M.Sc. Thesis
Master of Science in Engineering

DTU Compute
Department of Applied Mathematics and Computer Science

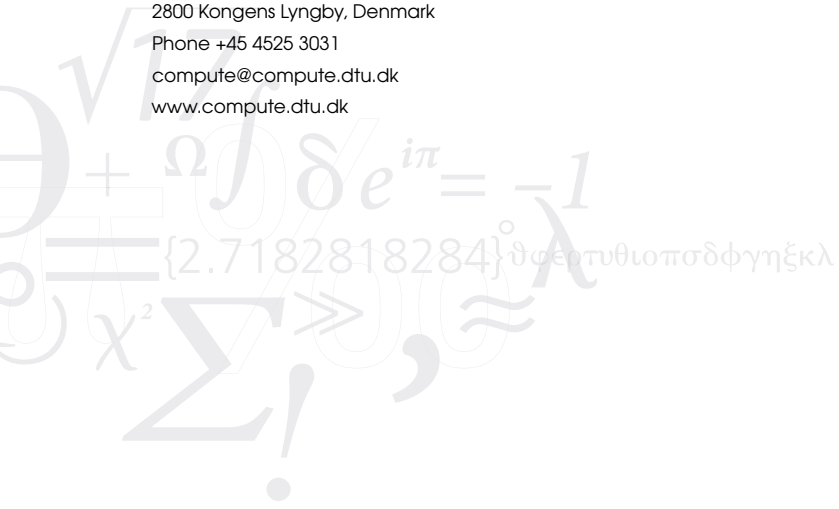
Regime-Based Asset Allocation

Do Profitable Strategies Exist?

Peter Nystrup

Kongens Lyngby
July 2014

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Matematiktorvet, Building 303B
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk



Short Contents

Abstract

Regime shifts present a big challenge to traditional strategic asset allocation, demanding a more adaptive approach. In the presence of time-varying investment opportunities, portfolio weights should be adjusted as new information arrives. Regime-switching models can match the tendency of financial markets to change their behavior abruptly and the phenomenon that the new behavior often persists for several periods after a change. They are well suited to capture the stylized behavior of many financial series including skewness, leptokurtosis, volatility persistence, and time-varying correlations.

This thesis builds on this empirical evidence to develop a quantitative framework for regime-based asset allocation. It investigates whether regime-based investing can effectively respond to changes in financial regimes at the portfolio level in an effort to provide better long-term results when compared to more static approaches. The thesis extends previous work by considering both discrete-time and continuous-time models, models with different numbers of states, different univariate and multivariate state-dependent distributions, and different sojourn time distributions. Out-of-sample success depends on developing a way to model the non-linear and non-stationary behavior of asset returns.

Dynamic asset allocation strategies are shown to add value over strategies based on rebalancing to static weights with rebalancing in itself adding value compared to buy-and-hold strategies in an asset universe consisting of a global stock index, a global government bond index, and a commodity index. The tested strategies based on an adaptively estimated two-state Gaussian hidden Markov model outperform a rebalancing strategy out of sample after accounting for transaction costs, assuming no knowledge of future returns, and with a realistic delay between the identification of a regime change and the portfolio adjustment.

Keywords: Regime switching; Markov-switching mixtures; Non-linear Time Series Modeling; Daily Returns; Adaptivity; Leptokurtic distributions; Volatility clustering; Dynamic asset allocation.

Resumé

Regimeskift udgør en stor udfordring for traditionel strategisk aktivallokering, da de sætter krav til en mere adaptiv tilgang. Ved tilstedeværelsen af tidsvarierende investeringsmuligheder bør porteføljevægte opdateres, efterhånden som ny information kommer til. Regimeskiftsmodeller kan matche finansielle markeders tendens til pludseligt at skifte opførsel og det fænomen, at den nye opførsel ofte varer ved længe efter et skift. De er velegnede til at fange den stiliserede opførsel, der er kendetegnende for mange finansielle serier, herunder skævhed, leptokurtosis, volatilitetsklumpning og tidsvarierende korrelationer.

Formålet med denne afhandling er at udvikle en kvantitativ ramme for regimebaseret aktivallokering. Det undersøges, hvorvidt regimebaseret investering kan reagere på effektiv vis på ændringer i finansielle regimer på porteføljeniveau med det formål at skabe bedre langsigtede resultater sammenlignet med mere statiske tilgange. Afhandlingen udvider tidligere studier ved at inkludere modeller i både diskret og kontinuert tid, modeller med forskellige antal regimer, forskellige univariate og multivariate regimebetingede fordelinger og forskellige opholdstidsfordelinger. Out-of-sample succes afhænger af udviklingen af en model, der beskriver finansielle afkasts ikke-lineære og ikke-stationære opførsel.

