determined by index i's log return behavior. That is, index j's and index i's log returns are only related through index i's states.

We consider the following index i's log return specification, which is a Markov switching regression model[8].

$$R_t^i = \begin{cases} \alpha_0^i + \sigma_0^i \epsilon_t^i & \text{if } S_t^i = 0 \\ \alpha_1^i + \sigma_1^i \epsilon_t^i & \text{if } S_t^i = 1 \end{cases} \qquad (t = 1, ..., T), \tag{3.1}$$

where $\epsilon_t^i \overset{\text{i.i.d.}}{\sim} N(0,1)$.

The discrete hidden state variable S_t^i is governed by a first-order ergodic irreducible Markov chain with transition probability matrix

$$P^i = \begin{pmatrix} p_{00}^i & p_{01}^i \\ p_{10}^i & p_{11}^i \end{pmatrix}.$$

We are interested in the following index j's log return specification

$$R_t^j = \begin{cases} \alpha_0^{j|i} + \sigma_0^{j|i} \epsilon_t^{j|i} & \text{if } S_t^i = 0 \\ \alpha_1^{j|i} + \sigma_1^{j|i} \epsilon_t^{j|i} & \text{if } S_t^i = 1 \end{cases} \qquad (t = 1, ..., T), \tag{3.2}$$

where $\epsilon_t^{j|i} \overset{\text{i.i.d.}}{\sim} N(0,1)$.

Index j's log returns are distributed according to a mixture of two Gaussian distributions. The mixing weights are given by $\mathbb{P}\{S_t^i = 0\}$ and $\mathbb{P}\{S_t^i = 1\}$, respectively.

[8] In fact, we consider the Markov switching regression model of Section 2 under a restricted parameter space.

In a traditional mixture model, the estimation of the mixing weights are part of the estimation procedure. However, in this setup, the mixing weights are determined by index i's log return dynamics. For now, we presume the mixing weights to be known with certainty (in Section 3.3, we relax this assumption).

We first derive maximum likelihood parameter estimates. Given each observation R_t^j, we know that it is generated from a Gaussian distribution parameterized by either $(\alpha_0^{j|i}, \sigma_0^{j|i})$ or $(\alpha_1^{j|i}, \sigma_1^{j|i})$. The density of the tth observation is given by

$$f(R_t^j; \theta^{j|i}) = \sum_{s=0}^{1} \mathbb{P}\{S_t^i = s\} f(R_t^j | S_t^i = s; \theta^{j|i}), \tag{3.3}$$

where the density function under each regime is specified by

$$f(R_t^j | S_t^i = s; \theta^{j|i}) = \frac{1}{\sqrt{2\pi}\sigma_s^{j|i}} \exp\left[\frac{-(R_t^j - \alpha_s^{j|i})^2}{2(\sigma_s^{j|i})^2}\right].$$

It is well-known from Markov theory that the *unconditional* probability $\mathbb{P}\{S_t^i = s\}$ is given for each $s \in \{0,1\}$ and $t \in \{1, ..., T\}$ by

$$\mathbb{P}\{S_t^i = s\} = \frac{1 - p_{(s-1)(s-1)}}{(1 - p_{00}) + (1 - p_{11})}.$$

The likelihood function is simply the product of the individual densities

$$L(\theta^{j|i} | \mathcal{R}_T^j) = \prod_{t=1}^{T} f(R_t^j; \theta^{j|i}).$$

$L(\theta^{j|i} | \mathcal{R}_T^j)$ increases without bound if we set e.g. $\alpha_0^{j|i} = R_t^j$ for any t and let $\sigma_0^{j|i} \to 0$ (this was first noticed by Kiefer and Wolfowitz [32]).

Kiefer [30] argued (by verifying a set of conditions) that there exists a unique root of the likelihood function that is consistent and asymptotically efficient. Kiefer's theorem reads as follows (which is given here as a lemma)

Lemma 3.1. *Let the parameter space be given by* $-\infty < \mu_0^{j|i}, \mu_1^{j|i} < \infty$, $0 < \sigma_0^{j|i} < \infty$, $0 \leq \sigma_1^{j|i} < \infty$. *If* R_t^j *is distributed according to (3.3), then for large* T *there exists a unique consistent root* $\hat{\theta}^{j|i}$ *of the likelihood function and* $\sqrt{n}(\hat{\theta}^{j|i} - \theta_0^{j|i})$ *is asymptotically normal distributed with mean zero and variance* $I(\theta_0^{j|i})^{-1}$, *where* I *is the Fisher information matrix.*

Proof. See Kiefer [30]. Q.E.D.

Hathaway [24] considered the restricted parameter space

$$\Theta_C^{j|i} = \{\theta^{j|i} \in \Theta^{j|i} : \frac{\sigma_1^{j|i}}{\sigma_0^{j|i}} \geq C > 0, \frac{\sigma_0^{j|i}}{\sigma_1^{j|i}} \geq C > 0\},$$

where $\Theta^{j|i}$ is the unrestricted parameter space and $C \in (0, 1]$.

He then showed that there exists a global maximizer of $L(\theta^{j|i}|\mathcal{R}_T^j)$ over $\Theta_C^{j|i}$ and that, provided $\theta_0^{j|i} \in \Theta_C^{j|i}$, this global maximizer is strongly consistent.

We now apply Bayes' rule to define the following set of probabilities for each $s \in \{0, 1\}$ and $t \in \{1, ..., T\}$

$$\mathbb{P}\{S_t^i = s, R_t^j; \theta^{j|i}\} = \mathbb{P}\{S_t^i = s\} f(R_t^j|S_t^i = s; \theta^{j|i}), \tag{3.4a}$$

$$\mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\} = \frac{\mathbb{P}\{S_t^i = s, R_t^j; \theta^{j|i}\}}{f(R_t^j; \theta^{j|i})}. \tag{3.4b}$$

(3.4b) reflects the posterior probability that index j's log return at time t is generated by state $S_t^i = s$ given that R_t^j is known (also called membership probability). This probability plays a crucial role in the EM algorithm, as considered in Section 3.2.

Computing derivates of the log likelihood function $l(\theta^{j|i}|\mathcal{R}_T^j) = \log \prod_{t=1}^T f(R_t^j; \theta^{j|i})$ yields for each $s \in \{0, 1\}$

$$\begin{aligned}
\frac{\partial l(\theta^{j|i}|\mathcal{R}_T^j)}{\partial \alpha_s^{j|i}} &= \sum_{t=1}^T \frac{\mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\}}{(\sigma_s^{j|i})^2} (\alpha_s^{j|i} - R_t^j), \\
\frac{\partial l(\theta^{j|i}|\mathcal{R}_T^j)}{\partial \sigma_s^{j|i}} &= \sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\} \Big[-\frac{1}{\sigma_s^{j|i}} + \frac{(R_t^j - \alpha_s^{j|i})^2}{(\sigma_s^{j|i})^3} \Big].
\end{aligned}$$

Setting the derivatives to zero, we find the following set of equations for each $s \in \{0, 1\}$

$$\begin{aligned}
\alpha_s^{j|i} &= \frac{\sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\} R_t^j}{\sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\}}, \\
\sigma_s^{j|i} &= \left[\frac{\sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\} (R_t^j - \alpha_s^{j|i})^2}{\sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\}} \right]^{\frac{1}{2}}.
\end{aligned}$$

The conditional probability $\mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\}$ appears in each equation. This probability in turn depends on $\theta^{j|i}$ through (3.4a) - (3.4b). We conclude that these equations are not easily solvable.

3.1.2 Model with Time-varying Mixing Weights

We extend index j's log return specification to an alternative setting. In particular,

$$R_t^j = \begin{cases} \beta_0^{j|i} R_t^i + \sigma_0^{j|i} \epsilon_t^{j|i} & \text{if } S_t^i = 0 \\ \beta_1^{j|i} R_t^i + \sigma_1^{j|i} \epsilon_t^{j|i} & \text{if } S_t^i = 1 \end{cases} \qquad (t = 1, ..., T), \tag{3.5}$$

where $\epsilon_t^{j|i} \overset{\text{i.i.d.}}{\sim} N(0,1)$.

Index i's log returns are again assumed to a follow a Markov switching process according to (3.1).

The density function under each regime is given by

$$f(R_t^j|S_t^i = s, \mathcal{R}_t^i; \theta^{j|i}) = \frac{1}{\sqrt{2\pi}\sigma_s^{j|i}} \exp\left[\frac{-(R_t^j - \beta_s^{j|i} R_t^i)^2}{2(\sigma_s^{j|i})^2}\right].$$

We can now specify the density of the tth observation

$$f(R_t^j|\mathcal{R}_t^i; \theta^{j|i}) = \sum_{s=0}^{1} \mathbb{P}\{S_t^i = s|\mathcal{R}_t^i\} f(R_t^j|S_t^i = s, \mathcal{R}_t^i; \theta^{j|i}).$$

The likelihood function is again the product of the individual densities

$$L(\theta^{j|i}|\mathcal{R}_T^j, \mathcal{R}_T^i) = \prod_{t=1}^{T} f(R_t^j|\mathcal{R}_t^i; \theta^{j|i}).$$

Maximum likelihood estimation yields the following set of equations for each $s \in \{0, 1\}$

$$\beta_s^{j|i} = \frac{\sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j, \mathcal{R}_t^i; \theta^{j|i}\} R_t^j R_t^i}{\sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j, \mathcal{R}_t^i; \theta^{j|i}\} R_t^i R_t^i},$$

$$\sigma_s^{j|i} = \left[\frac{\sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j, \mathcal{R}_t^i; \theta^{j|i}\}(R_t^j - \beta_s^{j|i} R_t^i)^2}{\sum_{t=1}^T \mathbb{P}\{S_t^i = s|R_t^j, \mathcal{R}_t^i; \theta^{j|i}\}}\right]^{\frac{1}{2}},$$

where the posterior probabilities are defined as follows

$$\mathbb{P}\{S_t^i = s|R_t^j, \mathcal{R}_t^i; \theta^{j|i}\} = \frac{\mathbb{P}\{S_t^i = s, R_t^j|\mathcal{R}_t^i; \theta^{j|i}\}}{f(R_t^j|\mathcal{R}_t^i; \theta^{j|i})}, \tag{3.6a}$$

$$\mathbb{P}\{S_t^i = s, R_t^j|\mathcal{R}_t^i; \theta^{j|i}\} = \mathbb{P}\{S_t^i = s|\mathcal{R}_t^i\} f(R_t^j|S_t^i = s, \mathcal{R}_t^i; \theta^{j|i}), \tag{3.6b}$$

for each $s \in \{0, 1\}$ and $t \in \{1, ..., T\}$.

3.2 EM algorithm

We outline the general principles underlying the expectation-maximization (EM) algorithm, which is an iterative method for obtaining maximum likelihood parameter estimates (Dempster et al. [11]). The general principle is as follows: given current parameter estimates $\hat{\theta}_{(n)}^{j|i}$, the conditional distribution of the latent variables is computed to determine the expected log likelihood function (a.k.a. expectation step); the expected log-likelihood function is then maximized to obtain new parameter estimates $\hat{\theta}_{(n+1)}^{j|i}$ (a.k.a. maximization step). These new parameter estimates are then used to compute the conditional distribution of the latent variables.

By way of motivation, we illustrate the general principles of the EM algorithm for the finite mixture model. The log likelihood function is then specified as follows

$$\begin{aligned} l(\theta^{j|i}|\mathcal{R}_T^j) &= \textstyle\sum_{t=1}^T \log f(R_t^j; \theta^{j|i}) \\ &= \textstyle\sum_{t=1}^T \log \sum_{s=0}^1 \mathbb{P}\{S_t^i = s\} f(R_t^j|S_t^i = s; \theta^{j|i}) \\ &= \textstyle\sum_{t=1}^T \log \sum_{s=0}^1 \mathbb{P}\{S_t^i = s\} f(R_t^j|S_t^i = s; \theta^{j|i}) \cdot \frac{\mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\}}{\mathbb{P}\{S_t^i = s|R_t^j; \theta^{j|i}\}}. \end{aligned}$$

In the expectation step, the probability $\mathbb{P}\{S_t^i = s|R_t^j; \hat{\theta}_{(n)}^{j|i}\}$ is computed for each $s \in \{0, 1\}$ and $t \in \{1, ..., T\}$. These probabilities are then used to specify an expected log likelihood function, which is a lower bound on the true log likelihood function. The construction follows via Jensen's inequality. We briefly rephrase Jensen's inequality.

Lemma 3.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a $\mathbb{P}$-integrable real-valued random variable and $\phi : \mathbb{R} \mapsto \mathbb{R}$ a concave function. Then*

$$\phi(\mathbb{E}X) = \phi\Big(\int_\Omega x d\mathbb{P}\Big) \geq \int_\Omega \phi(x) d\mathbb{P} = \mathbb{E}\phi(X).$$

Proof. See for example Yeh [48]. Q.E.D.

By virtue of Jensen's inequality, the log likelihood function can then be bounded as follows

$$l(\theta^{j|i}|\mathcal{R}_T^j) = \sum_{t=1}^{T} \log f(R_t^j;\theta^{j|i}) \geq \sum_{t=1}^{T}\sum_{s=0}^{1} \mathbb{P}\{S_t^i = s|R_t^j;\theta^{j|i}\} \log \frac{\mathbb{P}\{S_t^i = s\} f(R_t^j|S_t^i = s;\theta^{j|i})}{\mathbb{P}\{S_t^i = s|R_t^j;\theta^{j|i}\}}.$$

We introduce the following object, which can be viewed as an expected log likelihood function.

$$Q(\theta^{j|i}|\hat{\theta}^{j|i}_{(n)},\mathcal{R}_T^j) = \sum_{t=1}^{T}\sum_{s=0}^{1} \mathbb{P}\{S_t^i = s|R_t^j;\hat{\theta}^{j|i}_{(n)}\} \log \frac{\mathbb{P}\{S_t^i = s\} f(R_t^j|S_t^i = s;\theta^{j|i})}{\mathbb{P}\{S_t^i = s|R_t^j;\hat{\theta}^{j|i}_{(n)}\}}.$$

The EM algorithm is defined as follows

Start with an initial guess $\hat{\theta}^{j|i}_{(0)}$. Execute the following iterative scheme until convergence[9].

1. **Expectation Step**: given current parameter estimates $\hat{\theta}^{j|i}_{(n)}$, compute the probability $\mathbb{P}\{S_t^i = s|R_t^j;\hat{\theta}^{j|i}_{(n)}\}$ for each $s \in \{0,1\}$ and $t \in \{1,...,T\}$. Then construct the expected log likelihood function $Q(\theta^{j|i}|\hat{\theta}^{j|i}_{(n)},\mathcal{R}_T^j)$.

2. **Maximization Step**: Maximize $Q(\theta^{j|i}|\hat{\theta}^{j|i}_{(n)},\mathcal{R}_T^j)$. Put $\hat{\theta}^{j|i}_{(n+1)} = \arg\max_{\theta^{j|i}} Q(\theta^{j|i}|\hat{\theta}^{j|i}_{(n)},\mathcal{R}_T^j)$.

The likelihood function may have multiple (local) maxima. It could therefore be wise to start with a grid of initial values to find the largest maximum or to apply the restricted EM, as proposed by Hathaway [25].

Finally, we mention two well-known properties of the EM algorithm.

