Another Look at Stock Return Comovement: Some New Evidence and Test

Kaihua Deng

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington
2015

Reading Committee:

Eric Zivot, Chair

Yanqin Fan

Chang-Jin Kim

Program Authorized to Offer Degree:

Department of Economics

University of Washington

Abstract

Another Look at Stock Return Comovement: Some New Evidence and Test

Kaihua Deng

Chair of the Supervisory Committee:
Eric Zivot, Richards Professor of Economics
Department of Economics

The study of the comovement between asset returns reflects an ongoing effort by economists to understand investment risk in financial markets. Building on previous findings, in the current thesis I provide some new evidence on this topic with a focus on large-cap stocks and highlight an innovative way to evaluate the statistical significance of comovement asymmetry.

In the first part of the thesis, I revisit the question of how large-cap stock return comovement varies with volatility and market returns. I propose the use of an eigenvalue-based measure of comovement in a multivariate semi-Markov-switching framework. I conduct various model evaluation checks and compare the new results with that based on a benchmark. I estimate

models with two to four regimes and consider the impact of sample selection and outlier reduction. Contrary to the sweeping sentiment that comovement is highest when market is down and volatile, I illustrate the significance of comovement differential across states and find in most case studies evidence that suggests otherwise.

In the second part, I propose a test of asymmetric stock return comovement across states. The test can be viewed as a variation of Kendall's τ conditional on the state and has an asymptotic χ^2-distribution. A refined version of the test is derived based on the Markov chain theory of regenerative cycles which substantially improves finite sample size and power. I show that the test has power against local alternatives, which is nonetheless compromised due to a finite sample convergence bound put on the implied local alternative data generating process. I evaluate the new test against traditional correlation-based measures and demonstrate power attrition due to nuisance parameters when states are ignored. I find that asymmetric tail dependence becomes much less significant when considered state by state. A list of related tests is given as an extension at the end.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2

LIST OF TABLES

Chapter 2

Chapter 3

ACKNOWLEDGEMENTS

There are many people to whom I am indebted. First of all, I would like to thank my thesis advisor, Prof. Eric Zivot who provided constant academic support for my work and answered all my boring questions. I would also like to extend my thanks to Prof. Chang-Jin Kim and Prof. Ji-Hyung Lee whose insightful comments during my random excursions to their offices have always been a source of inspiration. I owe special thanks to Prof. Yanqin Fan who sets a high standard for me and young aspiring scholars. Moreover, I thank the Economics department as well as American Institute for Economic Research for their profound recognition and financial support.

Thanks are also due especially to Prof. Jon Wellner who introduced me to modern empirical process theories and inequalities, to Prof. Jon Wakefield who shared statistical programming skills with his students, and to Prof. Jarrad Harford who taught me corporate finance and replied to my emails in a timely manner.

I also wish to thank the department office staff, in particular Simon Reeve-Parker, Chris Fendrich, Nick Giese and Laura Pflum for their excellent assistance in all administrative matters. In addition, I appreciate the help from our senior computer specialists, Jonathan Mitchell and Will Stroud, for keeping the softwares and scanners up to date in the graduate lab.

On a more personal note, I am grateful to my parents for their endless support and understanding. Last but not least, I would very much like to thank Dr. Sam Kuang, Dr. Kuan Deng and Dr. Walker Todd for all the correspondences and encouragement along the way.

To my parents.

Chapter 1. INTRODUCTION

1.1 OBJECTIVES

In the field of asset allocation and risk management, it is often said that models become less useful when they matter the most in the face of market downturns or extreme events. Hence understanding the patterns of return comovement becomes even more crucial.

While conventional wisdom based on pairwise correlation measures suggests that stock return comovement is highest in bad times, recent authors have found that such asymmetry is often not significant for large-caps. Another problem with focusing on comovement *solely* is that some aspects of the data generating process may not be adequately captured, which in turn can make inference sensitive to volatile periods or abrupt changes in the market. In this respect, to say that stock returns become more synchronized without due respect to the time-varying nature of their data generating process seems to be an unfair account of all the hedging endeavors that have been going on for decades in the industry.

Against this background, three questions can be asked: Is (average) correlation or covariance an appropriate measure of comovement, especially for more than two assets? Is it possible to measure state-dependent comovement in a unified framework that also accounts for many other aspects of the data? And how to model and test the nonlinear aspect of comovement as opposed to Pearson's linear measure? To address the first two questions, I propose an eigenvalue-based measure of comovement and introduce the use of a versatile alternative approach, the *multivariate semi-Markov-switching* model. By doing so, I focus on large-cap stocks and shed some new light on the degree of comovement asymmetry depending on different industries. To address the third question, I derive a state-dependent test of symmetric comovement based on Kendall's τ, also known as the coefficient of concordance. This measure exploits the rank relations of returns and is *coherent* in certain precise sense.

1.2 STRUCTURE

In Chapter 2, I revisit the question of how large-cap stock return comovement varies with volatility and market returns. I propose the use of an eigenvalue-based measure of comovement in a multivariate semi-Markov-switching framework. The eigenvalues are extracted from the state-dependent correlation matrix, hence a reliable estimate of the variance-covariance matrix becomes very crucial in its implementation. By incorporating a duration parameter, the new model inherits all the nice properties of a standard Markov-switching model and improves significantly upon the modeling of nonlinear volatility dynamics. I conduct various model evaluation checks and compare the new results with that based on a benchmark model. I estimate models with two to four regimes and consider the impact of sample selection and outlier reduction. Contrary to the sweeping sentiment that comovement is highest when market is down and volatile, I illustrate the significance of comovement differential across states and find in most case studies evidence that suggests otherwise.

In Chapter 3, I propose a test of asymmetric stock return comovement across states. The test can be viewed as a variation of Kendall's τ conditional on the state and has an asymptotic χ^2-distribution. The test procedure has two steps: the goal of step one is to identify the states after which step two is model-free. A refined version of the test is derived based on the Markov chain theory of regenerative cycles which substantially improves finite sample size and power. I show that the test has power against local alternatives, which is nonetheless compromised due to a finite sample convergence bound put on the implied local alternative data generating process. I evaluate the new test against traditional correlation-based measures and demonstrate power attrition due to nuisance parameters when states are ignored. I find that asymmetric tail dependence becomes much less significant when considered state by state. A list of related tests is given as an extension at the end.

As will become clear in Chapter 3, the starting point (step one of the test) does not have to be an "all-encompassing" model such as the multivariate semi-Markov setup advocated in Chapter 2. The test can accommodate different specifications in the first step as long as they can satisfactorily identify the high volatility and low volatility states (negligible misclassification rate in large samples). In fact, I illustrate the use of the test based on states identified by a plain Gaussian Markov-switching model which has been shown to be a powerful benchmark. This is convenient but there seems to be a small discrepancy here—why not use a more elaborate model from the very beginning? The answer is twofold. The eigenvalue-based measures in Chapter 2 rely heavily on the proper estimation of the variance-covariance matrix, which makes the modeling of nonlinear features such as volatility clustering indispensable. To this end, a semi-Markov model comes handy. In contrast, the test given in Chapter 3 only depends on the accurate timing of states and thus does not require modeling all the stylized facts of returns. One may then use a quasi-ML argument to show that under certain assumptions states are consistently identified even if the true model deviates from a Gaussian benchmark.

Of course the above discussion raises another concern: to what extent are the asymptotic results in Chapter 3 affected by errors in state identification? To the best of my knowledge, this is still an open question and cannot be resolved simply by invoking a quasi-ML argument. This is different from the asymptotic impact of errors in conventional nuisance parameter estimation, which has received lots of attention in the past two decades. In order to evaluate the influence of errors in state estimation, I carry out simulation studies to assess the size and power properties of the test and check how they differ from the prediction of theory. When the sample size of daily returns is smaller than 1000 and the two states (high volatility vs. low volatility) are close to each other, indeed there is some noticeable size distortion and the power is not compelling. However, the impact of such errors vanishes as the sample size grows large. On balance, this suggests that the issue, if there is any, does not seem to weigh down the performance of the test in the limit.

The two approaches in this dissertation should be viewed as complementary perspectives. While the eigenvalue-based measure can be used to summarize comovement for multiple asset returns, the state-dependent test is pairwise. To apply the test to multiple hypotheses, one can use, say, a Bonferroni correction. The conclusions drawn from these two methods highlight the need to examine comovement differential across states on a case-by-case basis. Rather than subscribing to the common lore which says that comovement is highest in bad times, I find that, with a focus on large-caps, this is often not the case.

One last thought: As the notion of comovement can be very broad, the researcher should always bear in mind what aspect of or proxy for comovement is being used so that she does not run the risk of overgeneralizing the results.

1.3 NOTATION AND NOMENCLATURE

K:	The number of parameters
N, m:	The number of states
$C(j)$, C^h, C^l:	Coefficient of concordance conditional on the state
Cm, Cm_{diff}, Cm_{avg}:	The proposed comovement measures
$d_j(u)$:	Occupancy probability starting from state j
z_t:	Standard normal distribution
$P(\cdot \mid s_t)$:	Probability conditional on the state at time t
P_{ij}:	Transition probability from state i to state j
Γ:	Transition probability matrix
s_t, X_t:	State at time t
x_t, r_t, R_t:	Asset return
c:	Tail threshold of returns
σ:	Standard error of returns
μ:	Mean return
$\boldsymbol{\pi}$, π_j:	Initial states distribution
$\boldsymbol{\theta}$:	Vector of natural parameters
$\boldsymbol{\phi}$:	Vector of working parameters
$L(\cdot)$, $L_T(\cdot)$:	Likelihood function
$L_C(\cdot)$:	Likelihood function for one realization of states
G:	Gradient of the likelihood function
$\alpha_t^{(k)}(j)$:	Forward probability in k-th iteration
$\beta_t^{(k)}(j)$:	Backward probability in k-th iteration
$L_t^{(k)}(j)$, $G_t^{(k)}(j)$, $F_t^{(k)}(j)$:	Smoothed probabilities for state j in k-th iteration
acf:	Autocorrelation function

MSwC:	Markov switching order selection criterion
Figure #:	The figure number within each chapter
Section #:	The section number within each chapter
Table #:	The table number within each chapter

Chapter 2. LARGE-CAP STOCK RETURN COMOVEMENT: A SEMI-MARKOV SWITCHING APPROACH

2.1 INTRODUCTION

In the wake of the 2008 Great Recession, there appeared to be a reviving interest in the study of stock return comovement which, economists seem to believe, is highest under market distress, or is it? Over the past two decades, various models have been proposed to study asset return comovement, its time-varying properties and how it is related to market returns and volatility. The issue at stake is this: Based on certain comovement measure, does portfolio diversification break down in bad times?

For a group of economic variables, previous authors have tried to quantify the notion of comovement using popular models such as rolling window cross-sectional average correlation (Preis *et al*. 2012), dynamic correlation and GARCH-type models (Syllignakis and Kouretas, 2011), factor pricing models (Bekaert *et al.,* 2008), stochastic matrix theory (Reigneron *et al*. 2011), spectral analysis (Croux *et al*. 2001), and wavelet analysis (Rua 2010). Many of these models, however, are unable to incorporate explicitly a state-dependent correlation structure and are silent on many other aspects of the data, so it is hard to justify its use on a broader basis. Some of them also require very specific techniques subject to stringent assumptions (e.g., stochastic matrix theory) and the results can be rather irregular for the researcher to draw any simple conclusions (e.g., spectral and wavelet analysis). More importantly, to study comovement between more than two series, one needs to find some averaging scheme which is arbitrary and inefficient as suggested by the stochastic matrix literature. For non-switching models, it is not clear how comovement can be linked back to state-dependent volatility and returns. Against this backdrop, one of my goals is to model the regime-switching correlation matrix directly in a unified and robust framework.

The aforementioned contributions are mostly concerned with stock return covariance or correlation in different historical episodes. This raises three questions: Is (average) correlation or covariance an appropriate measure of comovement, especially for more than two series? Is the high volatility low return state effectively distinguished from the low volatility high return state? And, is it possible to measure state-dependent comovement in a unified framework that also accounts for many other aspects of the data? To address these questions, I propose an eigenvalue-based measure of comovement and suggest the use of a versatile alternative approach, the *multivariate semi-Markov-switching* model.

In a study of the covariance matrix of S&P 500 stocks, Laloux *et al.* (2000) fit a density for eigenvalues based on stochastic matrix theory and show that among the 406 available eigenvalues, only a fraction (6%) of them carry some information above and beyond that of a *purely random matrix*. The rest of the eigenvalues (94%) accounts for 74% of the total variance but does not seem to contain any information in terms of decreasing the entropy. These results suggest that stock return comovement can be measured in the factor space, for which a good proxy is the top eigenvalue since, for a given correlation matrix, the sum of all eigenvalues is always equal to its dimension.

When correlation is concentrated in a subgroup of variables, the smallest eigenvalue will be very small; when the overall correlation is high, the largest eigenvalue will be very large. This suggests another potentially informative measure which is just the difference between the top and the bottom eigenvalues which accounts for both overall and block comovement. In Section 4 and 5, it is shown that this version has little impact on the magnitude of comovement measures and the rankings are preserved in all cases. For a third measure, I use the average of all state-dependent correlation pairs, for which a Wald test of parameter equality can be conducted.

A closer look at the frequency-domain methods reveals that the benefits can be partially gained in the current semi-Markov-switching setup. It is shown in Section 4 that daily returns filtered through a semi-Markov model resemble white noise up to the fourth moment. Hence from a spectral density integral point of view, the frequency domain construction of comovement (e.g., Croux *et al.*, 2001) can be approximated by the state-dependent correlation matrix. Recall that the spectral density of white noise is constant for all frequencies. Thus I measure state-dependent comovement in three ways: the top eigenvalue as a fraction of the sum of all eigenvalues (Cm), the largest eigenvalue minus the smallest eigenvalue as a fraction of the sum of all eigenvalues (Cm_{diff}), and the average pairwise correlation (Cm_{avg}).

As a precursor to eigenvalue estimation, the state-dependent correlation matrix plays a central role, so an important aspect of the model is how to estimate this matrix. In this respect, one strand of the literature is focused on combining regime-switching and a multivariate GARCH-type structure, e.g., Edwards and Susmel (2003) and other multivariate generalizations to the models of Gray (1996) and Hamilton and Susmel (1994). Many of these models suffer from the curse of dimensionality and have severe estimation difficulties. By contrast, Pelletier (2006) takes a different path and solves the problem by decomposing the covariance matrix into an ordinary GARCH component for standard deviation and an independently switching correlation component. The model thus has the advantage of directly estimating the state-dependent correlation matrix instead of imputing it from the covariance matrix; see also Billio and Caporin (2005) for a Markov-switching DCC.

The decoupling of variance and correlation does not add more unknown parameters but is nontrivial as it is now possible to conduct hypothesis tests directly on the correlation parameters. In Pelletier (2006), one could test the significance of correlation differential across states, yet it is hard to relate the state of correlation to the state of returns or volatilities since the latter are non-switching—allowing all of them to switch in a multivariate GARCH model is still computationally very challenging.

In view of the above difficulties, I take a step back and ask the question: Can the justification for the use of a Markov-switching GARCH model be regained in a different setting that is reasonably easy to estimate and numerically stable? (It is known that Markov-switching GARCH models are powerful largely because they provide a relatively parsimonious way to capture, among others, time-varying volatility dynamics.) On the one hand, there are plenty of reasons why regime-switching should be preferred as even the simplest mixture-of-normal models can explain most of the stylized facts of financial time series (Ang and Chen (2002), Timmermann (2000)); on the other hand, as Rydén *et al.* (1998) point out, the very slow decay of autocorrelation of absolute or squared returns cannot be adequately reproduced solely by a

Markov-switching structure. Recently, this issue was addressed by Bulla and Bulla (2006) in a univariate semi-Markov-switching (sMSw) framework (Ferguson (1980), Guédon (2003)). The new model inherits all the nice properties of a standard Markov-switching model and improves significantly upon the modeling of nonlinear volatility dynamics. In this chapter, I extend the model to a multivariate setting and study how stock return comovement is related to volatilities and market returns. Efficient algorithms based on forward-backward probabilities are developed in Guédon (2003), an adapted derivation of which is given in Section 2.

In Section 4, I conduct a preliminary study for model evaluation purposes. The impact of outlier reduction, Student's t returns and VAR(1) pre-whitening are considered. To justify the use of a semi-Markov-switching model, I focus on state identification and the fit of ARCH effect. In Section 5 I apply the new model to full samples of a pool of 10 largest stocks by market capitalization in six major industries in the U.S. and one major foreign index, FTSE100. A close-up is given to the financial sector because of the recent crisis. I estimate models with two to four regimes and consider the impact of sample selection and outlier reduction. Finally, I use multivariate scaling to visualize the state-dependent stock return comovement.

Combing results of earlier works, one tends to conclude that,

$$Cm(\text{low return, high vol}) \;>\; Cm(\text{high return, low vol})$$

Through a battery of case studies, in most cases I find the following relationship:

$$Cm(\text{high return, low vol}) \geq\; Cm(\text{low return, high vol})$$

As will be clear in Section 5, the financial intermediation industry over the recent crisis period is an exception, which nonetheless does not easily agree with some other commonly recognized regularities of asset returns either. Extending samples to prior to 2008, the ordering becomes insignificant. In sum, my findings challenge the prevalent view in that, given my eigenvalue-based measures, stock return comovement is, in most case studies, lower during periods of high market volatility than in periods of low market volatility.

2.2 METHODOLOGY

Consider the n-variate process $\{x_1, x_2, x_3, \ldots, x_T\}$ generated by $x_t = \sigma_{s_t} z_t + \mu_{s_t}$, where $z_t \sim \text{iid } N(0,1)$ and $\{s_t\}$ is an m-state Markov chain governed by the *transition probability matrix* Γ. Let $P(x_t|s_t)$ denote the conditional probability of observing x_t given state s_t, let $\boldsymbol{\pi}$ denote the initial state distribution, and let $\boldsymbol{x}^{(t)}$, $\boldsymbol{s}^{(t)}$ denote the histories of observations and states up to time t. This is the benchmark model well understood in the literature (MacDonald and Zucchini, 1997) and can be estimated using EM algorithm or direct ML, a brief discussion of which is relegated to Appendix A.2. In what follows, I focus on the less familiar semi-Markov-switching model.

2.2.1 *The semi-Markov specification*

An implicit restriction on the standard Markov-switching model is that duration (occupancy) is necessarily geometric distributed by construction. This can be very unrealistic in two aspects. First, the overall shape of the distribution dictates that the duration probability must start to decline geometrically right away. A second drawback has to do with failure rate (or hazard function in the continuous case). In many applications, one often finds that the probability of staying in a certain state depends on how long the process has occupied that state. The failure rate for a geometric distribution is just $1 - p_{ii}$, a constant, whereas a semi-Markov model can in theory accommodate a full range of hazard functions.

For the rest of this section, I follow the discussion in Guédon (2003). A semi-Markov model can be summarized by four groups of parameters, i.e., initial distribution $\boldsymbol{\pi}$, transition probability matrix Γ, parameters of the state-dependent observation density, and parameters of the duration density. Let s_1^t and $\boldsymbol{x}_1^t$ denote the history of states and observed data between period 1 and t. Define the duration probability starting from a given state j to be

$$d_j(u) = P(s_{t+2} = j, \ldots, s_{t+u} = j, s_{t+u+1} \neq j | s_{t+1} = j, s_t \neq j), \sum_u d_j(u) = 1.$$

Also denote the semi-Markov transition probabilities as

$$P_{ij} = P(s_{t+1} = j | s_t = i, s_{t+1} \neq i), \sum_j P_{ij} = 1, P_{ii} = 0,$$

The ground work of semi-Markov-switching models was first laid out in Ferguson (1980). It was soon found that Ferguson's algorithm entails multiplication and concentration of many probabilities (e.g., transition, occupancy, or observation probabilities) and is quite prone to underflow errors. Guédon (2003) largely improves the algorithm by developing an efficient and numerically stable procedure based on backward-forward updating probabilities. He also replaces the original technical assumption that the process must leave the last in-sample state to a *different* state with the more realistic right-censoring assumption, in which an *out-of-sample* occupancy probability is introduced and modeled as a survival function that measures the probability of staying in the last visited state for at least u periods:

$$d_j^T(u) = \sum_{v \geq u} d_j(v).$$

Let R be the number of times the process hits a new state different from the current one, including the initial state, the complete data likelihood function for one realization then becomes

$$\begin{aligned} L_C(s_1^{T+u}, \boldsymbol{x}_1^{\mathrm{T}} | \theta) &= P(s_1^T, s_{T+1} = s_{T+2} = \cdots = s_{T+u-1} = s_T, s_{T+u} \neq s_T, \boldsymbol{x}_1^T | \boldsymbol{\theta}) \\ &= \pi_{s_1} d_{s_1}(u_1) \prod_{r=2}^{R} P_{s_{r-1}s_r} d_{s_r}(u_r) \prod_{t=1}^{T} P_{s_t}(x_t), \end{aligned} \tag{2.1}$$

$$d_{s_R}(u_R) \equiv d_{s_R}^T(u_R) = \sum_{v \geq u_R} d_{s_R}(v) = 1 - \sum_{v < u_R} d_{s_R}(v),$$

$$\sum\nolimits_{r=1}^{R-1} u_r < T \leq \sum\nolimits_{r=1}^{R} u_r.$$

The *complete likelihood function* for all realizations can be built from Eq. (2.1) as

$$L(\boldsymbol{\theta}) = \sum\nolimits_{s_1, s_2, \ldots, s_T} \sum\nolimits_{u_R} L_C\left(s_1^{T+u_R}, \boldsymbol{x}_1^{\mathrm{T}} | \boldsymbol{\theta}\right). \tag{2.2}$$

As the objective function is recast into an incomplete data problem, a natural way to proceed is by means of the EM algorithm. Let $\boldsymbol{\theta}^{(k)}$ denote the current value of $\boldsymbol{\theta}$ at iteration k. The algorithm then iterates between the conditional expectation of the complete-data log-likelihood:

$$L(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \equiv E\left(log\, L_C\left(S_1^{T+u_R}, \boldsymbol{X}_1^T|\boldsymbol{\theta}\right)|\boldsymbol{X}_1^T = \boldsymbol{x}_1^T, \boldsymbol{\theta}^{(k)}\right), \tag{2.3}$$

and the maximization step:

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$$

In general the algorithm will generate a sequence of $\{\boldsymbol{\theta}^{(k)}\}$ that converge to a local maximum of $L(\boldsymbol{\theta})$.

2.2.2 *EM algorithm*

From Section 2.1, Eq. (2.1) can be written out as

$$\begin{aligned} L_C(s_1^{T+u}, \boldsymbol{x}_1^{\mathrm{T}}|\theta) &= P(s_1^T, s_{T+1} = s_{T+2} = \cdots = s_{T+u-1} = s_T, s_{T+u} \neq s_T, \boldsymbol{x}_1^T|\theta) \\ &= \pi_{S_1} d_{S_1} \prod_{r=2}^{R} P_{S_{r-1}S_r} d_{S_r}(u_r) \prod_{t=1}^{T} P_{S_t}(x_t), \\ d_{S_R}(u_R) &= D_{S_R}^T(u_R) = \sum_{v \geq u} d_{S_R}(v) = 1 - \sum_{v<u} d_{S_R}(v). \end{aligned} \tag{2.4}$$

Combining Eq. (2.1)-(2.3), the E step for the complete likelihood function for *all realizations* becomes

$$\begin{aligned} L(\theta|\theta^{(k)}) &= E\left(log\, L_C\left(S_1^{T+u_R}, \boldsymbol{X}_1^T|\theta\right)|\boldsymbol{X}_1^T = \boldsymbol{x}_1^T, \theta^{(k)}\right) \\ &= \sum\nolimits_{s_1, s_2, \ldots, s_T} \sum\nolimits_{u_R} \log L_C\left(s_1^{T+u_R}, \boldsymbol{x}_1^{\mathrm{T}}|\theta\right) P\left(s_1^{T+u_R}|\boldsymbol{x}_1^T, \theta^{(k)}\right). \end{aligned}$$

The next critical step is to simplify the M step by splitting terms in the above expression and concentrating out irrelevant states and observed data by marginalization. Substituting Eq. (2.4) for L_C one gets

$$L(\theta|\theta^{(k)}) = \sum\nolimits_{S_1^T} \sum\nolimits_{u_R} \Big(\log \pi_{S_1} + \sum\nolimits_{r=2}^{R} \log P_{S_{r-1}S_r} + \sum\nolimits_{r=1}^{R} \log d_{S_r}(u_r) + \sum\nolimits_{t=1}^{T} \log P_{S_t}(\boldsymbol{x}_t) \Big) P\big(s_1^{T+u_R} \big| \boldsymbol{x}_1^T, \theta^{(k)}\big). \tag{2.5}$$

The first term in the bracket reduces to

$$\sum\nolimits_{j=1}^{m} \log \pi_j \, P\big(s_1 = j \big| \boldsymbol{x}_1^T, \theta^{(k)}\big).$$

The fourth term reduces to

$$\sum\nolimits_{j=1}^{m} \sum\nolimits_{t=1}^{T} \log P_j(\boldsymbol{x}_t) \, P\big(s_1 = j \big| \boldsymbol{x}_1^T, \theta^{(k)}\big).$$

The third term can be written as

$$\sum_{j=1}^{m} \sum_{u=1,2,\dots} \Bigg(P\big(s_1 = j, \dots, s_u = j, s_{u+1} \neq j \big| \boldsymbol{x}_1^T, \theta^{(k)}\big) + \sum_{t=1}^{T-1} P\big(s_t \neq j, s_{t+1} = j, \dots, s_{t+u} = j, s_{t+u+1} \neq j \big| \boldsymbol{x}_1^T, \theta^{(k)}\big) \Bigg) \log d_j(u). \tag{2.6}$$

If we define

$$\eta_{ju}^{(k)} = P\big(s_1 = j, \dots, s_u = j, s_{u+1} \neq j \big| \boldsymbol{x}_1^T, \theta^{(k)}\big) + \sum_{t=1}^{T-1} P\big(s_t \neq j, s_{t+1} = j, \dots, s_{t+u} = j, s_{t+u+1} \neq j \big| \boldsymbol{x}_1^T, \theta^{(k)}\big),$$

then Eq. (2.6) can be rewritten compactly as

$$\sum_{j=1}^{m} \sum_{u=1}^{M} \Big(\eta_{ju}^{(k)} \log d_j(u) \Big). \tag{2.7}$$

In Eq. (2.7), η_{ju} has a very intuitive interpretation which is the expected number of periods a process stays in state j. M is some large number chosen to be the upper bound of the occupancy time just to make the iteration feasible. For example M can be 500 meaning that a certain state is practically impossible to last longer than 500 days. In my study, this parameter ceases to matter once it is larger than, say, 300.