Det bliver vist, at dynamiske aktivallokeringsstrategier tilfører værdi sammenlignet med strategier baseret på rebalancering til statiske vægte. Rebalancering i sig selv tilfører værdi sammenlignet med køb-og-hold strategier i et aktivunivers, der består af et globalt aktieindeks, et globalt statsobligationsindeks og et råvareindeks. De testede strategier, baseret på en adaptivt estimeret gaussisk skjult Markov model med to regimer, outperformer en rebalanceringsstrategi out-of-sample efter handelsomkostninger, uden kendskab til fremtidige afkast og med en realistisk forsinkelse mellem identifikationen af et regimeskift og implementeringen af portføljetilpasningen.

Nøgleord: Regimeskift; Markov-skiftende miksturer; Ikke-lineær tidsrækkemodellering; Daglige afkst; Adaptivitet; Leptokurtiske fordelinger; Volatilitetsklumpning; Dynamisk aktivallokering.

Preface

This thesis was prepared at the Department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU) in partial fulfillment of the requirements for acquiring the M.Sc. degree in engineering with honors.

The thesis deals with different aspects of mathematical modeling of the stylized behavior of financial returns using regime-switching models with the aim of developing a quantitative framework for regime-based asset allocation. The developed strategies are tested out of sample under realistic assumptions to ensure that the conclusions have practical relevance.

The process that led to this thesis began in the spring of 2013 when I attended a class on hidden Markov models at DTU. In the summer of 2013, I began a special course on regime modeling of financial data with Henrik Madsen at DTU. Erik Lindström later joined the special course that led to the preparation of an article (Nystrup et al. 2014). I got the idea for the thesis when I read the article on Markov-switching asset allocation by Bulla et al. (2011). In continuation of the work on the article there was an initial focus on the estimation of continuous-time models, but the strategies ended up being based on adaptively estimated discrete-time hidden Markov models.

Copenhagen, July 2014

Peter Nystrup

Acknowledgements

I want to thank my supervisors Professor Henrik Madsen (Technical University of Denmark) and Associate Professor Erik Lindström (Lund University, Sweden) for the many interesting discussions. Henrik's calendar is always fully booked, but he spared time to share his vast experience with mathematical modeling and supervise the project anyway. Fortunately, Erik was visiting DTU on a regular basis in the fall of 2013 and the beginning of 2014, so he was able to co-supervise the project. Erik always made time to answer questions and to research for new ideas that could contribute to the project.

I also want to thank Bo William Hansen, Henrik Olejasz Larsen, and the rest of my colleagues at Sampension for their contributions to the project. Bo and Henrik showed a lot of interest in the project right from the beginning and encouraged me to work on the topic. We have had many interesting discussions and I have drawn on their background in economics and practical experience with asset allocation.

MSCI Inc., J.P. Morgan, and Standard & Poor's should be mentioned as the sources of the index data that was accessed through Bloomberg.

Acronyms

ACF	Autocorrelation function
AIC	Akaike information criterion
BIC	Bayesian information criterion
CAPM	Capital Asset Pricing Model
CDLL	Complete-data log-likelihood
CPI	Consumer price index
CTHMM	Continuous-time hidden Markov model
CVaR	Conditional Value-at-Risk
DAA	Dynamic asset allocation
DM	Developed market
EM	Emerging market
EM	Expectation–maximization
ER	Expected return
ETF	Exchange-traded fund
GDP	Gross domestic product
GFC	Global finacial crisis
GLRT	Generalized likelihood ratio test
HMM	Hidden Markov model
HSMM	Hidden semi-Markov model

JPM GBI	J.P. Morgan Government Bond Index
MCMC	Markov chain Monte Carlo
MDD	Maximum drawdown
ML	Maximum likelihood
MSCI ACWI	Morgan Stanley Capital International All Country World Index
RR	Realized return
S&P GSCI	Standard & Poor's Goldman Sachs Commodity Index
SAA	Strategic asset allocation
SD	Standard deviation
SR	Sharpe ratio
TAA	Tactical asset allocation
VaR	Value-at-Risk
VIX	Chicago Board Options Exchange Market Volatility Index

Contents

CHAPTER 1

Introduction

The behavior of financial markets changes abruptly. While some changes may be transitory, the changed behavior often persists for many periods. The mean, volatility, and correlation patterns in stock returns, for example, changed dramatically at the start of, and continued through the global financial crisis (GFC) of 2007–2008. Similar regime changes, some of which can be recurring (recessions versus expansions) and some of which can be permanent (structural breaks), are prevalent across a wide range of financial markets and in the behavior of many macro variables (Ang and Timmermann 2011).