Lemma 3.2. *The log likelihood function increases at each iteration step.*

Proof. To show the assertion of the lemma, note first that

$$\begin{aligned}
Q(\hat{\theta}^{j|i}_{(n)}|\hat{\theta}^{j|i}_{(n)},\mathcal{R}_T^j) &= \textstyle\sum_{t=1}^{T}\sum_{s=0}^{1} \mathbb{P}\{S_t^i = s|R_t^j;\hat{\theta}^{j|i}_{(n)}\} \log \frac{\mathbb{P}\{S_t^i=s\} f(R_t^j|S_t^i=s;\hat{\theta}^{j|i}_{(n)})}{\mathbb{P}\{S_t^i=s|R_t^j;\hat{\theta}^{j|i}_{(n)}\}} \\
&= \textstyle\sum_{t=1}^{T}\sum_{s=0}^{1} \mathbb{P}\{S_t^i = s|R_t^j;\hat{\theta}^{j|i}_{(n)}\} \log \frac{\mathbb{P}\{S_t^i=s,R_t^j;\hat{\theta}^{j|i}_{(n)}\}}{\mathbb{P}\{S_t^i=s|R_t^j;\hat{\theta}^{j|i}_{(n)}\} f(R_t^j;\hat{\theta}^{j|i}_{(n)})} + l(\hat{\theta}^{j|i}_{(n)}|\mathcal{R}_T^j) \\
&= \textstyle\sum_{t=1}^{T}\sum_{s=0}^{1} \mathbb{P}\{S_t^i = s|R_t^j;\hat{\theta}^{j|i}_{(n)}\} \log \frac{\mathbb{P}\{S_t^i=s,R_t^j;\hat{\theta}^{j|i}_{(n)}\}}{\mathbb{P}\{S_t^i=s,R_t^j;\hat{\theta}^{j|i}_{(n)}\}} + l(\hat{\theta}^{j|i}_{(n)}|\mathcal{R}_T^j) \\
&= l(\hat{\theta}^{j|i}_{(n)}|\mathcal{R}_T^j).
\end{aligned}$$

[9]Different convergence criteria can be used. For instance, stop the algorithm if $||\hat{\theta}^{j|i}_{(n+1)} - \hat{\theta}^{j|i}_{(n)}|| < 10^{-8}$, where $||\cdot||$ is an appropriate norm.

Combined with Jensen's inequality and the fact that $Q(\hat{\theta}^{j|i}_{(n+1)}|\hat{\theta}^{j|i}_{(n)}, \mathcal{R}^j_T) \geq Q(\hat{\theta}^{j|i}_{(n)}|\hat{\theta}^{j|i}_{(n)}, \mathcal{R}^j_T)$ (i.e. by definition of a maximum), we have that

$$l(\hat{\theta}^{j|i}_{(n+1)}|\mathcal{R}^j_T) \geq Q(\hat{\theta}^{j|i}_{(n+1)}|\hat{\theta}^{j|i}_{(n)}, \mathcal{R}^j_T) \geq Q(\hat{\theta}^{j|i}_{(n)}|\hat{\theta}^{j|i}_{(n)}, \mathcal{R}^j_T) = l(\hat{\theta}^{j|i}_{(n)}|\mathcal{R}^j_T).$$

Q.E.D.

Lemma 3.3. *The sequence* $\{\hat{\theta}^{j|i}_{(n)}\}$ *converges in probability to a local maximum likelihood estimate. In particular, if*

$$\frac{\partial Q(\theta^{j|i}|\hat{\theta}^{j|i}_{(n)}, \mathcal{R}^j_T)}{\partial \theta^{j|i}}\Big|_{\theta^{j|i}=\hat{\theta}^{j|i}_{(n)}} = 0,$$

then

$$\frac{\partial l(\theta^{j|i}|\mathcal{R}^j_T)}{\partial \theta^{j|i}}\Big|_{\theta^{j|i}=\hat{\theta}^{j|i}_{(n)}} = 0.$$

Proof. We have that

$$\begin{aligned}
\frac{\partial Q(\theta^{j|i}|\hat{\theta}^{j|i}_{(n)}, \mathcal{R}^j_T)}{\partial \theta^{j|i}}\Big|_{\theta=\hat{\theta}^{j|i}_{(n)}} &= \sum_{t=1}^T \sum_{s=0}^1 \frac{\mathbb{P}\{S^i_t=s|R^j_t;\hat{\theta}^{j|i}_{(n)}\}}{\mathbb{P}\{S^i_t=s\}f(R^j_t|S^i_t=s;\theta^{j|i})}\mathbb{P}\{S^i_t=s\}\frac{\partial f(R^j_t|S^i_t=s;\theta^{j|i})}{\partial\theta^{j|i}}\Big|_{\theta^{j|i}=\hat{\theta}^{j|i}_{(n)}} \\
&= \sum_{t=1}^T \sum_{s=0}^1 \frac{\mathbb{P}\{S^i_t=s|R^j_t;\hat{\theta}^{j|i}_{(n)}\}}{\mathbb{P}\{S^i_t=s\}f(R^j_t|S^i_t=s;\hat{\theta}^{j|i}_{(n)})}\mathbb{P}\{S^i_t=s\}\frac{\partial f(R^j_t|S^i_t=s;\theta^{j|i})}{\partial\theta^{j|i}}\Big|_{\theta^{j|i}=\hat{\theta}^{j|i}_{(n)}} \\
&= \sum_{t=1}^T \sum_{s=0}^1 \frac{1}{f(R^j_t;\hat{\theta}^{j|i}_{(n)})}\mathbb{P}\{S^i_t=s\}\frac{\partial f(R^j_t|S^i_t=s;\theta^{j|i})}{\partial\theta^{j|i}}\Big|_{\theta^{j|i}=\hat{\theta}^{j|i}_{(n)}} \\
&= \sum_{t=1}^T \frac{\partial \log f(R^j_t;\theta^{j|i})}{\partial\theta^{j|i}}\Big|_{\theta^{j|i}=\hat{\theta}^{j|i}_{(n)}} \\
&= \frac{\partial l(\theta^{j|i}|R^j)}{\partial\theta^{j|i}}\Big|_{\theta^{j|i}=\hat{\theta}^{j|i}_{(n)}}.
\end{aligned}$$

Q.E.D.

3.3 Particular form of EM algorithm

3.3.1 Finite Mixture Model

In the expectation step, the probability $\mathbb{P}\{S^i_t = s|R^j_t;\hat{\theta}^{j|i}_{(n)}\}$ is computed for each $s \in \{0, 1\}$ and $t \in \{1, ..., T\}$. However, these probabilities depend on the transition probabilities, which are unknown. An obvious candidate to estimate the posterior probability is the estimator

$$\mathbb{P}\{S^i_t = s|R^j_t;\hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\} = \frac{\mathbb{P}\{S^i_t = s, R^j_t;\hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\}}{f(R^j_t;\hat{\theta}^{j|i}_{(n)})}, \tag{3.7}$$

where

$$\mathbb{P}\{S_t^i = s, R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\} = \frac{1-\hat{p}_{(1-s)(1-s)}^i}{2-\hat{p}_{00}^i - \hat{p}_{11}^i} f(R_t^j | S_t^i = s; \hat{\theta}_{(n)}^{j|i}).$$

We observe that the estimated posterior probability is a function of the estimated transition probabilities. These transition probabilities are estimated on the basis of a sample of length T. The following lemma establishes that $\hat{p}_{ss}^i \xrightarrow{P} p_{ss}^i$ as $T \to \infty$ for each $s \in \{0, 1\}$. More generally,

Lemma 3.4. *The MLE of an (i.i.d.) Markov switching regression model are consistent and asymptotically normal distributed.*

Proof. See Kiefer [30]. Q.E.D.

The EM algorithm takes now the following form

1. **Expectation Step** Given current estimates $\hat{\theta}_{(n)}^{j|i}$, estimate $\mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}\}$ for each $s \in \{0, 1\}$ and $t \in \{1, ..., T\}$ by (3.7). Construct $Q(\theta^{j|i} | \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i, \mathcal{R}_T^j)$ as follows

$$Q(\theta^{j|i} | \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i, \mathcal{R}_T^j) = \sum_{t=1}^{T} \sum_{s=0}^{1} \mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\} \log \frac{\mathbb{P}\{S_t^i = s, R_t^j; \theta^{j|i}\}}{\mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\}}.$$

2. **Maximization Step** Maximize $Q(\theta^{j|i} | \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i, \mathcal{R}_T^j)$, or equivalently,

$$\sum_{t=1}^{T} \sum_{s=0}^{1} \mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\} \log \mathbb{P}\{S_t^i = s, R_t^j; \theta^{j|i}\}.$$

   We obtain the following estimates for each $s \in \{0, 1\}$

$$(\hat{\alpha}_s^{j|i})_{(n+1)} = \frac{\sum_{t=1}^T \mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\} R_t^j}{\sum_{t=1}^T \mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\}},$$

$$(\hat{\sigma}_s^{j|i})_{(n+1)} = \left[\frac{\sum_{t=1}^T \mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\} (R_t^j - (\hat{\alpha}_s^{j|i})_{(n+1)})^2}{\sum_{t=1}^T \mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\}} \right]^{\frac{1}{2}}.$$

We now argue that $\hat{\theta}^{j|i}$ is locally consistent, i.e. $\hat{\theta}^{j|i} = \theta_0^{j|i} + o_P(1)$.

Lemma 3.5. *Given fixed parameter estimates $\hat{\theta}_{(n)}^{j|i}$ and R_t^j,*

$$\mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}, \hat{\theta}^i\} \xrightarrow{P} \mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}_{(n)}^{j|i}\}$$

as $T \to \infty$.

Proof. By Lemma (3.4), $\hat{p}^i_{ss} \xrightarrow{P} p^i_{ss}$ as $T \to \infty$ for each $s \in \{0,1\}$. Let $g : [0,1]^2 \mapsto [0,1]$ be defined by

$$g(x,y) = \frac{1-x}{2-x-y}.$$

Since g is a continuous map and the Markov chain is ergodic, we have for each $s \in \{0,1\}$

$$\mathbb{P}\{S^i_t = s; \hat{\theta}^i\} = \frac{1-\hat{p}^i_{(s-1)(s-1)}}{2-\hat{p}^i_{00}-\hat{p}^i_{11}} \xrightarrow{P} \frac{1-p^i_{(s-1)(s-1)}}{2-p^i_{00}-p^i_{11}} = \mathbb{P}\{S^i_t = s\}.$$

By the continuous mapping theorem, for each $s \in \{0,1\}$ and $t = \{1,...,T\}$,

$$\begin{aligned}
\mathbb{P}\{S^i_t = s, R^j_t; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\} &\xrightarrow{P} \mathbb{P}\{S^i_t = s, R^j_t; \hat{\theta}^{j|i}_{(n)}\}, \\
\mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\} &\xrightarrow{P} \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(n)}\}.
\end{aligned}$$

Q.E.D.

Theorem 3.1. *$\hat{\theta}^{j|i}$ is locally consistent.*

Proof. Define

$$\hat{\theta}^{j|i}_{(\infty)} = \lim_{n\to\infty} \hat{\theta}^{j|i}_{(n)}.$$

By Lemma 3.5, $\mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}, \hat{\theta}^i\} = \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\} + o_P(1)$. We show the assertion for $(\alpha^{j|i}_s)_{(\infty)}$. For $(\sigma^{j|i}_s)_{(\infty)}$, the proof goes analogously. We have the following decomposition

$$\begin{aligned}
\frac{\sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}, \hat{\theta}^i\} R^j_t}{\sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}, \hat{\theta}^i\}} &= \frac{\sum_{t=1}^T (\mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\} + o_P(1)) R^j_t}{\sum_{t=1}^T (\mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\} + o_P(1))} \\
&= \frac{\sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\} R^j_t}{\sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\}} + ... \\
&\quad \frac{T^{-1} \sum_{t=1}^T o_P(1) R^j_t T^{-1} \sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\}}{[T^{-1} \sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\}]^2 + T^{-1} \sum_{t=1}^T o_P(1) T^{-1} \sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\}} - ... \\
&\quad \frac{T^{-1} \sum_{t=1}^T o_P(1) T^{-1} \sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\} R^j_t}{[T^{-1} \sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\}]^2 + T^{-1} \sum_{t=1}^T o_P(1) T^{-1} \sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\}}.
\end{aligned}$$

We show that the last two terms on the right-hand side converge to zero as $T \to \infty$. R^j_t is independent and identically distributed (i.i.d.). In particular,

$$f(R^j_t; \hat{\theta}^{j|i}_{(\infty)}) = \mathbb{P}\{S^i_t = 0\} f(R^j_t | S^i_t = 0; \hat{\theta}^{j|i}_{(\infty)}) + \mathbb{P}\{S^i_t = 1\} f(R^j_t | S^i_t = 1; \hat{\theta}^{j|i}_{(\infty)}).$$

It follows that $\mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\}$ is also an i.i.d. sequence and hence, by the weak law of large numbers,

$$\frac{1}{T} \sum_{t=1}^T \mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\} \xrightarrow{P} \mathbb{E}\mathbb{P}\{S^i_t = s | R^j_t; \hat{\theta}^{j|i}_{(\infty)}\} > 0.$$

Obviously,

$$\frac{1}{T}\sum_{t=1}^{T} o_P(1) \xrightarrow{P} 0.$$

$$\frac{1}{T}\sum_{t=1}^{T} o_P(1) R_t^j = \frac{1}{T}\sum_{t=1}^{T} o_P(1) O_P(1) \xrightarrow{P} 0.$$

Hence, by an application of the continuous mapping theorem, the last term on the right-hand side converges to

$$\frac{0 \cdot \mathbb{E}\mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}^{j|i}_{(\infty)}\}}{[\mathbb{E}\mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}^{j|i}_{(\infty)}\}]^2} - \frac{0 \cdot \mathbb{E}\{\mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}^{j|i}_{(\infty)}\} R_t^j\}}{[\mathbb{E}\mathbb{P}\{S_t^i = s | R_t^j; \hat{\theta}^{j|i}_{(\infty)}\}]^2} = 0.$$

The first term on the right-hand side converges to the true parameter value and hence it follows that $\theta_0^{j|i} = \theta^{j|i}_{(\infty)} + o_P(1)$. Q.E.D.