These are the easy ones. The not so obvious part is the second term in the bracket of Eq. (2.5). First note that it collects all the different combinations of two distinct but adjacent states in a combinatorial number of places from 1 to T where this can occur. By definition, the same combination can reappear multiple times at different places as long as they are separated apart in time. This gives the following expression,

$$\sum_{j=1}^{m}\sum_{j\neq i}\sum_{t=1}^{T-1}\left(\log P_{ij}\, P\left(s_{t+1}=j, s_t=i \middle| \boldsymbol{x}_1^T, \theta^{(k)}\right)\right).$$

Since only the third term, i.e., Eq. (2.7), has to do with the occupancy distribution, by the first order condition one can easily derive that

$$d_j^{(k)}(u)=\frac{\eta_{ju}^{(k)}}{\sum_u \eta_{ju}^{(k)}}. \tag{2.8}$$

This can be seen as a discretized nonparametric estimate of the duration probabilities. A parsimonious way to model a wide range of possible occupancy patterns is to use some parametric specification, for which I choose the gamma distribution which is versatile enough to mimic a wide range of shapes. (When I use the Poisson distribution, parameter estimates are extremely unstable because it does not allow for excess volatility.) To estimate the gamma distribution parameters, for state j,

$$\frac{\partial}{\partial\alpha_j}\sum_{u=1}^{M}\eta_{ju}\log\left(\frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}u^{\alpha_j-1}e^{-\beta_j u}\right)=0 \Rightarrow$$

$$\sum_{u=1}^{M} \eta_{ju}\big(\log \beta_j - \psi(\alpha_j) + \log u\big) = 0. \quad (2.9)$$

Divide both sides by $\sum_u \eta_{ju}^{(k)}$, treat $d_j^{(k)}(u)$ as pseudo-sample weights and use the fact that $E(u) = \alpha_j/\beta_j$ in a Gamma distribution, one arrives at

$$\log(\alpha_j) - \log E(u) - \psi(\alpha_j) + \overline{\log(u)} = 0.$$

The sample analog is to solve numerically for α_j in

$$\log(\alpha_j) - \log(\bar{u}) - \psi(\alpha_j) + \overline{\log(u)} = 0, \quad (2.10)$$

where $\bar{u} = \sum_u u\, d_j^{(k)}(u)$, $\overline{\log(u)} = \sum_u \log(u) d_j^{(k)}(u)$, and $\psi(\alpha_j)$ is the digamma function with respect to α_j. $d_j^{(k)}(u)$ can be estimated from (1) and $\hat{\beta}_j = \hat{\alpha}_j/\bar{u}$.

2.2.3 *Re-estimating equations and updating algorithm*

Let s_1^t and $\boldsymbol{x}_1^t$ summarize the history of states and returns between time 1 and t. Define

$$\text{Forward probability: } \alpha_t^{(k)}(j) = P(\boldsymbol{x}_1^t, s_t = j, \theta^{(k)})$$

$$\text{Backward probability: } \beta_t^{(k)}(j) = P(\boldsymbol{x}_{t+1}^T | s_t = j)$$

$$\text{Likelihood function: } L_T^{(k)} = \boldsymbol{\alpha}_t^{(k)\prime} \boldsymbol{\beta}_t^{(k)}$$

$$\text{Smoothed probability: } L_t^{(k)}(j) = P\big(s_t = j \big| \boldsymbol{x}_1^T, \theta^{(k)}\big) = \frac{\alpha_t^{(k)}(j)\beta_t^{(k)}(j)}{L_T^{(k)}}$$

$$G_{t+1}^{(k)}(j) = \frac{P\big(\boldsymbol{x}_{t+1}^T \big| s_{t+1} = j, s_t \neq j, \theta^{(k)}\big)}{P(\boldsymbol{x}_{t+1}^T | \boldsymbol{x}_1^t, \theta^{(k)})},$$

$$F_t^{(k)}(j) = P\big(s_{t+1} \neq j, s_t = j \big| \boldsymbol{x}_1^t, \theta^{(k)}\big),$$

Then using techniques developed in Peel and McLachlan (2000), one can derive

$$\mu_{\mathrm{j}}^{(k+1)} = \frac{\sum_1^T L_t^{(k)}(i)\, x_t}{\sum_1^T L_t^{(k)}(i)},$$

$$\Sigma_j^{(k+1)} = \frac{\sum_1^T L_t^{(k)}(i) \left(x_t - \mu_j^{(k+1)}\right)\left(x_t - \mu_j^{(k+1)}\right)'}{\sum_1^T L_t^{(k)}(i)},$$

$$\pi_{\mathrm{j}}^{(k+1)} = L_{t=1}^{(k)}(j),$$

$$p_{ij}^{(k+1)} = \frac{\sum_1^{T-1} P\left(s_{t+1} = j, s_t = i \middle| \boldsymbol{x}_1^T, \theta^{(k)}\right)}{\sum_1^{T-1} P(s_{t+1} \neq i, s_t = i | \boldsymbol{x}_1^T, \theta^{(k)})} = \frac{\sum_1^{T-1} G_{t+1}^{(k)}(j)\, F_t^{(k)}(i)}{\sum_1^{T-1} P(s_{t+1} \neq i, s_t = i | \boldsymbol{x}_1^T, \theta^{(k)})}.$$

The conditional probabilities used to update $L_t(j)$, $G_t(j)$, $F_t(j)$, η_{ju}, p_{ij}, $P(j|\boldsymbol{x}_t)$ can be calculated using the forward-backward algorithm outlined in Guédon (2003).

2.2.4 *Order selection*

Testing for the order of switching is a hard problem; see Cho and White (2007) for a survey of recent results. To check the possibility of over-fitting the order of switching, I compare traditional model selection criteria (AIC, BIC, AIC_c) with the one proposed by Smith *et al.* (2005). The new criterion, which I denote as MSwC, is a more elaborate approximation of KLIC tailored to the Markov-switching model. Their paper also demonstrates the severe limitations of AIC which is confirmed in Section 4.

A feasible version of MSwC in the current context takes the following form:

$$\text{MSwC} = -2\text{mllk} + \sum_{i=1}^{N} \frac{\hat{T}_i\left(\hat{T}_i + NK\right)}{\hat{T}_i - NK - 2},$$

where mllk is the maximized log-likelihood, N is the number of states, K is the number of explanatory variables, and $\hat{T}_i = \sum_{t=1}^{T} \hat{P}(s_t = i | X_1^T, \widehat{\boldsymbol{\theta}})$ the sum of smoothed probabilities of being in state i. When there is only one state the criterion reduces to AIC_c, $-2\text{mllk} + T(T+K)/(T-K-2)$. Since smoothed probabilities at any point in time always sum to one across all N states, $\hat{T}_i$ for a particular state can be interpreted as the expected number of arrivals at that state. Therefore, that $\hat{T}_i$ is so small as to make the denominator negative is strong evidence that the number of regimes is too large for the sample at hand. In cases when the denominator returns negative value(s) for certain state(s), I retain and add up only the positive ones. In most cases studied in Section 4 and 5, MSwC has one negative component when $m = 3$, and two or even three negative components when $m = 4$.

2.3 DATA

All historical daily adjusted closing prices are obtained from Yahoo Finance data service and double checked against the Center for Research in Security Prices (for U.S.-listed stocks) and Thomson Reuters financial datastream (for European listed stocks), both of which are accessed through Wharton Research Data Services. I proxy for market returns and volatilities using CRSP value-weighted data and all returns are computed as log-differences. Using S&P 500 returns does not change any conclusion of this chapter. I study short four-year samples and full samples in Section 4 and 5 respectively.

In Section 4, I study the four largest (market capitalization as of June 15, 2013) stocks in the U.S. banking sector, the Deutscher Aktien Index (DAX 30), and the Dow Jones Industrial Average index (DJIA). I choose a sample that stretches across the recent financial crisis from 1 January 2007 to 31 December 2010. In Section 5, I study comovement of the ten largest stocks for which data are available for at least 15 years from 1 January 1998 to 31 December 2012. These stocks come from six different industries and one foreign index, namely, NYSE Financial, NYSE Energy, NYSE Health Care, DJ Industrial Average 30, DJ Transportation 20, DJ Utility 15 and FTSE 100. I also study the longest-living stocks in DJIA 30 by studying a sample of the ten largest stocks with data available for at least 30 contiguous years. A 15-year pre-crisis sample from the financial intermediation industry is examined in Section 5. In Table C, data are multiplied by 100 before all summary statistics are reported.

Table C.1-3 in Appendix C report the summary statistics for, respectively, the four-year, the 15-year and the 30-year samples. In the recent financial crisis, Wells Fargo (WFC) was arguably an outstanding contranatant performer, whereas in 2006, BP was involved in an oil spill accident and dragged down FTSE 100. Descriptive statistics for the 15-year pre-crisis sample from the financial industry are given in Table C.3, the last row of which also reports the CRSP value-weighted returns over a similar period. Kurtosis in Table 1 is excess kurtosis which is high for all data series; not all stocks are negatively skewed over the sample period. Using adjusted closing prices, stocks that historically had large dividend payments or splits tend to have very large skewness and kurtosis. The Jarque-Bera test firmly rejects normality in all series (p-value $< e^{-15}$). In terms of autocorrelation, a handful of stocks such as BAS, BAC, BA, NSC, etc. manage to pass the Ljung-Box test with p-values larger than 10%. In most cases, the L-B test tends to reject less often the null hypothesis of significant autocorrelation when data are prewhitened by a VAR(1). But this is not always the case, e.g., BA in the bottom panel of Table C.1. The last column reports the number of outliers defined as observations lying outside the interval of $\pm 4\hat{\sigma}^2$ about the sample mean. These outliers, many of which are caused by large dividend payments or stock splits, are then set to be equal to the closest interval boundary.

2.4 Model Evaluation

To justify the use of the semi-Markov-switching model, in this section I focus on model validation through various stability and modeling checks. In particular, I consider the effect of prewhitening and outlier reduction and compare results based on the benchmark specification with that of EM-sMSw. I show that, in a multivariate context, the semi-Markov-switching framework is numerically stable, identifies states consistent with the market data, captures duration dependence (discussed in Section 5), and considerably reproduces the ARCH effect. These findings together with the fact that a semi-Markov-switching model has been documented to account for other stock market stylized facts provide economic ground for its use in the full sample study of Section 5.

2.4.1 *Basic results*

In Table 1, I present a preliminary study for model evaluation purposes. To see if the result is robust to simpler model specifications, I compare the plain EM-MSw model with the EM-sMSw model. The three measures of stock return comovement are arranged in a way so that they run from the low state to the high state based on the estimated state-dependent volatilities. Since individual stock returns are often noisy, I also match the ordering of comovement with market data by computing the state-dependent market returns and volatilities over the same sample period. The ordering of state-dependent mean returns for individual stocks (not reported) are mostly in line with that of the market mean returns but exceptions are not rare: when the market is down, not all stock returns are low at the same time.

Estimates based on direct ML are given in Appendix B, where I report the impact of outlier reduction and Student's t errors on estimation. In a nutshell, compared to the Gaussian case, use of Student's t distribution tends to smooth out the ordering of comovement measure, while outlier reduction tends to retain it. In fact, the tension between t-distribution and outlier reduction tells a compelling story: while the former appears to be accommodating too much to outliers, outlier reduction fixes this problem only imperfectly by introducing a mild clustering of high volatility states at the cutoff points. As another way of looking at the problem, when I simulate data from models estimated with a t-distribution, extreme returns occur just too often compared to the frequency found in real data. Another undesirable result of using t-distribution is that the estimated transition matrix is much more concentrated along the diagonal for all states, with the off-diagonal probabilities being extremely small and thus creating spuriously persistent states across the board. By contrast, Gaussian returns are much more robust to one-time big swings and the estimated transition matrix is well-behaved such that the probability of staying in one state is never insanely high especially for the high volatility state.

In this regard, the semi-Markov model is helpful as it fills the gap of modeling switching probability and duration probability separately in a unified framework and is more robust to outliers. For $\mathrm{m} = 2$ or $\mathrm{m} = 3$, the comovement measures and market data are all neatly aligned such that the high volatility state always corresponds to a low market mean return and *vice versa*. Even when $\mathrm{m} = 4$, EM-sMSw is able to correctly identify the states such that the state-dependent market returns and volatilities are all monotonically ordered—a degree of stability not seen in ordinary Markov-switching models (Table B.1). As a check of the estimated transition probabilities, the off-diagonal elements are never very close to the boundary. In addition, Bulla and Bulla (2006) show that, on average, a semi-Markov-switching model with Student's t returns performs worse in terms of explaining stylized facts than that with Gaussian returns and both improve upon a standard Markov-switching model. Outlier reduction (with Gaussian returns) slightly affects the probability transition matrix and the correlation matrix, but the ordering of comovement measures is largely preserved. For these reasons and because incorporating a multivariate t-distribution into a semi-Markov model is computationally demanding, I use Gaussian errors throughout.[1]

Estimation of EM-sMSw models, like the EM-MSw benchmark, is not sensitive to starting values of the conditional mean or the transition probability matrix, but is somewhat influenced by starting values of the variance-covariance matrix and the choice of duration distributions. For the latter, I use the versatile gamma distribution because its scale and shape parameters can vary independently and the result corroborates an early finding on duration dependence of states (Section 5). Of practical concern is the computational cost, at which direct ML scores very badly, whereas EM-MSw and EM-sMSw are much faster. In the case of two regimes, the latter converges very fast and starting values do not seem to matter. When the order of switching is greater than two, to guarantee proper convergence I use the EM-MSw estimates as starting values.

In addition to EM-sMSw, I use $\mathrm{VAR}(1)$ prewhitening/recoloring to check the potential influence of residual serial correlation and spurious ARCH effect. In the case of $\mathrm{m} = 3$ and $\mathrm{m} = 2$ (not reported), the results are stable across different model specifications and estimation techniques. It is found that $\mathrm{VAR}(1)$-prewhitening does not appear to yield noticeable differences and that, for a given model, conclusion is the same for all three measures of comovement. In terms of order selection, MSwC and BIC pick either $\mathrm{m} = 2$ or $\mathrm{m} = 3$, whereas AIC and $\mathrm{AIC_c}$ always pick the highest order of switching.

[1] Estimation of the degrees of freedom requires numerically solving a digamma function at each step, which adds another layer of complexity in terms of updating a weighted average for the conditional mean and covariance matrix. Analytic solutions do not exist; see the appendix in Bulla and Bulla (2006).

Table 1

Preliminary estimates, EM-sMSw model, Jan. 1, 2007-Dec. 31, 2010.

	m=3		EM-sMSw			
	EM-MSw	EM-VAR(1)	m=2	m=3	m=4	VAR(1), m=3
			NYSE Financial Top 4			
Cm	[0.73, 0.70, 0.72]	[0.71, 0.70, 0.72]	[0.72, 0.71]	[0.73, 0.70, 0.72]	[0.76, …, 0.73]	[0.73, 0.70, 0.73]
Cm_{diff}	[0.68, 0.67, 0.68]	[0.68, 0.67, 0.69]	[0.68, 0.67]	[0.69, 0.67, 0.68]	[0.72, …, 0.69]	[0.70, 0.67, 0.69]
Cm_{avg}	[0.62, 0.59, 0.62]	[0.61, 0.59, 0.62]	[0.62, 0.60]	[0.62, 0.59, 0.62]	[0.67, …, 0.63]	[0.62, 0.59, 0.62]
r_{CRSP}	[high, medium, low]				[high, ..., low]	
σ_{CRSP}	[low, medium, high]				[low, ..., high]	
σ	[low, medium, high]				[low, ..., high]	
BIC			No	Yes	No	
MSwC			Yes	No	No	
			DJIA Top 4			
Cm	[0.68, 0.63, 0.78]	[0.69, 0.64, 0.77]	[0.67, 0.72]	[0.68, 0.63, 0.78]	[0.42, …, 0.78]	[0.69, 0.63, 0.77]
Cm_{diff}	[0.60, 0.54, 0.73]	[0.61, 0.55, 0.73]	[0.58, 0.65]	[0.60, 0.54, 0.73]	[0.26, ..., 0.79]	[0.60, 0.54, 0.73]
Cm_{avg}	[0.58, 0.50, 0.70]*	[0.58, 0.51, 0.70]*	[0.55, 0.63]*	[0.58, 0.50, 0.70]*	[0.22, …, 0.70]**	[0.58, 0.50, 0.70]*
r_{CRSP}	[high, medium, low]				[high, ..., low]	
σ_{CRSP}	[low, medium, high]				[low, ..., high]	
σ	[low, medium, high]				[low, ..., high]	
BIC			No	Yes	No	
MSwC			Yes	No	No	
			DAX30 Top 4			
Cm	[0.69, 0.66, 0.57]	[0.69, 0.66, 0.56]	[0.65, 0.60]	[0.69, 0.67, 0.56]	[0.71, …, 0.36]	[0.69, 0.67, 0.55]
Cm_{diff}	[0.62, 0.59, 0.49]	[0.62, 0.59, 0.49]	[0.58, 0.52]	[0.62, 0.60, 0.48]	[0.64, …, 0.20]	[0.62, 0.59, 0.48]
Cm_{avg}	[0.58, 0.55, 0.41]*	[0.58, 0.54, 0.41]*	[0.53, 0.47]*	[0.58, 0.56, 0.40]**	[0.60, …, 0.13]**	[0.58, 0.56, 0.41]**
r_{CRSP}	[high, medium, low]				[high, ..., low]	
σ_{CRSP}	[low, medium, high]				[low, ..., high]	
σ	[low, medium, high]				[low, ..., high]	
BIC			Yes	No	No	
MSwC			Yes	No	No	

Note: p-values are computed from a Wald test of equal Cm_{avg} across states based on the Gaussian specification. * and ** denote the 10% and 5% significance level. AIC and AICc always pick the four-regime model. For DJIA Top 4, the comovement measure based on the four-regime EM-sMSw model is suspicious, e.g., the full estimate of Cm_{avg} is [0.22, 0.77, 0.48, 0.70]. Apparently the low volatility state is split into two due to over-fitting. Similarly for DAX30 Top 4, the four-regime EM-sMSw

2.4.2 *Model evaluation*

Since the four-regime model is not supported by BIC or MSwC, I only focus on two and three regimes in this section. As the chapter is primarily concerned with implied comovement based on the correlation matrix, to save some space, below I only report on correlation parameters and transition probabilities. In Table 1, I have conducted a Wald test of parameter equality on the average pairwise correlation measure of comovement (Cm_{avg}) in the Gaussian specification. The ordering of comovement is statistically insignificant for NYSE Fin Top 4, whereas for DJIA Top 4 and DAX 30 Top 4, significant but opposite conclusions are found.

For lack of a simple relationship between the model parameters and the top eigenvalue of the state-dependent correlation matrix, approximate confidence intervals for Cm and Cm_{diff} cannot be directly imputed using, say, the second-order delta method. So I conduct a bootstrap experiment to see whether eigenvalue-based comovement differentials are statistically significant. I limit the case to two regimes selected by MSwC. Figure 1 shows the distribution of Cm based on 2000 parametric bootstrap samples generated from the two-state Gaussian NYSE Fin Top 4 model.

Table 2

Correlation and transition probability estimates, NYSE Fin Top 4, m = 2, Jan. 1, 2007-Dec. 31, 2010.

	low volatility state						high volatility state						p_{11}	p_{22}
NYSE Fin Top 4	0.756	0.487	0.845	0.448	0.725	0.442	0.707	0.475	0.841	0.455	0.718	0.432	0.969	0.908
	(0.029)	(0.044)	(0.022)	(0.038)	(0.042)	(0.041)	(0.01)	(0.017)	(0.007)	(0.016)	(0.01)	(0.019)	(0.021)	(0.022)
DAX 30 Top 4	0.621	0.482	0.695	0.402	0.563	0.494	0.378	0.354	0.474	0.427	0.477	0.654	0.938	0.561
	(0.022)	(0.029)	(0.019)	(0.032)	(0.025)	(0.027)	(0.077)	(0.078)	(0.069)	(0.073)	(0.069)	(0.052)	(0.01)	(0.058)
DJIA Top 4	0.543	0.614	0.49	0.599	0.489	0.601	0.641	0.582	0.568	0.672	0.681	0.589	0.937	0.766
	(0.028)	(0.026)	(0.031)	(0.026)	(0.031)	(0.027)	(0.041)	(0.046)	(0.048)	(0.038)	(0.038)	(0.045)	(0.013)	(0.043)

Note: Standard errors are in parenthesis.

The small standard errors in Table 2 lead to the conjecture that the implied sampling distribution of Cm and Cm_{diff} should also have small standard errors, as the latter is fully determined by the correlation matrix. To see this more clearly, two normal distributions centered around the bootstrap means are superimposed onto Figure 1, both of which have a standard deviation of 0.02. This, together with the bootstrap distribution, suggests that a difference of 0.05~0.1 or more can be viewed as solid evidence for comovement differential across states. For example, in the case of $\mathrm{m} = 3$, NYSE Financial Top 4 does not deliver a clear message, whereas DAX30 Top 4 yields much more significant result and does *not* accord with the conventional wisdom of low return-high comovement, at least in the 4-year sample.

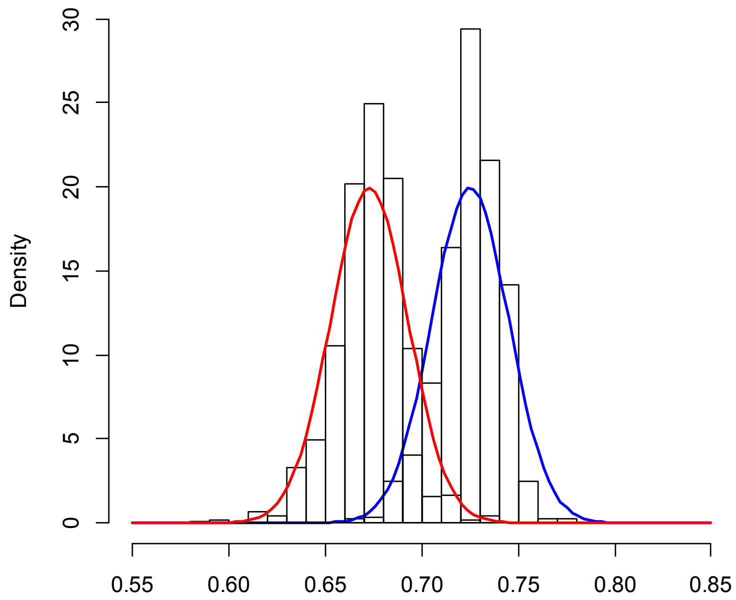

Figure 2.1. Bootstrap distribution of Cm from the Gaussian NYSE Fin Top 4 model, 2000 samples, $m = 2$.

To highlight the semi-Markov-switching model's ability to capture the slow decay of volatility dynamics, Figure 2 illustrates the model-implied autocorrelations for squared returns and the sample ACF for squared residuals. In the Gaussian Markov-switching case, exact autocorrelation functions for squared returns can be derived and is plotted in Figure 2-b; see Appendix A.1 for the expressions. In the semi-Markov case, I approximate the model-implied ACF for squared returns by simulating a sample of 40000 data points. A polynomial is fitted and superimposed onto each graph. Take DJIA Top 4, Figure 2-a shows that the semi-Markov model can explain a substantial amount of the ARCH effect as the residuals are very much whitened.

In Figure 2-b, it is clear that in a standard Markov-switching model, the model-implied ACF for squared returns decreases too fast due to the geometric declining nature of the duration probability. Bulla and Bulla (2006) find the same result in their univariate analysis.

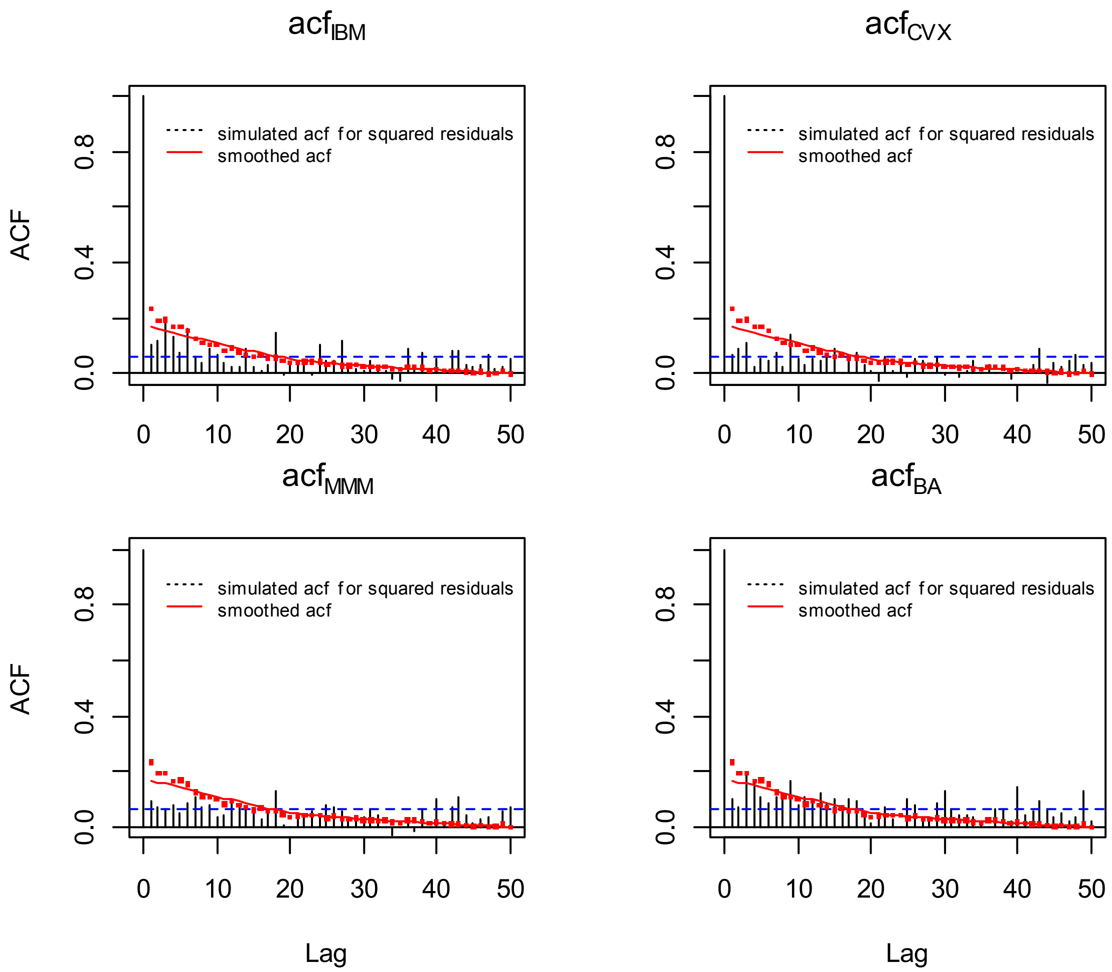

Figure 2-a. Model-implied ACF of squared returns and sample ACF of semi-Markov model residuals, DJIA Top 4, m=3

One remaining question is how well the states are identified since the high and low volatility states may be poorly separated thus confounding the comovement measures across states. Three commonly used methods in this area are filtering, smoothing and Viterbi's algorithm. I take the filtered (smoothed) state to be the one that has the highest filtered (smoothed) probability; in Viterbi's algorithm, a single most probable state is produced at each point in time as a byproduct of the global maximum likelihood. In the following simulation study, I compare filtered states, smoothed states and global most probable path inference by computing the percentage of correct matches for the case of $m = 3$. Table 3 is based on 1000 simulated samples of size 1000 from the corresponding estimated models. A match is said to occur if the estimated state coincides with the true state.