The observed regimes in financial markets are closely related to the phases of the business cycle (see e.g. Campbell 1998, Cochrane 2005). In the long run, the gross domestic product (GDP) tends to follow a trend path (growth). In the short to medium term, the GDP fluctuates around the long-term trend. The recurring pattern of recession and expansion is called the business cycle, although the length and severity of the cycles are irregular. The duration of a full cycle can be anything from one year to ten or twelve years (Burns and Mitchell 1946).

Regime changes present a big challenge to traditional strategic asset allocation (SAA), demanding a more adaptive approach. Moreover, the time-varying behavior of risk premiums, volatilities, and correlations have important implications for risk management. Understanding the correlations and their evolution through time is a prerequisite to evaluate the undiversifiable risk associated with the financial cycle.

This thesis examines whether regime-based asset allocation can effectively respond to financial regimes in an effort to provide better long-term results when

compared to static approaches. The first chapter introduces the concept of dynamic asset allocation (DAA), the importance of the time horizon, the use of Markov-switching mixture models to capture the time-varying behavior of financial returns, and their applicability to regime-based asset allocation.

1.1 Dynamic Asset Allocation

Asset allocation is the decision of how to divide a portfolio among the different major asset classes. It is the most important determinant of portfolio performance (see e.g. Brinson et al. 1986, Ibbotson and Kaplan 2000). Although asset class behavior can vary significantly over shifting economic scenarios—no single asset class dominates under all economic conditions—traditional SAA approaches make no effort to adapt to such shifts. Traditional approaches instead seek to develop static "all-weather" portfolios that optimize efficiency across a range of economic scenarios (Sheikh and Sun 2012). If economic conditions are persistent and strongly linked to asset class performance, then a DAA strategy should add value over static weights. The purpose of a dynamic strategy is to take advantage of positive economic regimes, as well as withstand adverse economic regimes and reduce potential drawdowns.

Regime-based investing is distinct from tactical asset allocation (TAA). While the latter is shorter term, higher frequency (i.e., weekly or monthly), and driven primarily by valuation considerations, regime-based investing targets a longer time horizon (i.e., a year or more) and is driven by changing economic fundamentals. A regime-based approach has the flexibility to adapt to changing economic conditions within a benchmark-based investment policy which can involve more than one rebalancing within a year. It straddles a middle ground between strategic and tactical (Sheikh and Sun 2012).

Strategic asset allocation is long-term in nature and based on long-term views of asset class performance although it is not unusual to update the forecasts every year and, hence, rebalance annually (Dahlquist and Harvey 2001). Strategic investors, such as pension plans, that invest with a long time horizon typically face constraints on the size of possible deviations from their benchmark allocation. Although the possible tilts might be small, strategic investors can still benefit from reacting to significant regime shifts (see e.g. Kritzman et al. 2012).

DAA is more restricted than SAA in terms of the size of the investment opportunity set as it is difficult to invest dynamically in direct real estate, forestry, private equity etc. Investments in illiquid asset classes require a long time horizon. This is worth mentioning given that unlisted asset classes have gradually become a larger part of professional investors' portfolios in recent years (Edwards and Manjrekar 2014).

The goal of DAA is not to predict the next regime shift or future market movements. The intention is to identify as fast as possible with as high credibility as

possible when a regime shift has occurred and then benefit from the persistence in equilibrium returns and volatilities. As a market participant, it is difficult to see through the daily noise and short-term volatility, therefore a mathematical model that captures the underlying structures in the market data can be useful.

A dynamic strategy can be profitable even if markets are efficient. A capital market is said to be efficient if prices in the market fully reflect available information. When this condition is satisfied, market participants cannot earn a riskless profit on the basis of available information, i.e. there are no arbitrage opportunities (Fama 1970). The term "fully reflect" implies the existence of an equilibrium model which might be stated either in terms of equilibrium prices or equilibrium expected returns. Trends can be present in efficient markets if the equilibrium expected return changes over time.

The existence of a business cycle where the expected rate of return on capital changes over time is one example. The business cycle could result from any set of factors, but as long as the business cycle is not a deterministic phenomenon, asset prices need not follow a random walk with a constant or deterministic trend (Levich 2001). In some periods, no significant events will take place to cause prices to change, thus returns will essentially reflect noise. In other periods, several important events will influence returns. A focus on identifying rather than predicting regime changes is indeed consistent with a belief in efficient markets.

1.2 Time Horizon

It is a common belief that time diversification reduces risk and therefore long-term investors should have a higher proportion of risky assets in their portfolio than short-term investors (see e.g. Siegel 2007, Guidolin and Timmermann 2007). With a regime-based approach it does not matter whether the horizon is one, two, or ten years because the aim is to rebalance whenever a regime shift has occurred. The horizon, however, is assumed to be long enough that it makes sense to defray the cost of rebalancing. The portfolio is optimized conditional on being in the current regime and asset returns are assumed to follow a random walk within each regime. The proportion of risky asset will depend on the current regime and the level of risk aversion, but not the time horizon.