3.3.2 Model with Time-varying Mixing Weights

We consider the following estimator of $\mathbb{P}\{S_t^i = s | R_t^j, \mathcal{R}_t^i; \hat{\theta}^{j|i}_{(n)}\}$

$$\mathbb{P}\{S_t^i = s | R_t^j, \mathcal{R}_t^i; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\} = \frac{\mathbb{P}\{S^i = s, R_t^j | \mathcal{R}_t^i; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\}}{f(R_t^j | \mathcal{R}_t^i; \hat{\theta}^{j|i}_{(n)})}, \tag{3.8}$$

where

$$\mathbb{P}\{S_t^i = s, R_t^j | \mathcal{R}_t^i; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\} = \mathbb{P}\{S_t^i = s | \mathcal{R}_t^i, \hat{\theta}^i\} f(R_t^j | S_t^i = s; \hat{\theta}^{j|i}_{(n)}).$$

The filtered probabilities are estimated recursively

$$\begin{aligned}
\mathbb{P}\{S_0^i = s\} &= \tfrac{1}{2}. \\
\mathbb{P}\{S_t^i = s | \mathcal{R}_t^i, \hat{\theta}^i\} &= \frac{\sum_{x=0}^{1} \mathbb{P}\{S_{t-1}^i = x | \mathcal{R}_{t-1}^i, \hat{\theta}^i\} \hat{p}_{xs}^i f(R_t^i | S_t^i = s; \hat{\theta}^i)}{\sum_{x=0}^{1}\sum_{y=0}^{1} \mathbb{P}\{S_{t-1}^i = x | \mathcal{R}_{t-1}^i, \hat{\theta}^i\} \hat{p}_{xy}^i f(R_t^i | S_t^i = y; \hat{\theta}^i)}.
\end{aligned}$$

The EM algorithm takes the following form

1. **Expectation Step** Given current estimates $\hat{\theta}^{j|i}_{(n)}$, estimate $\mathbb{P}\{S_t^i = s | R_t^j, \mathcal{R}_t^i; \hat{\theta}^{j|i}_{(n)}\}$ for each $s \in \{0, 1\}$ and $t \in \{1, ..., T\}$ by (3.8). Construct $Q(\theta^{j|i} | \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i, \mathcal{R}_T^j, \mathcal{R}_T^i,)$ as follows

$$Q(\theta^{j|i} | \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i, \mathcal{R}_T^j, \mathcal{R}_T^i) = \sum_{t=1}^{T}\sum_{s=0}^{1} \mathbb{P}\{S_t^i = s | R_t^j, \mathcal{R}_t^i; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\} \log \frac{\mathbb{P}\{S_t^i = s, R_t^j | \mathcal{R}_t^i; \theta^{j|i}\}}{\mathbb{P}\{S_t^i = s | R_t^j, \mathcal{R}_t^i; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\}}.$$

2. **Maximization Step** Maximize $Q(\theta^{j|i}|\hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i, \mathcal{R}^j_T, \mathcal{R}^i_T)$, or equivalently,

$$\sum_{t=1}^{T}\sum_{s=0}^{1}\mathbb{P}\{S^i_t = s|R^j_t, \mathcal{R}^i_t; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\}\log\mathbb{P}\{S^i_t = s, R^j_t|\mathcal{R}^i_t; \theta^{j|i}\}.$$

We obtain the following estimates for each $s \in \{0,1\}$

$$(\hat{\beta}^{j|i}_s)_{(n+1)} = \frac{\sum_{t=1}^{T}\mathbb{P}\{S^i_t = s|R^j_t, \mathcal{R}^i_t; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\}R^j_t}{\sum_{t=1}^{T}\mathbb{P}\{S^i_t = s|R^j_t, \mathcal{R}^i_t; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\}},$$

$$(\hat{\sigma}^{j|i}_s)_{(n+1)} = \left[\frac{\sum_{t=1}^{T}\mathbb{P}\{S^i_t = s|R^j_t, \mathcal{R}^i_t; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\}(R^j_t - (\hat{\beta}^{j|i}_s)_{(n+1)})^2}{\sum_{t=1}^{T}\mathbb{P}\{S^i_t = s|R^j_t, \mathcal{R}^i_t; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\}}\right]^{\frac{1}{2}}.$$

Lemma 3.6. *Given fixed parameter estimates $\hat{\theta}^{j|i}_{(n)}$, R^j_t and $\mathcal{R}^i_t$,*

$$\mathbb{P}\{S^i_t = s|R^j_t, \mathcal{R}^i_t; \hat{\theta}^{j|i}_{(n)}, \hat{\theta}^i\} \xrightarrow{P} \mathbb{P}\{S^i_t = s|R^j_t, \mathcal{R}^i_t; \hat{\theta}^{j|i}_{(n)}\}$$

as $T \to \infty$.

Proof. By Lemma (3.4), $\hat{\theta}^i \xrightarrow{P} \theta^i$ as $T \to \infty$. By the continuous mapping theorem, for each $s \in \{0,1\}$,

$$f(R^i_t|S^i_t = s; \hat{\theta}^i) \xrightarrow{P} f(R^i_t|S^i_t = s; \theta^i)$$

as $\hat{\theta}^i \xrightarrow{P} \theta^i$.

By another application of the continuous mapping theorem, it follows that

$$\mathbb{P}\{S^i_t = s|\mathcal{R}^i_t, \hat{\theta}^i\} \xrightarrow{P} \mathbb{P}\{S^i_t = s|\mathcal{R}^i_t\}$$

as $T \to \infty$. Q.E.D.

Theorem 3.2. *$\hat{\theta}^{j|i}$ is locally consistent.*

Proof. The proof is analogous to the one of Theorem (3.1). Q.E.D.

3.4 Monte Carlo Study: Finite Sample Behavior.

In this section, we conduct a Monte Carlo study to determine the finite sample distributions of the estimators.

One should be aware of the fact that we are estimating a Markov switching regression model and that the accuracy of the estimators strongly depends on the underlying transition matrix. It could happen that no switch occurs in any finite time period. We consider a (small) setting that mimics that of the empirical study conducted in Section 5. In particular, we choose a sample of length $T = 1,000$ and use the (fictitious) parameter values as presented in Table 2.

Model 1		**Model 2**	
α_0^i	0	α_0^i	0
α_1^i	0	α_1^i	0
σ_0^i	0.03	σ_0^i	0.03
σ_1^i	0.06	σ_1^i	0.06
p_{00}^i	0.95	p_{00}^i	0.95
p_{11}^i	0.85	p_{11}^i	0.85
$\alpha_0^{j\|i}$	-0.03	$\beta_0^{j\|i}$	0.30
$\alpha_1^{j\|i}$	0.03	$\beta_1^{j\|i}$	0.60
$\sigma_0^{j\|i}$	0.02	$\sigma_0^{j\|i}$	0.02
$\sigma_1^{j\|i}$	0.04	$\sigma_1^{j\|i}$	0.04

Table 2: **(Fictitious) parameter values**

The Monte Carlo study proceeds as follows. For a sample of length $T = 1,000$

1. With the help of a random number generator, simulate a Markov chain (i.e. index i's true states) with transition probability matrix P^i and a sequence $\{\epsilon_t^i\}$ of independent and identically distributed errors.
2. Calculate sample observations for R_t^i according to (3.1).
3. Estimate the transition probabilities and filtered probabilities.
4. Generate a sequence $\{\epsilon_t^{j|i}\}$ of independent and identically distributed errors.
5. Calculate sample observations for R_t^j according to (3.2) and (3.5).
6. Apply EM algorithm to obtain parameter estimates.

The above steps are repeated. The number of replications is set at 1,000. The variability in the parameter estimates is summarized by the figures and table below.

The variability in the estimators strongly depends on p_{ss}^i. Here, $p_{11}^i < p_{00}^i$ and hence the variability in the parameter estimates in state 1 is larger than the variability in the parameter estimates in state 0.

By inspection of the following figures, we observe that the estimators are concentrated around their true values. In particular, the finite sample distributions seem to be normal.

Model 1	10%	25%	50%	75%	90%
$\sigma_0^{j\|i}$	0.0187	0.0193	0.0200	0.0206	0.0212
$\sigma_1^{j\|i}$	0.0329	0.0364	0.0398	0.0425	0.0444
$\alpha_0^{j\|i}$	-0.0313	-0.0306	-0.0299	-0.0290	-0.0283
$\alpha_1^{j\|i}$	0.0186	0.0238	0.0298	0.0371	0.0435
Model 2	10%	25%	50%	75%	90%
$\sigma_0^{j\|i}$	0.0189	0.0194	0.0200	0.0206	0.0211
$\sigma_1^{j\|i}$	0.0365	0.0380	0.0396	0.0414	0.0433
$\beta_0^{j\|i}$	0.2607	0.2795	0.3017	0.3222	0.3416
$\beta_1^{j\|i}$	0.5387	0.5679	0.6005	0.6328	0.6602

Table 3: **Monte Carlo results**.

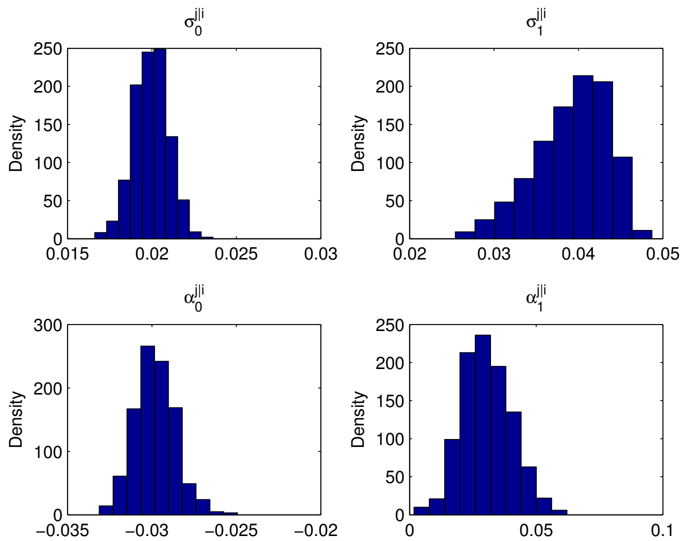

Figure 6: **Finite Sample Distributions**
This figure depicts the finite sample distributions of the estimators of the finite mixture model.

Standard errors

In the empirical study, we obtain standard errors by employing (parametric) bootstrap, which yields consistent estimates (i.e. consistency is guaranteed by the fact that the fitted parameters are consistent). Specifically, we simulate M samples $(S_{1,m}^i, ..., S_{T,m}^i)$, $(R_{1,m}^i, ..., R_{T,m}^i)$ and $(R_{1,m}^j, ..., R_{T,m}^j)$ from the corresponding fitted parametric distribution. For each simulated sequence $(R_{1,m}^j, ..., R_{T,m}^j)$, we apply the EM algorithm to obtain parameter estimates. The standard error can be found by computing the standard deviation of the finite sample distribution.

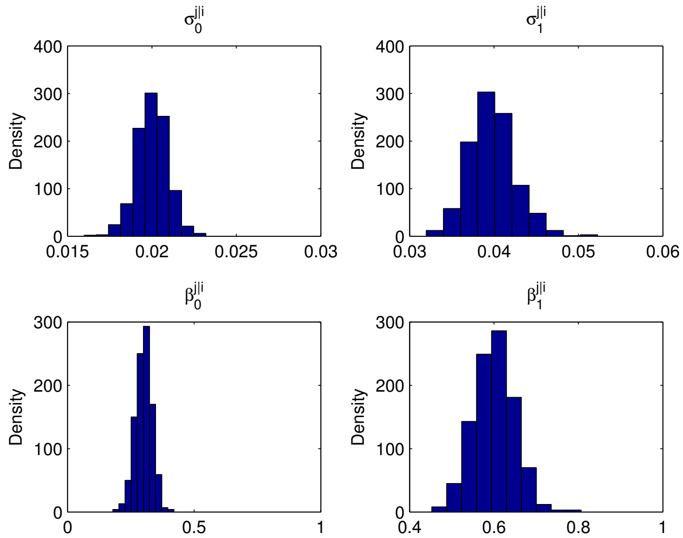

Figure 7: **Finite Sample Distributions**
This figure depicts the finite sample distributions of the estimators of the model with time-varying mixing weights.

4 Time-varying Transition Probabilities

In this section, we propose an alternative return specification. As in Section 2, we infer index i's states by using a classifier. Index j's log return behavior is now described by a Markov switching regression model with time-varying transition probabilities. It could well happen that transition probabilities vary with some $(1 \times k)$ vector Z_{t-1} of exogenous variables. In Section 5, we examine the extent to which index i's regimes affect index j's transition probabilities.

Consider the following index j's log return specification

$$R_t^j = \begin{cases} \alpha_0^j + \sigma_0^j \epsilon_t^j & \text{if } S_t^j = 0 \\ \alpha_1^j + \sigma_1^j \epsilon_t^j & \text{if } S_t^j = 1 \end{cases} \qquad (t = 1, ..., T),$$

where $\epsilon_t^j \overset{\text{i.i.d.}}{\sim} N(0,1)$.

The discrete hidden state variable S_t^j is governed by a first-order Markov chain with transition probability matrix

$$P_t^j = \begin{pmatrix} p_{00,t}^j & p_{01,t}^j \\ p_{10,t}^j & p_{11,t}^j \end{pmatrix},$$

where $p^j_{xy,t} = \mathbb{P}\{S^j_t = y | S^j_{t-1} = x, Z_{t-1}; \kappa^j_x\}$ denotes the probability of moving from state x to state y at time t.

We specify the following functional (i.e. logistic) form of the transition probability, for each $s \in \{0, 1\}$,

$$\mathbb{P}\{S^j_t = s | S^j_{t-1} = s, Z_{t-1}; \kappa^j_s\} = \frac{\exp\{Z_{t-1}\kappa^j_s\}}{1 + \exp\{Z_{t-1}\kappa^j_s\}},$$

where κ^j_s is a $(k \times 1)$ vector of parameters.

The parameter estimates $\hat{\theta}^j = (\hat{\alpha}^j_0, \hat{\alpha}^j_1, \hat{\sigma}^j_0, \hat{\sigma}^j_1, \hat{\kappa}^j_0, \hat{\kappa}^j_1)$ are obtained by applying the EM algorithm. Diebold et al. [13] has derived the EM procedure for a Markov switching specification with time-varying transition probabilities.

Denote by $\mathcal{Z}_t = (Z_s : s \in \{1, ..., t\})$. In the expectation step, the smoothed state probabilities $\mathbb{P}\{S^j_t = s | \mathcal{R}^j_T, \mathcal{Z}_{T-1}; \hat{\theta}^j_{(n)}\}$ and $\mathbb{P}\{S^j_t = y, S^j_{t-1} = x | \mathcal{R}^j_T, \mathcal{Z}_{T-1}; \hat{\theta}^j_{(n)}\}$ are computed. These probabilities are then used to construct the expected complete-data log likelihood function. Maximization of the expected log likelihood function yields the following updated parameter estimates

$$\begin{aligned}
(\alpha^j_s)_{(n+1)} &= \frac{\sum_{t=1}^T \mathbb{P}\{S^j_t = s | \mathcal{R}^j_T, \mathcal{Z}_{T-1}; \hat{\theta}^j_{(n)}\} R^j_t}{\sum_{t=1}^T \mathbb{P}\{S^j_t = s | \mathcal{R}^j_T, \mathcal{Z}_{T-1}; \hat{\theta}^j_{(n)}\}}, \\
(\sigma^j_s)_{(n+1)} &= \Big[\frac{\sum_{t=1}^T \mathbb{P}\{S^j_t = s | \mathcal{R}^j_T, \mathcal{Z}_{T-1}; \hat{\theta}^j_{(n)}\}(R^j_t - (\alpha^j_s)_{(n+1)})^2}{\sum_{t=1}^T \mathbb{P}\{S^j_t = s | \mathcal{R}^j_T, \mathcal{Z}_{T-1}; \hat{\theta}^j_{(n)}\}}\Big]^{\frac{1}{2}}.
\end{aligned}$$

Due to the logistic specification, the remaining first order conditions are nonlinear in κ^j_s. Diebold et al. [13] has deduced closed form solutions by using a first order Taylor series expansion. In particular, for $k = 2$,

$$(\kappa^j_s)_{(n+1)} =$$

$$A_s \times \begin{bmatrix} \sum_{t=2}^T Z_{1,t-1}\big(\mathbb{P}\{S^j_t = s, S^j_{t-1} = s | \mathcal{R}^j_T, \mathcal{Z}_{t-1}; \hat{\theta}^j_{(n)}\} - \mathbb{P}\{S^j_{t-1} = s | \mathcal{R}^j_T, \mathcal{Z}_{T-1}; \hat{\theta}^j_{(n)}\} B_s\big) \\ \sum_{t=2}^T Z_{2,t-1}\big(\mathbb{P}\{S^j_t = s, S^j_{t-1} = s | \mathcal{R}^j_T, \mathcal{Z}_{t-1}; \hat{\theta}^j_{(n)}\} - \mathbb{P}\{S^j_{t-1} = s | \mathcal{R}^j_T, \mathcal{Z}_{T-1}; \hat{\theta}^j_{(n)}\} B_s\big) \end{bmatrix},$$

where

$$A_s = \begin{bmatrix} \sum_{t=2}^{T} Z_{1,t-1}\mathbb{P}\{S_{t-1}^j = s|\mathcal{R}_T^j, \mathcal{Z}_{T-1}; \hat{\theta}_{(n)}^j\} \frac{\partial p_{ss,t}^j}{\partial \kappa_{1,s}^j} & \sum_{t=2}^{T} Z_{1,t-1}\mathbb{P}\{S_{t-1}^j = s|\mathcal{R}_T^j, \mathcal{Z}_{T-1}; \hat{\theta}_{(n)}^j\} \frac{\partial p_{ss,t}^j}{\partial \kappa_{2,s}^j} \\ \sum_{t=2}^{T} Z_{2,t-1}\mathbb{P}\{S_{t-1}^j = s|\mathcal{R}_T^j, \mathcal{Z}_{T-1}; \hat{\theta}_{(n)}^j\} \frac{\partial p_{ss,t}^j}{\partial \kappa_{1,s}^j} & \sum_{t=2}^{T} Z_{2,t-1}\mathbb{P}\{S_{t-1}^j = s|\mathcal{R}_T^j, \mathcal{Z}_{T-1}; \hat{\theta}_{(n)}^j\} \frac{\partial p_{ss,t}^j}{\partial \kappa_{2,s}^j} \end{bmatrix}^{-1},$$

$$B_s = p_{ss,t} - \begin{bmatrix} \frac{\partial p_{ss,t}^j}{\partial \kappa_{1,s}^j} & \frac{\partial p_{ss,t}^j}{\partial \kappa_{2,s}^j} \end{bmatrix} (\kappa_s^j)_{(n)}.$$

The partials are understood to be evaluated at $(\kappa_s^j)_{(n)}$.