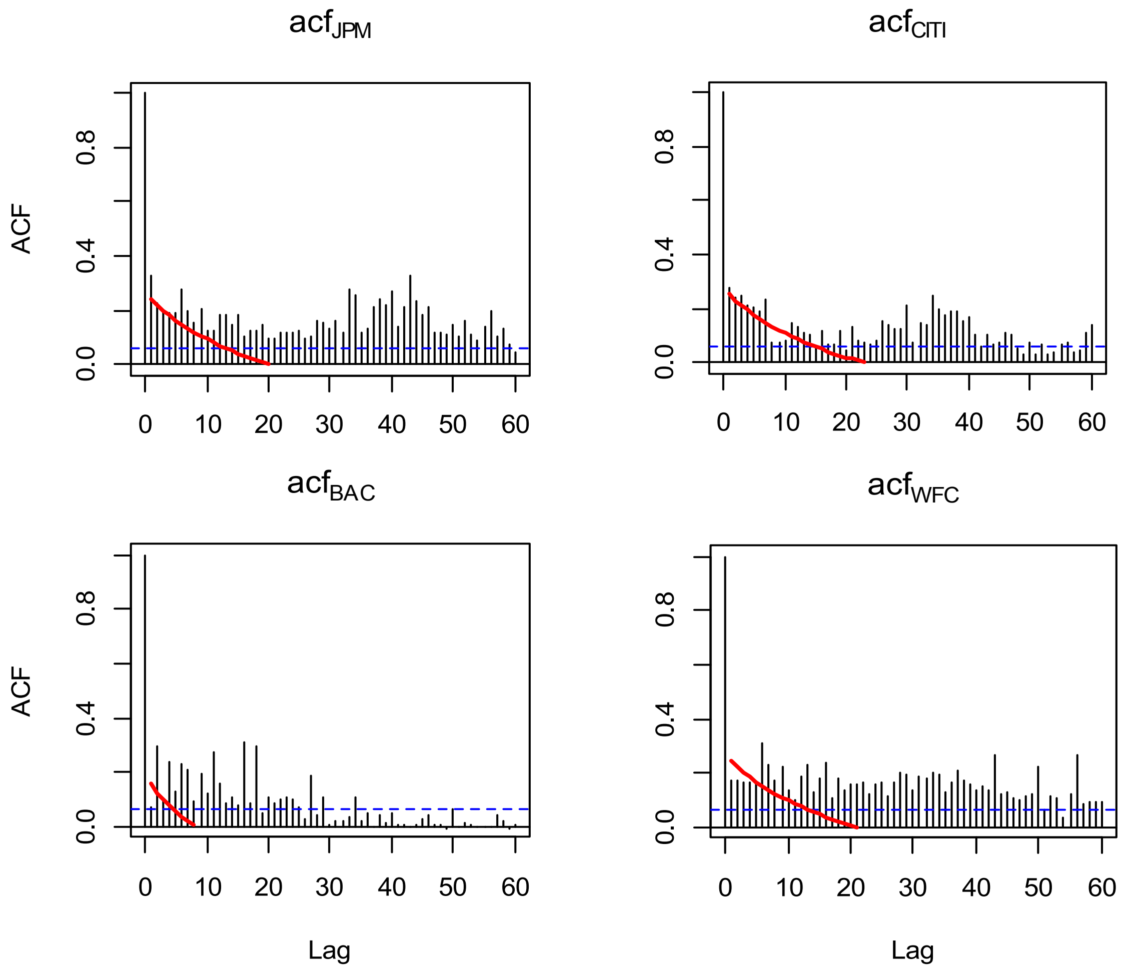

Figure 2-b. Sample ACF and exact ACF of squared returns of the Markov-switching model, NYSE Fin Top 4, m=3

It can be seen from Table 3 that the medium volatility state is the hardest to estimate. All three methods are good at estimating the low volatility state, while smoothed inference and Viterbi's algorithm perform better for the high and medium volatility states. The pattern changes little regardless of the model specification used and using smoothed or Viterbi states does not make any noticeable difference. In Section 5, results are based on global most probable states using Viterbi's algorithm.

Table 3

Percentages of state matching for the EM-sMSw model, m=3.

	low vol. state			medium vol. state			high vol. state		
	filter	smth	Vit	filter	smth	Vit	filter	smth	Vit
NYSE Fin Top 4	0.953	0.958	0.958	0.788	0.827	0.827	0.939	0.954	0.953
DAX 30 Top 4	0.948	0.949	0.951	0.767	0.816	0.819	0.917	0.93	0.93
DJIA Top 4	0.95	0.954	0.955	0.801	0.844	0.842	0.933	0.947	0.947

2.5 FULL SAMPLE ANALYSIS

Thus far the findings are mixed: while for DJIA Top 4 the result stands in line with the prevailing view, that for NYSE Fin Top 4 and DAX 30 Top 4 does not. To study this issue further, I conduct a full sample analysis in this section. My motivation is two-fold. First, in Section 5, it is shown that $Cm(\text{low vol, high return})$ and $Cm(\text{high vol, low return})$ for NYSE Fin Top 4 over the crisis period are very close to each other. Most likely, the contranatant bank, e.g., Wells Fargo, played an important role in mitigating part of the comovement, which I now investigate in a sample of 15 years and 10 major stocks. Second, it is important to check whether earlier results are robust to different samples and can be generalized to other industries.

The results are collected in Table 4-a, b and c. MSwC picks m = 2 in most cases and occasionally m = 3; BIC picks either m = 3 or m = 4; AIC and AIC_c always pick m = 4. As in Table 1, the estimated state-dependent mean returns for individual stocks are not always ordered from high to low because the former is known to be noisy and might be temporarily countercyclical as opposed to their market rivals. This result is in accordance with the finding by some authors that the leverage effect is more salient for market or portfolio returns than for individual stocks. The state-dependent market mean returns are computed by matching the global most probable path estimated from the model with the CRSP value-weighted returns. In all cases, the use of a semi-Markov-switching enables one to order market returns and volatilities in a way that is consistent with the leverage effect. Table 4-a is based on five different indices that can be largely summarized as $\mathrm{Cm(low\ vol)} \geq \mathrm{Cm(\ high\ vol)}$. On account of Figure 1 and Wald tests of parameter equality in Section 5, the ordering is very significant for DJT Top 10 and FTSE Top 10 but insignificant for NYSE Energy Top 20.

The recent crisis is unlike most other post-war recessions in the U.S. as it was triggered not by supply or monetary/fiscal policy shocks, but by a decade-long deleveraging. For this reason and because certain contranatant banks may have exerted a large impact on the dynamics of stock returns, the financial sector over the recent crisis period deserves a close-up. To see whether the pattern of comovement and returns in the financial sector can be generalized to other times, I report in Table 4-b results based on two samples, one of which is constructed prior to 2008.

Table 4-a

Parameter estimates of the full sample EM-sMSw model, Jan. 1, 1998-Dec. 31, 2012

	EM-sMSw			EM-sMSw (Out. Red.)		
	m=2	m=3	m=4	m=2	m=3	m=4
NYSE Energy Top 10						
Cm	[0.71, 0.69]	[0.71, 0.66, 0.73]	[0.71, …, 0.77]	[0.71, 0.67]	[0.71, 0.69, 0.67]	[0.71, …, 0.71]
Cm_{diff}	[0.69, 0.67]	[0.70, 0.63, 0.72]	[0.69, …, 0.76]	[0.69, 065]	[0.69, 0.67, 0.65]	[0.70, …, 0.70]
Cm_{avg}	[0.67, 0.65]	[0.68, 0.62, 0.70]	[0.68, …, 0.74]	[0.68, 0.64]	[0.68, 0.65, 0.63]	[0.68, …, 0.68]
r_{CRSP}		[high, ..., low]			[high, ..., low]	
σ_{CRSP}		[low, ..., high]			[low, ..., high]	
σ		[low, ..., high]			[low, ..., high]	
BIC	No	No	Yes	No	Yes	No
MSwC	Yes	No	No	Yes	No	No
NYSE Health Care Top 10						
Cm	[0.47, 0.43]	[0.48, 0.40, 0.48]	[0.50, …, 0.49]	[0.47, 0.44]	[0.50, 0.41, 0.46]	[0.50, …, 0.55]
Cm_{diff}	[0.43, 0.38]	[0.45, 0.36, 0.44]	[0.47, …, 0.46]	[0.43, 0.40]	[0.47, 0.37, 0.42]	[0.47, …, 0.53]
Cm_{avg}	[0.40, 0.35]	[0.42, 0.32, 0.41]	[0.44, …, 0.43]	[0.40, 0.37]	[0.44, 0.33, 0.38]	[0.44, …, 0.50]
r_{CRSP}	[high, ..., low]				[high, ..., low]	
σ_{CRSP}	[low, ..., high]				[low, ..., high]	
σ	[low, ..., high]				[low, ..., high]	
BIC	No	No	Yes	No	Yes	No
MSwC	Yes	No	No	Yes	No	No
DJT Top 10						
Cm	[0.56, 0.40]	[0.60, 0.43, 0.46]	[0.63, …, 0.48]	[0.59, 0.45]	[0.63, 0.52, 0.43]	[0.69, …, 0.64]
Cm_{diff}	[0.54, 0.37]	[0.58, 0.39, 0.44]	[0.62, …, 0.48]	[0.57, 0.42]	[0.61, 0.50, 0.40]	[0.67, …, 0.63]
Cm_{avg}	[0.51, 0.32]**	[0.55, 0.36, 0.38]**	[0.59, …, 0.38]**	[0.54, 0.38]**	[0.58, 0.47, 0.37]**	[0.65, …, 0.60]
r_{CRSP}		[high, ..., low]			[high, ..., low]	
σ_{CRSP}		[low, ..., high]			[low, ..., high]	
σ		[low, ..., high]			[low, ..., high]	
BIC	No	No	Yes	No	No	Yes
MSwC	No	Yes	No	Yes	No	No
DJU Top 10						
Cm	[0.62, 0.56]	[0.64, 0.56, 0.59]	[0.64, …, 0.62]	[0.63, 0.58]	[0.65, 0.58, 0.59]	[0.63, …, 0.60]
Cm_{diff}	[0.59, 0.53]	[0.61, 0.53, 0.56]	[0.62, …, 0.60]	[0.60, 0.55]	[0.63, 0.55, 0.56]	[0.61, …, 0.58]
Cm_{avg}	[0.57, 0.50]*	[0.59, 0.50, 0.52]*	[0.60, …, 0.55]	[0.58, 0.52]*	[0.60, 0.52, 0.53]*	[0.58, …, 0.55]
r_{CRSP}		[high, ..., low]			[high, ..., low]	
σ_{CRSP}		[low, ..., high]			[low, ..., high]	
σ		[low, ..., high]			[low, ..., high]	
BIC	No	No	Yes	No	Yes	No
MSwC	No	Yes	No	Yes	No	No
FTSE Top 10						
Cm	[0.45, 0.32]	[0.46, 0.37, 0.32]	[0.46, …, 0.30]	[0.46, 0.39]	[0.45, 0.40, 0.40]	[0.41, …, 0.40]
Cm_{diff}	[0.44, 0.29]	[0.45, 0.34, 0.29]	[0.44, …, 0.30]	[0.45, 0.36]	[0.44, 0.37, 0.38]	[0.40, …, 0.38]
Cm_{avg}	[0.38, 0.24]**	[0.39, 0.29, 0.22]**	[0.38, …, 0.11]**	[0.39, 0.32]*	[0.38, 0.32, 0.33]*	[0.34, …, 0.33]
r_{CRSP}		[high, ..., low]			[high, ..., low]	
σ_{CRSP}		[low, ..., high]			[low, ..., high]	
σ		[low, ..., high]			[low, ..., high]	
BIC	No	No	Yes	No	No	Yes
MSwC	No	Yes	No	Yes	No	No

Note: * and ** denote the 10% and 5% significance level respectively. AIC and AIC_c always pick the four-regime model.

For the original sample covering the period from 1 January 1998 to 31 December 2012, comovement is found to be significantly larger in the low return and high volatility state; for the sample prior to 2008, the orderings become insignificant in both the raw data and the outlier-reduced sample. The result given in the top panel of Table 4-b is in line with the traditional view and stands in sharp contrast to the rest of the study. However, it does not easily fit into the existent literature because this is the only sample in which there are five stocks (WFC, JPM, BRKB, RY, TD) whose high volatility state returns turn out to be the highest and much higher than the low volatility state returns. So the common wisdom of leverage effect breaks down miserably for these stocks. Apparently, the high performers as a block managed increase the overall comovement over this period.

Table 4-b

Parameter estimates in the pre-crisis vs. crisis sample, NYSE Fin Top 10.

	EM-sMSw			EM-sMSw (Out. Red.)		
	m=2	m=3	m=4	m=2	m=3	m=4
			NYSE Financial Top 10: Jan. 1, 1998-Dec. 31, 2012			
Cm	[0.54, 0.61]	[0.49, 0.56, 0.64]	[0.46, …, 0.66]	[0.52, 0.60]	[0.46, 0.57, 0.62]	[0.46, …, 0.67]
Cm_{diff}	[0.51, 0.59]	[0.47, 0.53, 0.62]	[0.43, …, 0.66]	[0.50, 0.58]	[0.43, 0.55, 0.60]	[0.43, …, 0.66]
Cm_{avg}	[0.47, 0.56]*	[0.42, 0.50, 0.59]**	[0.36, …, 0.62]**	[0.45, 0.55]*	[0.37, 0.51, 0.57]**	[0.36, …, 0.63]**
r_{CRSP}		[high, ..., low]			[high, ..., low]	
σ_{CRSP}		[low, ..., high]			[low, ..., high]	
σ		[low, ..., high]			[low, ..., high]	
BIC	No	No	Yes	No	No	Yes
MSwC	No	Yes	No	Yes	No	No
			NYSE Financial Top 10: Jan. 1, 1993-Dec. 31, 2007			
Cm	[0.52, 0.49]	[0.54, 0.47, 0.52]	[0.54, …, 0.53]	[0.53, 0.50]	[0.54, 0.47, 0.53]	[0.52, …, 0.53]
Cm_{diff}	[0.49, 0.46]	[0.51, 0.43, 0.50]	[0.51, … 0.51]	[0.50, 0.47]	[0.51, 0.43, 0.51]	[0.50, …, 0.50]
Cm_{avg}	[0.46, 0.41]	[0.47, 0.38, 0.45]	[0.47, …, 0.47]	[0.46, 0.42]	[0.47, 0.39, 0.45]	[0.45, …, 0.45]
r_{CRSP}		[high, ..., low]			[high, ..., low]	
σ_{CRSP}		[low, ..., high]			[low, ..., high]	
σ		[low, ..., high]			[low, ..., high]	
BIC	No	Yes	No	No	No	Yes
MSwC	Yes	No	No	Yes	No	No

Note: * and ** denote the 10% and 5% significance level respectively. AIC and AIC_c always pick the four-regime model.

As a last experiment, Table 4-c shows the effect of extending the sample size to 30 years. As this inevitably introduces more outliers due to historical episodes of big swings in the market, one should expect that outlier reduction may give rise to different results. This conjecture is verified in the bottom panel of Table 4-c: the hypothesized order tends to be preserved once outlier reduction is used. Overall, compared to the significant ordering in the top panel which says comovement is highest in good times, the differential becomes insignificant when the sample is extended to 30 years.

Table 4-c
Parameter estimates in the pre-crisis vs. crisis sample, DJIA Top 10.

	EM-sMSw			EM-sMSw (Out. Red.)		
	m=2	m=3	m=4	m=2	m=3	m=4
			DJI Top 10: Jan. 1, 1998-Dec. 31, 2012			
Cm	[0.51, 0.43]	[0.54, 0.43, 0.44]	[0.54, …, 0.51]	[0.52, 0.44]	[0.54, 0.47, 0.43]	[0.64, …, 0.43]
Cm_{diff}	[0.48, 0.39]	[0.51, 0.39, 0.40]	[0.51, …, 0.48]	[0.48, 0.39]	[0.51, 0.43, 0.39]	[0.62, …, 0.39]
Cm_{avg}	[0.45, 0.36]*	[0.49, 0.36, 0.37]**	[0.48, …, 0.44]	[0.46, 0.37]*	[0.48, 0.40, 0.36]**	[0.60, …, 0.37]**
r_{CRSP}		[high, ..., low]			[high, ..., low]	
σ_{CRSP}		[low, ..., high]			[low, ..., high]	
σ		[low, ..., high]			[low, ..., high]	
BIC	No	No	Yes	No	No	Yes
MSwC	Yes	No	No	Yes	No	No
			DJI Top 10: Jan. 1, 1983-Dec. 31, 2012			
Cm	[0.45, 0.49]	[0.52, 0.42, 0.53]	[0.58, …, 0.60]	[0.47, 0.46]	[0.55, 0.41, 0.50]	[0.59, …, 0.55]
Cm_{diff}	[0.41, 0.46]	[0.48, 0.37, 0.49]	[0.54, …, 0.59]	[0.44, 0.41]	[0.51, 0.36, 0.46]	[0.56, …, 0.52]
Cm_{avg}	[0.39, 0.43]	[0.46, 0.35, 0.47]	[0.52, …, 0.55]	[0.41, 0.39]	[0.50, 0.34, 0.44]*	[0.54, …, 0.49]
r_{CRSP}		[high, ..., low]			[high, ..., low]	
σ_{CRSP}		[low, ..., high]			[low, ..., high]	
σ		[low, ..., high]			[low, ..., high]	
BIC	No	No	Yes	No	No	Yes
MSwC	Yes	No	No	Yes	No	No

Note: * and ** denote the 10% and 5% significance level respectively. AIC and AIC_c always pick the four-regime model.

It is worth noting that all models in this section have satisfactorily identified the high volatility periods of the three most recent economic downturns in the data, including the two mild and brief recessions of 1990-1991 and 2001. Other than these historical episodes, the sample period is largely occupied by low and medium volatility states corresponding to what is commonly known as the *Great Moderation*. In a study of market returns using duration-dependent Markov-switching models, Maheu and McCurdy (2000) find that the hazard function of a duration-dependent transition probability is declining in both the bull (low volatility state) and bear (high volatility) market and is declining faster for the latter (see their Fig. 6). The same result is found in the current chapter: the estimates of the shape parameter of the occupancy distribution (gamma) are smaller than one in every two-state semi-Markov-switching model, which corresponds to a declining hazard function by the property of gamma distribution. Besides, the estimate of the shape parameter in the high volatility state is much smaller than that in the low volatility state, which implies faster decline of the hazard function in the former. In a state-dependent asset allocation setting (stock and bond), Guidolin and Timmermann (2007) also find evidence that the bull market has significantly longer duration than the crash or recovery state. Alternatively, one can allow for duration dependence in a time-varying transition probability model; see Durland and McC urdy (1994) for a univariate algorithm.

Taken as a whole, a balanced view of Table 4 is to focus on m = 2 and m = 3 and to look at Gaussian models using both the raw data and outlier-reduced data. In some cases the result is insignificant, while in others, I find $Cm(\text{high return}, \text{low vol}) > Cm(\text{low return}, \text{high vol})$. In Figure 3, I project the state-dependent correlation matrix onto a two-dimensional space using multivariate scaling, in which the polygons are such that the smaller the area the higher the comovement, e.g., perfect comovement is represented by one concentrated point. From these projections, one can also identify clusters of firms that are very close to each other, while others might be effectively detracting from comovement as a group.

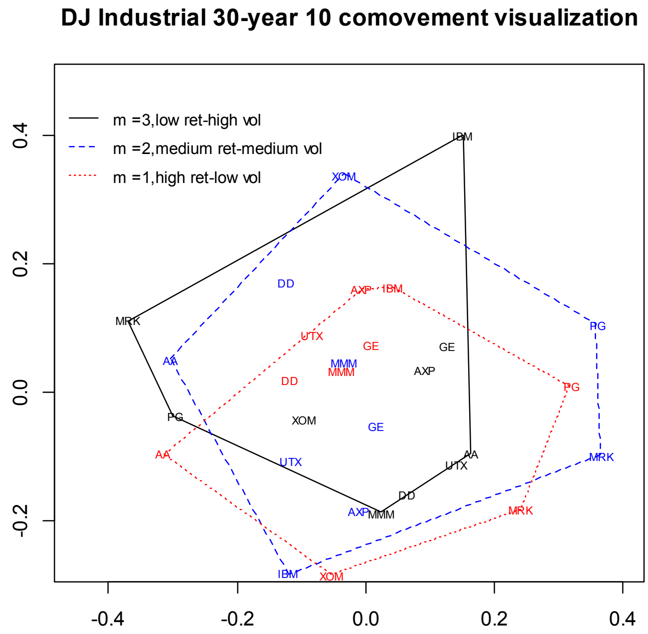

Figure 2-a. Comovement visualization using multidimensional scaling for DJIA 30-yr, m = 3

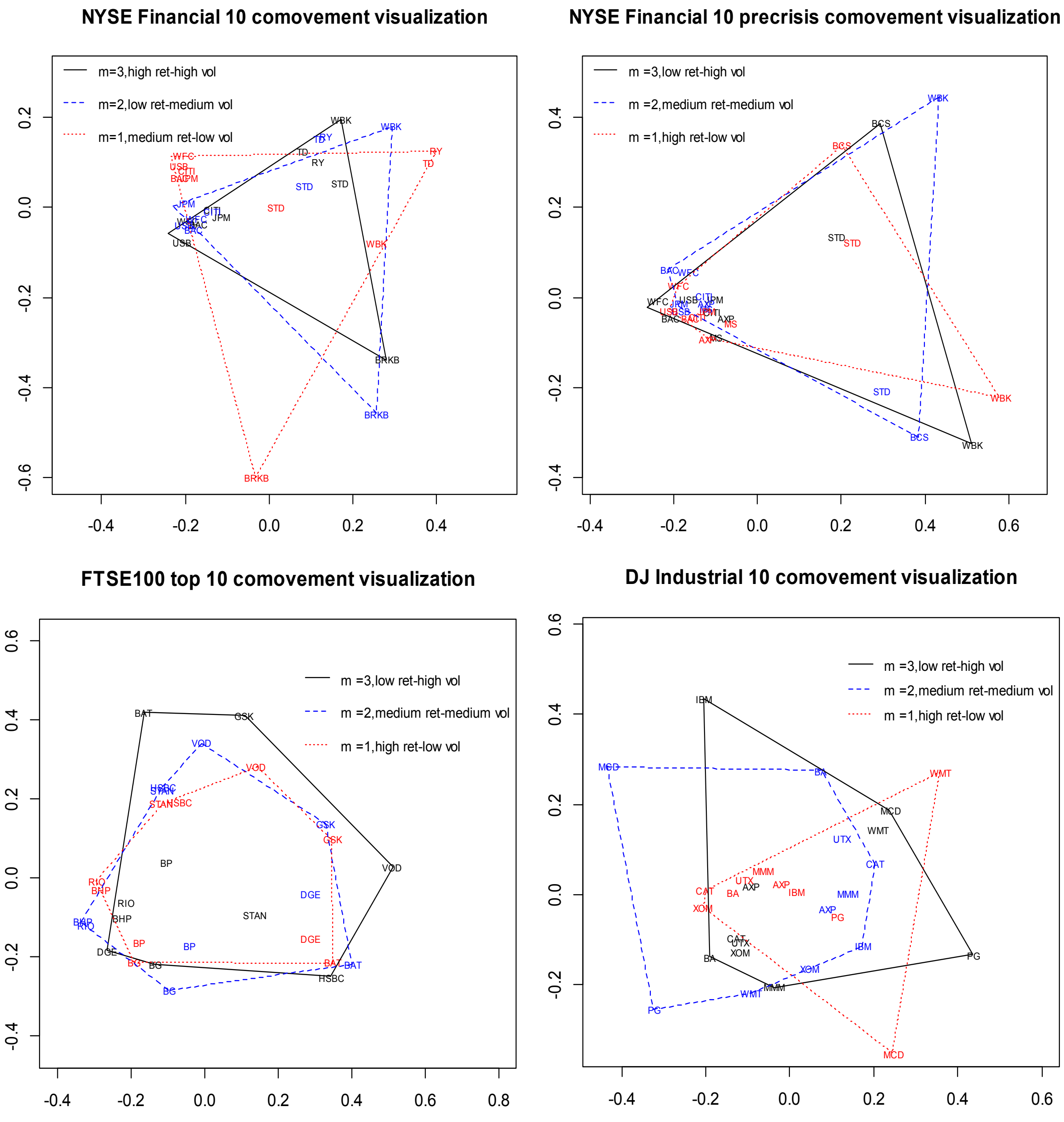

Figure 3-b. Comovement visualization using multidimensional scaling for selected models, m = 3.

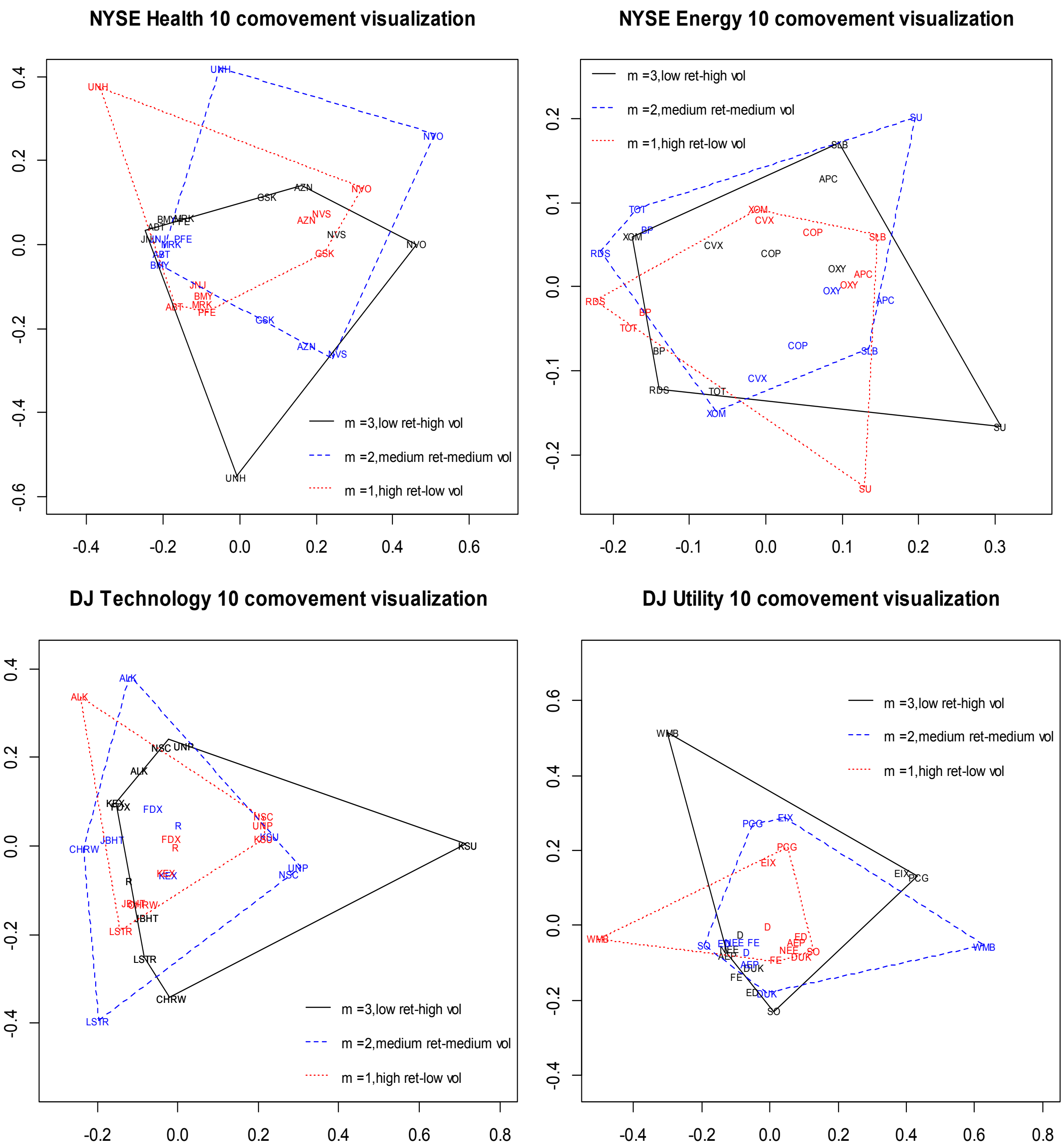

Figure 3-c. Comovement visualization using multidimensional scaling for selected models, m = 3.