Oftentimes the frequency of the analyzed data is chosen to reflect the length of the investment horizon, meaning monthly or annual data for long-term investment decisions. Bulla et al. (2011) were among the first to consider regime-switching asset allocation based on daily returns. By using daily rather than monthly data the impact of wrong regime forecasts reduces from an entire month to a single trading day and the delay in identifying and addressing regime shifts is shorter. With the shorter response time it is feasible to wait until the same

regime has been decoded for a number of days before changing the allocation in order to minimize the costs of trading on spurious signals.

Bulla et al. (2011) further argued that the use of daily data increases the amount of data available for markets with a short history, but for many financial indices monthly data is available with a much longer history than daily data. The optimal length of the data period is debatable, but it should span at least the time needed for a financial cycle to unfold. It would, presumably, be favorable if the period included more than one cycle. On the other hand, the longer the data horizon the more questionable it is whether stationarity of the data-generating process can be assumed.

The use of daily data presents some challenges. Daily returns contain a lot of noise and extreme observations that are evened out on a monthly basis. For example, if returns are conditionally normal, conditional on a variance parameter which is itself random, then the Central Limit Theorem applies to the resulting heavy-tailed unconditional distribution as it has a finite variance and finite higher order moments. Consequently, long-horizon returns will tend to be closer to the normal distribution than short-horizon returns (Campbell et al. 1997). It complicates the modeling significantly and is likely part of the reason why it is more popular to consider monthly data. In addition, the use of daily data makes the link to macroeconomic data more difficult. In spite of the complications, the arguments for considering daily rather than monthly data are compelling.

1.3 Stylized Facts and Markov-Switching Mixtures

The normal distribution is a poor fit to most financial returns. Mixtures of normal distributions provide a better fit as they are able to reproduce both the skewness and leptokurtosis often observed (see e.g. Cont 2001). An extension to a Markov switching mixture model, also referred to as a hidden Markov model (HMM), is often used to capture both the distributional and temporal properties of financial returns. The class of Markov-switching models was introduced to financial econometrics by Hamilton (1989, 1990). In an HMM, the distribution that generates an observation depends on the state of an underlying and unobserved Markov chain. Although the states are identified through a statistical filtering procedure, they can often be interpreted in terms of economic cycles (see e.g. Ahlgren and Dahl 2010).

Rydén et al. (1998) showed the ability of an HMM to reproduce most of the stylized facts of daily return series introduced by Granger and Ding (1995a,b) using daily returns of the S&P 500 stock index from 1928 to 1991. The one stylized fact that they could not reproduce was the persistence of the squared daily returns (i.e., volatility).

According to Bulla and Bulla (2006), the lack of flexibility of an HMM to model this temporal higher order dependence can be explained by the implicit assump-

tion of geometrically distributed sojourn times in the hidden states. Silvestrov and Stenberg (2004), among others, argued that the memoryless property of the geometric distribution is inadequate from an empirical perspective, although it is consistent with the no-arbitrage principle.

Bulla and Bulla (2006) considered hidden semi-Markov models (HSMMs) in which the sojourn time distribution is modeled explicitly for each hidden state so that the Markov property is transferred to the imbedded first-order Markov chain. They showed that HSMMs with negative binomial sojourn time distributions reproduced most of the stylized facts comparably well, and often better, than the HMM for eighteen series of daily stock market sector returns from 1987 to 2005. Specifically, they found HSMMs to reproduce the long-memory property of squared daily returns much better than the HMM. They did not, however, consider the complicated problem of selecting the most appropriate sojourn time distributions and they only considered models with two hidden states.

Bulla (2011) later showed that HMMs with t-distributed components reproduce most of the stylized facts as well or better than the Gaussian HMM, while at the same time increasing the persistence of the visited states and the robustness to outliers. Bulla (2011) also found that models with three states provided a better fit than models with two states to daily returns of the S&P 500 index from 1928 to 2007.

The fact that increasing the number of states leads to a much better fit to the empirical moments and the persistence of squared daily returns was confirmed in Nystrup et al. (2014). The quadratic increase in the number of parameters with the number of states is a major limitation of discrete-time HMMs and HSMMs. In Nystrup et al. (2014) it was shown how a continuous-time formulation leads to a linear rather than quadratic growth in the number of parameters. A continuous-time hidden Markov model (CTHMM) with four states was found to provide a better fit to daily returns of the S&P 500 index than the discrete-time models with three states with a similar number of parameters. There was no indication that the memoryless property of the sojourn time distribution was inconsistent with the long-memory property of the squared daily returns.