5 Empirical Study

In this section, we conduct an empirical study to capture the extent to which index i's performance contributes to index j's performance. We are interested in the systemic risk exposure (i.e. j = BHC, i = Bank Index) and the systemic risk contribution (i.e. j = Bank Index, i = BHC) of 25 US BHCs. These US BHCs are a subset of the 50 largest US BHCs ranked by total asset value (as at 03/31/2011). Table 7 in Appendix A.3 provides a list of the BHCs and their abbreviations. Throughout, we extensively use these abbreviations.

This section is structured as follows. In section 5.1, we describe the data. Section 5.2 provides summary statistics and examines the extent to which a Markov switching regression model is capable of mimicking the observed log return behavior. Finally, in Section 5.3, we present the empirical results. The main tables are gathered in Appendix A.3.

5.1 Data Description

The adjusted weekly closing equity prices are obtained from Yahoo!Finance. We begin our data sample on January 1, 1986 and end it on December 31, 2010 ($T = 1,303$). For some BHCs, the sample starts at a later period (see Table 9 for details). To define the financial system, we use the WORLD-DS Banks index (Code: BANKSWD) which is provided by Datastream.

To control for the state-of-the economy, we include a vector M_{t-1} of lagged macro-economic fundamentals in index i's and index j's log return specification. Following Adrian and Brunnermeier [2] and Hautsch et al. [26], we confine ourselves to a small set of macro-economic fundamentals that are well-known to capture time-variation in conditional moments of log returns.

Specifically,

1. the change in the **implied volatility index** (lagged), VXO, as computed by the Chicago Options Board Exchange (abbr. VXO),

2. the change in the **short term interest rate** (lagged), which is defined as the 13-week Treasury bill rate (abbr. TBR),

3. the change in the **yield spread** (lagged), which is defined as the difference between the 30-year Treasury yield and the short term interest rate (abbr. YSP),

4. the change in the **credit spread** (lagged), which is defined as the difference between the interest rate on Moody's BAA rated bonds and the 30-year Treasury yield (abbr. CSP),

5. the **market return** (lagged), which is defined as the S&P 500 index log return (abbr. S&P).

Except for a liquidity spread and a cumulative real sector return, we use the same set of variables as suggested by Adrian and Brunnermeier [2] and Hautsch et al. [26]. The data of the fundamentals are extracted from Yahoo!Finance.

Figure 8 depicts the evolution of the macro-economic fundamentals over time.

The top panel of Figure 8 shows the evolution of the VXO, which is a gauge of S&P 100 at-the-money index options experiencing one month out, over the period 1986 - 2010. In October, 1987 and late 2008, it skyrocketed to very high levels (e.g. on October 22, 2008 the VXO closed at 70.4%). Late 2008 was experienced as a fearful time, which can also be concluded from the observed levels of the credit spread. Particularly, the credit risk associated with Moody's BAA rated bonds was relatively high during late 2008.

At the end of 2009, the interest rate on 13-week Treasury bills reached an all-time low of one basis point. We generally observe that the short term interest rate shows a cyclical pattern, while the yield spread tends to be countercyclical. The logic behind this is that during economic good times, central bankers will increase short term interest rates to moderate inflation and to slow down the economy.

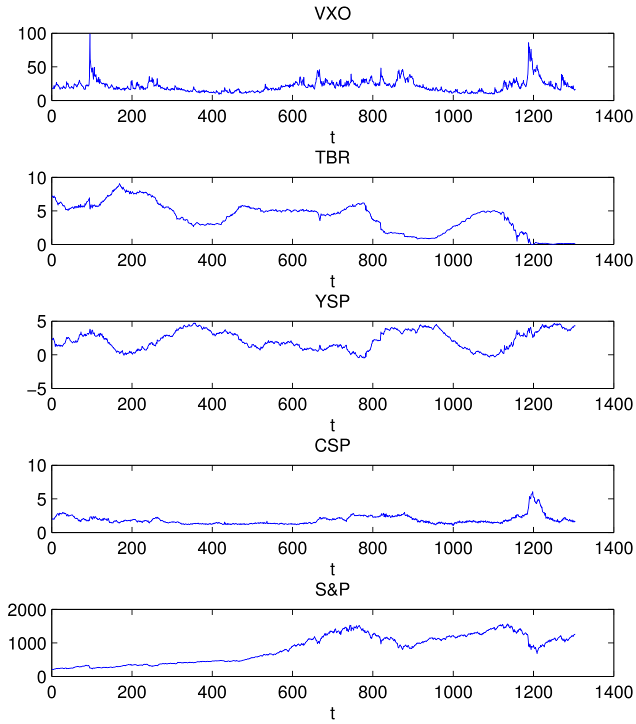

Figure 8: **Fundamentals**
Evolution of the macro-economic fundamentals over time.

5.2 Summary statistics and Model performance

Table 8 and Table 9 report the sample size, mean, standard deviation, skewness, 5%-quantile, median, 95%-quantile and the first, second and third order autocorrelation coefficients (denoted by ρ_1, ρ_2 and ρ_3, respectively) of the macro-economic fundamentals, the Bank Index (log return) and the BHCs (log return). The BHC expected weekly log return ranges from 0.1% to 0.3%. In general, the expected weekly return is slightly less than the median weekly return, which implies a left-skewed empirical return distribution. The volatility of log returns for the Bank Index is 3.1%, while the volatility of log returns for the BHC ranges from 3.8% to 7.1%. We also observe that the first order autocorrelation coefficients are negative, except for one BHC. This supports the evidence for mean reversion in equity prices. However, higher order autocorrelation coefficients have no clear sign. The first order autocorrelation coefficient for the VXO, the 13-week Treasury bill rate, the 30-year Treasury yield and the interest rate on Moody's BAA rated bonds is close to one, which means a high level of persistence in these variables. We also observe that 56% (14 out of 25) of the empirical return distributions are negatively skewed. Many authors have also found varying degrees of skewness (see e.g. Jondeau and Rockinger [29] and Singleton and Wingender [45]).

	Markov switching			Empirical		
BHC	Std.	Skew.	Kurt.	Std.	Skew.	Kurt.
BAC	0.0591	−0.2859	11.4709	0.0592	−0.3166	27.5513
JPM	0.0544	−0.0999	5.6378	0.0542	−0.1860	8.5902
C	0.0705	−0.6521	18.4582	0.0701	−1.5168	49.2842
WFC	0.0483	−0.1342	12.4204	0.0486	0.2473	17.1952
GS	0.0581	−0.0303	6.4053	0.0578	0.0787	7.7355
MS	0.0716	−0.0958	10.9480	0.0722	−1.0029	37.6688
MET	0.0625	−0.2183	13.9241	0.0619	−0.3471	16.5165
HBC	0.0411	−0.4474	11.2470	0.0410	−1.2750	11.3710
PNC	0.0492	−0.1508	7.6330	0.0492	−0.1494	13.9337
BK	0.0463	−0.1231	4.5547	0.0464	0.0296	3.0496

Table 4: **Model Performance**
The last three columns report the empirically observed standard deviation, skewness and excess kurtosis. The first three columns report the standard deviation, skewness and excess kurtosis generated by a Markov switching regression model, assuming the steady state.

Table 4 reports the moments of a Markov switching regression model under the condition that the steady state probabilities apply. That is, it is assumed that the returns have converged towards their steady state. The results should therefore be interpreted with care. Appendix A.2 shows the derivation of the moments of a Markov switching regression model, which is based on a paper of Timmermann [46]. Not surprisingly, the empirically observed standard deviation matches almost exactly the standard deviation generated by the Markov switching regression model. As can be seen from the observed levels of the excess kurtosis[10], the empirical return distributions are all leptokurtic (i.e. have fatter tails than a Gaussian distribution). A Markov switching regression model is capable - to some extent - of generating fat tails. In general, we observe that a Markov switching regression model underestimates the excess kurtosis coefficient (except for BK). The table shows mixed results regarding the skewness coefficient. In particular, for BAC and PNC, the skewness coefficient generated by a Markov switching regression model is almost equal to the empirically observed skewness coefficient.

So far, we have assumed that the $(k \times 1)$ parameter vector γ^i is non-switching. We now examine the implication of this assumption. The parameter estimates of a Markov switching regression model with switching coefficients can be found in Table 10. Except for S&P, the point estimates do not differ much between state 0 and state 1.

[10]This is equal to the kurtosis parameter minus 3, which is the kurtosis of a Gaussian distribution.

To test whether $\gamma_0^i - \gamma_1^i \neq 0$, we can use the sum of the corresponding standard errors to approximate the standard error of $\gamma_0^i - \gamma_1^i$. Except for a few cases[11], we found that γ_0^i is not significantly different from γ_1^i. The last two columns of Table 10 provide information on the absolute deviation between the smoothed probabilities of a Markov switching regression model with switching coefficients and the smoothed probabilities of a Markov switching regression model with non-switching coefficients. We observe that the mean is, in general, close to zero, while the standard deviation ranges from 0.014 to 0.031. We conclude that a specification with switching parameters has little consequences for the inference regarding index i's different regimes.

5.3 Empirical Results

We present the main empirical results for three different model specifications. Model A and Model B are versions of the models presented in Section 2. They only differ in the calibration of the classifier. More precisely,

Model A: $\tau_0 = 0.5$ and $\tau_1 = 0.5$.

Model B: $\tau_0 = 0.97$ and $\tau_1 = 0.97$.

Depending on the realization of the smoothed state probability, index i is labeled to be in state $S_t^i = s$ at time t. By way of illustration, Figure 9 depicts the smoothed state probability of being in crisis at time t.

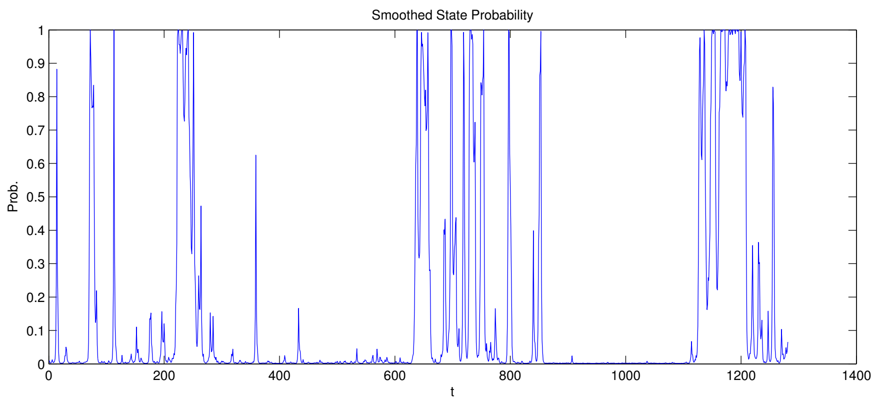

Figure 9: **Smoothed State Probability BAC**
This figure depicts the smoothed state probability of being in state $S_t^i = 1$ at time t.

[11]The parameter estimates of TBR and S&P on HBC and CSP on BK are significantly different between state 0 and state 1.

The third model specification (henceforth called Model C) reads as follows

$$R_t^i = \begin{cases} \alpha_0^i + M_{t-1}\gamma^i + \sigma_0^i \epsilon_t^i & \text{if } S_t^i = 0 \\ \alpha_1^i + M_{t-1}\gamma^i + \sigma_1^i \epsilon_t^i & \text{if } S_t^i = 1 \end{cases} \qquad (t = 1, ..., T),$$

where $\epsilon_t^i \overset{\text{i.i.d.}}{\sim} N(0,1)$.

$$R_t^j = \begin{cases} \alpha_0^{j|i} + M_{t-1}\gamma^{j|i} + \beta_0^{j|i} R_t^i + \sigma^{j|i} \epsilon_t^{j|i} & \text{if } S_t^i = 0 \\ \alpha_1^{j|i} + M_{t-1}\gamma^{j|i} + \beta_1^{j|i} R_t^i + \sigma^{j|i} \epsilon_t^{j|i} & \text{if } S_t^i = 1 \end{cases} \qquad (t = 1, ..., T),$$

where $\epsilon_t^{j|i} \overset{\text{i.i.d.}}{\sim} N(0,1)$.

Parameter estimates are obtained by employing the EM algorithm (cf. Section 3) and are reported in Table 17.

Table 20 and Table 21 provide a ranking among the first ten BHCs.

5.3.1 Estimated correlation coefficient

We estimate Pearson correlation coefficients on the basis of model specification A. The results can be found in Table 11 and are reported for $j =$ Bank Index, $i =$ BHC and $j =$ BHC, $i =$ Bank Index. The results show that the estimated change in correlation is, in general, positive. That is, correlations of returns increase during times of crises. This observation has been noticed by many authors. The change in correlation between the Bank Index and C, BAC, HBC, MET and PNC is relatively large, while the change in correlation between the Bank Index and GS and MS is relatively small. In Table 12, we also report Spearman's rho to capture non-linear responses. These results confirm the fact that correlations of returns increase during times of crises. As noted in Section 2, heteroskedasticity in returns affects estimates of correlation coefficients (i.e. correlation coefficients are biased upwards). $\Delta\rho^{j|i}_{\Delta\beta^{j|i}}$ captures the change in correlation that is not ascribed to heteroskedasticity in index i's returns (Forbes and Rigobon [18] call this the "contagion effect"). The table shows mixed results regarding $\Delta\rho^{j|i}_{\Delta\beta^{j|i}}$. In particular, for $j =$ Bank Index and $i =$ BHC, 3 out of 10 BHCs feature a negative $\Delta\rho^{j|i}_{\Delta\beta^{j|i}}$. By way of illustration, Figure 10 depicts $\Delta\rho^{j|i}_{\Delta\beta^{j|i}}$ versus $\Delta\rho^{j|i}$.