2.6 Concluding Remarks

In retrospect, previous authors have found somewhat similar results but with a different focus. Ang and Chen (2002) find more prevailing asymmetric comovements in portfolios sorted by small stocks, value stocks and past loss stocks. Later authors such as Hong, Tu, and Zhou (2007) derive statistical tests for asymmetric correlation in portfolios. However, papers along this line are mostly based on pairwise correlation comparisons and it is found that such asymmetry is not significant for larger size stocks. In this respect, I have set up a unified modeling environment to provide a more detailed look at the patterns of comovement in different industries. The study makes use of an eigenvalue-based measure which can be applied to a group of assets instead of just two.

The semi-Markov approach is evaluated against the standard Markov-switching benchmark and is superior in terms of capturing the ARCH effect, estimation stability and modeling many other stylized facts. It also paves the way for a full-fledged analysis of financial data and for forecasting purposed as well. After accounting for the financial sector over the recent crisis period, in both the two-regime and three-regime models, the finding of this chapter uniformly casts doubt on the common lore which says comovement is highest when market is down and volatile. Rather, the ordering is found to be either insignificant or significant in the other way round. Except for a very short event window at the beginning of a large market meltdown—which, understandably, catches everyone off guard—all three measures proposed in this study suggest that, for large-caps, evidence in favor of higher comovement in bad times may not be as obvious as previously thought and should be evaluated case by case.

2.7 Appendix A: Moments and Benchmark Model

2.7.1 *Exact marginal moments of Markov-switching models*

Let P be the transition matrix, $p_{ij}(k)$ the (i,j) element of P^k. Denote the stationary distribution and conditional mean vector as $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_m)'$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_m)'$.

$$\mu = \sum_{i=1}^{m} \mu_i \pi_i,$$

$$Var(X_t) = \sum_{i=1}^{m} (\sigma_i^2 + \mu_i^2)\pi_i - \left(\sum_{i=1}^{m} \mu_i \pi_i\right)^2,$$

$$\begin{aligned} Cov(X_t X_{t+k}) &= E(X_t X_{t+k}) - \mu^2 = \sum_{i,j=1}^{m} E(X_t|i)\pi_i p_{ij}(k) E(X_{t+k}|j) - \mu^2 \\ &= \boldsymbol{\pi}' diag(\boldsymbol{\mu}) P^k \boldsymbol{\mu} - (\boldsymbol{\pi}'\boldsymbol{\mu})^2, \end{aligned}$$

$$Kurtosis = \frac{E(X_t - \mu)^4}{Var^2(X_t)} = \frac{E(X_t^4) - 4\mu E(X_t^3) + 6\mu^2 E(X_t^2) - 3\mu^4}{(\sum_{i=1}^{m}(\sigma_i^2 + \mu_i^2)\pi_i - (\sum_{i=1}^{m} \mu_i \pi_i)^2)^2}$$

$$= \frac{\sum_{i=1}^{m} \pi_i(\mu_i^4 + 6\mu_i^2\pi_i^2 + 3\sigma_i^4) - 4\mu \sum_{i=1}^{m} \pi_i(\mu_i^3 + 3\mu_i\sigma_i^2) + 6\mu^2 \sum_{i=1}^{m} \pi_i(\sigma_i^2 + \mu_i^2) - 3\mu^4}{(\sum_{i=1}^{m}(\sigma_i^2 + \mu_i^2)\pi_i - (\boldsymbol{\pi}'\boldsymbol{\mu})^2)^2},$$

$$\rho_k = \frac{Cov(X_t X_{t+k})}{Var(X_t)} = \frac{\boldsymbol{\pi}' diag(\boldsymbol{\mu}) P^k \boldsymbol{\mu} - (\boldsymbol{\pi}'\boldsymbol{\mu})^2}{\sum_{i=1}^{m}(\sigma_i^2 + \mu_i^2)\pi_i - (\boldsymbol{\pi}'\boldsymbol{\mu})^2},$$

$$\rho_k^{sqr} = \frac{Cov(X_t^2 X_{t+k}^2)}{Var(X_t^2)} = \frac{E(X_t^2 X_{t+k}^2) - E^2(X_t^2)}{E(X_t^4) - E^2(X_t^2)},$$

$$E(X_t^2 X_{t+k}^2) = \sum_{i,j=1}^{m} E(X_t^2|i)\pi_i p_{ij}(k) E(X_{t+k}^2|j).$$

The last expression can be computed from the conditional second moment of the state-dependent distribution.

2.7.2 *Estimating the benchmark model*

Using notation of Section 2.1 and treating the unobserved states as missing data one gets,

$$L(T) = P\left(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}\right) = \sum_{s_1, s_2, \ldots, s_T \in \{1,2\}} P(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, \boldsymbol{S}^{(T)} = \boldsymbol{s}^{(T)}),$$

and by the strong Markov property, a particular realization of states contributes to the likelihood by as much as

$$P\left(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, \boldsymbol{S}^{(T)} = \boldsymbol{s}^{(T)}\right) = \pi(s_1)P(x_1|s_1)p(s_2|s_1)P(x_2|s_2) \dots p(s_T|s_{T-1})P(x_T|s_T).$$

A full permutation of the histories of states thus yields

$$\begin{aligned} L(T) &= \sum_{s_1, s_2, \dots, s_T \in \{1,2\}} \pi(s_1)P(x_1|s_1)p(s_2|s_1) \dots p(s_T|s_{T-1})P(x_T|s_T) \\ &= \boldsymbol{\pi}' P(x_1)\Gamma P(x_2) \dots \Gamma P(x_n)\mathbf{1}, \end{aligned}$$

in which Γ is the $m \times m$ transition matrix, $P(\cdot)$ is the diagonal conditional probability matrix of observation x_t given each possible state s_t. For a general homogenous and irreducible m-state conditionally serially uncorrelated Markov-switching model, the likelihood function can be written conveniently in matrix form as

$$L(\boldsymbol{\theta}) = \boldsymbol{\pi}' P(x_1)\Gamma P(x_2) \dots \Gamma P(x_n)\mathbf{1}.$$

Thus a model of n assets has in total $m(m - 1 + n(n + 3)/2)$ parameters: $m^2 - m$ to determine Γ, mn for the state-dependent means, and $mn(n + 1)/2$ for the state-dependent variances and covariances. These parameters are collected in $\boldsymbol{\theta}$. The transition probability matrix dictates that $p_{ii}^{u-1}(1 - p_{ii})$ is the duration probability of staying in state i for u periods, $u = 1,2, \dots$

A variation on the ML-MSw model is the EM algorithm based on the reknowned Baum-Welch forward-backward procedure. The EM algorithm can be used to check the influence of initial state on subsequent estimates. It does not evaluate the likelihood function directly and is fastest when the E-step can be broken down into a collection of terms involving subsets of parameters of interest, each of which is simple to maximize either numerically, or even better, analytically. This is true when stationarity is not imposed on the initial state but not otherwise. In the latter case, the E-step cannot be neatly seperated and no analytic solution exists. The E step and M step in the basic model are simplified versions of the semi-Markov-switching algorithm and can be found in MacDonald and Zucchini (1997). To implement direct ML-MSw, I assume that the process starts from its stationary distribution, while to implement the EM algorithm, I do not impose this restriction.

To avoid boundary problems, I transform natural parameters into working ones before each run of estimation. Thus the algorithm returns a Hessian in terms of working parameters. I then recover the variance-covariance matrix with respect to natural parameters using the sandwich estimator:

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\theta}}) = \widehat{\mathrm{H}}^{-1}(\widehat{\boldsymbol{\theta}})\hat{I}_{BHHH}(\widehat{\boldsymbol{\theta}})\widehat{\mathrm{H}}^{-1}(\widehat{\boldsymbol{\theta}}),$$

Let $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ denote the vector of working parameters and natural parameters respectively, then

$$I = E(G'G) = \sum_{t=1}^{n} E(G_t G_t') = E(\boldsymbol{g}\boldsymbol{g}'),$$

$$\hat{I}_{BHHH} = G'(\widehat{\boldsymbol{\theta}})G(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{\partial l_1}{\partial \theta_1} & \cdots & \frac{\partial l_T}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial l_1}{\partial \theta_n} & \cdots & \frac{\partial l_T}{\partial \theta_n} \end{pmatrix}_{\widehat{\boldsymbol{\theta}}}' \begin{pmatrix} \frac{\partial l_1}{\partial \theta_1} & \cdots & \frac{\partial l_T}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial l_1}{\partial \theta_n} & \cdots & \frac{\partial l_T}{\partial \theta_n} \end{pmatrix}_{\widehat{\boldsymbol{\theta}}},$$

$$\frac{\partial \boldsymbol{L}_{1:T}}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\phi}'}{\partial \boldsymbol{\theta}}\frac{\partial \boldsymbol{L}_{1:T}}{\partial \boldsymbol{\phi}} = \begin{pmatrix} \frac{\partial \phi_1}{\partial \theta_1} & \cdots & \frac{\partial \phi_n}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_1}{\partial \theta_n} & \cdots & \frac{\partial \phi_n}{\partial \theta_n} \end{pmatrix} \begin{pmatrix} \frac{\partial l_1}{\partial \phi_1} & \cdots & \frac{\partial l_T}{\partial \phi_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial l_1}{\partial \phi_n} & \cdots & \frac{\partial l_T}{\partial \phi_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial l_1}{\partial \theta_1} & \cdots & \frac{\partial l_T}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial l_1}{\partial \theta_n} & \cdots & \frac{\partial l_T}{\partial \theta_n} \end{pmatrix},$$

$$\widehat{\mathrm{H}}(\widehat{\boldsymbol{\theta}}) = \left(\frac{\partial \boldsymbol{\phi}'}{\partial \boldsymbol{\theta}}\widehat{\mathrm{H}}(\widehat{\boldsymbol{\phi}})\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\theta}}\right)_{\widehat{\boldsymbol{\theta}}}.$$

As an example, in the case of bivariate normal MSw model with two states, $m = 2$, the natural parameters are

$$\boldsymbol{\theta} = (\mu_1^1,\ \mu_1^2,\ \mu_2^1,\ \mu_2^2,\ \sigma_1^1,\ \sigma_1^2,\ \sigma_2^1,\ \sigma_2^2,\ \rho^1,\ \rho^2,\ p_{21},\ p_{12})',$$

and the working parameters become

$$\boldsymbol{\phi} = \left(\ldots, \log\sigma_1^1, \log\sigma_1^2, \log\sigma_2^1, \log\sigma_2^2, \log\frac{(1+\rho^1)}{(1-\rho^1)}, \log\frac{(1+\rho^2)}{(1-\rho^2)}, \log\frac{p_{21}}{(1-p_{21})}, \log\frac{p_{12}}{(1-p_{12})}\right)'.$$

2.8 Appendix B: ML-based Results

Table B.1 collects results of the ML-based estimates, where I report three variations for each choice of m and study the impact of outlier reduction and t-distribution. In the case of multivariate t-distribution, the estimated degrees of freedom (not switching to be parsimonious) is larger than four in all cases and larger than six for outlier-reduced data.

In all cases, MSwC picks m = 2; BIC picks m = 3 (at least marginally); AIC and AIC_c pick m = 4. For NYSE Fin Top 4, different estimation methods and outlier reduction do not have a noticeable impact on the result. In the case of m = 4 for DJIA Top 4 and DAX 30 Top 4, when outlier reduction is used, the highest volatility state does not always correspond to the lowest return and neither does the lowest volatility state to the highest return.

Table B.1

Parameter estimates of ML-MSw models, Jan. 1, 2007-Dec. 31, 2010.

	m=2			m=3			m=4		
	Normal	Out. Red.	t and Red.	Normal	Out. Red.	t and Red.	Normal	Out. Red.	t and Red.
					NYSE Financial Top 4				
Cm	[0.72, 0.71]	[0.72, 0.72]	[0.72, 0.72]	[0.73, 0.70, 0.72]	[0.73, 0.71, 0.73]	[0.74, 0.71, 0.73]	[0.76, …, 0.73]	[0.77, …, 0.73]	[0.76, …, 0.74]
Cm_{diff}	[0.68, 0.67]	[0.68, 0.68]	[0.68, 0.68]	[0.68, 0.67, 0.68]	[0.69, 0.67, 0.69]	[0.69, 0.68, 0.69]	[0.72, …, 0.69]	[0.73, …, 0.70]	[0.71, …, 0.70]
Cm_{avg}	[0.62, 0.61]	[0.62, 0.61]	[0.61, 0.61]	[0.62, 0.59, 0.62]	[0.62, 0.60, 0.62]	[0.64, 0.60, 0.62]	[0.67, …, 0.62]	[0.69, …, 0.64]	[0.67, …, 0.64]
r_{CRSP}		[high, low]			[high, medium, low]		[high, ..., low]	[high, …, low, ...]	[high, ..., low]
σ_{CRSP}		[low, high]			[low, medium, high]		[low, ..., high]	[low, ..., high]	[low, ..., high]
σ		[low, high]			[low, medium, high]		[low, ..., high]	[low, ..., high]	[low, ..., high]
p-value (Cm_{avg})		0.634			0.958			0.523	
BIC		No			Yes			No	
MSwC		Yes			No			No	
					DJIA Top 4				
Cm	[0.67, 0.72]	[0.67, 0.71]	[0.64, 0.76]	[0.68, 0.63, 0.78]	[0.74, 0.60, 0.78]	[0.69, 0.60, 0.82]	[0.72, …, 0.82]	0.72, …, 0.80, …	[0.74, …, 0.82]
Cm_{diff}	[0.58, 0.65]	[0.58, 0.63]	[0.55, 0.69]	[0.60, 0.54, 0.73]	[0.67, 0.50, 0.73]	[0.61, 0.51, 0.78]	[0.65, …, 0.78]	0.65, …, 0.75, …	[0.67, …, 0.77]
Cm_{avg}	[0.56, 0.62]	[0.56, 0.61]	[0.52, 0.68]	[0.58, 0.50, 0.70]	[0.65, 0.46, 0.70]	[0.58, 0.47, 0.76]	[0.63, …, 0.76]	0.63, …, 0.73, …	[0.66, …, 0.75]
r_{CRSP}		[high, low]			[high, medium, low]		[high, ..., low]	[..., low, high]	[high, ..., low]
σ_{CRSP}		[low, high]			[low, medium, high]		[low, ..., high]	[low, ..., high, ...]	[low, ..., high]
σ		[low, high]			[low, medium, high]		[low, ..., high]	[low, ..., high, ...]	[low, ..., high]
p-value (Cm_{avg})		0.107			0.006			0.044	
BIC		No			Yes			No	
MSwC		Yes			No			No	
					DAX 30 Top 4				
Cm	[0.66, 0.60]	[0.67, 0.66]	[0.65, 0.76]	[0.69, 0.66, 0.57]	[0.72, 0.61, 0.72]	[0.61, 0.74, 0.77]	[0.72, …, 0.51]	[0.72, …, 0.75, ...	[0.63, …, 0.85]
Cm_{diff}	[0.59, 0.52]	[0.59, 0.59]	[0.57, 0.72]	[0.62, 0.59, 0.49]	[0.66, 0.53, 0.67]	[0.53, 0.70, 0.72]	[0.65, …, 0.57]	[0.66, …, 0.71, ...	[0.55, …, 0.83]
Cm_{avg}	[0.54, 0.46]	[0.55, 0.54]	[0.53, 0.68]	[0.59, 0.55, 0.41]	[0.62, 0.47, 0.63]	[0.48, 0.65, 0.69]	[0.62, …, 0.48]	[0.62, …, 0.67, ...	[0.51, …, 0.80]
r_{CRSP}		[high, low]			[high, medium, low]		[high, ..., low]	[..., low, high]	[high, ..., low]
σ_{CRSP}		[low, high]			[low, medium, high]		[low, ..., high]	[low, ..., high, ...]	[low, ..., high]
aa		[low, high]			[low, medium, high]		[low, ..., high]	[low, ..., high, ...]	[low, ..., high]
p-value (Cm_{avg})		0.083			0.072			0.101	
BIC		No			Yes			No	
MSwC		Yes			No			No	

Note: p-value is from a Wald test of equal Cm_{avg} across states using the Normal specification. AIC and AIC_c always pick the four-regime model.

2.9 APPENDIX C: DATA SUMMARY

In this appendix I report the ticker symbols and summary statistics for the four-year, the 15-year and the 30-year samples. Data are multiplied by 100.

AA	Alcoa Inc.	AEP	American Electric Power Co.	ABT	Abbott Laboratories
ALK	Alaska Air Group, Inc.	AXP	American Express Co.	AZN	AstraZeneca PLC
AXP	American Express Co.	APC	Anadarko Petroleum Corp.	BA	The Boeing Co.
BAC	Bank of America Corp.	BP	BP PLC	BATS.L	British American Tobacco PLC
BCS	Barclays PLC	CVX	Chevron Corp.	BG.L	BG Group PLC
BRK-B	Berkshire Hathaway Inc.	COP	ConocoPhillips	BLT.L	BHP Billiton PLC
C	Citigroup Inc.	ED	Consolidated Edison Inc.	BMY	Bristol-Myers Squibb Co.
CAT	Caterpillar Inc.	D	Dominion Resources, Inc.	BP.L	BP PLC
CHRW	CH Robinson Worldwide Inc.	DUK	Duke Energy Corporation	DGE.L	Diageo PLC
FDX	FedEx Corporation	DD	E.I. du Pont de Nemours & Co.	GSK	GlaxoSmithKline PLC
GE	General Electric Company	EIX	Edison International	GSK.L	GlaxoSmithKline PLC
JBHT	JB Hunt Transp Services Inc.	XOM	Exxon Mobil Corp.	HSBA.L	HSBC Holdings PLC
JPM	JPMorgan Chase & Co.	FE	FirstEnergy Corp.	JNJ	Johnson & Johnson
KEX	Kirby Corporation	IBM	IBM	MCD	McDonald's Corp.
KSU	Kansas City Southern	MS	Morgan Stanley	MMM	3M Company
LSTR	Landstar System Inc.	NEE	NextEra Energy, Inc.	MRK	Merck & Co. Inc.
MRK	Merck & Co. Inc.	OXY	Occidental Petroleum Corp.	NVO	Novo Nordisk A/S
NSC	Norfolk Southern Corp.	PCG	PG&E Corp.	NVS	Novartis AG
R	Ryder System, Inc.	RDS-B	Royal Dutch Shell PLC B	PFE	Pfizer Inc.
RY	Royal Bank of Canada	SLB	Schlumberger Ltd.	RIO.L	Rio Tinto PLC
SAN	Banco Santander S.A.	SO	Southern Company	STAN.L	Standard Chartered PLC
TD	Toronto-Dominion Bank	SU	Suncor Energy Inc.	UNH	UnitedHealth Group Inc.
UNP	Union Pacific Corporation	PG	The Procter & Gamble Co.	UTX	United Technologies Corp.
USB	U.S. Bancorp	TOT	Total S.A.	VOD.L	Vodafone Group PLC
WBK	Westpac Banking Corp.	WMT	Wal-Mart Stores Inc.	XOM	Exxon Mobil Corp.
WFC	Wells Fargo & Co.	WMB	Williams Companies, Inc.		

Table C.1

Summary statistics of the four-year samples, January 1, 2007-December 31, 2010.

	# of Obs.	Mean	S.D.	Skewness	Kurtosis	J-B p-val	L-B p-val	L-B* p-val	# of outliers
					NYSE Financial Top 4				
JPM	1007	-0.0045	3.86	0.328	7.75	0.000	0.002	0.010	2
C	1007	-0.2341	5.58	-0.277	16.5	0.000	0.001	0.007	12
BAC	1007	0.0611	3.56	0.483	5.60	0.000	0.091	0.094	10
WFC	1007	-0.0036	4.36	0.703	9.84	0.000	0.000	0.001	2
					DAX 30 Top 4				
BAS	1022	0.1244	3.17	10.0	217	0.000	0.341	0.387	3
BAYN	1022	0.0383	2.21	1.34	22.1	0.000	0.001	0.008	9
SAP	1022	-0.0001	1.85	-0.506	15.9	0.000	0.001	0.000	10
SIE	1022	0.0307	2.61	-0.113	10.9	0.000	0.074	0.051	9
					DJIA Top 4				
IBM	1007	0.0478	1.67	0.171	4.18	0.000	0.007	0.009	2
CVX	1007	0.0381	2.19	0.220	12.2	0.000	0.000	0.000	12
MMM	1007	0.0206	1.74	-0.060	4.49	0.000	0.045	0.102	10
BA	1007	-0.0210	2.27	0.221	3.60	0.000	0.100	0.070	2

Table C.2

Summary statistics of the 15-yr samples, January 1, 1993-December 31,2007.

	# of Obs.	Mean	S.D.	Skewness	Kurtosis	J-B p-val	L-B p-val	L-B* p-val	# of outliers
					NYSE Financial Top 10				
WFC	3772	0.0255	2.69	0.796	21.2	0.000	0.000	0.000	38
JPM	3772	0.0162	2.82	0.250	9.84	0.000	0.004	0.236	30
C	3772	-0.0407	3.50	-0.422	32.0	0.000	0.000	0.000	38
BAC	3772	-0.0128	3.27	-0.291	21.3	0.000	0.000	0.000	37
BRK-B	3772	0.0285	1.58	0.772	9.81	0.000	0.000	0.000	28
RY	3772	0.0584	1.76	0.009	8.33	0.000	0.236	0.219	25
TD	3772	0.0568	1.86	-0.073	5.51	0.000	0.030	0.311	20
SAN	3772	0.0216	2.72	-0.101	5.58	0.000	0.009	0.370	24
WBK	3772	0.0623	2.16	0.065	11.7	0.000	0.000	0.001	36
USB	3772	0.0149	2.54	-0.534	15.9	0.000	0.000	0.000	30
					NYSE Energy Top 10				
XOM	3772	0.0362	1.67	0.038	8.64	0.000	0.000	0.000	19
CVX	3772	0.0400	1.73	0.090	9.46	0.000	0.000	0.000	16
BP	3772	0.0163	1.90	-0.355	8.74	0.000	0.000	0.000	23
RDS-B	3772	0.0267	1.87	-0.006	6.37	0.000	0.000	0.002	20
TOT	3772	0.0321	1.91	-0.133	4.40	0.000	0.000	0.000	19
SLB	3772	0.0225	2.55	-0.275	4.21	0.000	0.003	0.010	13
COP	3772	0.0425	1.92	-0.344	5.70	0.000	0.000	0.002	23
OXY	3772	0.0551	2.22	-0.220	7.96	0.000	0.000	0.003	19
SU	3772	0.0580	2.55	-0.566	6.95	0.000	0.034	0.028	24
APC	3772	0.0456	2.68	-0.463	7.16	0.000	0.004	0.002	21
					NYSE Health Care Top 10				
JNJ	3772	0.0293	1.36	-0.322	11.7	0.000	0.000	0.000	17
PFE	3772	0.0115	1.84	-0.227	3.98	0.000	0.000	0.000	21
NVS	3772	0.0199	1.45	0.159	3.85	0.000	0.002	0.009	13
MRK	3772	0.0076	1.92	-1.39	23.7	0.000	0.075	0.053	25

GSK	3772	0.0125	1.70	0.022	5.68	0.000	0.008	0.083	17
ABT	3772	0.0292	1.65	-0.233	7.68	0.000	0.000	0.000	17
AZN	3772	0.0213	1.84	-0.221	5.90	0.000	0.000	0.000	17
UNH	3772	0.0587	2.40	-0.704	23.6	0.000	0.013	0.008	22
NVO	3772	0.0707	2.03	-0.287	8.11	0.000	0.001	0.001	14
BMY	3772	0.0064	1.91	-0.844	13.4	0.000	0.001	0.001	19
					DJI Top 10				
IBM	3772	0.0387	1.86	-0.125	7.5	0.000	0.005	0.024	20
MMM	3772	0.0308	1.63	0.050	3.85	0.000	0	0.000	20
BA	3772	0.0191	2.12	-0.370	6.36	0.000	0.083	0.063	14
MCD	3772	0.0427	1.68	-0.057	5.12	0.000	0.028	0.021	21
UTX	3772	0.0470	1.9	-1.05	20	0.000	0.000	0.000	12
XOM	3772	0.0362	1.67	0.038	8.64	0.000	0.000	0.000	19
CAT	3772	0.0443	2.26	-0.083	3.77	0.000	0.504	0.421	19
PG	3772	0.0228	1.57	-3.37	76.6	0.000	0.000	0.001	21
WMT	3772	0.0383	1.75	-0.103	3.86	0.000	0.000	0.000	23
AXP	3772	0.0260	2.57	0.018	7.11	0.000	0.000	0.004	29
					DJT Top 10				
ALK	3772	0.0214	3.12	-0.110	9.70	0.000	0.120	0.611	22
CHRW	3772	0.0687	2.30	0.182	6.57	0.000	0.000	0.000	19
FDX	3772	0.0306	2.15	-0.035	3.93	0.000	0.031	0.020	14
JBHT	3772	0.0718	2.80	-0.474	10.3	0.000	0.184	0.441	19
KEX	3772	0.0494	2.15	-0.023	4.87	0.000	0.007	0.015	20
KSU	3772	0.0257	5.31	-36.4	1880	0.000	0.475	0.853	1
NSC	3772	0.0296	2.25	-0.034	3.07	0.000	0.429	0.567	18
R	3772	0.0215	2.37	-0.378	5.20	0.000	0.031	0.100	20
UNP	3772	0.0438	1.94	-0.223	3.46	0.000	0.000	0.001	11
LSTR	3772	0.0741	2.40	0.006	4.63	0.000	0.000	0.010	17
					DJU Top 10				
AEP	3772	0.0154	1.69	-0.463	24.30	0.000	0.010	0.006	29
D	3772	0.0449	1.41	-0.589	9.71	0.000	0.019	0.012	17
DUK	3772	0.0255	1.67	-0.173	10.20	0.000	0.077	0.310	19
ED	3772	0.0287	1.22	0.171	4.82	0.000	0.274	0.671	20
EIX	3772	0.0242	2.39	-1.52	59.30	0.000	0.016	0.016	27
FE	3772	0.0282	1.62	0.133	10.70	0.000	0.000	0.004	27
NEE	3772	0.0371	1.51	0.200	8.77	0.000	0.003	0.008	24
PCG	3772	0.0215	2.42	-3.73	108.00	0.000	0.000	0.000	26
SO	3772	0.0488	1.31	0.278	5.67	0.000	0.000	0.030	20
WMB	3772	0.0202	3.98	-2.96	125.00	0.000	0.000	0.000	22
					FTSE 100 Top 10				
HSBA.L	3905	0.0529	2.67	19.6	843	0.000	0.334	0.321	10
BP.L	3905	0.0338	2.17	8.97	298	0.000	0.001	0.006	14
VOD.L	3905	0.0493	2.94	13.9	536	0.000	0.161	0.132	7
GSK.L	3905	0.0121	1.68	0.178	8.13	0.000	0.000	0.000	20
BATS.L	3905	0.0638	1.91	0.562	12.4	0.000	0.000	0.000	33
STAN.L	3905	0.0350	3.05	-0.517	119	0.000	0.000	0.000	22
RIO.L	3905	0.0515	2.87	-1.32	26.5	0.000	0.001	0.001	27
DGE.L	3905	0.0355	1.69	-1.13	24.9	0.000	0.000	0.000	21
BLT.L	3905	0.0812	2.66	0.043	6.78	0.000	0.000	0.000	21
BG.L	3905	0.0560	2.21	1.04	24.8	0.000	0.000	0.001	13

Table C.3

Summary statistics of the 30-yr samples and the pre-crisis sample

	# of Obs.	Mean	S.D.	Skewness	Kurtosis	J-B p-val	L-B p-val	L-B* p-val	# of outliers
				DJI Top 30 (January 1, 1983-December 31,2012)					
IBM	7565	0.0363	1.76	-0.453	14	0.000	0.211	0.570	42
MMM	7565	0.0427	1.5	-0.547	10.9	0.000	0.000	0.001	36
UTX	7565	0.0519	1.74	-0.786	16.5	0.000	0.000	0.000	28
XOM	7565	0.0568	1.54	-0.538	20.5	0.000	0.000	0.000	30
PG	7565	0.0490	1.54	-2.63	69.2	0.000	0.000	0.000	29
AXP	7565	0.0440	2.32	-0.296	10.6	0.000	0.000	0.000	45
DD	7565	0.0411	1.77	-0.284	5.94	0.000	0.011	0.006	35
MRK	7565	0.0502	1.74	-1.03	19.2	0.000	0.001	0.001	38
GE	7565	0.0429	1.81	-0.131	8.56	0.000	0.000	0.000	46
AA	7565	0.0206	2.34	-0.289	9.81	0.000	0.000	0.001	46
				NYSE Financial Top 10 (pre-crisis, January 1, 1993-December 31, 2007)					
WFC	3777	0.0563	1.64	0.145	2.99	0.000	0.004	0.008	11
JPM	3777	0.0453	2.11	0.12	5.85	0.000	0.597	0.523	15
C	3777	0.0627	2.09	0.057	4.98	0.000	0.012	0.026	17
BAC	3777	0.0453	1.78	-0.15	3.42	0.000	0.214	0.200	17
SAN	3777	0.0695	2.09	0.014	5.25	0.000	0.000	0.000	22
WBK	3777	0.0894	1.56	-0.044	1.81	0.000	0.013	0.054	7
USB	3777	0.0527	1.88	-1.33	27.1	0.000	0.000	0.000	20
AXP	3777	0.0609	2.00	-0.052	3.30	0.000	0.002	0.001	16
BCS	3777	0.0660	2.04	0.094	2.69	0.000	0.002	0.004	16
MS	3777	0.0616	2.46	0.012	3.53	0.000	0.060	0.085	16
				CRSP value-weighted (January 1, 1998-December 31, 2012)					
CRSP	3772	0.0287	1.33	-0.089	6.52	0.000	0.000	0.000	

Note: J-B is the Jarque-Bera test of normality. L-B is the ljung-Box test of autocorrelation using 12 lags; the L-B* column reports results for VAR(1)-prewhitened data. Outliers are observations that are more than four times the sample standard deviation away from the sample mean. A number of stocks have extremely large excess kurtosis and skewness, which is due to large dividend payments or stock splits or both. E.g., on July 13, 2000, KSU had a 1:2 stock split and paid a dividend of 90.803 dollars per share, which translates into a 273% daily loss in log returns. And on July 5, 1999, HSBC had a 3:1 stock split. With outlier reduction, the excess kurtosis of KSU becomes 4.47. There are minor differences in the effective sample size simply because different countries have different stock market and capital market holidays. The financial intermediation sector has the largest number of outliers.