Kritzman and Li (2010) used a different approach based on discriminant analysis to show that financial turbulence has been highly persistent and risk-adjusted returns have been substantially lower during turbulent periods, irrespective of the source of turbulence. Their study included monthly returns from 1980 to 2009 of six asset-class indices including US and non-US stocks and bonds, commodities, and listed real estate. Following Chow et al. (1999), they defined financial market turbulence as a condition in which asset prices behave in an uncharacteristic fashion given their historical pattern of behavior. Turbulent periods were not necessarily characterized only by low or negative returns. They applied the squared Mahalanobis distance (Mahalanobis 1936) to identify tur-

bulent periods as being returns outside the 75-percent quantile of a multivariate normal distribution. The Mahalanobis distance takes into account the correlation convergence that is part of turbulent periods.

The same measure could be used for a cluster analysis rather than just discriminating between turbulent or nonturbulent periods. The Mahalanobis distance does, however, not take the serial correlation (autocorrelation) into account. Given the observed persistence, it should be beneficial to consider the serial correlation in the classification. An HMM takes the autocorrelation into account and with a multivariate conditional distribution it is possible to include the correlation between the assets in the classification. A Markov-switching mixture model that captures the stylized behavior of financial returns should be useful as foundation for a regime-based asset allocation strategy.

1.4 Regime-Switching Asset Allocation

Ang and Bekaert (2002) were among the first to consider the impact of regime shifts on asset allocation. They modeled monthly equity returns from Germany, the UK, and the US from the period 1970 to 1997 as a multivariate regime-switching process with two states. The costs of ignoring regime switching were small for all-equity portfolios, but much higher when a risk-free asset could be held. Their main finding was that international diversification was still valuable in the presence of regime changes despite the increasing correlations and volatilities in bear markets.

In a subsequent study, Ang and Bekaert (2004) extended the analysis by including further equity indices from around the world. Their sample included monthly returns from 1975 through 2000. A regime-switching strategy was found to dominate static strategies out of sample for global equity portfolios. They also considered market timing based on a regime-switching model in which the transition probabilities depended on a short-term interest rate.[1] With an asset universe consisting of a stock index, cash, and a ten-year constant-maturity bond, the main hedge for volatility was found to be the risk-free asset and not the bond investment.

Bauer et al. (2004) studied monthly returns from 1976 to 2002 of a six-asset portfolio consisting of equities, bonds, commodities, and real estate using the multivariate outlier approach of Chow et al. (1999). They observed changing correlations and volatilities among the assets and demonstrated, under the assumption of perfect foresight with regard to the prevailing regime, a significant information gain by using a regime-switching strategy instead of the standard

[1] The interest rate had a statistically significant influence on the transition probabilities as the probability of switching to the high-volatility regime and the probability of staying in the high-volatility regime both increased when the interest rate rose.

mean–variance optimization strategy. After accounting for transaction costs, however, a substantial part of the positive excess return disappeared.

Ammann and Verhofen (2006) estimated a multivariate regime-switching model, similar to that of Ang and Bekaert (2002), for the four-factor model of Carhart (1997) using monthly data for the four equity risk factors from 1927 to 2004. They found two clearly separable regimes with different mean returns, volatilities, and correlations. One of their key findings was that value stocks provided high returns in the high-variance state, whereas momentum stocks and the market portfolio performed better in the low-variance state.

Guidolin and Timmermann (2007) estimated a four-state Markov-switching autoregressive model to monthly returns on stocks, bonds, and T-bills from 1954 to 1991. The optimal asset allocation varied significantly across the regimes. Stock allocations were found to be monotonically increasing as the investment horizon got longer in only one of the four regimes. In the other regimes, a downward sloping allocation to stocks was observed. They confirmed the economic importance of accounting for the presence of regimes in asset returns in out-of-sample forecasting experiments.

Bulla et al. (2011) fitted two-state hidden Markov models to daily returns of stock indices from Germany, Japan, and the US using data from 1985 (1976 for some of the indices) to 2006. A strategy of switching to cash in the high-variance regime led to a significant variance reduction when tested out of sample. In addition, all strategies outperformed their respective index in terms of annual return after accounting for transaction costs.

Kritzman et al. (2012) applied a two-state HMM to forecast regimes in market turbulence (as defined by Chow et al. 1999), inflation, and economic growth. A DAA strategy based on the forecasted regimes was shown to reduce downside risk and improve the ratio of return to Value-at-Risk (VaR) relative to a static strategy out of sample when applied to stocks, bonds, and cash. They considered monthly returns from 1973 to 2009 in the out-of-sample analysis. Rather than making an assumption about transaction costs, the authors reported the break-even transaction cost that would offset the advantage of the dynamic strategy.