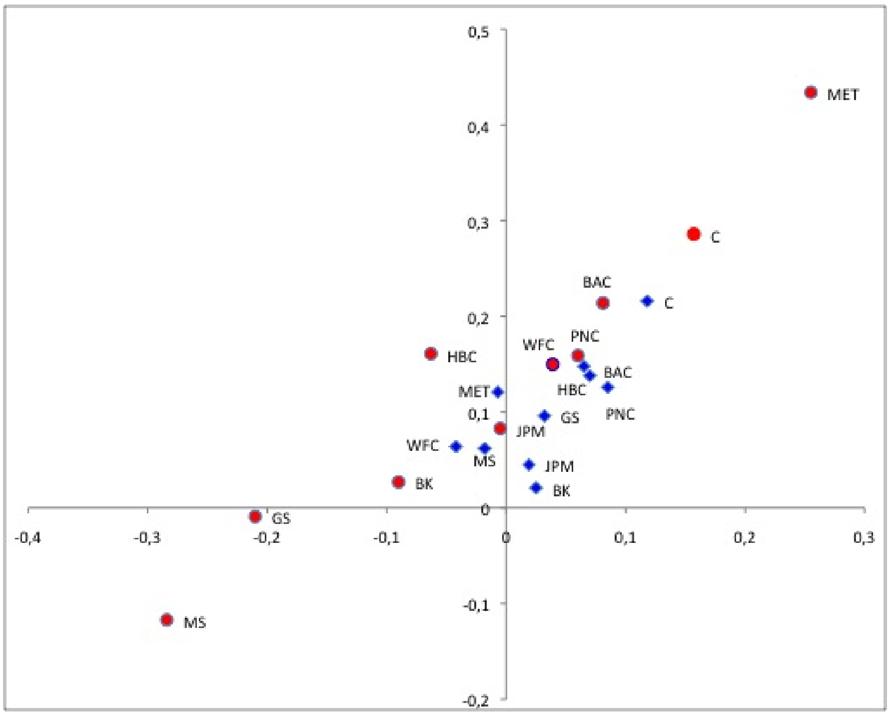

Figure 10: **Estimated change in correlation**
This figure depicts the estimated change in correlation during times of turmoil (vertical axis) versus the estimated change in correlation during times of turmoil after controlling for a change in the variance-covariance matrix of M and a change in the volatility of idiosyncratic shocks (horizontal axis). We have j = Bank Index, i = BHC (square markers) and j = Bank Index, i = Bank Index (circular markers).

If we assess the systemic risk relevance on the basis of $\Delta\rho^{j|i}_{\Delta\beta^{j|i}}$, then the BHCs located in the northeast corner (square markers) can be regarded as systemically relevant banks.

5.3.2 Parameter estimates

Tables 13 through Table 17 provide information on parameter estimates. The parameter estimates differ somewhat among the model specifications, but have the same order of magnitude. Table 13, Table 14 and the first two columns of Table 17 consider the performance of the sytem conditional on the BHC's performance. In particular, the $\hat{\beta}^{j|i}_0$ and $\hat{\beta}^{j|i}_1$ parameter estimates represent the fraction of shocks that are (on average) instantenously transmitted from the BHC to the Bank Index during regime 0 and regime 1, respectively. As can be observed, the tables reveal mixed results and in general it cannot be concluded that $\beta^{j|i}_0$ is smaller than $\beta^{j|i}_1$. For model A, the following table provides information on the significance of the individual regression coefficients. One should be ware of the fact that an insignificant coefficient means that either the data provide insufficient information to reject the null (crises are rare occurrences) or the regression coefficient is negligibly close to zero.

$\beta_0^{j\|i}$ not significant	$\beta_0^{j\|i}$ significant	$\beta_1^{j\|i}$ not significant	$\beta_1^{j\|i}$ significant
JPM, PNC, BK, STT, BBT, KEY, NTRS, MTB, ZION, MIC, BPOP, FHN	BAC, C, WFC, GS, MS, MET, HBC, STI, AXP, RF, FITB, SNV, HTBAN	JPM, WFC, MS, BK, BBT, RF, KEY, NTRS, MTB, SNV, HT-BAN, MIC, BPOP, FHN	BAC, C, GS, MET, HBC, PNC, STI, STT, AXP, FITB, HTBAN

Table 5: **Significance of parameter estimates**
Significance level is 10%.

Comparison of Table 5 with Figure 10 reveals that $\beta_1^{j|i}$ is significant for BAC, C, GS, HBC and PNC (i.e. the BHCs located in the northeast corner of Figure 10) and insignificant for JPM, WFC, MS and BK.

We have also conducted a direct test to determine whether $\beta_0^{j|i}$ is significantly smaller than $\beta_1^{j|i}$. Only for STT and ZION, we are able to reject the null hypothesis. If we consider the null hypothesis that $\beta_0^{j|i}$ is smaller than $\beta_1^{j|i}$, then we cannot reject the null for any BHC at a significance level of 10%.

Table 15, Table 16 and the last two columns of Table 17 consider the BHC's performance conditional on the performance of the system. We have again performed a test to determine whether the regression coefficients are significantly different from zero at a significance level of 10%. The results are summarized below.

$\beta_0^{j\|i}$ not significant	$\beta_0^{j\|i}$ significant	$\beta_1^{j\|i}$ not significant	$\beta_1^{j\|i}$ significant
JPM, MTB, HT-BAN	BAC, C, WFC, GS, MET, HBC, STT, FITB, MS, PNC, BK, STI, BBT, AXP, RF, KEY, NTRS, ZION, MIC, BPOP, SNV, FHN	JPM, MTB, HT-BAN, MC, PNC, BK, STI, BBT, AXP, RF, KEY, NTRS, ZION, MIC, BPOP, SNV, FHN	BAC, C, WFC, GS, MET, HBC, STT, FITB

Table 6: **Significance of parameter estimates**
Significance level is 10%.

We observe that $\beta_0^{j|i}$ is significant for most BHCs, while $\beta_1^{j|i}$ is insignificant for most BHCs. This observation does, however, not imply that $\beta_1^{j|i}$ is zero for most BHCs. The t-values are relatively large (in absolute value) for BAC, C, GS and HBC. We have also conducted a test to determine whether $\beta_0^{j|i}$ is significantly smaller than $\beta_1^{j|i}$.

We are not able to reject any null hypothesis. Only for STT, we found a t-value smaller than minus one.

5.3.3 Implications for supervision

We demonstrate one possible implication for supervision. Regulatory frameworks are often concerned with the financial institution's risk taking behavior seen in isolation ("micro prudential aspect of regulation"). A commonly used measure to assess the institutions's risk taking behavior is VaR. VaR is computed unconditionally and provides insight about the institution's performance under "normal market conditions". Instead of employing a model for the conditional mean of R_t^j, we can also employ a model for the conditional quantiles of R_t^j.

In what follows, we consider $\tau_0 = \tau_1 = \frac{1}{2}$.

We estimate the "unconditional VaR" (i.e. based on the entire sample) for index j at level q, which is implicitly defined as

$$\mathbb{P}\{R_t^j \leq V\tilde{a}R(q)_t^{j|i}\} = q.$$

We impose the following parametric structure

$$V\tilde{a}R(q)_t^{j|i} = \tilde{\alpha}^{j|i} + M_{t-1}\tilde{\gamma}^{j|i} + \tilde{\beta}^{j|i}R_t^i + \tilde{\sigma}^{j|i}\tilde{\epsilon}_t^{j|i},$$

where $\tilde{\epsilon}_t^{j|i} \overset{\text{i.i.d.}}{\sim} N(0,1)$.

We also examine the "conditional VaR" for index j at level q, which is implicitly defined as

$$\mathbb{P}\{R_t^j \leq V\bar{a}R(q)_t^{j|i}\} = q.$$

We impose the following parametric structure

$$V\bar{a}R(q)_t^{j|i} = \bar{\alpha}_0^{j|i} + (\bar{\alpha}_1^{j|i} - \bar{\alpha}_0^{j|i})\hat{S}_t^i + M_{t-1}\bar{\gamma}^{j|i} + \bar{\beta}_0^{j|i}R_t^i(1-\hat{S}_t^i) + \bar{\beta}_1^{j|i}R_t^i\hat{S}_t^i + \bar{\sigma}^{j|i}\epsilon_t^{j|i},$$

where $\epsilon_t^{j|i} \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $\hat{S}_t^i$ is index i's inferred state at time t.

Estimates are obtained by employing the technique of quantile regressions (Koenker and Bassett [33]). We are especially interested in the systemic risk exposure. Figure 11 depicts the estimated $V\tilde{a}R(0.05)_t^{j|i}$ (green line) and the estimated $V\bar{a}R(0.05)_t^{j|i}$ (blue line) for j = BAC, JPM and C and i = Bank Index.

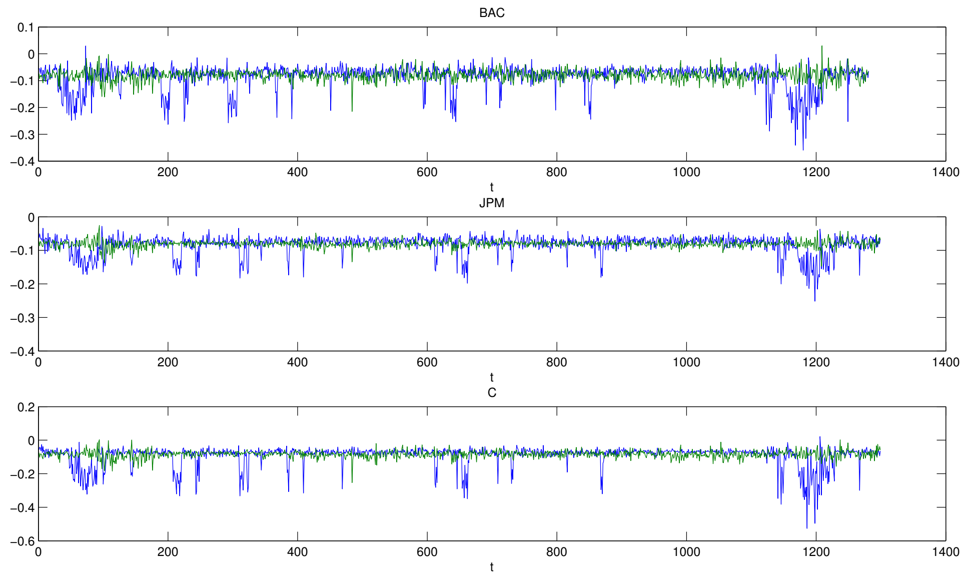

Figure 11: **VaR - time series**
This figure depicts the estimated "unconditional VaR" (green line) and the estimated "conditional VaR" (blue line) for j = BAC, JPM and C and i = Bank Index.

We observe that especially during the recent financial crisis, the estimated "conditional VaR" is larger in absolute value than the estimated "unconditional VaR". Expressed differently, conditional on the system being under financial distress, the "unconditional VaR" does not appropriately reflect the "conditional VaR". During times of turmoil, the "unconditional VaR" may not be the right device for assessing the institution's risk taking behavior.

Analogous to Δ CoVaR (Adrian and Brunnermeier [2]), we can also define the difference

$$\Delta VaR(q)_t^{j|i} = V\bar{a}R(q)_t^{j|i} - V\tilde{a}R(q)_t^{j|i}.$$

Figure 12 depicts the histogram of $\Delta VaR(q)_t^{j|i}$ for a few BHCs.

The (left) tail thickness differ somewhat among the BHCs. We observe that C, WFC and MS feature relatively thick left tails. That is, conditional on the system being under financial distress, they have a relatively large exposure to systemic risk.

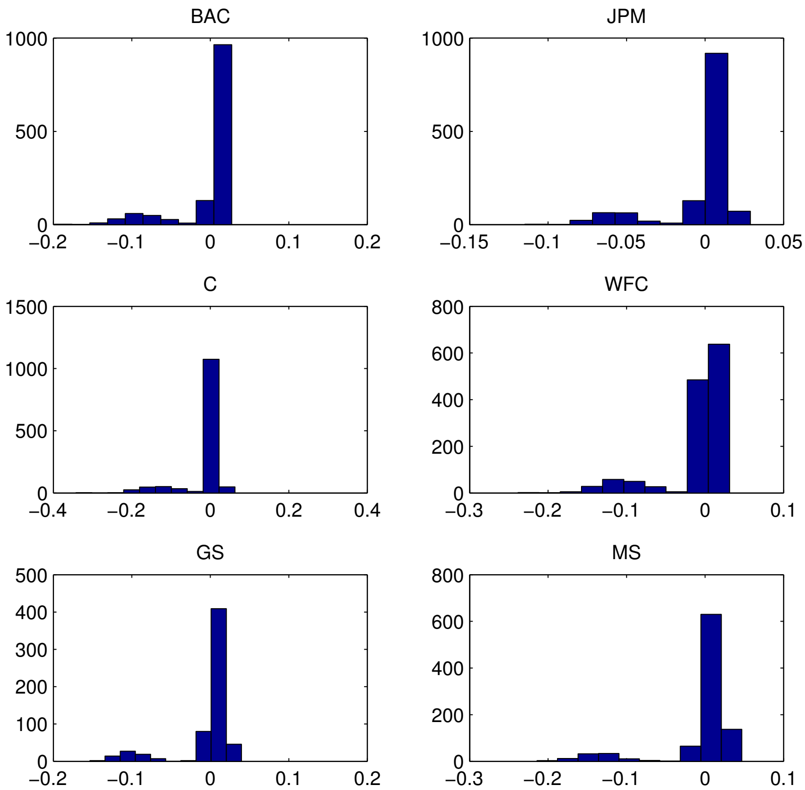

Figure 12: **VaR - histograms**
This figure depicts the histogram of $\Delta VaR(q)_t^{j|i}$ for j = BAC, JPM, C, WFC, GS and MS and i = Bank Index.

5.3.4 Measure of regime co-movement

In Table 18, we report the measure of regime co-movement. This measure can be interpreted as a correlation coefficient in the sense that a reported measure of one represents perfect positive correlation and a reported measure of minus one represents perfect negative correlation. The reported measures range from 0.2509 to 0.8245. We observe that the regime co-movement between the Bank Index and MS, MET and HBC is relatively large, while the regime co-movement between the Bank Index and BK is relatively small. As noted by Billio et al. [7], under the assumption of positive dependence, one should be aware of the fact that the reported measure is a conservative measure of the "true" regime co-movement. The reported measure is an average measure, but we can also define the time-varying measure

$$\hat{\Phi}_t^{i,j} = 1 - 2[\mathbb{P}\{S_t^i = 0|\mathcal{R}_T^i; \hat{\theta}^i\}\mathbb{P}\{S_t^j = 1|\mathcal{R}_T^j; \hat{\theta}^j\} + \mathbb{P}\{S_t^i = 1|\mathcal{R}_T^i; \hat{\theta}^i\}\mathbb{P}\{S_t^j = 0|\mathcal{R}_T^j; \hat{\theta}^j\}].$$

Figure 13 depicts $\hat{\Phi}_t^{i,j}$ between Bank of America and the Bank Index. The red ellipses indicate the periods where Bank of America and the Bank Index are tightly coupled.

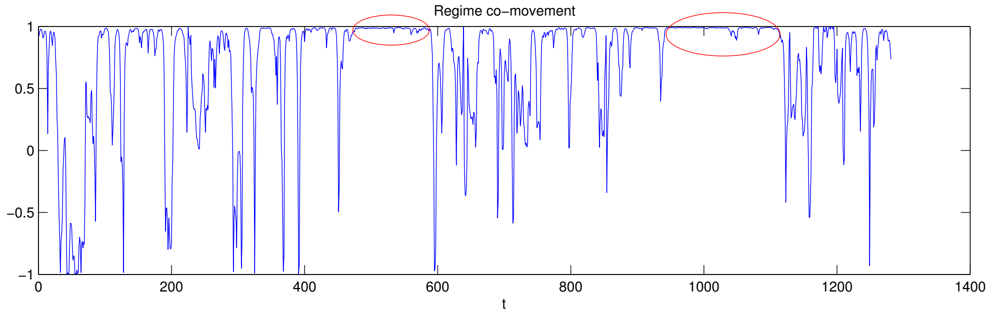

Figure 13: **Regime co-movement**
This figure depicts the regime co-movement between Bank of America and the Bank Index.