2.10 APPENDIX D: STABILITY CHECK

Before getting to the results of this section I introduce yet a third technique called discrete Maximum Likelihood. It can be used to check whether the sample size is too small given the granularity of the data and can answer questions like: For a given sample size, is this sample a "good" sample? It is possible to have a sample that accidentally has some irregular local clusters, in which case the sample can be said to lack certain "smoothness" property and is a bad representation of the true data generating process. The discr-ML can be implemented by replacing the probability density function-based likelihood,

$$\prod_{j=1}^{n}\sum_{i}^{m}\pi_i p_i(x_j,\theta_i),$$

with the discretized likelihood,

$$\prod_{j=1}^{n}\sum_{i}^{m}\pi_i\int_{a_j}^{b_j}p_i(x,\theta_i)dx,$$

where m is the number of states, $\{(a_j,b_j)\}_{j=1}^{n}$ is a collection of intervals that break the domain of observations into small "silos." The intervals do not have to be articulated, neither do they have to cover the whole real line. All x's falling into the same interval are treated the same which simply amounts to recognizing explicitly the interval nature of all continuous observations. An advantage of discrete MLE is that one does not have to worry about explosive likelihood since the probability integral is always bounded. A practical issue is how to optimally choose the intervals, on which there seems to be no clear guidance. In the empirical study of Section 5, I first rank the absolute differences of consecutive daily returns, I then pick the 5^{th} lower quantile as the length of the silo and string them one after the other covering the range of the observed data (a small fraction of the absolute differences is zero or very close to zero anyway).

Next I present results of some alternative ways of estimation. I examine the effect of the assumption of a stationary initial distribution, i.e., limited memory of initial state in Markov switching-type models; I also use discrete-ML to check the granularity of the samples. It is known that, for daily returns, a VAR(1) prewhitening and recoloring is sometimes useful in reducing residual serial correlation and spurious ARCH effect, so I also study the effect of prewhitening on comovement measures. Finally, the EM-based multivariate Gaussian semi-Markov switching model is compared to all previous results. For reasons given in section 5.2 and because incorporating a multivariate t-distribution into a sMSw model is computationally demanding, this variation is not pursued here.[2]

In the case of $\mathrm{m}=3$ and $\mathrm{m}=2$ (not reported), all results are consistent with those from the basic Gaussian models in Table 2. In particular, VAR prewhitening does not appear to matter in any material way and discrete ML also yields results very similar to continuous ML models. The maximized likelihood of EM-based models is slightly larger than that of the exact ML because EM uses unconstrained initial state as opposed to a stationary distribution; see section 3. Estimates of the transition probability matrix and covariance matrix are also quite similar. Their major difference lies in the conditional mean of certain states and certain stocks depending on which initial state the EM algorithm picks. As for EM-sMSw, the results largely agree with those from Table 2-a, with the state-dependent market returns and volatilities all monotonically ordered this time. Once again MSwC and BIC pick either $\mathrm{m}=2$ or $\mathrm{m}=3$, whereas AIC and $\mathrm{AIC_c}$ always pick the highest order of switching. (Note that BIC picks $\mathrm{m}=2$ in DAX 30 Top 4.) EM-MSw and EM-sMSw give very similar estimates for all commonly shared parameters, including many of those not reported here, which is yet another evidence that sMSw builds upon, not deviates from, MSw by refining the sojourn structure.

[2] Estimation of the degrees of freedom requires numerically solving a digamma function at each step, which adds another layer of complexity in terms of updating a weighted average for the conditional mean and covariance matrix. Analytic solutions do not exist; see the appendix in Bulla and Bulla 2007.

Table D.1

Stability checks using alternative methods of estimation, Jan. 1, 2007-Dec. 31, 2010.

	m=3			EM-sMSw			
	EM-MSw	EM-VAR(1)	discr-ML	m=2	m=3	m=4	VAR(1), m=3
				NYSE Financial Top 4			
Cm	[0.73, 0.70, 0.72]	[0.71, 0.70, 0.72]	[0.73, 0.70, 0.71]	[0.72, 0.71]	[0.73, 0.70, 0.72]	[0.76, …, 0.73]	[0.73, 0.70, 0.73]
Cm_{diff}	[0.68, 0.67, 0.68]	[0.68, 0.67, 0.69]	[0.68, 0.67, 0.67]	[0.68, 0.67]	[0.69, 0.67, 0.68]	[0.72, …, 0.69]	[0.70, 0.67, 0.69]
Cm_{avg}	[0.62, 0.59, 0.62]	[0.61, 0.59, 0.62]	[0.62, 0.58, 0.62]	[0.62, 0.60]	[0.62, 0.59, 0.62]	[0.67, …, 0.63]	[0.62, 0.59, 0.62]
r_{CRSP}		[high, medium, low]				[high, ..., low]	
σ_{CRSP}		[low, medium, high]				[low, ..., high]	
σ		[low, medium, high]				[low, ..., high]	
BIC				No	Yes	No	
AIC/AIC_C				No	No	Yes	
MSwC				Yes	No	No	
				DJIA Top 4			
Cm	[0.68, 0.63, 0.78]	[0.69, 0.64, 0.77]	[0.72, 0.60, 0.78]	[0.67, 0.72]	[0.68, 0.63, 0.78]	[0.42, …, 0.78]	[0.69, 0.63, 0.77]
Cm_{diff}	[0.60, 0.54, 0.73]	[0.61, 0.55, 0.73]	[0.66, 0.51, 0.73]	[0.58, 0.65]	[0.60, 0.54, 0.73]	[0.26, ..., 0.79]	[0.60, 0.54, 0.73]
Cm_{avg}	[0.58, 0.50, 0.70]*	[0.58, 0.51, 0.70]*	[0.64, 0.49, 0.71]*	[0.55, 0.63]*	[0.58, 0.50, 0.70]*	[0.22, …, 0.70]**	[0.58, 0.50, 0.70]*
r_{CRSP}		[high, medium, low]				[high, ..., low]	
σ_{CRSP}		[low, medium, high]				[low, ..., high]	
σ		[low, medium, high]				[low, ..., high]	
BIC				No	Yes	No	
AIC/AIC_C				No	No	Yes	
MSwC				Yes	No	No	
				DAX30 Top 4			
Cm	[0.69, 0.66, 0.57]	[0.69, 0.66, 0.56]	[0.69, 0.66, 0.57]	[0.65, 0.60]	[0.69, 0.67, 0.56]	[0.71, …, 0.36]	[0.69, 0.67, 0.55]
Cm_{diff}	[0.62, 0.59, 0.49]	[0.62, 0.59, 0.49]	[0.62, 0.59, 0.48]	[0.58, 0.52]	[0.62, 0.60, 0.48]	[0.64, …, 0.20]	[0.62, 0.59, 0.48]
Cm_{avg}	[0.58, 0.55, 0.41]**	[0.58, 0.54, 0.41]**	[0.59, 0.55, 0.41]**	[0.53, 0.47]*	[0.58, 0.56, 0.40]**	[0.60, …, 0.13]**	[0.58, 0.56, 0.41]**
r_{CRSP}		[high, medium, low]				[high, ..., low]	
σ_{CRSP}		[low, medium, high]				[low, ..., high]	
σ		[low, medium, high]				[low, ..., high]	
BIC				Yes	No	No	
AIC/AIC_C				No	No	Yes	
MSwC				Yes	No	No	

Note: * and ** denote the 10% and 5% significance level. For DJIA Top 4, the comovement measure based on the four-regime EM-sMSw model is spurious, e.g., the full estimate of Cm_{avg} is [0.22, 0.77, 0.48, 0.70]. Apparently the low volatility state is split into two due to over-fitting. Similarly for DAX30 Top 4, the four-regime EM-sMSw model splits the high volatility state into two.

A few words about numerical issues are in place. Like the benchmark model specifications, the sMSw algorithm is not sensitive to the starting values of conditional mean, transition probability matrix, and sojourn probabilities, but is somewhat influenced by starting values of variance-covariance matrix and the choice of sojourn distributions. For the latter, I use the versatile gamma distribution because its scale and shape parameters can vary independently and the result matches an early finding about duration dependence of state occupancy. When the order of switching is greater than two, the algorithm frequently reaches a local minimum which nevertheless turns out to be close to the pseudo global minimum. An arbitrary grid search is obvious not efficient, so I use a method similar to the bisection approach and various starting values are tried based on the unconditional and conditional variance-covariance matrix. I find that using the EM-MSw estimates as starting values gives the fastest convergence in all cases, with the improvement being most salient for four or more states. This is quite intuitive: since a sMSw model refines the occupancy distribution, state-dependent distributions for returns may thus be affected only to a limited extent. In the case of two states, the algorithm converges very fast and starting values do not seem to matter. Speaking of discrete ML, it is little affected by the choice of discretization knots until the silos start to get very coarse (e.g., 10^{th} lower percentile of the sequence of absolute differences), since some efficiency might be lost as more and more observations that fall into the same silo are treated as the same.

2.11 Appendix E: Software Accuracy

As many authors use GAUSS for Markov-switching models, it is helpful to note the following. Zeileis and Kleiber (2005) attempt to adapt the GAUSS code for Bai and Perron's (2003) structural change models to R. In resolving a discrepancy between the GAUSS and R output, they find inaccuracies in the tails of the GAUSS function for the log of the normal cumulative distribution function. Yalta (2007) finds that GAUSS 8.0 still has serious problems, especially with statistical distributions and random number generation. Contrariwise, McCullough (2009) provides evidence that the R programming language is very accurate with respect to statistical distribution and random number generation. Similar studies in an econometric context for the latest version of GAUSS or Matlab do not seem to exist. Since the normal and Student-*t* distributions are used a lot in this thesis, I prefer to code in R.

Zeileis A. and Kleiber C., 2005. Validating multiple structural change models–A case study. *Journal of Applied Econometrics* 20(5): 685-690.

Yalta, A.T., 2007. The numerical reliability of GAUSS 8.0. *The American Statistician* 61(3): 262-268.

McCullough B.D., 2009. The accuracy of econometric software. In *Handbook of Computational Econometrics*, Belsley D.A. and Kontoghiorghes E.J. (eds). Wiley: Chichester, 55-80.

Chapter 3. A TEST OF ASYMMETRIC STOCK RETURN COMOVEMNT ACROSS STATES

3.1 INTRODUCTION

Ever since excess kurtosis and sporadic peaks in stock returns were recognized in the economic profession five decades ago, a central theme in financial econometrics has been the study of the temporal and distributional characteristics of asset returns. One late comer that has recently turned quite a few heads is on stock return comovement, in which a shared feeling is that comovement is asymmetric conditional on volatilities and returns.

One strand of research has focused on tail comovement, or more generally, the characterization of upside and downside dependence structure. Theoretical contributions made in this direction have used copulas (Patton 2006; Chollete *et al.* 2009; Garcia and Tsafack 2011) and multivariate extreme value theory (Stărică 1998; Longin and Solnik 2001). On the empirical front, this topic has received close attention in recent years in the international financial market literature; see Ramchand and Susmel (1998), Ang and Chen (2002), and Syllignakis and Kouretas (2011). As comovement in the tail does not measure overall comovement, another intensively studied topic in this area is comovement differential across different states of interest. The constituent states are most easily seen through the lens of vastly different volatilities and mean returns, so an interesting question to ask is whether comovement is significantly higher in the volatile and low return periods.

A somewhat overlooked fact in the literature is that comovement is not necessarily higher conditional on (large) negative returns, e.g., Karolyi and Stulz (1996) speak in terms of large "absolute returns" and Patton (2006) finds that in the post-euro era, the dependence structure between two exchange rate returns (DM-USD and Yen-SD) has gone from significantly asymmetric in the negative direction to weakly asymmetric in the positive direction. In this respect, economic theory does not give a unified answer either. Ribeiro and Veronesi (2002) model news in bad times as more informative about the true state of the economy, which then predicts that during volatile periods of more uncertainty about the future, cross-market correlation tends to be higher. By contrast, Simsek (2013) studies the role of financial innovation in portfolio risk diversification and finds that when large belief disagreements exist among traders, financial innovation tends to increase average portfolio volatility and decrease average portfolio comovement. In view of regime-switching models, this suggests that when states are ignored, comovement asymmetry conditional on, say, exceeding certain threshold of return, may be easily confounded by the state because a state-free test is effectively pooling all the returns coming from states of potentially very different degrees of comovement. For example, Longin and Solnik (1995) fit a threshold GARCH model to monthly market returns and find that conditional correlation tends to increase in periods of high turbulence but is no more sensitive to negative than to positive shocks. Another possibility is that tail comovement asymmetry is a result of regime switching. In this chapter, I study all these issues based on a test of comovement differential across states.

Previous authors have used popular models such as the DCC, the GARCH-M, the regime-switching GARCH, or the Gaussian regime-switching model, yet results are mixed and existent tests are mostly based on pairwise correlation which is inappropriate for nonlinear dependence structures. In copula-based models, researchers have frequently relied on Goodness-of-fit-type tests for lack of a direct test of asymmetric comovement. The test I propose takes a probabilistic form to which the closest predecessor is the state-free tail-dependence test of Li (2013). They differ, however, in the following important aspects. First, I construct a test of symmetric state-dependent comovement and prove the asymptotic results by exploiting the structure of a Markov chain. Second, I show in Section 5.3 that a test ignoring states is subject to severe power attrition due to nuisance parameters. Third, I examine the local power property of the test and map it into local data generating processes. Fourth, I show how a refined version of the test can be constructed based on the regenerative cycle theory and compare results based on different HAC estimators. Further, to facilitate comparison with a Wald test of correlation equality, I estimate the variance and correlation separately in models that allow both of them to switch.

In the next section, I first lay out the model specification and give a summary of the test procedure. In Section 3, I present the main asymptotic results for the test, the proofs of which are postponed until the appendix. I then analyze the power of the test in Section 4 where it is shown that the test has nonnegligible power against local alternatives and is invariant in the so-called meta-elliptical family. Numerical studies are given in Section 5 followed by a list of extensions. Section 7 concludes.

3.2 Model and Test Procedure

In a study of asymmetric comovement, Ang and Chen (2002) find that a mixture of normal model is very good at matching the empirical correlation asymmetries in the data. In terms of stylized facts and higher order distributional characteristics, Rydén et al. (1998) and Timmermann (2000) show that a Gaussian Markov-switching model can significantly expand the model's scope for asymmetry, fat tails, volatility dynamics and infrequent breaks in the conditional mean or variance. See also Diebold and Inoue (2001) for the near observational equivalence of the Markov switching model and long memory process. For these reasons, I consider the data generating process to be a multivariate normal distribution selected from one of m Markovian regimes.

The current DGP is a useful way to capture the centerpiece of the test which can easily accommodate many other interesting variations with a switching structure. This is because the test is invariant in a much larger family of DGP's (Section 4.3) and enjoys certain robustness to the specification of state-dependent variances and means (see Remark 1.). The procedure contains two separable steps: the goal of step one is to identify the true state and to estimate the stationary state distribution, after which step two is model-free. As such, it can be used for DGP's with switching in the mean or variance, switching GARCH in the variance, duration in the mean or variance (e.g., Maheu and McCurdy (2000)), time-varying transition probability that only depends on current information, or a combination of these. On the other hand, if one wants to extend the test to models with switching AR components (e.g. Garcia and Perron (1996)) or general time-varying parameter models that depend on past information (e.g., Diebold (1993)), the very interpretation of state-dependent comovement would become less compelling because in these models—which also require a very large sample to estimate—returns in one state frequently depend on the immediate past of other states.

3.2.1 *Model specification*

Consider the n -variate process $\{r_1, r_2, \dots, r_n\}$ generated by $r_t = \sigma_{s_t} z_t + \mu_{s_t}$, where z_t~iid $N(0,1)$ and $\{x_t\}$ is an m-state Markov chain governed by the *transition probability matrix* Γ. Let $P(r_t|x_t)$ denote the conditional probability of observing r_t given state x_t, let $\boldsymbol{\pi}$ denote the initial state distribution, and let $\boldsymbol{r}^{(t)}$, $\boldsymbol{x}^{(t)}$ denote the histories of observations and states up to time t. Treating the unobserved states as missing data one gets,

$$L(T) = P\left(\boldsymbol{R}^{(\boldsymbol{T})} = \boldsymbol{r}^{(T)}\right) = \sum_{x_1, x_2, \dots, x_n \in \{1, \dots, m\}} P(\boldsymbol{R}^{(\boldsymbol{T})} = \boldsymbol{r}^{(T)}, \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}),$$

and by the strong Markov property, a particular realization of states contributes to the likelihood by as much as

$$P\left(\boldsymbol{R}^{(\boldsymbol{T})} = \boldsymbol{r}^{(T)}, \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}\right) = \pi(x_1)P(r_1|x_1)p(x_2|x_1)P(r_2|x_2) \dots p(x_n|x_{n-1})P(r_n|x_n).$$

A full permutation of the histories of states thus yields

$$\begin{aligned} L(T) &= \sum_{x_1, x_2, \dots, x_n \in \{1, \dots, m\}} \pi(x_1)P(r_1|x_1)p(x_2|x_1)P(r_2|x_2) \dots p(x_n|x_{n-1})P(r_n|x_n) \\ &= \boldsymbol{\pi}' P(r_1)\Gamma P(r_2) \dots \Gamma P(r_n)\mathbf{1}, \end{aligned} \tag{3.1}$$

where $P(\cdot)$ is the $m \times m$ diagonal conditional probability matrix of observation r_t given each possible state x_t. Thus a model of n stocks has in total $m(m - 1 + n(n + 3)/2)$ parameters: $m^2 - m$ to determine Γ, mn for the state-dependent means, and $mn(n+1)/2$ for the state-dependent variances and covariances. These parameters are collected in $\boldsymbol{\theta}$.

3.2.2 *Summary of the Test Procedure*

Let $\delta(u, v)$ be some comovement measure of two time series. The so-called coherent measure of comovement is thought superior because it has the following good features:

$\delta(u, v) = \delta(u, v)$; $-1 \leq \delta(u, v) \leq 1$; $\delta(u, v) = 1$ if x and y are co-monotonic and

$$\delta(T(u),v)=\begin{cases}\delta(u,v), & \text{for increasing transformation } T\\ -\delta(u,v), & \text{for decreasing transformation } T\end{cases}.$$

Let $(\bar{R}_t^1,\bar{R}_t^2)$ and $(\tilde{R}_t^1,\tilde{R}_t^2)$ be two independent draws from the bivariate joint distribution of log-differenced returns of two stocks at time t, then a natural candidate satisfying all these properties is Kendall's τ, or the *coefficient of concordance:*

$$\begin{aligned}\tau\equiv\tau\left(\bar{R}_t^1,\bar{R}_t^2,\tilde{R}_t^1,\tilde{R}_t^2\right)&=P(\text{concordant pairs})-P(\text{discordant pairs})\\ &=2P\left(\left(\bar{R}_t^1-\tilde{R}_t^1\right)\left(\bar{R}_t^2-\tilde{R}_t^2\right)>0\right)-1.\end{aligned} \tag{3.2}$$

In a time series setting, R_{t+1}^1 and R_t^1 may be correlated and can come from potentially different distributions, in which case the definition of Eq. (3.2) would not apply. However, one can exploit the conditional independence property of Markov-switching models and define a new *state-dependent* measure of comovement as

$$C(i)\equiv C(R_{t+1}^1,R_t^1,R_{t+1}^2,R_t^2|i)=2P\left((R_{t+1}^1-R_t^1)(R_{t+1}^2-R_t^2)>0|i\right)-1 \text{ for } i=1,2,\dots,m,$$

where m is the number of states and i denotes the state.

Of particular interest is the comovement differential between the high- and low-volatility states, so I use superscript h (l) for high (low) volatility regimes (e.g., $C^h\equiv C(i=high)$, $C^l\equiv C(i=low)$). The null and alternative hypotheses then take the following form:

$$\mathrm{H}_0\colon C^h=C^l,$$
$$\mathrm{H}_1\colon C^h\neq C^l.$$

By constructing an appropriate empirical process, I show in Section 3 that, under the null, an asymptotically pivotal statistic exists and can be summarized as

$$\sqrt{n}\left(\hat{C}^h-\hat{C}^l\right)\xrightarrow{d}N(0,\sigma^2),$$

where $\hat{C}^h$ and $\hat{C}^l$ are consistent estimates of C^h and C^l defined in Section 3.

To identify the state at each time t, I conduct state inference based on global most probable path (Viterbi's algorithm) which yields very good result for the high- and low-volatility state. (State inference based on smoothed probabilities yields almost the same result; see the appendix.) This being said, errors in state estimation are likely to have an impact on the finite sample performance of the test, so I investigate the issue in Section 5.1. Now to build a feasible test from $\{R_t^1, R_t^2\}_1^{n+1}$, define

$$\begin{aligned} D_t &= I(\text{concordant pairs at } t) - I(\text{discordant pairs at } t) \\ &= I\big((R_{t+1}^1 - R_t^1)(R_{t+1}^2 - R_t^2) > 0\big) \\ &\quad - I\big((R_{t+1}^1 - R_t^1)(R_{t+1}^2 - R_t^2) < 0\big), \end{aligned} \tag{3.3}$$

and

$$W_t(i) = \frac{\big(D_t - C(i)\big)I(x_t = i)}{P(x_t = i)}.$$

Note that $E(W_t(i)) = 0$. Let W_t^h and W_t^l denote $W_t(i = high)$ and $W_t(i = low)$ and define the estimate of $W_t(i)$ as

$$\widehat{W}_t(i) = \frac{\big(D_t - \hat{C}(i)\big)\, I(x_t = i)}{\hat{P}(i)}, \tag{3.4}$$

where $\hat{P}(i)$, defined in Section 3, is a consistent estimate of the unconditional probability of hitting state i, $P(x_t = i)$.

The test statistic will be shown to have an asymptotic normal distribution,

$$\sqrt{n}\big(\,\hat{C}^h - \,\hat{C}^l\big) \xrightarrow{d} \nu \sim \mathrm{N}(0, \sigma^2),$$

where

$$\nu \,=\, \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(W_t^h - W_t^l\big).$$

In Section 5.1, I show that a simple Bartlett kernel-HAC estimator for σ^2 has the best power performance.

Remark 1. By construction this measure is invariant to standardization and is thus robust to breaks or changes in the first two moments:

$$P\big((R_{t+1}^1 - R_t^1)(R_{t+1}^2 - R_t^2) > 0|i\big) = P\left(\left(\frac{(R_{t+1}^1 - \mu^1)}{\sigma_1} - \frac{(R_t^1 - \mu^1)}{\sigma_1}\right)\left(\frac{(R_{t+1}^2 - \mu^2)}{\sigma_2} - \frac{(R_t^2 - \mu^2)}{\sigma_2}\right) > 0|i\right). \quad (3.5)$$

This robustness is nontrivial as the asymptotic distribution would be contaminated by errors in parameter estimation even in the limit. The problem may be solved using techniques such as Khmaladze's martingale transformation as in Bai and Ng (2001).

3.3 Asymptotic Results

In this section, I establish the asymptotic properties of the test statistic and derive a correction based on the theory of regenerative cycles. I use notations that are either self-explanatory or standard in the literature. Detailed proofs can be found in the online appendix. To set up the stage, I first state the global assumptions needed to prove the main result.

Assumption 1. (Markov) The latent Markov chain $\{X_k\}_{k=0}^{\infty}$ on a measurable hidden state space $(\mathrm{X},\ \mathcal{X} \equiv \mathcal{B}(\mathrm{X}))$ is homogenous, irreducible, aperiodic, and has a finite number of states.

Assumption 2. (Consistency) Let Θ be a compact parameter space that parameterizes the transition kernel of the Markov chain and the conditional distribution of the observations given the states. For all $\theta \in \Theta$, $sup_{\theta\in\Theta}|l_n(\theta)/n - l(\theta)| \overset{p}{\to} 0$ *a.s.* as $n \to \infty$, where $l_n(\theta)$ is either the log-likelihood conditional on an initial state or the stationary log-likelihood given the first n observations, and $l(\theta)$ is a continuous deterministic function with a unique global maximum at θ_0.

Remark 2. For most practically relevant applications, Assumption 1 is harmless. Then one can show that irreducibility and finite order of switching imply positive recurrence; see Brémaud (1999, page 118) for a proof. In addition the process is aperiodic, so the (finite state) transition probability matrix only has one eigenvalue lying on the unit circle with all others inside the unit circle, i.e., the Markov chain is ergodic. Under Assumption 1, there exists a unique invariant measure and an induced stationary distribution.