Zakamulin (2014) tested two DAA strategies based on unexpected volatility; unexpected volatility being the difference between the forecasted volatility (one month ahead) using a GARCH(1,1) model and the realized volatility. The author referred to previous studies that had focused on implied volatility using the CBOE Market Volatility Index (VIX). The data included daily and monthly returns of the S&P 500 and the Dow Jones Industrial Average index from 1950 through 2012. Unexpected volatility was shown to be negatively related to expected future returns and positively related to expected future volatility. In the first strategy, the weight of stocks relative to cash was changed gradually on a monthly basis based on the level of unexpected volatility, whereas the second strategy was either all in stocks or all in cash depending on whether the

unexpected volatility was below or above its historical average. Both strategies were found to outperform static strategies out of sample.

It is important to consider transaction costs when comparing the performance of dynamic and static strategies. Frequent rebalancing can offset the potential excess return of a dynamic strategy as described by Bauer et al. (2004). Ang and Bekaert (2002, 2004), Guidolin and Timmermann (2007), and Zakamulin (2014) did not account for transaction costs. Reporting the break-even transaction, as done by Kritzman et al. (2012), is the most meaningful approach as the transaction costs faced by private investors are likely to exceed those of professionals who can implement dynamic strategies in a cost-efficient way using financial derivatives like futures or swaps.

Another issue neglected in many studies is out-of-sample testing. Testing a model on the same data that it was fitted to does not reveal its actual potential. As noted by Bauer et al. (2004), the out-of-sample potential is likely to be lower (than the in-sample performance), as investors do not have perfect foresight. It is not unusual that non-linear techniques provide a good in-sample fit, yet get outperformed by a random walk when used for out-of-sample forecasting. Dacco and Satchell (1999) showed that it only takes a small estimation error to lose any advantage from knowing the correct model specification. Thus, a good in-sample fit but no outperformance over a random walk in terms of mean squared error out of sample does not necessarily imply that a model is overfitting. Dacco and Satchell (1999) argued that the performance should instead be evaluated by methods appropriate for the particular problem, in this case economic profit or excess return. An example is Ammann and Verhofen (2006), who found a regime-switching strategy to be profitable out of sample although the forecasting ability of the underlying model was weak compared to a random walk.

A poor out-of-sample performance can also be an indication that the data-generating process is non-stationary. Rydén et al. (1998) found that the parameters of the estimated HMMs varied considerably through the 63-year data period they studied. This can be addressed by applying an adaptive estimation technique that allows the parameters of the model to be gradually changing through the sample period by assigning more weight to the most recent observations. This is increasingly important the longer the data period is. Adaptivity is often used within other areas for automatic regulation (see e.g. Krstic et al. 1995) and modeling and forecasting (see e.g. Pinson and Madsen 2012), but it has not received the same attention within empirical finance.

Of the referenced studies, Kritzman et al. (2012) were the only ones that did not identify regimes in asset prices, but instead forecasted regimes in important drivers of asset returns and then reallocated assets accordingly. If financial markets are efficient, the outlook for the economy should be reflected in asset prices to the extent that it can be predicted. The use of macroeconomic data in this connection is troublesome due to the delay in availability, the low frequency

of the data, and the fact that the data is often revised subsequently. The arguments for modeling financial returns directly are strong as the financial markets should be the first to react to changes to the economic outlook with market data being available real time. Interestingly, only Bulla et al. (2011) considered daily rather than monthly returns.

The majority of the studies included a risk-free asset. The holding of a risk-free asset, of course, yields a volatility reduction, but it raises the question of what return to expect on a risk-free asset. A lower return on cash would dilute the performance of the dynamic switching strategies. Furthermore, if stocks and other risky assets underperform in turbulent periods, then it is worth considering a negative exposure to these assets in the most turbulent regimes. The referenced studies, with the exception of Ang and Bekaert (2002, 2004) and Ammann and Verhofen (2006), considered long-only strategies. In practice, there can be restrictions on short positions that prevent the implementation of other strategies, but it is interesting to establish the potential if not only to know the cost of a long-only investment policy.

The list of studies referenced is not exhaustive, yet it includes various interesting approaches to regime-switching asset allocation. Surprisingly, none of the studies considered model selection in depth. Guidolin (2011) found in his review of the literature on applications of Markov-switching models in empirical finance that roughly half the studies selected a Markov-switching model based on economic motivations rather than statistical reasoning. In addition, half the studies did not consider it a possibility that the number of states could exceed two and there was an overweight of studies based on Gaussian mixtures in which the underlying Markov chain was assumed to be time-homogeneous. A study is missing that combines the economic intuition and application with a statistical analysis of the model class, the number of states, the type of marginal distributions as well as the need for time-heterogeneity/adaptivity.

1.5 Thesis Statement

The purpose of this thesis is to compare the performance of a regime-based asset allocation strategy under realistic assumptions to a strategy based on rebalancing to static weights. It will be examined whether the volatility reduction found in previous studies on dynamic asset allocation can be achieved when there is no risk-free asset, but rather the possibility for diversification by holding a portfolio of assets which may include short positions.