We can also focus on the regime 1 co-movement. Specifically, the following measure represents the "correlation" between $S_t^i = 1$ and $S_t^j = 1$.

$$\hat{\Phi}_t^{i,j} - 2\mathbb{P}\{S_t^i = 0|\mathcal{R}_T^i; \hat{\theta}^i\}\mathbb{P}\{S_t^j = 0|\mathcal{R}_T^j; \hat{\theta}^j\}.$$

Figure 14 depicts the regime 1 co-movement between Bank of America and the Bank Index. We observe that during the recent financial crisis, the regime 1 co-movement is almost equal to one.

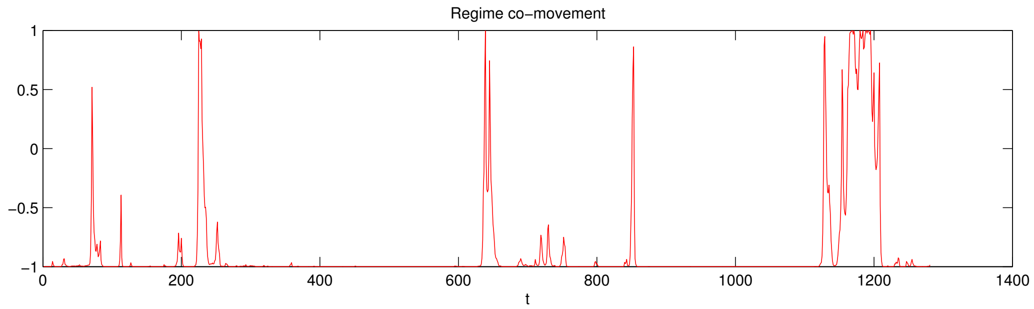

Figure 14: **Regime 1 co-movement**
This figure depicts the regime 1 co-movement between Bank of America and the Bank Index.

5.3.5 Time-varying transition probabilities

We examine the model specification of Section 4 with $Z_{t-1}^i = (1, \hat{S}_{t-1}^i)$, where $\hat{S}_{t-1}^i$ is index i's inferred state at time $t-1$. Parameter estimates can be found in Table 19. MET and C have a relatively large impact on the transition probabilities of the Bank Index.

Figure 15 depicts $\mathbb{P}\{S_t^j = 1 | S_{t-1}^j = 0, Z_{t-1}^i\}$.

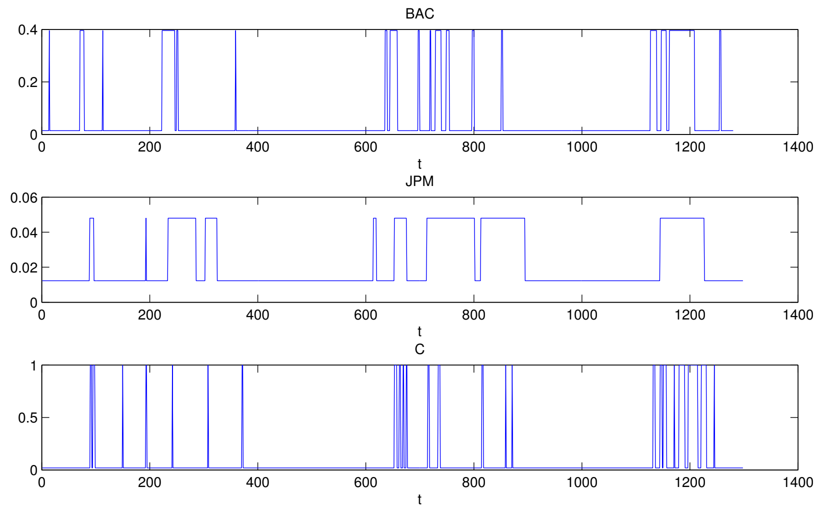

Figure 15: **Time-varying transition probabilities Bank Index**
This figure depicts the time-varying transition probability $\mathbb{P}\{S_t^j = 1 | S_{t-1}^j = 0, Z_{t-1}^i\}$. We have j = Bank Index and i = BHC.

The probability of moving from state 0 to state 1 is larger during a period of turmoil (i.e. $\hat{S}_{t-1}^i = 1$). We observe that C has a large effect on the transition probability, while JPM has a negligible effect on the transition probability. This observation seems consistent with the estimated change in correlation during times of turmoil (cf. Figure 10). Figure 16 depicts $\mathbb{P}\{S_t^j = 1 | S_{t-1}^j = 1, Z_{t-1}^i\}$.

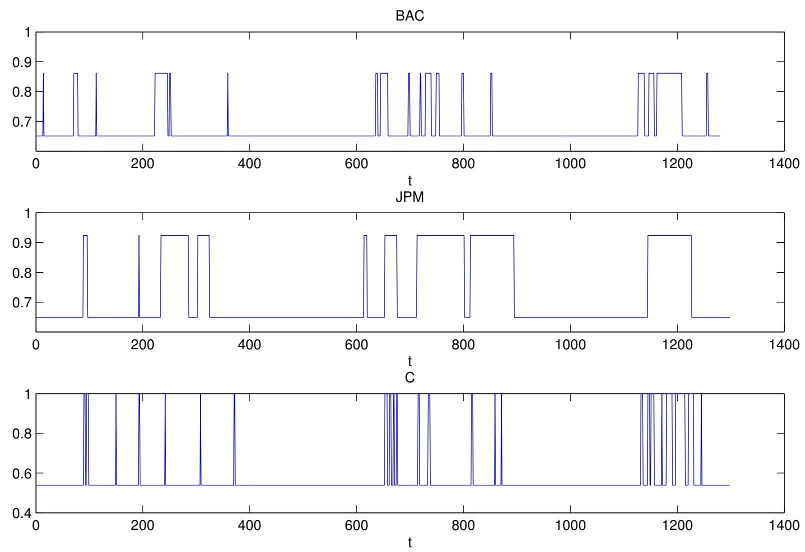

Figure 16: **Time-varying transition probabilities Bank Index**
This figure depicts the time-varying transition probability $\mathbb{P}\{S_t^j = 1 | S_{t-1}^j = 1, Z_{t-1}^i\}$. We have j = Bank Index and i = BHC.

6 Concluding remarks

We have introduced a classifier that infers index i's regimes on the basis of a Markov switching regression specification. Conditional on index i being in a particular regime, we have estimated index j's conditional mean as an affine function of macro-economic state variables and index i's log returns. Due to the prediction error associated with index i's inferred states, OLS estimates are generally inconsistent and biased. In Section 3, we have developed an alternative consistent estimation procedure. We have also briefly considered a Markov switching regression model with time-varying transition probabilities.

The empirical study shows that correlations of returns, in general, increase during times of market turmoil. The results are mixed regarding the change in the fraction of shocks that are on average transmitted form index i to index j. We have also demonstrated a possible implication for supervision. Specifically, we have modeled the conditional quantile instead of the conditional mean. The results indicate that "unconditional Value-at-Risk" might not be the right device for assessing risk taking behavior during times of turmoil.

In future research we intend to incorporate firm-specific characteristics into the different model specifications. We also intend to extend the Markov switching regression specification with time-varying transition probabilities.

References

[1] Acharya, V., L.H. Pedersen, T. Philippon and M. P. Richardson (2010), "Measuring Systemic Risk", AFA 2011 Denver Meetings Paper.

[2] Adrian, T. and M. Brunnermeier (2010). "CoVaR", working paper, Federal Reserve Bank of New York

[3] Ang, A. and G. Bekaert (2002). "International Asset Allocation With Regime Shifts", The Review of Financial Studies 15, 4, 1137-1187.

[4] Ang, A. and G. Bekaert (2002). "Regime Switches in Interest Rates", Journal of Business and Economic Statistics, 20, 2, 163-182.

[5] Barberis, N., A. Shleifer and J. Wurgler (2002). "Comovement", NBERWorking Paper No. 8884.

[6] Bekaert, G. and R.J. Hodrick (1993). "On biases in the measurement of foreign exchange risk premiums", Journal of International Money and Finance 12, 115-138.

[7] Billio, M., M. Getmansky, A. Lo and L. Pelizzon (2010). "Measuring Systemic Risk in the Finance and Insurance Sectors", MIT Working Paper.

[8] Cai, J. (1994). "A markov model of unconditional variance in ARCH", Journal of Business and Economic Statistics 12, 3, 309-316.

[9] Calvo, S. and C. Reinhart (1996). "Capital Flows to Latin America: Is There Evidence of Contagion Effects?", in: *Private Capital Flows to Emerging Markets*, Guillermo Calvo, Morris Goldstein, and Eduard Hochreiter, eds., (Washington, DC: Institute for International Economics).

[10] De Bandt, O. and P. Hartmann (2000). "Systemic Risk: A Survey", European Central Bank Working Paper 35.

[11] Dempster, A. P., N.M. Laird and D.B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society, ser. B, 39, 1-38.

[12] Diamond, D. W. and P. H. Dybvig (1983). "Bank Runs, Deposit Insurance, and Liquidity", Journal of Political Economy, 91 (3), 401-419.

[13] Diebold, F.X., J.H. Lee and G. Weinbach (1994). "Regime Switching with Time-Varying Transition Probabilities", in C. Hargreaves (ed.), Nonstationary Time Series Analysis and Cointegration. (Advanced Texts in Econometrics, C.W.J. Granger and G. Mizon, eds.), 283-302. Oxford: Oxford University Press.

[14] Duffie, D. (2010). "How Big Banks Fail and What to Do about It", ebook.

[15] Engel, C. and J.D. Hamilton (1990). "Long swings in the dollar: are they in the data and do markets know it?" American Economic Review 80, 4, 689-713.

[16] Engle, F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation", Econometrica 50: 987-1008, 1982.

[17] Filardo A.J. (1994). "Business-cycle phases and their transitional dynamics." Journal of Business and Economic Statistics 12: 299-308.

[18] Forbes, K. and R. Rigobon (2001). "No Contagion, Only Interdependence: Measuring Stock Market Co-movements", Journal of Finance 57, 5, 2223-2261.

[19] Friedman, M. and A.J.A. Schwartz (1963). "Monetary History of the United States, 1867-1960". Princeton, N.J.: Princeton Univ. Press.

[20] Goldfeld, S. M. and R. M. Quandt (1973). "A Markov model for switching regressions", Journal of Econometrics 1, 3-16.

[21] Gray, S.F. (1996). "Modeling the conditional distribution of interest rates as a regime switching process", Journal of Financial Economics 42, 27-62.

[22] Hamilton, J. D. (1989). "A new approach to the economic analysis of non-stationary time series and the business cycle", Econometrica 57, 357-384.

[23] Hamilton, J. D. (1990). "Analysis of Time Series Subject to Changes in Regime", Journal of Econometrics.

[24] Hathaway, R. J. (1985). "A constrained formulation of maximun likelihood estimation for normal mixture distributions". Ann. Statist. 13, 795-800.

[25] Hathaway, R.J. (1986). "A constrained EM algorithm for univariate normal mixtures". J. Statist. Comput. Simulation 23, 211-230.

[26] Hautsch, N., J. Schaumburg and M. Schienle (2011). "Quantifying Time-Varying Marginal Systemic Risk Contributions", Working paper.

[27] Huang, X., H. Zhou and H. Zhu (2009). "A framework for assessing the systemic risk of major financial institutions", Journal of Banking and Finance 33: 2036-2049.

[28] Hull, J., and A. White (2004). "Valuation of a CDO and an n-th to default CDS without Monte Carlo simulation", Journal of Derivatives 12, 8-23.

[29] Jondeau, E. and M. Rockinger (2003). "Conditional Volatility, Skewness and Kurtosis: Existence, Persistence and Co-movements", Journal of Economic Dynamics and Control 27(10), 1699-1737.

[30] Kiefer, N. M. (1978). "Discrete parameter variation: Efficient estimation of a switching regression model". Econometrica 6, 427-434.

[31] Kiefer, J. and J. Wolfowitz (1956). "Consistency of the maximum-likelihood estimator in the presence of infinitely many incidental parameters". Ann. Math. Statist. 27 888-906.

[32] King, M. and S. Wadhwani (1990). "Transmission of Volatility Between Stock Markets", Review of Financial Studies.

[33] Koenker, R. and G. Bassett Jr. (1987). "Regression Quantiles", Econometrica, Vol. 46, No. 1. (Jan., 1978), pp. 33-50.

[34] Kraus, A. and R. Litzenberger (1976). "Skewness preference and the valuation of risk assets", Journal of Finance 31(4):1085-1100.

[35] Kritzman, M., Y. Li, S. Page and R. Rigobon (2010), "Principal Components as a Measure of Systemic Risk", MIT Sloan Research Paper No. 4785-10.

[36] Mandelbrot, B.B. (1963). "The variation of certain speculative prices", Journal of Business XXXVI, 392-417.

[37] Owyang M.T., J. Piger and H. J. Wall (2005). "Business Cycle Phases in U.S. States". The Review of Economics and Statistics, Vol. 87, No. 4, pp. 604-616.

[38] Perlin, M. (2010), "MS Regress - The MATLAB Package for Markov Regime Switching Models". Available at SSRN: http://ssrn.com/abstract=1714016

[39] Quandt, R. E. (1958). "The estimation of parameters of linear regression system obeying two separate regimes", Journal of the American Statistical Association 55, 873-880.

[40] Richardson, M. and T. Smith (1994). "A Direct Test of the Mixture of Distributions Hypothesis: Measuring the Daily Flow of Information", Journal of Financial and Quantitative Analysis, 101-116.

[41] Roll, R. (1988). "R^2", Journal of Finance, 43: 541-566.

[42] Rubinstein, M. (1973). "The fundamental theorem of parameter-preference security valuation", Journal of Financial and Quantitative Analysis 8(1):61-69.

[43] Schumpeter, J. (1934), "The Theory of Economic Development", Cambridge, Mass.: Harvard University Press.

[44] Schwert, G. (1989). "Why Does Stock Market Volatility Change Over Time?", Journal of Finance 44, 1115-1153.

[45] Singleton, J. and J. Wingender (1986). "Skewness Persistence in Common Stock Returns", Journal of Financial and Quantitative Analysis 13, 335-341.

[46] Timmermann A. (2000). "Moments of Markov switching models". Journal of Econometrics, 96(1), 75-111.

[47] White H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", Econometrica 48 (4): 817-838.

[48] Yeh, J. (2006). "Real analysis: theory of measure and integration, 2nd ed.", World Scientific Publishing Co.

A Appendix

A.1 Correlation coefficient

Assume that R_t^i is described by a Markov switching regression model, i.e.

$$R_t^i = \begin{cases} \alpha_0^i + M_{t-1}\gamma^i + \sigma_0^i \epsilon_t^i & \text{if } S_t^i = 0 \\ \alpha_1^i + M_{t-1}\gamma^i + \sigma_1^i \epsilon_t^i & \text{if } S_t^i = 1 \end{cases} \qquad (t = 1, ..., T),$$

where $\epsilon_t^i \overset{\text{i.i.d.}}{\sim} N(0,1)$ and S_t^i is governed by a first-order Markov chain.

Suppose that R_t^j is described by the following linear regression model

$$R_t^j = \begin{cases} \alpha_0^{j|i} + M_{t-1}\gamma^{j|i} + \beta_0^{j|i} R_t^i + \sigma^{j|i}\epsilon_t^{j|i} & \text{if } S_t^i = 0 \\ \alpha_1^{j|i} + M_{t-1}\gamma^{j|i} + \beta_1^{j|i} R_t^i + \sigma^{j|i}\epsilon_t^{j|i} & \text{if } S_t^i = 1 \end{cases} \qquad (t = 1, ..., T),$$

where $\epsilon_t^{j|i} \overset{\text{i.i.d.}}{\sim} N(0,1)$

Assume the following

(A1) The usual OLS assumptions hold.

(A2) $\mathbb{E}\{\epsilon_t^i \epsilon_t^{j|i}\} = 0$.