Remark 3. Assumption 2 is a high-level condition that guarantees consistency of the ML estimates. The uniform convergence and the identifiability constraints ensure that the estimates converge to the true parameters almost surely. Most works on the MLE of Markov switching models have focused on numerical aspects suitable for approximating the exact maximum likelihood. Asymptotic properties for such models with finite states and observations taking values in a general space have only recently been worked out. The proof requires certain stationarity and conditional mixing assumptions; see Bickel *et al.*(1998), Jensen and Petersen (1999), and Douc *et al.* (2004).

Define the (random) hitting time as $\tau_\alpha = \inf\{k \geq 1, x_k \in \alpha\}$, with $\alpha \in \mathcal{X}$ being an accessible state of the σ- field on the hidden state space $(\mathrm{X}, \mathcal{X})$. Let $\eta_n^i = \sum_{t=1}^{n} I_{[i]}(x_t)$ be a counting function of the number of times the process hits state i including the initial state. For a certain state i^* that is different from the initial state, denote the above counting function by η_n^*. Let π be the initial state distribution.

Lemma 1. Let $f\colon \mathbf{X} \to \mathbb{R}$ be a bounded function such that $P(x_t \in \mathbf{X}) = 1, P(\mathrm{X} \in \mathbf{X}) = 1$ and $E_\alpha(|f|) < \infty$, define

$$S_j(f) = \sum_{t=\tau_\alpha^{(j-1)}+1}^{\tau_\alpha^{(j)}} f(x_t),\ j \geq 1, \tau_\alpha^{(0)} = 0,$$

where $\tau_\alpha^{(j)}$ is the j^{th} (cumulative) hitting time $(\mathrm{e.g.}, \tau_\alpha^{(1)} = \tau_\alpha, \tau_\alpha^{(n+1)} = \inf\ \{k > \tau_\alpha^{(n)}, x_k \in \alpha\})$ and x_t is the state. Then for any initial state distribution π,

$$\frac{1}{\eta_n^*}\sum_{j=1}^{\eta_n^*} S_j(f) \xrightarrow{a.s.} \mu_{IM}(f).$$

$\mu_{IM}(f) = E_\alpha(\sum_{t=1}^{\tau_\alpha} f(x_t)) \equiv E(\sum_{t=\tau_\alpha^{(i)}+1}^{\tau_\alpha^{(i+1)}} f(x_t))$ is called the *invariant measure*.

Note that the process can start from any state. $E_\alpha(\cdot)$ is expectation with respect to the Markov chain starting from state α at $t = 0$; $f(x_t) \equiv f(x_t; \theta \in \Theta)$. In the form of the current test statistic, define $f^h(x_t) = D_t I(x_t = h)/P(x_t = h)$, $g^h(x_t) = C^h I(x_k = h)/P(x_t = h)$, $W_t^h = f^h(x_t) - g^h(x_t)$, and similarly for the low-volatility state. The functions f^h, f^l, g^h and g^l are all measureable on the hidden state space $(\mathrm{X}, \mathcal{X})$. Lemma 1 can then be used to prove the following Theorem.

Theorem 1. Let $r_t = (R_{t+1}^1 - R_t^1)(R_{t+1}^2 - R_t^2)$ and consider a bounded function $f\colon \mathbf{X} \times \mathbb{R} \to \mathbb{R}$ such that $P\big((x_t, r_t) \in \mathbf{X} \times \mathbb{R}\big) = 1$, $\mathrm{P}(\mathrm{X} \times \mathbb{R} \in \mathbf{X} \times \mathbb{R}) = 1$ and $E_\alpha(|f|) < \infty$. Define

$$\hat{f}^i(x_t, r_t) = \frac{D_t I(x_t = i)}{1/\left(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha\right)\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} I\,(x_t = i)} = \frac{I(r_t > 0, x_t = i) - I(r_t < 0, x_t = i)}{1/\left(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha\right)\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} I\,(x_t = i)},$$

$$i = 1,2, \dots, m.$$

Then

$$\frac{\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \hat{f}^i(x_t, r_t)}{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha} \xrightarrow{a.s.} 2P(r_t > 0|i) - 1. \tag{3.6}$$

Replacing $\tau_\alpha^{(\eta_n^*)}$ with n and τ_α with 0, one gets the uncorrected version of the theorem.

By the next theorem, the limiting behavior of the test statistic, $\sqrt{n}\big(\hat{C}^h - \hat{C}^l\big)$, has the same asymptotic distribution as $n^{-1/2}\sum_{t=1}^{n}\big(W_t^h - W_t^l\big)$

Theorem 2 (Main result). Let W_t^h and W_t^l be defined as in Section 2.2 and recall from Eq. (3.4) that

$$\widehat{W}_t(i) = \frac{\left(D_t - \hat{C}(i)\right) I(s_t = i)}{\hat{P}(i)},$$

where

$$\hat{P}(x_t = i) = \frac{1}{\left(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha\right)} \sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} I\,(x_t = i), \tag{3.7}$$

$$\hat{C}^h = \frac{\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \hat{f}^h(x_t, r_t)}{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha}. \tag{3.8}$$

Then under the null

$$n^{1/2}\left(\hat{C}^h - \hat{C}^l\right) = n^{-\frac{1}{2}} \sum_{t=1}^{n} \left(W_t^h - W_t^l\right) + o_p(1) \xrightarrow{d} N(0, \sigma^2), \tag{3.9}$$

and

$$\widehat{W}_t^h \xrightarrow{a.s.} W_t^h, \qquad \widehat{W}_t^l \xrightarrow{a.s.} W_t^l.$$

Eq. (3.9) holds up to the first order. To eliminate the impact of finite-sample dependence on the initial state of a persistent Markov chain, a regenerative cycle-corrected version of the test can be achieved with little modification:

$$\left(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha\right)^{\frac{1}{2}} \left(\hat{C}^h - \hat{C}^l\right) \xrightarrow{a.s.} \left(\frac{1}{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha}\right)^{\frac{1}{2}} \sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \left(W_t^h - W_t^l\right) \xrightarrow{d} N\left(0, \sigma^{*2}\right). \tag{3.10}$$

In the appendix, I show how the test can be further modified with a second-order bias-in-mean correction. The result can be given here: Under the null, the correction term automatically vanishes.

3.4 POWER ANALYSIS

In this section, I show that the test has power against local alternatives and diverge at the rate of $\sqrt{n}$ for fixed alternatives. I then show that the local alternatives can be mapped one to one into the data generating process. Although the test has many good properties inherited from Kendall's τ, it is likely to have limited power for small samples. For a very versatile family of distributions known as the *meta-elliptical* family, the test is shown to be invariant to alternative DGP's sharing the same *core correlation matrix*

3.4.1 *Local alternatives*

Let $C^h = C^l + \delta/\sqrt{n}$, then

$$\hat{C}^h = \frac{1/n\sum_{t=1}^{n}(D_t - C^h + C^h)I(x_t = i)}{\left(1/n\sum_{t=1}^{n} I\,(x_t = i)\right)} = \left(\frac{P(x_t = h)}{\left((1/n)\sum_{t=1}^{n} I\,(x_t = h)\right)}\right)\frac{1}{n}\sum_{t=1}^{n} W_t^h + C^h,$$

and

$$\hat{C}^l = \left(\frac{P(x_t = h)}{\left((1/n)\sum_{t=1}^{n} I\,(x_t = h)\right)}\right)\frac{1}{n}\sum_{t=1}^{n} W_t^l + C^l.$$

The test statistic now becomes

$$\begin{aligned}\sqrt{n}\left(\hat{C}^h - \hat{C}^l\right) &= \left(n^{-\frac{1}{2}}\sum_{t=1}^{n}\left(W_t^h - W_t^l\right)\right)\left(\frac{P(x_t = h)}{\left((1/n)\sum_{t=1}^{n} I\,(x_t = h)\right)}\right) + \delta \\ &= n^{-\frac{1}{2}}\sum_{t=1}^{n}\left(W_t^h - W_t^l\right) + \delta + o_p(1) \xrightarrow{d} N(\delta, \sigma^2),\end{aligned} \tag{3.11}$$

the last line of which follows from Theorem 1 and 2.

3.4.2 *Local data generating process*

Let $\boldsymbol{X^1}, \boldsymbol{X^2}, \dots, \boldsymbol{X^n} = (X_1^n, X_2^n, \dots, X_k^n)' \in \mathbb{R}^k$ be a sequence of independent random vectors defined on $L^{p\geq 2}(\Omega, dF)$, where Ω is the support for the sigma-finite probability measure dF. Define the coefficient of concordance between element i and j as

$$\tau_{ij} = P\left((X_i^1 - X_i^2)(X_j^1 - X_j^2) > 0\right) - P\left((X_i^1 - X_i^2)(X_j^1 - X_j^2) < 0\right)$$
$$\equiv E\left(sign\{(X_i^1 - X_i^2)(X_j^1 - X_j^2)\}\right),$$

an unbiased estimator of which is

$$\hat{\tau}_{ij} = \frac{2}{n(n-1)} \sum\sum_{p<q} \{I(X_i^p - X_i^q) I(X_j^p - X_j^q)\}. \tag{1}$$

Eq. (3.6) is in fact a special case of the more general one-sample second-order U-statistic:

$$U_n \equiv \frac{2}{n(n-1)} \sum\sum_{p<q} g(X^p, X^q),$$

where the function g is called the kernel of the U-statistic such that $Eg(X^i, X^j) \equiv E(U_n) = \mu_U$ and $E(g^2) < \infty$. Suppose further $g(X^p, X^q)$ is bounded by $[a, b]$, then by Hoeffding's inequality, $P(U_n - \mu_U \geq \epsilon) \leq e^{-2\lfloor n/2 \rfloor \epsilon^2/(b-a)^2}$ for $\epsilon > 0$.

Below I will focus on the case of $k = 2$ and represent the high- and low-volatility states by the correlation matrix Σ^h and Σ^l. Define the element of the *implied* sample correlation matrix, $\tilde{\Sigma}$, by $\tilde{\rho}_{ij} = \sin((\pi/2)\hat{\tau}_{ij})$, then in current notation, $\hat{C}^h \equiv \hat{\tau}_{12}^h$, $\tilde{\rho}^h \equiv \tilde{\Sigma}_{12}^h$ and similarly for the low-volatility state. The next theorem puts a probabilistic bound on the test statistic.

Theorem 3. Let $|C^h - C^l| = \delta/\sqrt{n}$ for some fixed positive δ, under conditions given above

$$P\left(|\hat{C}^h - \hat{C}^l| \geq \frac{\Delta}{\sqrt{n}}\right) \begin{cases} < 4e^{-(\Delta-\delta)^2/16}\left(e^{(\Delta-\delta)^2/32} - 1\right) \leq 1, & \text{for } \Delta \geq 4\sqrt{2\log 2} + \delta \\ \leq 1, & \text{for } \delta < \Delta < 4\sqrt{2\log 2} + \delta \end{cases}, \tag{3.13}$$

$$P\left(|\tilde{\rho}^h - \tilde{\rho}^l| \geq \frac{\Delta}{\sqrt{n}}\right)\begin{cases} < 4e^{-(\Delta-\delta)^2/4\pi^2}\left(e^{(\Delta-\delta)^2/2} - 1\right) \leq 1, & \text{for } \Delta \geq 2\pi\sqrt{2\log 2} + \delta \\ \leq 1, & \text{for } \delta < \Delta < 2\pi\sqrt{2\log 2} + \delta \end{cases}.$$

Corollary 1. Let $C^h - C^l = \delta/\sqrt{n}$ for some fixed positive number δ, then with a little modification to the proof of Theorem 3,

$$P\left(\hat{C}^h - \hat{C}^l \geq \frac{\Delta}{\sqrt{n}}\right) < 2e^{-(\Delta-\delta)^2/16}\left(e^{(\Delta-\delta)^2/32} - \frac{1}{2}\right) < 1, \text{for } \Delta > \delta, \tag{3.14}$$

$$P\left(\tilde{\rho}^h - \tilde{\rho}^l \geq \frac{\Delta}{\sqrt{n}}\right) < 2e^{-(\Delta-\delta)^2/4\pi^2}\left(e^{(\Delta-\delta)^2/2} - \frac{1}{2}\right) < 1, \text{for } \Delta > \delta.$$

Ideally one would like to see the upper bound in Eq. (3.12) and Eq. (3.13) to be sharp and as large as possible because then local alternatives can generate sizable differences in comovement measures hence making the test more powerful. But as Theorem 3 shows, when Δ gets mildly larger, the bounds can be quite small. Because of long-memory sampling errors in a persistent Markov switching process, the implied correlation differential and the estimated comovement differential may not be highly significant for local alternative DGP's characterized by correlation matrices that are close to each other.

3.4.3 *Invariance in the meta-elliptical family*

It is known that Kendall's τ is invariant under strictly increasing transformations to the marginal variables, so one might wonder whether the proposed test is robust to distributions of potentially unobserved pricing factors that are not necessarily captured by a Markov switching model. The answer is affirmative which I now show by first introducing the elliptical and meta-elliptical distribution families.

Definition 1. (Elliptical distribution, Fang *et al.* (2002)). A continuous k-dimensional random vector $\boldsymbol{X}$ is said to have an elliptical distribution (ECD) with parameter $\boldsymbol{\mu}$ $(n \times 1)$ and Σ $(n \times n)$ if it has the stochastic representation $\boldsymbol{X} = \boldsymbol{\mu} + r\mathbf{A}\mathbf{u}$, where $r \geq 0$ is a random variable, $\mathbf{u}$ is uniformly distributed on the unit sphere in $\mathbf{R}^k$ and is independent of r, $\mathbf{A}$ is an $n \times n$ constant matrix such that $\Sigma \equiv \mathbf{A}\mathbf{A}'$. Denote this distribution by $\boldsymbol{X} \sim EC_k(\boldsymbol{\mu}, \Sigma, r)$.

Definition 2. (Meta-elliptical distribution). A continuous k-dimensiona random vector $\boldsymbol{X}$ is said to have a meta-elliptical distribution if there exist univariate strictly increasing functions $f_1, \dots, f_k$ such that $\left(f_1(X_1), \dots, f_k(X_k)\right)' \sim EC_k(\boldsymbol{\mu}, \Sigma, r)$. Denote this distribution by $\boldsymbol{X} \sim ME_k(\boldsymbol{\mu}, \Sigma, r; f_1, \dots, f_k)$. Σ is a semi-positive definite correlation matrix which may be called the core *correlation matrix*. (A more general definition can be found in Han and Liu (2013).)

The meta-elliptical family contains many useful symmetric or asymmetric distributions, including all elliptical distributions, multivariate t and asymmetric multivariate t; see Fang *et al.* (2002) for more examples. Based on the above definition, the following result shows that the proposed stock return comovement measure can also be interpreted as measuring dependence between pairs of unique pricing factors.

Theorem 4. Let $\boldsymbol{X} = (\boldsymbol{X}''_{1:k}, \boldsymbol{X}'_{k+1:p})', p > k$ be a continuous p-dimensional random vector of pricing factors or changes in time thereof. Let $\boldsymbol{X}_{1:k}|_{\boldsymbol{X}_{k+1:p}} \sim ME_k(\boldsymbol{\mu}, \Sigma, r; f_1, f_2, \dots, f_k)$ be a k-dimensional meta-elliptical random vector conditional on $\boldsymbol{X}_{k+1:p}$ and $\boldsymbol{R} \sim EC_k(\boldsymbol{\mu}, \Sigma, r)$ be a k-dimensional random vector of stock returns determined by

$$(R_1, R_2 \dots, R_k)' = \left(f_1\left(X_{1,}, \boldsymbol{X}'_{k+1:p}\right), f_2\left(X_{2,}, \boldsymbol{X}'_{k+1:p}\right), \dots, f_k\left(X_k, \boldsymbol{X}'_{k+1:p}\right)\right)',$$

where the functions $f_1, \dots, f_k$ are implied by the pricing formula and is strictly increasing in their first argument. Then conditional on $\boldsymbol{X}'_{k+1:p}$, for stock returns i and j and their unique pricing factors,

$$\rho_{ij} = \sin((\pi/2)\tau_{R_iR_j}) = \sin((\pi/2)\tau_{X_iX_j}).$$

The proof of this theorem combines the proof of Lemma A.5 and the fact that Kendall's τ is invariant under strictly increasing transformations to the marginal variables. It is an extension to the proof of Theorem 3.2 in Han and Liu (2013) and may be omitted without loss of concreteness. The theorem fits all linear pricing functions and can be augmented to incorporate multiple unique pricing factors in each asset. Theorem 4 makes it clear that within the meta-elliptical family, in order to generate a difference in the comovement measure, the DGP's must have different core correlation matrices, i.e., $\rho_{ij}^1 \neq \rho_{ij}^2$ for state 1 and 2.

3.5 Numerical Studies

In this section, standard assumptions on the kernel function and bandwidth are maintained as in Andrews (1991) and Newey and West (1994). The HAC estimators used are kernel-HAC, VARHAC and prewhitened HAC. I use the Bartlett kernel throughout the study. Data-dependent bandwidth selection is implemented in the R package NeweyWest. To the best of my knowledge, the effect of state estimation errors on inference is still an open question, so I compare the test's actual size and power based on estimated states with the true states. I only report results of the regenerative cycle-corrected test, Eq. (3.10), as the uncorrected version performs noticeably worse, especially for models with persistent states

3.5.1 *Size and power of the test*

I compute the actual rejection rate by drawing 1000 samples from the two-regime and three-regime models specified in Section 2.1. Based on empirical results cited at the beginning of Section 2, I specify the top correlation coefficient in the high-volatility state to be 0.75 and 0.6 and vary the magnitude of correlation differentials across states. (This corresponds to the conventional wisdom of high correlation when market is down and volatility is high. In normal times, the unconditional correlation of equity returns is typically in the range of 0.2-0.4.) In the three-regime model, an extra complication is the difference between the intermediate state correlation and that of the high and low states, so I study three reasonable correlation differentials: 0.1, 0.2, and 0.3. To evaluate the influence of persistent states for each choice of the order of switching, I compare two different transition probability matrices inspired by previous estimates in the literature. For example, in the two-state case a realistic and persistent transition structure is compared to an equally distributed benchmark with all entries equal to 0.5.

The choice of state-dependent means turns out to be inconsequential for the test, so to find a middle ground with respect to the real data, I set it to [0.05, 0.05; -0.5, -0.5] for m=2 and [0.1, 0.1; 0, 0; -0.5,-0.5] for m=3. On the other hand, the relative and absolute magnitudes of the high-and low-volatility state variances do affect size and power of the test through errors in state estimation. Intuitively, a large gap in state-dependent variances amounts to stretching and separating the data apart more thoroughly, and one would expect the states to be identified with more confidence. Empirical studies on international and domestic stock markets provide a plethora of evidence that, typically, this difference in magnitude across high- and low-volatility states is 5-10 fold for market indices and 10-20 fold or even higher for individual stocks; see Ang and Bekaert (2002), Maheu and McCurdy (2000), Ramchand and Susmel (1998) for applications using various Markov switching specifications. Table 1 collects results for two different sets of variances: the former represents a relatively stable situation such as broad-based market index returns; the latter represents, by contrast, volatile individual stock returns. One can also estimate a general multivariate model and conduct pairwise tests instead. I also report, in *italics*, the size and power of the test assuming that the true states are known; all other parameters, especially $P(h)$ and $P(l)$, are still estimated.

A possibly more efficient way to measure D_t defined by Eq. (3.3) is to sum over all possible pairs of returns involving R_t at each time t conditional on the state and then take the average. In the simulation studies, the marginal improvement using this estimator does not appear to be stable and significant especially when the sample size gets larger. So these are not reported.

Table 1

Size of the two-sided test of symmetric comovement across states, H_0: $C^h = C^l$ (asymp. nominal size=0.

			small volatility			large volatility		
			n=500	n=1000	n=2000	n=500	n=1000	n=2000
			ker-H	ker-H	ker-H	ker-H	ker-H	ker-H
m=2	Tran_1	*corr*=0.75	0.019	0.012	0.029	0.052	0.051	0.057
			0.014	*0.023*	*0.04*	*0.046*	*0.049*	*0.054*
			0.094	*0.076*	*0.06*	*0.081*	*0.054*	*0.06*
		corr=0.60	0.021	0.022	0.037	0.05	0.047	0.054
			0.015	*0.024*	*0.035*	*0.047*	*0.045*	*0.056*
			0.101	*0.094*	*0.07*	*0.094*	*0.088*	*0.078*
	Tran_2	*corr*=0.75	0.064	0.049	0.051	0.071	0.065	0.055
			0.017	*0.028*	*0.048*	*0.053*	*0.051*	*0.05*
			0.088	*0.071*	*0.055*	*0.074*	*0.056*	*0.051*
		corr=0.60	0.072	0.052	0.053	0.06	0.061	0.051
			0.036	*0.044*	*0.052*	*0.06*	*0.053*	*0.05*
			0.091	*0.074*	*0.061*	*0.089*	*0.071*	*0.052*
m=3	Tran_1	M1	0.033	0.041	0.047	0.012	0.023	0.033
		M2	0.047	0.052	0.052	0.022	0.033	0.041
		M3	0.072	0.066	0.056	0.023	0.034	0.052
		M1'	0.053	0.049	0.048	0.023	0.034	0.044
		M2'	0.065	0.056	0.054	0.024	0.041	0.05
	Tran_2	M1	0.038	0.049	0.052	0.030	0.035	0.035
		M2	0.127	0.102	0.076	0.033	0.042	0.047
		M3	0.235	0.162	0.113	0.039	0.051	0.052
		M1'	0.112	0.102	0.056	0.029	0.042	0.046
		M2'	0.122	0.105	0.077	0.039	0.048	0.05

Note: In the case of m = 2, actual size based on the true states is given by the second row in italics; actual size based on the true states and true probabilities is given by the third row in italics. In the small volatility case, for m=2, σ^h=[10, 15], σ^l= [3, 5]; for m=3, σ^h=[12, 20], σ^m=[7, 13], σ^l=[3, 5]. In the large volatility case, for m=2, σ^h=[30, 50], σ^l= [3, 5]; for m=3, σ^h=[60, 100], σ^m=[20, 30], σ^l=[3, 4]. Tran_1: [0.95,0.05; 0.15, 0.85] for m=2 and [0.95,0.03,0.02;0.1,0.80,0.1;0.12,0.03,0.85] for m=3; Tran_2: [0.5, 0.5; 0.5, 0.5] for m=2 and [0.86, 0.11, 0.03; 0.23, 0.72, 0.05; 0.11, 0.11, 0.78] for m=3. Correlation: M1: [0.75, 0.65, 0.75]; M2: [0.75, 0.55, 0.75]; M3: [0.75, 0.45, 0.75]; M1': [0.6, 0.5, 0.6]; M2': [0.6, 0.4, 0.6]. VARHAC and Prewhitened HAC yield very similar size distortions and are slightly more conservative than ker-HAC.

Table 2

Power of the two-sided test of symmetric comovement across states, H0: $C^h = C^l$.

			small volatility			large volatility		
			n=500	n=1000	n=2000	n=500	n=1000	n=2000
			ker-H	ker-H	ker-H	ker-H	ker-H	ker-H
m=2	$Tran_1$	[0.75, 0.65]	0.04	0.172	0.249	0.286	0.42	0.702
		[0.75, 0.55]	0.273	0.471	0.826	0.527	0.737	0.992
		[0.75, 0.45]	0.518	0.856	0.999	0.697	0.962	1.000
	$Tran_2$	[0.75, 0.65]	0.339	0.429	0.529	0.293	0.483	0.753
		[0.75, 0.55]	0.573	0.709	0.871	0.484	0.84	0.97
		[0.75, 0.45]	0.667	0.834	0.961	0.645	0.907	1.000
m=3	$Tran_1$	M1	0.325	0.639	0.92	0.279	0.501	0.911
			0.546	*0.728*	*0.88*	*0.31*	*0.598*	*0.898*
		M2	0.176	0.295	0.551	0.171	0.273	0.451
			0.399	*0.581*	*0.68*	*0.213*	*0.444*	*0.718*
		M3	0.186	0.28	0.526	0.163	0.227	0.541
			0.44	*0.576*	*0.67*	*0.21*	*0.434*	*0.725*
		M1'	0.118	0.136	0.194	0.123	0.18	0.225
			0.316	*0.435*	*0.529*	*0.204*	*0.308*	*0.526*
		M2'	0.324	0.495	0.854	0.257	0.471	0.774
			0.534	*0.782*	*0.839*	*0.283*	*0.524*	*0.836*
		M3'	0.366	0.467	0.685	0.152	0.306	0.518
			0.442	*0.59*	*0.694*	*0.17*	*0.442*	*0.712*
	$Tran_2$	M1	0.235	0.345	0.633	0.334	0.549	0.991
			0.222	*0.424*	*0.655*	*0.428*	*0.646*	*0.91*
		M2	0.121	0.23	0.351	0.13	0.304	0.654
			0.215	*0.374*	*0.559*	*0.225*	*0.408*	*0.676*
		M3	0.157	0.201	0.365	0.174	0.328	0.6
			0.193	*0.363*	*0.58*	*0.233*	*0.409*	*0.633*
		M1'	0.114	0.134	0.188	0.113	0.122	0.179
			0.147	*0.258*	*0.466*	*0.151*	*0.225*	*0.317*
		M2'	0.218	0.4	0.686	0.246	0.507	0.872
			0.248	*0.409*	*0.568*	*0.388*	*0.636*	*0.89*
		M3'	0.175	0.239	0.423	0.19	0.282	0.573
			0.184	*0.37*	*0.434*	*0.253*	*0.426*	*0.621*

Note: In the case of m=3, power based on the true states (but not the true probabilities) is given by the second row in italics. In the small volatility case, for m=2, σ^h=[10, 15], σ^l= [3, 5]; for m=3, σ^h=[12, 20], σ^m=[7, 13], σ^h=[3, 5]. In the large voltility case, for large volatility, m=2, σ^h=[30, 50], σ^l= [3, 5]; for m=3, σ^h=[60, 100], σ^m=[20, 30], σ^l=[3, 4]. $Tran_1$ and $Tran_2$ are the same as in Table 1. Correlation: M1: [0.75, 0.40,0. 45]; M2: [0.75, 0.45, 0.55]; M3: [0.75, 0.55, 0.55]; M1': [0.75, 0.55, 0.65]; M2': [0.75, 0.65, 0.45]; M3': [0.75, 0.65, 0.55]. VARHAC and Prewhitened HAC yield very similar results and are less powerful than ker-HAC.

Several patterns can be identified from Table 1. First, when the transition probability matrix is more persistent, i.e., Tran_1, the test tends to reject less often compared to Tran_2 and sometimes, as in $\mathrm{m} = 2$, Tran_1, it constantly under-rejects even in mildly large samples. Second, over-rejection is not uncommon in small and less persistent samples in the left panel of Table 1 but is largely eliminated when the data have large variances (right panel). Third, for $\mathrm{m} = 3$, the larger the correlation differential the larger the rejection rate which is due to some of the low correlation intermediate states being misclassified as the high correlation states. Fourth, as the sample size gets larger, size property improves. Lastly, when only the true states are used, size distortion becomes more severe. To evaluate the impact of estimation errors, I compare the actual size of the test with that based on the true states and that based on the true states and true transition probabilities. These are reported in the second and third rows of the top panel of Table 1 ($\mathrm{m} = 2$ case). It can be seen that as the sample size gets reasonably large, inference on the estimated model converges very fast to the truth.