As asset allocation is the most important determinant of portfolio performance it is clearly relevant whether a dynamic strategy can outperform a static strategy by taking advantage of favorable economic regimes and reducing potential drawdowns. The relevance is supported by the large amount of articles written on the subject. A recent survey by Mercer (Edwards and Manjrekar 2014) found

that professional investors are increasingly looking to incorporate some element of dynamic decision-making within portfolios, both for return enhancement and as a risk management tool.

The asset classes considered are limited to stocks, bonds, and commodities to keep it simple, yet complex enough for diversification possibilities to arise. The data includes a global equity index, a global government bond index, and a commodity index. Daily closing prices covering the 20-year period from 1994 to 2013 are considered. The data prior to 2009 will be used for in-sample analysis and estimation, while the five years from 2009 to 2013 will be used for out-of-sample testing. The 15-year in-sample data period is special in that it includes the build-up and burst of two major financial bubbles: the dot-com bubble around year 2000 and the US housing bubble that triggered the global financial crisis in 2007.

Everything is measured in USD as seen from a US investor's point of view. The stock index is global but denominated in USD, the government bond index is hedged to USD, and the commodity index is traded in USD. In this way there will be no need to consider currency risk.

The analysis will include the following steps:

1. Analysis of the distributional and temporal properties of the index data.
2. Estimation and selection of an appropriate time series model.
3. Evaluation of the performance of an optimized SAA portfolio for different levels of risk aversion with and without rebalancing in and out of sample.
4. Implementation and evaluation of the performance of different dynamic strategies in and out of sample.

The statistical software R (R Core Team 2013) will be used for all data analysis, modeling, and simulation. The approach that will be used is through data analysis to determine the necessary properties of a time series model that is able to describe the observed characteristics of the index data. Model selection will include model class, the number of states, the character of the marginal distributions, and the need for time-heterogeneity. The SAA portfolios will be optimized based on scenarios generated using a regime-switching model.

CHAPTER 2

Index Data

The featured indices are selected with the aim of keeping it simple and replicable, yet complex enough for diversification possibilities to arise. The indices include a global equity index (MSCI ACWI), a global government bond index (JPM GBI) with weight on developed countries, and a commodity index (S&P GSCI) with low correlation to the other two indices. The developed strategies can easily be implemented as similar indices are investable through exchange-traded funds (ETFs).

The 20-year data period goes back almost to the start of the JPM GBI hedged to USD in mid-1993. The total return version of the MSCI ACWI only goes back to 1999, but the data prior to 1999 can easily be reconstructed based on the price index that goes back to 1988. The S&P GSCI started trading in 1991, but reconstructed daily data is available back to 1970.

The optimal length of the data period is debatable. Other global government bond indices have been researched, but none were found to have daily data hedged to USD that goes back further. A 20-year sample is deemed reasonable with 15 years for in-sample estimation and 5 years for out-of-sample testing. It is questionable whether data that goes back much further than 20 years is representative of today's market.

It is emphasized that the purpose is not to accentuate these particular indices. It is possible to include many other indices and to over or underweight different regions or sectors compared to the featured indices. The indices are presented in the next three sections, the distribution of the index data is analyzed in section 2.4, and the temporal properties are considered in section 2.5. An in-sample adjustment of the data is discussed in section 2.6.

2.1 The MSCI ACWI

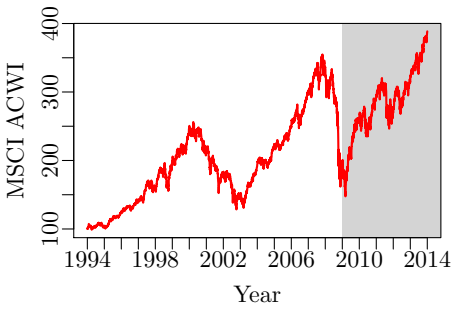
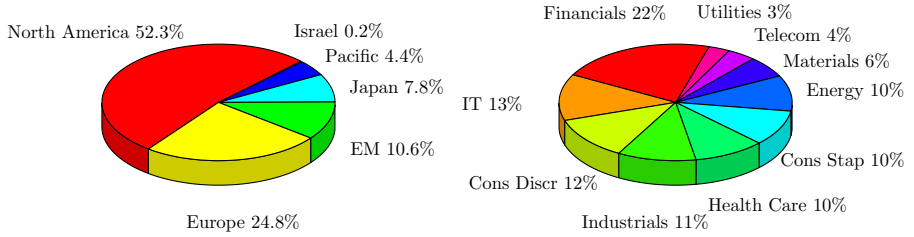


Figure 2.1: The development in the MSCI ACWI in-sample and out-of-sample.