We derive explicit expressions for the correlation coefficient under regime 0 and regime 1.

The correlation coefficient between index i and index j during state s is defined as follows

$$\rho_{(s)}^{j|i} = \frac{\mathbb{C}_{(s)}(R_t^i, R_t^j)}{[\mathbb{C}_{(s)}(R_t^i, R_t^i)\mathbb{C}_{(s)}(R_t^j, R_t^j)]^{\frac{1}{2}}}, \tag{A.1}$$

where $\mathbb{C}_{(s)}(\cdot,\cdot)$ denotes the covariance.

Then $\mathbb{C}_{(s)}(R_t^i, R_t^i)$ and $\mathbb{C}_{(s)}(R_t^j, R_t^j)$ are described by

$$\begin{aligned} \mathbb{C}_{(s)}(R_t^i, R_t^i) &= \gamma^{i\prime}\Sigma_{(s)}^M \gamma^i + (\sigma_s^i)^2 \\ \mathbb{C}_{(s)}(R_t^j, R_t^j) &= \gamma^{j|i\prime}\Sigma_{(s)}^M \gamma^{j|i} + (\beta_s^{j|i})^2(\gamma^{i\prime}\Sigma_{(s)}^M\gamma^i + (\sigma_s^i))^2 + \ldots \ , \\ & \quad 2\beta_s^{j|i}\gamma^{j|i\prime}\Sigma_{(s)}^M\gamma^i + (\sigma^{j|i})^2 \end{aligned}$$

where $\Sigma_{(s)}^M$ represents the variance-covariance matrix of M under regime s.

The covariance $\mathbb{C}_{(s)}(R_t^i, R_t^j)$ can be written as follows

$$\begin{aligned}\mathbb{C}_{(s)}(R_t^i, R_t^j) &= \mathbb{C}_{(s)}(M_{t-1}\gamma^i + \sigma_s^i\epsilon_t^i, M_{t-1}\gamma^{j|i} + \beta_s^{j|i}(M_{t-1}\gamma^i + \sigma_s^i\epsilon_t^i) + \sigma^{j|i}\epsilon_t^{j|i}) \\ &= \gamma^{i'}\Sigma_{(s)}^M\gamma^{j|i} + \beta_s^{j|i}\gamma^{i'}\Sigma_{(s)}^M\gamma^i + \beta_s^{j|i}(\sigma_s^i)^2.\end{aligned}$$

The correlation coefficient $\rho_{(s)}^{j|i}$ can then be expressed in terms of $\Sigma_{(s)}^M$ and the parameter coefficients in state s using (A.1). However, the states are not observed. We estimate S_t^i by (2.2) with $\tau_0 = \tau_1 = \frac{1}{2}$. The sample estimate of $\Sigma_{(s)}^M$ is specified as follows

$$\hat{\Sigma}_{(s)}^M = \frac{1}{|\mathcal{T}_s|}\sum_{t\in\mathcal{T}_s} M_{t-1}^{'}M_{t-1},$$

where $|\mathcal{T}_s|$ denotes the cardinality of $\mathcal{T}_s$.

The estimated change in correlation is then given by

$$\Delta\hat{\rho}^{j|i} = \hat{\rho}_{(1)}^{j|i} - \hat{\rho}_{(0)}^{j|i},$$

where $\hat{\rho}_{(s)}^{j|i} = \rho^{j|i}(\hat{\gamma}^i, \hat{\sigma}^{j|i}, \hat{\gamma}^{j|i}, \hat{\beta}_s^{j|i}, \hat{\Sigma}_{(s)}^M, \hat{\sigma}_s^i)$.

We propose the following decomposition

$$\begin{aligned}\Delta\hat{\rho}^{j|i} \quad &= \quad \rho^{j|i}(\hat{\gamma}^i, \hat{\sigma}^{j|i}, \hat{\gamma}^{j|i}, \hat{\beta}_1^{j|i}, \hat{\Sigma}_{(1)}^M, \hat{\sigma}_1^i) - \rho^{j|i}(\hat{\gamma}^i, \hat{\sigma}^{j|i}, \hat{\gamma}^{j|i}, \hat{\beta}_0^{j|i}, \hat{\Sigma}_{(1)}^M, \hat{\sigma}_1^i) + ... \\ &\quad\quad \rho^{j|i}(\hat{\gamma}^i, \hat{\sigma}^{j|i}, \hat{\gamma}^{j|i}, \hat{\beta}_0^{j|i}, \hat{\Sigma}_{(1)}^M, \hat{\sigma}_1^i) - \rho^{j|i}(\hat{\gamma}^i, \hat{\sigma}^{j|i}, \hat{\gamma}^{j|i}, \hat{\beta}_0^{j|i}, \hat{\Sigma}_{(0)}^M, \hat{\sigma}_1^i) + ... \\ &\quad\quad \rho^{j|i}(\hat{\gamma}^i, \hat{\sigma}^{j|i}, \hat{\gamma}^{j|i}, \hat{\beta}_0^{j|i}, \hat{\Sigma}_{(0)}^M, \hat{\sigma}_1^i) - \rho^{j|i}(\hat{\gamma}^i, \hat{\sigma}^{j|i}, \hat{\gamma}^{j|i}, \hat{\beta}_0^{j|i}, \hat{\Sigma}_{(0)}^M, \hat{\sigma}_0^i) \\ &= \quad \Delta\hat{\rho}_{\Delta\beta^{j|i}}^{j|i} + \Delta\hat{\rho}_{\Delta\Sigma^M}^{j|i} + \Delta\hat{\rho}_{\Delta\sigma^i}^{j|i}.\end{aligned}$$

A.2 Moments of a Markov switching regression model

The vector of steady state probabilities π solves the following equation

$$P\pi = \pi.$$

Assuming that the steady state probabilities apply, Timmermann [46] has shown that the central moments of a Markov switching regression model can be written as follows (use the Binomial Theorem)

$$\begin{aligned}
\mathbb{E}\{(R^i - \mathbb{E}\{R^i\})^n\} &= \textstyle\sum_{s=0}^{1} \pi_s \mathbb{E}\{(\alpha_s^i + M\gamma^i + \sigma_s^i \epsilon^i - \mathbb{E}\{R^i\})^n\} \\
&= \textstyle\sum_{s=0}^{1} \pi_s \sum_{k=0}^{n} \binom{n}{k} (\sigma_s^i)^k \mathbb{E}\{(\alpha_s^i + M\gamma^i - \mathbb{E}\{R^i\})^{n-k} (\epsilon^i)^k\} \\
&= \textstyle\sum_{s=0}^{1} \pi_s \sum_{k=0}^{n} \binom{n}{k} (\sigma_s^i)^k \mathbb{E}\{(\pi_{1-s}(\alpha_s^i - \alpha_{(1-s)}^i))^{n-k} (\epsilon^i)^k\} \\
&= \textstyle\sum_{s=0}^{1} \pi_s \sum_{k=0}^{n} \binom{n}{k} (\sigma_s^i)^k (\pi_{1-s}(\alpha_s^i - \alpha_{(1-s)}^i))^{n-k} \mathbb{E}\{(\epsilon^i)^k\}.
\end{aligned}$$

He has also derived explicit formulas for the variance, skewness and excess kurtosis. Specifically,

$$\begin{aligned}
\mathbb{E}\{(R^i - \mathbb{E}\{R^i\})^2\} &= \pi_0(\sigma_0^i)^2 + \pi_1(\sigma_1^i)^2 + \pi_0\pi_1(\alpha_0^i - \alpha_1^i)^2 \\
\mathbb{E}\{(R^i - \mathbb{E}\{R^i\})^3\} &= \pi_0\pi_1(\alpha_0^i - \alpha_1^i)\{3[(\sigma_0^i)^2 - (\sigma_1^i)^2] + (1 - 2\pi_0^2)(\alpha_0^i - \alpha_1^i)^2\} \\
\mathbb{E}\{(R^i - \mathbb{E}\{R^i\})^4\} &= \pi_1\{3(\sigma_1^i)^4 + \pi_0^4(\alpha_0^i - \alpha_1^i)^4 + 6\pi_0^2(\sigma_1^i)^2(\alpha_0^i - \alpha_1^i)^2\} + ... \\
&\quad \pi_0\{3(\sigma_0^i)^4 + \pi_1^4(\alpha_0^i - \alpha_1^i)^4 + 6\pi_1^2(\sigma_0^i)^2(\alpha_0^i - \alpha_1^i)^2\}.
\end{aligned}$$

The variance, skewness and excess kurtosis are functions of the second, third and fourth central moment.

A.3 Tables

Code	**BHC**
BAC	BANK OF AMERICA CORPORATION
JPM	JPMORGAN CHASE & CO.
C	CITIGROUP INC.
WFC	WELLS FARGO & COMPANY
GS	GOLDMAN SACHS GROUP, INC., THE
MS	MORGAN STANLEY
MET	METLIFE, INC.
HBC	HSBC NORTH AMERICA HOLDINGS INC.
PNC	PNC FINANCIAL SERVICES GROUP, INC., THE
BK	BANK OF NEW YORK MELLON CORP., THE
STI	SUNTRUST BANKS, INC.
STT	STATE STREET CORPORATION
BBT	BB&T CORPORATION
AXP	AMERICAN EXPRESS COMPANY
RF	REGIONS FINANCIAL CORPORATION
FITB	FIFTH THIRD BANCORP
KEY	KEYCORP
NTRS	NORTHERN TRUST CORPORATION
MTB	M&T BANK CORPORATION
HTBAN	HUNTINGTON BANCSHARES INCORPORATED
ZION	ZIONS BANCORPORATION
MIC	MARSHALL & ILSLEY CORPORATION
BPOP	POPULAR, INC.
SNV	SYNOVUS FINANCIAL CORP.
FHN	FIRST HORIZON NATIONAL CORPORATION

Table 7: **List of Banking Holding Companies**
This table reports a list of bank holding companies and their abbreviations.

					Quantile			Autocorrelation		
BHC	T	Mean	Std.	Skew.	5%	50%	95%	ρ_1	ρ_2	ρ_3
VXO	1304	21.375	9.128	2.296	11.204	19.920	37.482	0.916	0.872	0.825
TBR	1304	4.070	2.203	−0.224	0.140	4.670	7.703	0.996	0.993	0.989
YSP	1304	2.221	1.395	−0.004	0.060	2.080	4.303	0.994	0.987	0.981
CSP	1304	1.869	0.674	2.474	1.240	1.690	2.783	0.984	0.976	0.967
S&P	1304	826.248	416.551	−0.006	251.958	895.775	1441.393	0.997	0.995	0.992
Bank Index	1304	0.001	0.031	−0.566	−0.047	0.003	0.046	−0.023	−0.013	0.067

Table 8: **Summary statistics fundamentals and Bank Index**
This table reports summary statistics for the fundamentals and the return series of the Bank Index.

BHC	T	Mean	Std.	Skew.	Quantile			Autocorrelation		
					5%	50%	95%	ρ_1	ρ_2	ρ_3
BAC	1282	0.001	0.059	−0.317	−0.079	0.003	0.075	−0.035	0.047	−0.089
JPM	1303	0.002	0.054	−0.187	−0.082	0.003	0.084	−0.087	0.035	−0.037
C	1303	0.001	0.070	−1.520	−0.078	0.001	0.081	−0.075	0.028	−0.082
WFC	1303	0.003	0.049	0.245	−0.061	0.003	0.068	−0.158	0.048	−0.107
GS	607	0.002	0.058	0.079	−0.081	0.004	0.083	−0.156	0.069	0.022
MS	930	0.002	0.072	−1.003	−0.087	0.001	0.097	−0.236	0.054	0.010
MET	548	0.002	0.062	−0.347	−0.069	0.002	0.079	−0.044	−0.068	−0.069
HBC	597	0.001	0.041	−1.275	−0.056	0.003	0.063	−0.048	−0.000	0.004
PNC	1163	0.002	0.049	−0.149	−0.066	0.002	0.075	−0.155	0.045	−0.001
BK	1303	0.002	0.046	0.030	−0.068	0.002	0.076	−0.094	0.012	−0.027
STI	1199	0.002	0.054	−0.050	−0.067	0.003	0.067	−0.181	0.078	0.004
STT	1276	0.002	0.053	−1.254	−0.065	0.002	0.080	−0.065	−0.048	−0.030
BBT	1082	0.002	0.042	0.210	−0.063	0.001	0.074	−0.016	−0.013	−0.035
AXP	1303	0.002	0.048	−0.024	−0.073	0.001	0.072	−0.373	0.027	−0.049
RF	1082	0.001	0.058	0.795	−0.072	0.002	0.063	−0.160	0.063	0.036
FITB	1082	0.002	0.071	0.414	−0.069	0.001	0.080	−0.137	0.064	0.112
KEY	1207	0.001	0.055	−1.198	−0.069	0.003	0.069	−0.168	0.012	−0.074
NTRS	1082	0.003	0.042	−0.042	−0.062	0.003	0.067	−0.108	0.023	−0.021
MTB	1001	0.003	0.038	0.144	−0.054	0.002	0.058	−0.028	−0.042	−0.043
HTBAN	1082	0.001	0.063	1.392	−0.077	0.001	0.070	−0.105	0.066	0.070
ZION	1082	0.002	0.059	0.819	−0.080	0.002	0.076	−0.067	−0.034	0.007
MIC	1082	0.001	0.058	0.505	−0.077	0.002	0.071	−0.058	0.016	0.024
BPOP	1082	0.001	0.055	−0.056	−0.072	0.000	0.069	0.042	0.022	0.085
SNV	1093	0.001	0.058	−0.363	−0.075	0.000	0.079	−0.130	−0.002	0.049
FHN	1082	0.002	0.050	0.143	−0.070	0.002	0.071	−0.190	0.063	−0.069

Table 9: **Summary statistics**
This table reports summary statistics for the return series of various BHCs.