Similar findings carry over to Table 2. For instance, it is found that the larger the correlation differential the higher the power, that the relative magnitude of the intermediate state matters in terms of power when $\mathrm{m} = 3$. In the two-regime models, the test has higher power when volatility is large and power based on the true states is no larger than based on the estimated states. In the three-regime models, I also report power based on the true states (but not assuming true transition probabilities) in the second row of each panel. Once again, it is found that estimation error only has limited influence. When sample size is small, power in large volatility models (right panel of Table 2) appears to be smaller than in small volatility models (left panel). Tests based on the Bartlett kernel-HAC estimator have the highest power and is thus the one used in the empirical study.

3.5.2 *Empirical Application*

In this section I apply the test to stock and market index returns. The pool of individual stocks consists of the four largest components of each of three indices, namely, NYSE Financial, DJIA 30 and DAX 30. The market indices are chosen for financially highly integrated economies with a significant weight in the world index since the late 1990s: S&P 500 (U.S.), FTSE 100 (U.K.), DAX 30 (Germany) and Nikkei 225 (Japan); see Demirgüç-Kunt and Levine (1996), Lane and

Milesi-Ferretti (2003). Data in daily frequency are taken from Yahoo Finance and checked against the CRSP database.

Table 3-a
Pairwise two-sided comovement test for stock returns, m=2, H_0: $C^h = C^l$ (January 1, 2007-December 31, 2010).

	NYSE Financial Top 4 (1007 obs.)				DAX 30 Top 4 (1022 obs.)		
	C	BAC	WFC		BAYN	SAP	SIE
JPM	0.118 (0.906)	0.242 (0.809)	0.093 (0.926)	BAS	0.085 (0.932)	0.058 (0.954)	-0.299 (0.765)
	0.049 (0.115)	0.012 (0.797)	-0.007 (0.868)		0.243 (**0.002**)	0.129 (0.127)	-0.025 (0.754)
C	0	-0.452 (0.651)	0.062 (0.951)	BAYN	0	0.513 (0.608)	-0.171 (0.864)
	0	0.003 (0.886)	0.007 (0.876)		0	0.222 (**0.002**)	0.086 (0.251)
BAC		0	0.610 (0.542)	SAP		0	0.613 (0.540)
		0	0.010 (0.830)			0	-0.160 (**0.007**)
WFC			0	SIE			0
			0				0
	DJIA Top 4 (1007 obs.)				U.S.-U.K.-Ger-Jp (1 Jan. 1998-31 Dec. 2012, 3483 obs.)		
	CVX	MMM	BA		FTSE	DAX	NKY
IBM	0.547 (0.584)	-0.123 (0.902)	0.289 (0.773)	S&P	-1.095 (0.274)	-0.437 (0.662)	-1.097 (0.273)
	0.097 (**0.058**)	0.031 (0.565)	-0.073 (0.125)		0.059 (**0.032**)	0.009 (0.723)	0.001 (0.921)
CVX	0	0.375 (0.708)	0.904 (0.366)	FTSE	0	-0.126 (0.900)	1.687 (**0.091**)
	0	-0.079 (0.191)	-0.192 (**0.000**)		0	0.012 (0.733)	-0.124 (**0.000**)
MMM		0	0.709 (0.478)	DAX		0	-0.153 (0.878)
		0	0.012 (0.834)			0	-0.044 (0.218)
BA			0	NKY			0
			0				0

Note: p-values are in parenthesis. The test statistic is squared and compared to a $\chi^2(1)$ distribution. The second row of each entry corresponds to the correlation differential and a Wald test of correlation equality.

For each group of returns I estimate the model with two to three states based on the model selection criteria of Smith *et al.*(2005) and BIC (not reported). It is found that state estimation is more precise when a larger number of stocks provide more information which improves power and size upon the bivariate case as reported in Table 1; see the online appendix for the very high state identification rate. I then test the null hypothesis of $H_0: C^h = C^l$ for each pairs. Values in parenthesis correspond to a Wald test of correlation equality.

Table 3-b
Pairwise two-sided comovement test, m=3, H_0: $C^h = C^l$ (January 1, 2007-December 31, 2010).

	NYSE Financial Top 4 (1007 obs.)				DAX 30 Top 4 (1022 obs.)		
	C	BAC	WFC		BAYN	SAP	SIE
JPM	0.026 (0.979)	-0.558 (0.577)	0.164 (0.870)	BAS	0.355 (0.722)	-0.240 (0.810)	0.167 (0.867)
	0.098 (**0.030**)	-0.033 (0.673)	0.012 (0.872)		0.347 (**0.032**)	0.231 (0.153)	0.135 (0.398)
C	0	-0.449 (0.653)	-0.065 (0.948)	BAYN	0	-0.239 (0.594)	-0.834 (0.404)
	0	-0.016 (0.605)	0.048 (0.276)		0	0.315 (**0.037**)	0.178 (0.235)
BAC		0	0.432 (0.666)	SAP		0	-0.189 (0.850)
		0	-0.095 (0.253)			0	-0.154 (0.133)

Three observations emerge from Table 3. First, the majority of p-values fall into the range of [0.2, 1], so these tests do not strongly support a significantly higher degree of comovement in either state. Second, the test statistic can be very different in models with different numbers of states. This is not unexpected because precisely how one classifies "volatile" and "stable" has a large bearing on the measure of state-dependent comovement. In practice, the researcher might just want to presume a two-state volatile vs. stable dichotomy without referring to a rigorous model selection procedure, in which case Table 3-a shows that FTSE 100 seems to comove a lot with Nikkei 225 in volatile periods. By contrast, for the three-regime model in Table 3-b, the DAX/NKY pair concludes otherwise. Third, Table 3 also illustrates the risk of relying solely on the traditional linear measure of comovement. Correlation differential can be significantly different from the proposed nonlinear measure in either direction.

3.5.3 *Power attrition of the state-free test*

Adopting the same parameterization as in the bivariate mixture model of Ang and Bekaert's (2002) ($\mu_1 = (1.283, 1.304)$, $\mu_2 = (-1.288, -0.692), \rho_1 = 0.4455$, $\rho_2 = 0.6097, \sigma_1 = (3.7689, 5.2194), \sigma_2 = (7.0376, 13.7177), \Gamma = [0.9818, 0.0182; 0.1454, 0.8546]$), Table 4-a demonstrates power attrition of the test of Li (2013) due to nuisance parameters whenever the DGP has a switching structure. The first row of each sub-panel replicates the result in Table 8 of Li (2013).

Table 4-a

Power attrition in the state-free tail dependence test, tail threshold c=0, H_0: symmetric tail dependen

	corr = [0.45, 0.65]				*corr* = [0.4455, 0.6097]			
	VHAC	pre-HAC	ker-HAC	KVB	VHAC	pre-HAC	ker-HAC	KVB
T=250	0.297	0.316	0.297	0.24	0.293	0.305	0.29	0.229
prob =0.5	0.149	0.152	0.131	0.11	0.128	0.144	0.128	0.103
μ_{low}=0.5	0.094	0.109	0.096	0.075	0.092	0.098	0.091	0.066
T=500	0.433	0.449	0.428	0.327	0.425	0.435	0.425	0.334
prob =0.5	0.239	0.246	0.227	0.185	0.23	0.237	0.216	0.164
μ_{low}=0.5	0.111	0.116	0.104	0.086	0.09	0.097	0.092	0.094
T=1000	0.573	0.58	0.572	0.4	0.548	0.557	0.554	0.399
prob =0.5	0.403	0.419	0.399	0.295	0.413	0.423	0.411	0.311
μ_{low}=0.5	0.149	0.152	0.141	0.119	0.171	0.17	0.16	0.123

Note: μ_{low} sets the low volatility state mean return to 0.5. *prob* =0.5 sets every entry of the transition probability matrix to 0.5. The test studies whether comovement in the upper tail is the same as comovement in the lower tail and is given in Li (2013).

Notice how the cases of μ_{low} and *prob* =0.5 are able to decrease the power of the test. To fix ideas, suppose the stable state has a lower correlation and positive mean return and the conditioning tail threshold is 0. Since Li's test defines tail with respect to zero, the test statistic will stay away from zero even when the bivariate distribution is perfectly standard and symmetric about a non-zero mean, which is likely to confuse real asymmetric comovement in the "tail" and statistical comovement due to, say, a positive mean return. In addition, the power of the test also depends on the switching probabilities. If the low correlation state is more persistent than the high correlation state, then, *ceteris paribus*, this will necessarily make upper tail dependence less salient simply because the low correlation state has a positive mean. Finally, state-dependent means also matter. Imagine the low correlation state has a mean return of 1% compared to -0.5% in the high correlation state, then, *ceteris paribus*, conditional on positive returns, this will place more weight on the low correlation state.

Table 4-b

Pairwise state-free tail dependence test, H_0: symmetric tail dependence.

	DJIA Top 4				U.S.-U.K.-Ger-Jp		
			tail threshold c=0				
	CVX	MMM	BA		FTSE	DAX	NKY
IBM	-1.650 (**0.099**)	-1.328 (0.184)	0.819 (0.413)	S&P	0.852 (0.394)	0.917 (0.359)	1.199 (0.231)
CVX	0	-2.126 (**0.033**)	-1.731 (**0.084**)	FTSE	0	-0.507 (0.612)	1.115 (0.265)
MMM		0	-1.617 (**0.106**)	DAX		0	2.508 (**0.012**)
BA			0	NKY			0
			tail threshold c=1				
	CVX	MMM	BA		FTSE	DAX	NKY
IBM	1.285 (0.199)	0.257 (0.797)	1.339 (0.181)	S&P	1.740 (**0.082**)	0.745 (0.456)	1.749 (**0.080**)
CVX	0	-0.086 (0.931)	1.180 (0.238)	FTSE	0	-0.462 (0.644)	1.899 (**0.058**)
MMM		0	1.202 (0.229)	DAX		0	0.241 (0.809)
BA			0	NKY			0

Note: p-values are in parenthesis. Standard errors are based on the Newey-West HAC estimator. The test statistic is squared and compared to a $\chi^2(1)$ distribution.

As a last experiment, I apply the asymmetric tail dependence test of Li (2013) to two empirical models studied in Section 5.2. In Table 4-b, a positive (negative) test statistic corresponds to larger upper (lower) tail dependence. First note that lower tail dependence is not always stronger; more importantly, while my test does not support asymmetric comovement across states in almost all cases (Table 3), Li's test rejects symmetric tail comovement in many of them even when there is no tail dependence conditional on the state. Intuitively this is because the interaction between tails of the marginal distribution intimately depends on factors such as the tail threshold and the mixture probabilities. In the general case, a state-free tail comovement measure depends heavily on the magnitude of correlations, how often the process hits the high correlation state, and how the mean return of one state compares with the other.

3.6 Extensions

The method used to prove the main result of this chapter can now be extended to a family of tests based on the construction of C:

1. $C(i) = P\big((R_{t+1} - R_t) > c|i\big)$ or $C(i) = P\big((R_{t+1} - R_t) < -c|i\big)$;
2. $C(i) = P(\Delta R^1_{t+1} > c, \Delta R^2_{t+1} > c|i)$;
3. $C(i) = 2P\big((R^*_{t+1} - R^*_t)(R^M_{t+1} - R^M_t) > 0|i\big) - 1$;
4. $C(i|c) = P\left((R^1_{t+1} - R^1_t)(R^2_{t+1} - R^2_t) > 0|i, R^{1,2}_t > c \text{ (or} < -c)\right)$, fix c;
5. $C(i|c) =$ $P\left((R^1_{t+1} - R^1_t)(R^2_{t+1} - R^2_t) > 0|i, R^{1,2}_t > c\right) - P\left((R^1_{t+1} - R^1_t)(R^2_{t+1} - R^2_t) > 0|i, R^{1,2}_t < -c\right)$, fix c or i or both;
6. $C(i|c) = P\big((R_{t+1} - R_t) > c|i\big) - P\big((R_{t+1} - R_t) < -c|i\big)$, fix c or i or both;

In the first case, c is a nonnegative threshold and the test can be used to address the question of whether (big) positive or negative swings are more likely in one regime compared to the other. Case two is the bivariate version of case one. In the third case, $\{R^*_t\}$ can be a return series of one stock, the top principal component of a group of stocks, or a portfolio sorted by industry classification, size, book-to-market, or momentum; $\{R^M_t\}$ is the market performance. The test can be used to test comovement with the market in different regimes. Case four is an upper (lower) tail dependence test across regimes. Case five is a variation on the test given in Li (2013) which is now conditioned on the states. Finally, case six is a symmetry test of large swings.

3.7 CONCLUDING REMARKS

In this chapter, I have proposed a Kendall's τ-type test statistic to study state-dependent stock return comovement. The test is shown to have power against $\sqrt{n}$-local alternatives and can be mapped into local alternative DGP's that are allowed to take on a much richer distributional form in the meta-elliptical family. For two pairs of international market index returns, the new test does marginally support asymmetric state-dependent comovement in two opposite directions, whereas in all other cases, one still needs more evidence. I find that correlation-based comovement differential can be quite different from the proposed measure and that asymmetric tail dependence can be much less significant in regime switching models. It would be interesting to separate the two and study the leftover asymmetry conditional on the state.

The study nevertheless begs the following question: Why does or does not asymmetric comovement in asset returns arise in the first place? Suppose investors can use momentum, size, value or other strategies to exploit asymmetric comovement across stocks, portfolios, and markets, does it still exist? If yes, this might suggest that domestic and international portfolios tend to be less diversified in one state of the world compared to the other, and, measured by whether there is a surge in comovement, certain portfolios are more absorbent of large shocks than others; if not, one might say that, on average, shocks of different magnitudes do not seem to detract from portfolio diversification and risk propagation is largely diversified away except in the most extreme events. Based on my study, the sweeping sentiment that comovement is higher in volatile periods needs to be reevaluated case by case.

3.8 APPENDIX A: PROOFS AND TECHNICAL NOTES

3.8.1 *Long-run variance estimation*

One way to estimate the asymptotic variance is by means of a kernel-based HAC estimator, e.g., the Newey-West estimator. Based on the regularity conditions laid out in Andrews (1991) and Newey and West (1994), a consistent estimator of the asymptotic variance takes the form: $\hat{\sigma}^2_{\mathrm{NW}} = \sum_{l=1-h}^{h-1} k_{Bartlett}(l/h)\hat{\sigma}_l^2$, $\hat{\sigma}_l^2 = 1/n \sum_{t=|l|+1}^{n} \hat{d}_t \hat{d}_{t-|l|}$, $\hat{d}_t = \widehat{W}_{nt}^h - \widehat{W}_{nt}^l$. A plug-in lag selection procedure is given in Newey and West (1994).

Alternatively one can use the VARHAC procedure proposed in den Haan and Levin (1998). Their method is to first fit a finite order VAR selected by AIC and then compute the HAC estimate from the long run covariance matrix implied by the VAR. The estimator is positive semidefinite by construction and has a convergence rate faster than any positive semidefinite kernel-based HAC estimator for almost all autocovariance structures. Monte Carlo results in den Haan and Levin (1998) show that the VARHAC estimator (univariate or multivariate) compares favorably to kernel estimators when the data generating process is a finite order ARMA and is especially good in higher order VAR processes, whereas the first order prewhitened kernel-HAC estimator performs somewhat better in lower order MA models. den Haan and Levin (2000) find that the lag order chosen by BIC is typically much smaller than the true lag order, leading to substantial bias in the estimated spectral density at frequency zero, whereas the somewhat higher lag order chosen by AIC reduces this bias at little cost in terms of additional variance, especially for highly serially correlated data. As serial or cross-sectional independence conditional on the state does not imply unconditional independence, the presence of Markov switching induces both persistent serial dependence and cross-sectional dependence in the component series. So I use AIC to select the lag order for VARHAC. An extension to VARHAC is the *prewhitened* HAC, in which I apply the kernel HAC estimator to the residuals of a VAR selected by AIC before the recoloring step.

Proposition A.1 Under regularity conditions of den Haan and Levin (1998, Theorem 2), which is straightforward to verify for the current functional form, the variance of the limiting distribution of the test statistic can be estimated in two steps:

1. Fit a finite order AR model to $n^{-1/2}\ \sum_{t=1}^{n}\left(\widehat{W}_{nt}^{h}-\widehat{W}_{nt}^{l}\right)$ using AIC to choose the lag order.
2. Estimate its long run variance by $\hat{\sigma}_{LR}^{2}=\hat{\sigma}^{2}/\left(1-\sum_{i=1}^{p}\hat{\phi}_{i}\right)^{2}$, in which $\hat{\sigma}^{2}=1/(n-p)\sum_{i=p+1}^{n}\hat{e}_{i}^{2}$ is the sample variance of estimated AR residuals. Use $\hat{\sigma}_{LR}^{2}$ to estimate σ^{2} in Lemma 4.

A proof can be found in den Haan and Levin (1998).

LLN and refinement

Proof of Lemma 1.

By the strong Markov property, when the process returns to the original state α, it starts anew as if being regenerated independently; see Brémaud (1999, p. 126) for a proof of the independence of regenerative cycles. The waiting times are thus independent of each other and are also independent from the state. Hence $S_j(f), j=2,3,\ldots$ are i.i.d. and

$$\frac{1}{\eta_n^*}\sum_{j=1}^{\eta_n^*}S_j(f)=\frac{1}{\eta_n^*}S_1(f)+\left(\frac{\eta_n^*-1}{\eta_n^*}\right)\left(\frac{1}{\eta_n^*-1}\right)\sum_{j=2}^{\eta_n^*}S_j(f)$$

Proposition 14.2.9 in Cappé *et al.* (2005) shows that $P(\tau_\alpha<\infty)=1$ *iff* the accessible state α is recurrent. This result holds here given the assumptions about f and the Markov chain, in which case $\eta_n^*\to\infty$ as $n\to\infty$, whereas S_1 is finite almost surely. So the first term vanishes asymptotically and $(\eta_n^*-1)/\eta_n^*\to 1$.

At this point one would like to apply SLLN for i.i.d. series to the second term on the right, for which it remains to be shown that $\mu_{IM}(f)$ is finite. Note that $E_\alpha(\tau_\alpha)$ is finite by positive recurrence of the Markov chain and $E_\alpha(|f(x_t)|)$ is finite by integrability; hence the argument follows:

$$E_\alpha\left(\left|\sum\nolimits_{t=\tau_\alpha^{(j-1)}+1}^{\tau_\alpha^{(j)}}f(x_t)\right|\right)=E_\alpha\left(\left|\sum\nolimits_{t=\tau_\alpha+1}^{\tau_\alpha^{(2)}}f(x_t)\right|\right)\leq E_\alpha\left(\sum\nolimits_{t=\tau_\alpha+1}^{\tau_\alpha^{(2)}}|f(x_t)|\right)$$
$$=E_\alpha(\tau_\alpha)\,E_\alpha(|f(x_t)|)<\infty.$$

■

Bias-in-mean correction

From distribution theory, a further finite-sample bias-in-mean correction can be based on a second-order Taylor expansion. Let $Y_j = \left(S_j(f), {\tau_\alpha}^{(j)} - \tau_\alpha^{(j-1)}\right)$ denote an independent, identically distributed random vector. Assume that the first four cumulants of this distribution exist and are finite. It is straightforward to show that for a four-times differentiable function $f\colon \mathrm{Y} \to \mathbf{R}$, where Y is a convex subset of $\mathbb{R}^2$ such that $P(Y_j \in \mathrm{Y}) = 1$. Assume $\sup_{i,j,k,l=1,2} \sup_{y\in \mathrm{Y}} |f_{ijkl}^{(4)}(y)| < \infty$, then

$$E\left(\frac{1/(\eta_n^*-1)\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} f(x_t)}{1/(\eta_n^*-1)\left(\tau_\alpha^{(\eta_n^*)}-\tau_\alpha\right)} - \frac{\mu_{IM}(f)\left(\left(\tau_\alpha^{(\eta_n^*)}-\tau_\alpha\right)/(\eta_n^*-1)-\mu_{IM}(1)\right)^2}{\mu_{IM}^3(1)}\right)$$

$$\equiv E\left(\frac{\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} f(x_t)}{\left(\tau_\alpha^{(\eta_n^*)}-\tau_\alpha\right)}\right) - \frac{\mu_{IM}(f)\sigma_{\tau_\alpha}^2}{\mu_{IM}^2(1)(\eta_n^*-1)} = \frac{\mu_{IM}(f)}{\mu_{IM}(1)} + O\left({\eta_n^*}^{-2}\right) = \frac{\mu_{IM}(f)}{\mu_{IM}(1)} + O(n^{-2}).$$

The last equality comes from the fact that $\mu_{IM}(1)$ is almost surely finite. $\mu_{IM}(1)$ and $\mu_{IM}(f)$ can be either estimated by their sample counterparts or calculated exactly from estimated transition probabilities, e.g.,

$$E_\alpha(\tau_\alpha) = P(x_{t+1} = \alpha | x_t = \alpha) + 2P(x_{t+2} = \alpha, x_{t+1} \neq \alpha | x_t = \alpha) + \cdots$$
$$= p_{\alpha\alpha} + 2p_{\alpha\bar{\alpha}}p_{\bar{\alpha}\alpha} + 3p_{\alpha\bar{\alpha}}p_{\bar{\alpha}\bar{\alpha}}p_{\bar{\alpha}\alpha} + \cdots = p_{\alpha\alpha} + p_{\alpha\bar{\alpha}}p_{\bar{\alpha}\alpha}\frac{2-p_{\bar{\alpha}\bar{\alpha}}}{(1-p_{\bar{\alpha}\bar{\alpha}})^2},$$

where $p_{\alpha\bar{\alpha}} = 1 - p_{\alpha\alpha}$, $p_{\bar{\alpha}\bar{\alpha}} = 1 - p_{\bar{\alpha}\alpha} = 1 - p_{\alpha\bar{\alpha}}p_\alpha/p_{\bar{\alpha}}$, p_α is the stationary distribution.

To prove Theorem 1, I need the following two lemmas:

Lemma A.1 (Regenerative cycle correction) Under conditions of Lemma 1,

$$\frac{1}{n}\sum_{t=1}^{n} f(x_t) \xrightarrow{a.s.} \mu_f \equiv \frac{\mu_{IM}(f)}{\mu_{IM}(1)}.$$

A regenerative cycle-corrected estimate of μ_f is

$$\frac{\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} f(x_t)}{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha} \xrightarrow{a.s.} \frac{\mu_{IM}(f)}{\mu_{IM}(1)},$$

in which $1/(\eta_n^* - 1) \sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} f(x_t) \xrightarrow{a.s.} \mu_{IM}(f)$ and $1/(\eta_n^* - 1)\,(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha) \xrightarrow{a.s.} \mu_{IM}(1)$.

Lemma A.2 Let $r_t = (R_{t+1}^1 - R_t^1)(R_{t+1}^2 - R_t^2)$ and consider a bounded function $f: \mathbf{X} \times \mathbb{R} \to \mathbb{R}$ such that $P\big((x_t, r_t) \in \mathbf{X} \times \mathbb{R}\big) = 1$, $\mathrm{P}(\mathrm{X} \times \mathbb{R} \in \mathbf{X} \times \mathbb{R}) = 1$ and $E_\alpha(|f|) < \infty$. Define

$$f^i(x_t, r_t) = \frac{D_t I(x_t = i)}{P(x_t = i)} = \frac{I(r_t > 0, x_t = i) - I(r_t < 0, x_t = i)}{P(x_t = i)}, \quad i = 1,2, \dots, m.$$

Then,

$$\frac{\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} f^i(x_t, r_t)}{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha} \xrightarrow{a.s.} E\left(f^i(x_t, r_t)\right) = 2P(r_t > 0|i) - 1.$$

Proof of Lemma A.1

The proof of this lemma makes use of the observation that, conditional on the sample size n, $\tau_\alpha^{(j)}$ and $f(x_t)$ are asymptotically independent. Intuitively, when the states are persistent as is often the case in real data, the process may hit a long string of the same state at the beginning which, if included for the construction of the test statistic, will contribute a large yet very random component to the test. Unless the sample size is very large, this correction is non-negligible. By the definition of $S_j(f)$,

$$\sum_{j=1}^{\eta_n^*} S_j(f) \le \sum_{t=1}^{n} f(x_t) < \sum_{j=1}^{\eta_n^*+1} S_j(f)$$

which is equivalent to

$$\left(\frac{\eta_n^*}{n}\right)\left(\frac{1}{\eta_n^*}\right)\sum_{j=1}^{\eta_n^*} S_j(f) \le \frac{1}{n}\sum_{t=1}^{n} f(x_t) < \left(\frac{\eta_n^* + 1}{n}\right)\left(\frac{1}{\eta_n^* + 1}\right)\sum_{j=1}^{\eta_n^*+1} S_j(f).$$

Applying Lemma 1 to both sides yields

$$\left(\frac{n}{\eta_n^*}\right)\frac{1}{n}\sum_{t=1}^{n} f(x_t) \xrightarrow{a.s.} \mu_{IM}(f).$$

Applying Lemma 1 again to $f = 1$ and note that $\tau_\alpha^{(\eta_n^*)} \leq n < \tau_\alpha^{(\eta_n^*+1)}$ with $P\left(\tau_\alpha^{(\eta_n^*+1)} - \tau_\alpha^{(\eta_n^*)} < \infty\right) = 1$,

$$\frac{1}{\eta_n^*}\sum_{j=1}^{\eta_n^*} S_j(f = 1) = \frac{\tau_\alpha^{(\eta_n^*)}}{\eta_n^*} = \frac{n - (n - \tau_\alpha^{(\eta_n^*)})}{\eta_n^*} \xrightarrow{a.s.} \lim_{n\to\infty}\frac{n}{\eta_n^*} \xrightarrow{a.s.} \mu_{IM}(f = 1) \equiv E_\alpha(\tau_\alpha) < \infty.$$

By the Ergodic Theorem and asymptotic independence of $\tau_\alpha^{(j)}$ and $f(x_t)$, $\mu_f = \mu_{IM}(f)/\mu_{IM}(1)$. This proves the first part. To prove the second part, first note that

$$E\left(\frac{1}{\eta_n^*}\sum_{j=1}^{\eta_n^*} S_j(f)\right) = \frac{1}{\eta_n^*}E\big(S_1(f)\big) + \frac{\eta_n^* - 1}{n}\mu_{IM}(f),$$

$$E\left(\frac{1}{\eta_n^*}\sum_{j=1}^{\eta_n^*+1} S_j(f)\right) = \frac{1}{\eta_n^*}E\big(S_1(f)\big) + \mu_{IM}(f),$$

$$\frac{1}{\eta_n^*}\sum_{t=1}^{n} f(x_t) = \frac{1}{\eta_n^*}S_1(f) + \frac{1}{\eta_n^*}\sum_{j=2}^{\eta_n^*} S_j(f) + \frac{1}{\eta_n^*}\sum_{t=\tau_\alpha^{(\eta_n^*)}+1}^{n} f(x_t).$$

Combining the three equations and after some rearrangement one gets

$$E\left(\frac{1}{\eta_n^*-1}\left(\sum_{t=1}^{n} f(x_t) - S_1(f) - \sum_{t=\tau_\alpha^{(\eta_n^*)}+1}^{n} f(x_t)\right)\right)$$

$$= E\left(\frac{1}{\eta_n^*-1}\left(\sum_{\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} f(x_t)\right)\right) = \mu_{IM}(f).$$

Finally, by the independence of regenerative cycles,

$$E\left(\frac{1}{\eta_n^*-1}\left(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha\right)\right) = \mu_{IM}(1).$$

■

Proof of Lemma A.2

In view of Lemma A.1 it suffices to show that $Ef^i(x_t, r_t) = 2P(r > 0|x_t = i) - 1$, i.e., the theorem holds for (x_t, r_t) on the product measurable space $(\mathrm{X}, \mathcal{X}) \otimes (\mathbb{R}, \mathcal{B})$. For example, when there are three regimes, $\mathrm{X} = \{1,2,3\}$, $\mathcal{X} = \sigma[\mathrm{X}]$ and $r_t \in \mathcal{B}$. Define the product σ-field to be $\mathcal{X} \times \mathcal{B} \equiv \sigma[\mathcal{F}]$ with $\mathcal{F} = \left\{\sum_{j=1}^{k} X_j \times r_j : k \geq 1, X_j \in \mathcal{X}, r_j \in \mathcal{B}\right\}$, in which $\sum_{j=1}^{k}$ is a countable non-overlapping sum. Let $(\mathrm{X}, \mathcal{X}, \mu)$ and $(\mathbb{R}, \mathcal{B}, \nu)$ be two σ-finite probability measure spaces then by the existence theorem for product measures one can define a unique σ-finite product measure $\phi\left(\sum_{j=1}^{k} X_j \times r_j\right) = \sum_{j=1}^{k} \mu(X_j) \times \nu(r_j)$.