	Parameter estimates in regime 0					Parameter estimates in regime 1					Δ Prob.	
BHC	VIX	TBR	YSP	CSP	S&P	VIX	TBR	YSP	CSP	S&P	Mean	Std.
BAC	0.001	−0.027	−0.019	−0.027	−0.012	0.002	−0.051	−0.071	0.082	0.466	0.005	0.014
	(0.001)	(0.014)	(0.012)	(0.013)	(0.079)	(0.002)	(0.088)	(0.082)	(0.072)	(0.375)		
JPM	−0.000	−0.006	−0.005	−0.011	−0.170	0.001	−0.031	−0.010	0.065	0.323	0.006	0.019
	(0.001)	(0.015)	(0.013)	(0.014)	(0.096)	(0.001)	(0.048)	(0.050)	(0.041)	(0.206)		
C	0.002	−0.039	−0.022	−0.051	0.004	0.001	−0.085	−0.189	0.184	1.077	0.005	0.019
	(0.001)	(0.016)	(0.014)	(0.015)	(0.043)	(0.003)	(0.130)	(0.151)	(0.111)	(0.603)		
WFC	0.000	−0.025	−0.005	0.013	−0.068	0.000	0.015	−0.023	0.067	−0.160	0.001	0.004
	(0.000)	(0.013)	(0.012)	(0.012)	(0.067)	(0.003)	(0.158)	(0.160)	(0.089)	(0.545)		
GS	0.000	−0.027	−0.002	−0.011	0.079	0.004	−0.013	−0.085	−0.045	−0.037	0.008	0.023
	(0.001)	(0.031)	(0.024)	(0.018)	(0.132)	(0.004)	(0.086)	(0.073)	(0.060)	(0.550)		
MS	0.001	0.018	0.018	0.003	−0.054	0.004	0.054	−0.011	0.029	−0.227	0.004	0.016
	(0.001)	(0.024)	(0.021)	(0.026)	(0.138)	(0.008)	(0.270)	(0.300)	(0.160)	(1.221)		
MET	0.001	−0.030	−0.004	−0.015	−0.124	0.001	−0.084	−0.020	0.099	0.638	0.009	0.019
	(0.001)	(0.021)	(0.018)	(0.014)	(0.127)	(0.004)	(0.151)	(0.117)	(0.085)	(0.645)		
HBC	0.001	−0.022	−0.018	−0.026	0.110	0.006	−0.191	−0.065	−0.002	1.105	0.007	0.031
	(0.001)	(0.016)	(0.014)	(0.012)	(0.078)	(0.003)	(0.077)	(0.067)	(0.025)	(0.346)		
PNC	0.001	−0.009	−0.003	−0.002	−0.075	−0.000	−0.002	0.023	0.092	0.145	0.008	0.020
	(0.001)	(0.012)	(0.011)	(0.012)	(0.087)	(0.002)	(Inf)	(0.035)	(0.039)	(0.288)		
BK	0.001	−0.026	−0.021	−0.022	−0.124	0.002	0.021	0.028	0.052	0.169	0.005	0.014
	(0.001)	(0.014)	(0.012)	(0.013)	(0.089)	(0.001)	(0.029)	(0.027)	(0.026)	(0.134)		

Table 10: **Model comparison**

This table reports parameter estimates of Markov switching regression models with switching coefficients. Standard errors are in parentheses. The last two columns provide information on the absolute deviation between the smoothed probabilities of a Markov switching regression model with switching coefficients and the smoothed probabilities of a Markov switching regression model with non-switching coefficients.

	j = Bank Index, i = BHC				j = BHC, i = Bank Index			
BHC	$\Delta\rho^{j\|i}$	$\Delta\rho^{j\|i}_{\Delta\beta^{j\|i}}$	$\Delta\rho^{j\|i}_{\Delta\Sigma^M}$	$\Delta\rho^{j\|i}_{\Delta\sigma^i}$	$\Delta\rho^{j\|i}$	$\Delta\rho^{j\|i}_{\Delta\beta^{j\|i}}$	$\Delta\rho^{j\|i}_{\Delta\Sigma^M}$	$\Delta\rho^{j\|i}_{\Delta\sigma^i}$
BAC	0.148	0.065	−0.094	0.177	0.214	0.081	−0.030	0.163
JPM	0.045	0.019	−0.037	0.062	0.083	−0.005	−0.012	0.100
C	0.216	0.118	−0.148	0.246	0.286	0.157	−0.029	0.158
WFC	0.064	−0.042	−0.147	0.254	0.150	0.039	−0.070	0.181
GS	0.096	0.032	−0.052	0.117	−0.009	−0.210	−0.137	0.338
MS	0.062	−0.018	−0.126	0.207	−0.117	−0.284	−0.146	0.313
MET	0.121	−0.007	−0.197	0.325	0.434	0.255	−0.074	0.254
HBC	0.138	0.070	−0.128	0.197	0.161	−0.063	−0.074	0.298
PNC	0.126	0.085	−0.061	0.102	0.159	0.060	−0.057	0.156
BK	0.021	0.025	−0.063	0.058	0.027	−0.090	−0.048	0.166

Table 11: **Estimated change in correlation**
This table reports the estimated change in the correlation coefficient.

	j = Bank Index, i = BHC		j = BHC, i = Bank Index	
BHC	Crisis	Non-crisis	Crisis	Non-crisis
BAC	0.1650	0.0615	0.0899	0.0629
JPM	0.1250	0.0087	0.1075	0.0420
C	0.2743	0.0406	0.0988	0.0586
WFC	-0.0547	0.0559	0.0567	0.0286
GS	0.0688	0.0931	0.1140	0.0770
MS	0.0827	0.0496	0.0617	0.0554
MET	0.2539	-0.0306	0.2349	0.0091
HBC	0.2775	0.1232	0.1596	0.1662
PNC	0.1018	-0.0074	0.0334	0.0199
BK	0.0752	0.0094	0.0745	0.0294

Table 12: **Spearman's rho**
This table reports Spearman's rank correlation coefficient.

	Model A		Model B	
BHC	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$
BAC	0.048	0.069	0.034	0.079
	(0.018)	(0.028)	(0.023)	(0.031)
JPM	0.029	0.037	0.009	0.036
	(0.021)	(0.027)	(0.027)	(0.030)
C	0.041	0.069	0.044	0.072
	(0.019)	(0.022)	(0.020)	(0.024)
WFC	0.075	0.059	0.106	0.059
	(0.022)	(0.041)	(0.026)	(0.044)
GS	0.062	0.077	0.071	0.093
	(0.024)	(0.038)	(0.023)	(0.042)
MS	0.043	0.038	0.036	0.039
	(0.015)	(0.032)	(0.017)	(0.035)
MET	0.080	0.078	0.065	0.073
	(0.032)	(0.036)	(0.034)	(0.038)
HBC	0.098	0.141	0.135	0.143
	(0.036)	(0.071)	(0.043)	(0.074)
PNC	0.034	0.071	0.002	0.075
	(0.022)	(0.038)	(0.029)	(0.044)
BK	0.034	0.048	0.022	0.039
	(0.023)	(0.031)	(0.026)	(0.033)
STI	0.074	0.054	0.074	0.051
	(0.026)	(0.032)	(0.027)	(0.036)
STT	0.026	0.124	0.034	0.130
	(0.018)	(0.048)	(0.023)	(0.057)
BBT	0.017	0.059	-0.033	0.062
	(0.029)	(0.036)	(0.044)	(0.040)
AXP	0.045	0.066	0.036	0.072
	(0.019)	(0.039)	(0.035)	(0.052)
RF	0.082	0.032	0.070	0.029
	(0.028)	(0.035)	(0.032)	(0.036)

Table 13: **Regression coefficients**
This table reports the estimated regression coefficients for two different model specifications. The robust standard errors are in parentheses. We have j = Bank index and i = BHC.

	Model A		Model B	
BHC	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$
FITB	0.052	0.069	0.041	0.074
	(0.021)	(0.035)	(0.020)	(0.035)
KEY	0.030	0.032	0.052	0.028
	(0.024)	(0.029)	(0.029)	(0.032)
NTRS	0.026	0.033	0.030	0.033
	(0.021)	(0.044)	(0.025)	(0.055)
MTB	0.059	0.016	0.084	0.025
	(0.037)	(0.054)	(0.044)	(0.059)
HTBAN	0.070	0.036	0.018	0.032
	(0.026)	(0.030)	(0.033)	(0.032)
ZION	-0.024	0.059	-0.026	0.057
	(0.027)	(0.027)	(0.031)	(0.030)
MIC	0.040	0.051	0.039	0.052
	(0.027)	(0.033)	(0.028)	(0.034)
BPOP	0.032	0.054	0.037	0.052
	(0.030)	(0.041)	(0.034)	(0.044)
SNV	0.071	0.042	0.030	0.038
	(0.025)	(0.029)	(0.024)	(0.031)
FHN	0.027	0.023	0.024	0.015
	(0.024)	(0.043)	(0.026)	(0.046)

Table 14: **Regression coefficients – continued from previous page**
This table reports the estimated regression coefficients for two different model specifications. The robust standard errors are in parentheses. We have j = Bank index and i = BHC.

	Model A		Model B	
BHC	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$
BAC	0.270	0.363	0.344	0.360
	(0.111)	(0.172)	(0.200)	(0.206)
JPM	0.159	0.154	0.221	0.096
	(0.098)	(0.145)	(0.149)	(0.176)
C	0.318	0.534	0.423	0.531
	(0.126)	(0.261)	(0.208)	(0.319)
WFC	0.212	0.247	0.402	0.245
	(0.083)	(0.147)	(0.126)	(0.184)
GS	0.472	0.289	0.512	0.323
	(0.174)	(0.155)	(0.276)	(0.178)
MS	0.543	0.210	0.588	0.158
	(0.276)	(0.200)	(0.462)	(0.245)
MET	0.296	0.526	0.303	0.586
	(0.178)	(0.243)	(0.211)	(0.251)
HBC	0.321	0.279	0.313	0.326
	(0.107)	(0.135)	(0.163)	(0.148)
PNC	0.203	0.254	0.281	0.357
	(0.100)	(0.166)	(0.169)	(0.195)
BK	0.191	0.115	0.312	0.130
	(0.081)	(0.111)	(0.120)	(0.138)
STI	0.285	0.276	0.157	0.322
	(0.105)	(0.190)	(0.180)	(0.235)
STT	0.164	0.436	0.162	0.519
	(0.088)	(0.219)	(0.152)	(0.277)
BBT	0.208	0.087	0.282	0.140
	(0.078)	(0.115)	(0.126)	(0.133)
AXP	0.213	0.183	0.224	0.177
	(0.087)	(0.115)	(0.146)	(0.139)
RF	0.294	0.187	0.341	0.291
	(0.113)	(0.231)	(0.202)	(0.282)

Table 15: **Regression coefficients**
This table reports the estimated regression coefficients for two different model specifications. The robust standard errors are in parentheses. We have j = BHC and i = Bank Index.

	Model A		Model B	
BHC	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$
FITB	0.368	0.651	0.365	0.885
	(0.140)	(0.373)	(0.226)	(0.448)
KEY	0.313	0.057	0.486	0.108
	(0.139)	(0.168)	(0.288)	(0.209)
NTRS	0.160	0.028	0.142	0.039
	(0.085)	(0.114)	(0.140)	(0.140)
MTB	0.042	0.061	0.041	0.125
	(0.080)	(0.140)	(0.130)	(0.176)
HTBAN	0.187	0.333	0.166	0.454
	(0.127)	(0.253)	(0.222)	(0.308)
ZION	0.189	0.318	0.273	0.426
	(0.107)	(0.217)	(0.169)	(0.264)
MIC	0.268	0.261	0.198	0.299
	(0.117)	(0.213)	(0.206)	(0.259)
BPOP	0.274	0.240	0.191	0.362
	(0.111)	(0.268)	(0.164)	(0.327)
SNV	0.319	0.230	0.492	0.303
	(0.120)	(0.170)	(0.210)	(0.213)
FHN	0.250	0.004	0.305	-0.027
	(0.119)	(0.170)	(0.215)	(0.204)

Table 16: **Regression coefficients – continued from previous page**
This table reports the estimated regression coefficients for two different model specifications. The robust standard errors are in parentheses. We have j = BHC and i = Bank Index.

	j = Bank Index, i = BHC		j = BHC, i = Bank Index	
BHC	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$	$\beta_0^{j\|i}$	$\beta_1^{j\|i}$
BAC	0.061	0.063	0.220	0.400
	(0.017)	(0.029)	(0.097)	(0.152)
JPM	0.034	0.035	0.146	0.165
	(0.016)	(0.026)	(0.095)	(0.138)
C	0.071	0.057	0.197	0.616
	(0.021)	(0.033)	(0.119)	(0.222)
WFC	0.080	0.056	0.303	0.190
	(0.019)	(0.037)	(0.072)	(0.124)
GS	0.063	0.076	0.487	0.288
	(0.018)	(0.032)	(0.154)	(0.226)
MS	0.049	0.034	0.570	0.199
	(0.016)	(0.033)	(0.129)	(0.218)
MET	0.074	0.079	0.218	0.572
	(0.178)	(0.243)	(0.157)	(0.390)
HBC	0.140	0.113	0.266	0.314
	(0.026)	(0.081)	(0.099)	(0.184)
PNC	0.051	0.063	0.144	0.299
	(0.014)	(0.033)	(0.081)	(0.117)
BK	0.044	0.044	0.230	0.098
	(0.023)	(0.069)	(0.076)	(0.108)

Table 17: **Regression coefficients**
This table reports the estimated regression coefficients for Model C. The bootstrap standard errors are in parentheses.

	BAC	JPM	C	WFC	GS	MS	MET	HBC	PNC	BK	BIX
BAC	1	0.6135	0.7788	0.7694	0.6732	0.7833	0.7422	0.7366	0.7189	0.4682	0.6214
JPM		1	0.5191	0.5321	0.5452	0.5904	0.4220	0.3319	0.5934	0.7203	0.4504
C			1	0.8165	0.5781	0.7834	0.7482	0.7896	0.6510	0.3702	0.6686
WFC				1	0.6961	0.7779	0.7493	0.7386	0.6773	0.3700	0.6665
GS					1	0.7196	0.6903	0.5590	0.5966	0.4419	0.6425
MS						1	0.7966	0.7540	0.6776	0.4495	0.7608
MET							1	0.8245	0.6761	0.3412	0.7947
HBC								1	0.6163	0.2509	0.7763
PNC									1	0.5161	0.6023
BK										1	0.2939
BIX											1

Table 18: **Measure of regime co-movement**

This table reports the measure of regime co-movement between index i and index j. BIX is the abbreviation of Bank Index.

	Parameter estimates $p_{00,t}$		Parameter estimates $p_{11,t}$	
BHC	Constant	$\hat{S}^i_{t-1}$	Constant	$\hat{S}^i_{t-1}$
BAC	4.2526	-3.8274	0.6234	1.2065
JPM	4.3864	-1.3988	0.6165	1.8864
C	3.8881	-26.4341	0.1552	20.5399
WFC	4.2081	-3.0762	0.8758	1.6596
GS	4.8809	7.3440	2.6581	7.1963
MS	4.7408	-3.0770	0.9409	1.7877
MET	4.9701	-14.9760	-21.8856	24.9520
HBC	4.7216	-5.8315	0.9826	4.1194
PNC	6.0647	-4.3701	-1.9696	3.7343
BK	4.0275	-0.0700	1.0179	2.1498

Table 19: **Time-varying transition probabilities**
This table reports the parameter estimates of a Markov switching regression model with time-varying transition probabilities. We have j = Bank Index and i = BHC.

		$\beta_1^{j\|i} - \beta_0^{j\|i}$			$\beta_1^{j\|i}$		
$\Delta\rho^{j\|i}$	$\Delta\rho^{j\|i}_{\Delta\beta^{j\|i}}$	Model A	Model B	Model C	Model A	Model B	Model C
C	C	HBC	PNC	GS	HBC	HBC	HBC
BAC	PNC	PNC	BAC	PNC	MET	GS	MET
HBC	HBC	C	C	MET	GS	BAC	GS
PNC	BAC	BAC	JPM	BAC	PNC	PNC	BAC
MET	GS	GS	GS	JPM	C	MET	PNC
GS	BK	BK	BK	BK	BAC	C	C
WFC	JPM	JPM	HBC	C	WFC	WFC	WFC
MS	MET	MET	MET	MS	BK	MS	BK
JPM	MS	MS	MS	WFC	MS	BK	JPM
BK	WFC	WFC	WFC	HBC	JPM	JPM	MS

Table 20: **Ranking — Systemic Risk Contribution**

		$\beta_1^{j\|i} - \beta_0^{j\|i}$			$\beta_1^{j\|i}$		
$\Delta\rho^{j\|i}$	$\Delta\rho^{j\|i}_{\Delta\beta^{j\|i}}$	Model A	Model B	Model C	Model A	Model B	Model C
MET	MET	BAC	BAC	C	C	MET	C
C	C	GS	GS	MET	MET	C	MET
BAC	BAC	MS	HBC	BAC	BAC	BAC	BAC
HBC	PNC	WFC	MS	PNC	GS	PNC	HBC
PNC	WFC	MET	MET	HBC	HBC	HBC	PNC
WFC	JPM	HBC	WFC	JPM	PNC	GS	GS
JPM	HBC	PNC	C	WFC	WFC	WFC	MS
BK	BK	C	PNC	BK	MS	MS	WFC
GS	GS	BK	BK	GS	JPM	BK	JPM
MS	MS	JPM	JPM	MS	BK	JPM	BK

Table 21: **Ranking – Systemic Risk Exposure**