Now let $F = I(r_t > 0, X_t = i) = I_{[X_t=i]}(\omega) \times I_{[r_t>0]}(\omega'), \omega \in X, \omega' \in \mathbb{R}$. Since $F \geq 0$, is bounded and almost everywhere continuous, it is ϕ-integrable. Thus by Fubini's theorem, any r-section of F, $F_{r\in\mathcal{B}}\left(X_j(\omega)\right)$ is μ-integrable on X. Hence $E(F) = \int_{\mathrm{X}\times\mathbb{R}} F(\omega, \omega') d\phi(\omega, \omega') = P(x_t = i|r > 0)P(r > 0) = P(x_t = i, r > 0)$. It follows that $Ef^i(x_t, r_t) = 2P(r > 0|x_t = i) - 1$.

■

Proof of Theorem 1

It suffices to prove the first part. First note that

$$\frac{\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \hat{f}^i(x_t,r_t)}{\tau_\alpha^{(\eta_n^*)}-\tau_\alpha} = \frac{\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} f^i(x_t,r_t)}{\tau_\alpha^{(\eta_n^*)}-\tau_\alpha} \times \left(\frac{P(x_t=i)}{1/\left(\tau_\alpha^{(\eta_n^*)}-\tau_\alpha\right)\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} I\,(x_t=i)} \right).$$

Applying Lemma A.1 to $f = I(x_t = i)$ yields $1/(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha)\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} I\,(x_t = i) \xrightarrow{a.s.} P(x_t = i) > 0$. The result then follows by Lemma A.2.

■

3.8.2 *CLT and the test statistic*

Lemma A.3 Let $f\colon \mathbf{X} \to R$ be a real-valued, bounded function such that $P(x_t \in \mathbf{X}) = 1, t = 1,2,\ldots$, $\mathrm{P}(\mathrm{X} \in \mathbf{X}) = 1$ and $E_\alpha(f^2) < \infty$. Define $\bar{f}(x_t) = f(x_t) - E_\alpha(f_t)$, under Assumption 1 and 2,

$$\left(\frac{1}{\tau_\alpha^{(\eta_n^*)}-\tau_\alpha}\right)^{\frac{1}{2}} \sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \bar{f}(x_t) \xrightarrow{d} N\left(0,\sigma_{\bar{f}}^2\right).$$

Define S_j as in Lemma 1 then by the strong Markov property, $\{S_j(\bar{f})\}_{j=2}^{\eta_n^*}$ are i.i.d. $E(S_j(\bar{f})) = \mu_{IM}(f) - E_\alpha(\tau_\alpha)E_\alpha(f) = 0$ by Theorem 1. Next, $E(S_j(\bar{f})^2) < \infty$ because $E(\tau_\alpha) < \infty$ and $E(f^2) < \infty$ by assumption. For fixed η_n^* and τ_α , one can apply the Lindberg-Lévy central limit theorem for i.i.d. series to $S_j(\bar{f})$. Unfortunately this simple approach does not work here because, as $n \to \infty$ both η_n^* and τ_α are random variables. While a delta method is not directly applicable here, the correct theorem to use is the so-called Anscombe's theorem, a special case of which is given as follows.

Lemma A.4 (Anscombe, 1952) Suppose $X_1, X_2, \ldots, X_n, \ldots$ are independent and identically distributed random variables with mean zero and variance 1. Let S_n=$X_1 + X_2 + \cdots + X_n$. Let $n(t)$ denote a positive integer-valued random variable for any t>0 such that $n(t)/t$ converges in probability to a constant $c > 0$ as $t \to \infty$. Then the following holds:

$$\frac{S_{n(t)}}{\sqrt{n(t)}} \xrightarrow{d} N(0,1).$$

A proof may be based on Kolmogorov's inequality and can be found in Anscombe (1952).

Remark. It is tempting to conjecture an alternative noncentral F-test as follows:

$$\left(\frac{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha}{\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \bar{f}(x_t)}\right)^2 = \frac{1/(\eta_n^* - 1)\left(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha\right)^2}{1/(\eta_n^* - 1)\left(\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \bar{f}(x_t)\right)^2} \xrightarrow{d} \frac{N^2\left(\mu_{IM}(1), \sigma_{\tau_\alpha}^2\right)}{N^2\left(0, \sigma_{S(\bar{f})}^2\right)}$$

$$\sim \frac{\sigma_{\tau_\alpha}^2}{\mu_{IM}(\tau_\alpha)\sigma_{\bar{f}}^2} F\left(1,1; \lambda = \mu_{IM}^2(1)\right),$$

in which λ is the noncentral parameter and I have used $\sigma_{\bar{f}}^2 = \sigma_{S(\bar{f})}^2 / \mu_{IM}(\tau_\alpha)$ together with Anscombe's theorem. Monte Carlo results show that the actual size and power of this test are irreconcilable with theory and they are not converging as the sample size grows large. A closer look at the numerator and denominator reveals that $\tau_\alpha^{(j)}$ and $S_j(f)$ are not independent in general.

Proof of Lemma A.3

Recall from Theorem 1 that $n/\eta_n^* \xrightarrow{a.s.} \mu_{IM}(1) \equiv E_\alpha(\tau_\alpha) < \infty$, and by construction $\sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \bar{f}(x_t) = \sum_{j=2}^{\eta_n^*} S_j(\bar{f})$. So Anscombe's theorem is readily applicable:

$$\left(\frac{1}{\eta_n^* - 1}\right)^{\frac{1}{2}} \sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \bar{f}(x_t) = \left(\frac{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha}{\eta_n^* - 1}\right)^{\frac{1}{2}} \left(\frac{1}{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha}\right)^{\frac{1}{2}} \sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \bar{f}(x_t)$$

$$= \left(\frac{1}{\eta_n^* - 1}\right)^{\frac{1}{2}} \sum_{j=2}^{\eta_n^*} S_j(\bar{f})$$

$$\xrightarrow{a.s.} N\left(0, \sigma_{S(\bar{f})}^2\right).$$

Again from the proof of Theorem 1, $\left(\left(\tau_\alpha^{(\eta_n^*)} - \tau_\alpha\right)/(\eta_n^* - 1)\right)^{1/2} \xrightarrow{a.s.} E(\tau_\alpha)^{1/2} = \mu_{IM}(1)^{1/2} = constant.$ Hence

$$\left(\frac{1}{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha}\right)^{\frac{1}{2}} \sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \bar{f}(x_t) \xrightarrow{d} N\left(0, \sigma_{\bar{f}}^2\right),$$

where $\sigma_{\bar{f}}^2 = \sigma_{S(\bar{f})}^2 / E_\alpha(\tau_\alpha)$.

■

Corollary A.1 Under conditions of Lemma A.2,

$$\left(\frac{1}{\tau_\alpha^{(\eta_n^*)} - \tau_\alpha}\right)^{\frac{1}{2}} \sum_{t=\tau_\alpha+1}^{\tau_\alpha^{(\eta_n^*)}} \bar{f}(x_t, r_t) \xrightarrow{d} N\left(0, \sigma_{\bar{f}}^2\right),$$

with $f(x_t, r_t)$ defined as in Lemma A.1 and $\sigma_{\bar{f}}^2$ redefined in the obvious way.

Proof of Corollary A.1

The proof to this result is essentially the same as in Lemma A.2 and A.3 and the integrability condition is easy to verify. Another way to justify the use of Fubini's theorem is by Tonelli's theorem which only requires the function to be nonnegative (e.g., $I(r_t > 0, x_t = i) \geq 0$).

■

Proof of Theorem 2.

Only the first part of the Theorem needs to be proved. The second part can be justified by Lemma A.3. Recall that $W_t^h = (D_t - C^h)I(x_t = h)/P(x_t = h)$, so

$$\begin{aligned}
\frac{\left(D_t - \hat{C}^h\right)I(x_t = h)}{\left((1/n)\sum_{t=1}^n I\,(x_t = h)\right)} &= \frac{\left(D_t - C^h + \left(C^h - \hat{C}^h\right)\right)I(x_t = h)}{P(x_t = h)} \cdot \frac{P(x_t = h)}{\left((1/n)\sum_{t=1}^n I\,(x_t = h)\right)} \\
&= \left(W_t^h + \frac{\left(C^h - \hat{C}^h\right)I(x_t = h)}{P(x_t = h)}\right) \cdot \left(\frac{P(x_t = h)}{\left((1/n)\sum_{t=1}^n I\,(x_t = h)\right)}\right) \\
&\xrightarrow{a.s.} W_t^h.
\end{aligned}$$

Let $\hat{f}^h(x_t, r_t) = D_t I(x_t = h) / \left(1/n \sum_{t=1}^{n} I\,(x_t = h)\right)$ and apply Theorem 2, the last line follows from $\hat{C}^h - C^h = 1/n \sum_{t=1}^{n} \hat{f}^h(x_t, r_t) - C^h \xrightarrow{a.s.} 0$. The argument also holds for W_t^l, hence,

$$\widehat{W}_{nt}^{h,l} = W_t^{h,l} + o_p(1).$$

This relationship can in fact be further refined:

$$\hat{C}^h = \frac{1/n \sum_{t=1}^{n} (D_t - C^h + C^h) I(x_t = i)}{\left(1/n \sum_{t=1}^{n} I\,(x_t = i)\right)} = \left(\frac{P(x_t = h)}{\left((1/n) \sum_{t=1}^{n} I\,(x_t = h)\right)}\right) \frac{1}{n} \sum_{t=1}^{n} W_t^h + C^h;$$

under the null $C^h = C^l$,

$$\sqrt{n}\left(\hat{C}^h - \hat{C}^l\right) = \left(n^{-\frac{1}{2}} \sum_{t=1}^{n} \left(W_t^h - W_t^l\right)\right) \left(\frac{P(x_t = h)}{\left((1/n) \sum_{t=1}^{n} I\,(x_t = h)\right)}\right)$$

$$= n^{-\frac{1}{2}} \sum_{t=1}^{n} \left(W_t^h - W_t^l\right) + o_p(1).$$

The asymptotic normal distribution follows by defining $f(x_t, r_t) = f^h(x_t, r_t) - f^l(x_t, r_t) = W_t^h - W_t^l$ and applying Theorem 3; $\widehat{W}_{nt}^{h,l}$ has sample mean zero.

■

3.8.3 *Power analysis*

Lemma A.4 Under regularity conditions given in Section 4.2,

$$P\left(\left|\hat{\tau}_{ij} - \tau_{ij}\right| < \frac{\epsilon}{\sqrt{n}}\right) > 1 - 2e^{-\epsilon^2/8}.$$

Lemma A.5 Under conditions of Lemma A.4, plus the random vector $\boldsymbol{X^n}$ is multivariate normal, then the following holds

$$P\left(\left\|\tilde{\Sigma} - \Sigma\right\|_\infty \geq \frac{\epsilon}{\sqrt{n}}\right) < k(k-1)e^{-\epsilon^2/2\pi^2}, k \geq 2,$$

where $\tilde{\Sigma}$ is the *implied* sample correlation matrix defined element-wise as $\tilde{\rho}_{ij} = \sin((\pi/2)\hat{\tau}_{ij})$. Lemma A.5 corresponds to the Gaussian Markov switching model which shows that for large k and fixed ϵ, a large sample may be necessary for the researcher to get a probabilistically accurate estimate of Σ. Put another way, for DGPs with correlation matrices close to each other, the estimated Kendall's τ may not have strong distinguishing power, in finite samples, with respect to a non-negligible range of implied and estimated sample correlation matrices, hence detracting from the power of the test.

Proof of Lemma A.4

The proof of this lemma is to first observe that the kernel of the U-statistic, $\hat{\tau}_{ij}$, is $I(X_i^p - X_i^q)I(X_j^p - X_j^q)$ which is bounded in $[-1,1]$. The rest of the proof is an application of Hoeffding's inequality. The inequality works best for large ϵ and is only informative for $\epsilon, s.t. e^{-\epsilon^2/8} \leq 1/2$, i.e., the right hand side falls in the range of $[0, 1)$. Next I present the lemma that is used in the proof of Theorem 3. Apply Hoeffding's inequality to the current unbiased estimator of the coefficient of concordance, one gets

$$P(\hat{\tau}_{ij} - \tau_{ij} \geq \epsilon) \leq e^{-\lfloor n/2 \rfloor \epsilon^2/2} < e^{-n\epsilon^2/8},$$

or by symmetry and making use of a local root-n drift,

$$P\left(\left|\hat{\tau}_{ij} - \tau_{ij}\right| < \frac{\epsilon}{\sqrt{n}}\right) > 1 - 2e^{-\epsilon^2/8}.$$

■

Proof of Lemma A.5

As a slight extension to the result given in Kruskal (1958), when the random vector is multivariate normal, τ_{ij} is related to the Pearson's correlation coefficient in a very simple manner, $\rho_{ij} = \sin((\pi/2)\tau_{ij})$. Define the *implied* sample correlation as $\tilde{\rho}_{ij} = \sin((\pi/2)\hat{\tau}_{ij})$ which forms a one-to-one correspondence with $\hat{\tau}_{ij}$. Note that $\tilde{\rho}_{ij}$ is in general not the same as the *estimated* sample correlation $\hat{\rho}_{ij}$ based on, say, the EM algorithm. Nevertheless, it is a very useful indicator for local alternative data generating processes and can be used to determine the extent to which different DGPs result in different comovement measures of the test. Using essentially the same techniques in the proof of Theorem 4.2 in Liu *et al.* (2012), one gets

$$P\left(\tilde{\rho}_{ij} - \rho_{ij} \geq \epsilon\right) = P\left(\sin(\frac{\pi}{2}\hat{\tau}_{ij}) - \sin(\frac{\pi}{2}\tau_{ij}) \geq \epsilon\right) \leq P\left(\left|\hat{\tau}_{ij} - \tau_{ij}\right| \geq \frac{2}{\pi}\epsilon\right) < e^{-n\epsilon^2/2\pi^2},$$

or

$$P\left(\left|\tilde{\rho}_{ij} - \rho_{ij}\right| \geq \epsilon\right) < 2e^{-n\epsilon^2/2\pi^2}.$$

By symmetry of the implied correlation matrix $\tilde{\Sigma}$ and using the union bound yields

$$P\left(\left\|\tilde{\Sigma} - \Sigma\right\|_\infty \geq \epsilon\right) < \frac{k(k-1)}{2} \times 2e^{-n\epsilon^2/2\pi^2} = k(k-1)e^{-n\epsilon^2/2\pi^2}, k \geq 2.$$

In terms of local data generating processes, the above inequality can be rewritten as

$$P\left(\left\|\tilde{\Sigma} - \Sigma\right\|_\infty \geq \frac{\epsilon}{\sqrt{n}}\right) < \frac{k(k-1)}{2} \times 2e^{-\epsilon^2/2\pi^2} = k(k-1)e^{-\epsilon^2/2\pi^2}, k \geq 2.$$

■

Proof of Theorem 3.

Only the first part of the theorem needs a proof; the second part follows from Lemma A.5.

$$\begin{aligned} P\left(\left|\hat{C}^h - \hat{C}^l\right| \geq \frac{\Delta}{\sqrt{n}}\right) &= P\left(\left|\hat{C}^h - C^h + C^h - C^l + C^l - \hat{C}^l\right| \geq \frac{\Delta}{\sqrt{n}}\right) \\ &\leq P\left(\left|\hat{C}^h - C^h\right| + \left|C^h - C^l\right| + \left|C^l - \hat{C}^l\right| \geq \frac{\Delta}{\sqrt{n}}\right) \\ &= 1 - P\left(\left|\hat{C}^h - C^h\right| + \left|C^h - C^l\right| + \left|C^l - \hat{C}^l\right| < \frac{\Delta}{\sqrt{n}}\right) \\ &\leq 1 - P\left(\left\{\left|\hat{C}^h - C^h\right| < \frac{\Delta - \delta}{2\sqrt{n}}\right\} \cap \left\{\left|\hat{C}^l - C^l\right| < \frac{\Delta - \delta}{2\sqrt{n}}\right\}\right) \\ &\leq 1 - P\left(\left|\hat{C}^h - C^h\right| < \frac{\Delta - \delta}{2\sqrt{n}}\right) P\left(\left|\hat{C}^l - C^l\right| < \frac{\Delta - \delta}{2\sqrt{n}}\right) \\ &< 4e^{-(\Delta-\delta)^2/16}\left(e^{(\Delta-\delta)^2/32} - 1\right), \qquad \text{for } e^{-(\Delta-\delta)^2/32} \leq \frac{1}{2}. \end{aligned}$$

The fourth line holds from assumption MC2. The last line does not apply for $\delta < \Delta < 4\sqrt{2\log 2} + \delta$ *s.t.* $e^{-(\Delta-\delta)^2/32} > 1/2$ since the lower bound on $P\left(\left|\hat{C}^h - C^h\right| < (\Delta - \delta)/2\sqrt{n}\right)$ becomes some non-positive number, in which case, $P\left(\left|\hat{C}^h - \hat{C}^l\right| \geq \frac{\Delta}{\sqrt{n}}\right) \leq 1$.

Proof of Corollary 1.

The proof of this corollary is very similar to the proof of Theorem 3 and follows from $P\left(\hat{C}^h - C^h < (\Delta - \delta)/2\sqrt{n}\right) > 1 - e^{-(\Delta-\delta)^2/32}$ for $\Delta > \delta$.

■

3.9 APPENDIX B: STATE INFERENCE AND IDENTIFICATION

Three commonly used methods in state inference are filtering, smoothing and Viterbi's algorithm. I take the filtered (smoothed) state to be the one that has the highest filtered (smoothed) probability; in Viterbi's algorithm, a single most probable state is produced at each point in time as a byproduct of the global maximum likelihood. In the following simulation study, I compare filtered states, smoothed states and global most probable path inference by computing the percentage of correct matches for the case of $m = 3$. Table A.1 is based on 1000 simulated samples of size 1000 from the corresponding estimated models. A match is said to occur if the estimated state coincides with the true state.

Table B.1

Percentages of state matching for the estimated models, MLE, m=3.

	low vol. state			medium vol. state			high vol. state		
	filter	smth	Vit	Filter	smth	Vit	filter	smth	Vit
NYSE Fin Top 4	0.951	0.955	0.955	0.787	0.821	0.822	0.933	0.952	0.953
DAX 30 Top 4	0.942	0.946	0.947	0.768	0.819	0.821	0.911	0.935	0.935
DJIA Top 4	0.949	0.950	0.949	0.792	0.844	0.843	0.932	0.945	0.945

BIBLIOGRAPHY

Andrews, D. W. (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59(3), 817-858.

Ang, A., and Bekaert, G. (2002), International asset allocation with regime shifts. *Review of Financial studies* 15(4), 1137-1187.

Ang, A., and Chen, J. (2002), Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63(3): 443-494.

Anscombe, F. J. (1952), Large sample theory of sequential estimation. *Proceedings of the Cambridge Philosophical Society* 48(4), 600-607.

Bai, J., and Ng, S. (2001), A consistent test for conditional symmetry in time series models. *Journal of Econometrics* 103(1), 225-258.

Bekaert, G., Hodrick, R. J., and Zhang, X. (2008), *Stock return comovements*. ECB working paper 931, European Central Bank.

Bickel, P., Ritov, Y., and Rydén, T. (1998), Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *The Annals of Statistics* 26(4), 1614–1635.

Billio, M., and Caporin, M. (2005), Multivariate Markov switching dynamic conditional correlation GARCH representations for contagion analysis. *Statistical methods and applications* 14(2), 145-161.

Bouchaud, J. P., and Potters, M. (2001), More stylized facts of financial markets: leverage effect and downside correlations. *Physica A: Statistical Mechanics and its Applications* 299(1), 60-70.

Brémaud, P. (1999), *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, vol. 31. Springer Science & Business Media.

Bulla, J., and Bulla, I. (2006), Stylized facts of financial time series and hidden semi-Markov Models. *Computational Statistics & Data Analysis* 51(4), 2192-2209.

Chan, W. S. (1995), Time series outliers and spurious autocorrelations. *Journal of Applied Statistical. Science* 2(2), 153-162.

Cho, J. S., White, H. (2007), Testing for regime switching. *Econometrica* 75(6), 1671-720.

Croux, C., Forni, M., and Reichlin, L. (2001), Measure of comovemnet for economic variables: theory and empirics. *The Review of Economics and Statistics* 83(2), 232-241.

Demirgüç-Kunt, A., and Levine, R. (1996), Stock market development and financial intermediaries: Stylized facts. *The World Bank Economic Review* 10(2), 291-321.

den Haan J. W., and Levin, T. A. (1998), Vector autoregressive covariance matrix estimation. Manuscript, University of California, San Diego.

den Haan J. W., and Levin, T. A. (2000), Robust covariance matrix estimation with data-dependent VAR prewhitening order. *NBER Techinical Working Paper* 255.

Diebold, F. X., and Inoue, A. (2001), Long memory and regime switching. *Journal of Econometrics* 105(1), 131-159.

Douc, R., Moulines, E., and Rydén, T. (2004), Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *The Annals of Statistics* 32(5), 2254-304.

Durland, J. M., and McCurdy, T. H. (1994), Duration-dependent transitions in a Markov model of US GNP growth. *Journal of Business & Economic Statistics* 12(3), 279-288.

Edwards, S., and Susmel, R. (2003), Interest-rate volatility in emerging markets. *Review of Economics and Statistics* 85(2), 325–348.

Fang, H. B., Fang, K. T., and Kotz, S. (2002), The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis* 82(1), 1-16.

Ferguson, J. D. (1980), Variable Duration Models for Speech. In *Proceedings of the Symposium on the Applications of Hidden Markov Models to Text and Speech*, Ferguson J. D. (eds). Princeton: New Jersey, 143–179.

Gray, S. (1996), Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42(1), 27-62.

Guedon, Y. (2003), Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics* 12(3), 604-639.

Guidolin, M., and Timmermann, A. (2007), Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control* 31(11), 3503-3544.

Hamilton, J., Sumsel, R. (1994), Autoregressive conditional heteroskedasticity and change in regime. *Journal of Econometrics* 64(1), 307-333.

Han, F., and Liu, H. (2014), Scale-Invariant Sparse PCA on High Dimensional Meta-elliptical Data. *Journal of the American Statistical Association* 109(505), 275-287.

Hansen, B. E. (1992), The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics* 7(1), 561-582.

Hong, Y., Tu, J., and Zhou G. (2007), Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* 20(5), 1547-81.

Jensen, J. L., and Petersen, N. V. (1999), Asymptotic normality of the maximum likelihood estimator in state space models. *The Annals of Statistics* 27(2), 514–535.

Karolyi, G. A., and Stulz, R. M. (1996), Why do markets move together? an investigation of US-Japan stock return comovements. *The Journal of Finance* 51(3), 951-986.

Kruskal, W. H. (1958), Ordinal measures of association. *Journal of the American Statistical Association* 53(284), 814-861.

Laloux, L., Cizeau, P., Potters, M., and Bouchaud, J. P. (2000), Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance* 3(3), 391-397.

Lane, P. R., and Milesi-Ferretti, G. M. (2003), International financial integration. *IMF Staff Papers* 50, 82-113.

Li, F. (2014), Identifying asymmetric comovements of international stock market returns. *Journal of Financial Econometrics* 12(3), 507-543.

Longin, F., and Solnik, B. (1995), Is the correlation in international equity returns constant: 1960-1990?. *Journal of International Money and Finance* 14(1), 3-26.

MacDonald, I. L., and Zucchini, W. (1997), Hidden Markov and Other Models for Discrete-valued Time Series. Chapman & Hall: London.

Maheu, J. M., and McCurdy, T. H. (2000), Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics* 18(1), 100-112.

Newey, W. K., and West, K. D. (1994), Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* 61(4), 631-653.

O'Connell J., Tøgersen. F.A., Friggens N.C., Løvendahl P., and Højsgaard S. (2011), Combining Cattle Activity and Progesterone Measurements Using Hidden Semi-Markov Models. *Journal of Agricultural, Biological and Ecological Statistics* 16(1), 1-16.

Patton, A. J. (2006), Modeling asymmetric exchange rate dependence. *International Economic Review* 47(2), 527-556.

Peel, D., McLachlan, G. J. (2000), Robust mixture modelling using the t distribution. *Statistics and computing* 10(4), 339-348.

Pelletier, D. (2006), Regime switching for dynamic correlations. *Journal of Econometrics* 131(1), 445-473.

Preis, T., Kenett, D. Y., Stanley, H. E., Helbing, D., and Ben-Jacob, E. (2012), Quantifying the behavior of stock correlations under market stress. *Sientific Reports* 2(752), 1-5.

Ramchand, L., and Susmel, R. (1998), Volatility and cross correlation across major stock markets. *Journal of Empirical Finance*, 5(4), 397-416.

Ribeiro, R., and Veronesi, P. (2002), Excess comovement of international stock markets in bad times: a rational expectations equilibrium model. Manuscript, University of Chicago, Chicago.

Reigneron, P.A., Allez, R., and Bouchaud, J. P. (2011), Principal regression analysis and the index leverage effect. *Physica A* 390(17), 3026-35.

Rua, A., 2010. Measuring comovement in the time-frequency space. *Journal of Macroeconomics* 32(2), 685-91.

Rydén, T., Teräsvirta, T., and Asbrink, S. (1998), Stylized facts of daily return series and the hidden markov model. *Journal of Applied Econometrics* 13(3), 217-244.

Simsek, A. (2011), Speculation and risk sharing with new financial assets. *NBER working paper* 17506.

Smith, A., Naik, P. A., and Tsai, C. L., 2005. Markov-switching model selection using Kullback-Leibler divergence. *Journal of Econometrics* 134(2), 553-77.

Stărică, C. (1999), Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance* 6(5), 515-553.

Syllignakis, M. N., and Kouretas, G. P. (2011), Dynamic correlation analysis of financial contagion: Evidence from the Central and Eastern European markets. *International Review of Economics & Finance* 20(4), 717-732.

Timmermann, A., 2000. Moments of Markov switching models. *Journal of Econometrics* 96(1), 75-111.

VITA

Before starting graduate school, I received my B.S. in economics and B.S. in physics from Peking University. Outside the academia, I'm a chartered Sino-English conference interpreter and hold a certificate in bamboo flute and painting. Back in college, I was a runner and finished ANA Beijing International Half Marathon in 2005. While classics and humanities have always been the best of my companions, I repair mechanical watches and enjoy a variety of sports.