The Morgan Stanley Capital International All Country World Index[2] captures large and mid cap representation across 23 Developed Market (DM) and 21 Emerging Market (EM) countries.[3] The difference compared to the more well-known MSCI World Index is the weight on EM countries. The development in the net total return index, denominated in USD, is depicted in figure 2.1. The data prior to 1999, where the total return index began, has been reconstructed based on the price index[4] by adding the average daily net dividend return over the period from 1999 to 2013 of 0.007% to the price returns.

With 2,434 constituents, the free float-adjusted market capitalization weighted index covers approximately 85% of all global investable equities. The weights across regions and sectors are shown in figure 2.2 as of the end of 2013. Although it is a world index, North America makes up almost half the index. The financial



(a) Regional weights

(b) Sector weights

Figure 2.2: The weight of the different regions and sectors in the MSCI ACWI at the end of 2013.

[2]Bloomberg ticker: NDUEACWF Index.

[3]DM countries include: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, UK, and US. EM countries include: Brazil, Chile, China, Colombia, Czech Republic, Egypt, Greece, Hungary, India, Indonesia, Korea, Malaysia, Mexico, Peru, Philippines, Poland, Russia, South Africa, Taiwan, Thailand, and Turkey.

[4]Bloomberg ticker: MXWD Index.

Table 2.3
In-sample summary statistics for the daily MSCI ACWI log-returns.

Mean	SD	Skewness	Kurtosis	ACF(1)	Annual SR
0.00018	0.0095	-0.40	13	0.15	0.29

sector is by far the largest of the ten sectors in the index. It should be noted that the weights are constantly changing.

The first four central moments are shown in table 2.3 together with the first-order autocorrelation and the annual Sharpe ratio (SR). The SR is the excess return per unit risk, i.e. the excess return divided by the standard deviation (SD). The excess return is the mean return minus the risk-free rate, but the risk-free rate can be neglected, since it is the same for all three indices as they are all denominated in USD. The daily log-returns are left-skewed and highly leptokurtic. The annual SR of 0.29 would have been a lot higher had the in-sample data period been one year shorter or longer.

2.2 The JPM GBI

The global J.P. Morgan Government Bond Index[5] measures performance and quantifies risk across 13 developed fixed income bond markets.[6] The index measures the total return in each market hedged to USD. It includes only traded issues available to international investors, with liquidity guidelines ensuring that there is no price bias related to infrequently traded issues. The constituents are selected from all government bonds, excluding floating rate notes, perpetuals, bonds targeted at the domestic market for tax purposes, and bonds with less than one year to maturity.

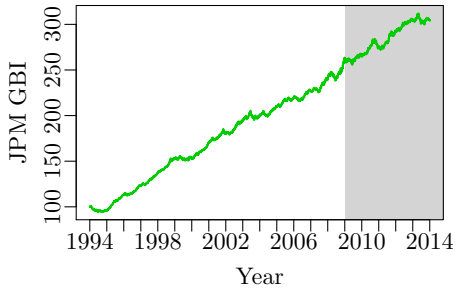
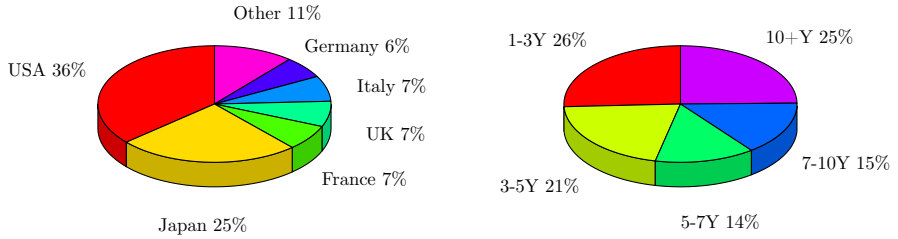


Figure 2.4: The development in the JPM GBI in-sample and out-of-sample.

The weights across countries and maturities are shown in figure 2.5 as of year end 2013. Only countries with a weight above 5% are shown. The U.S. and Japan are by far the heaviest constituents. Danish government bonds receive a tiny weight of 0.56% in the index. The index had a remaining maturity of 8.7

[5]Bloomberg ticker: JHDCGBIG Index.

[6]The 13 countries included have remained unchanged over time and are Australia, Belgium, Canada, Denmark, France, Germany, Italy, Japan, Netherlands, Spain, Sweden, UK, and US.



(a) Country weights (b) Maturity weights

Figure 2.5: The weight of the different countries and maturities in the JPM GBI at the end of 2013. Only countries with a weight above 5% are shown.

years and a modified duration of 6.8 at the end of 2013 resulting from a fairly even weight distribution across maturities.

Compared to the MSCI ACWI the daily log-returns of the JPM GBI are less left-skewed, less leptokurtic, exhibit similar first-order autocorrelation, and have a significantly higher Sharpe ratio due to the much lower standard deviation and higher mean return.

The 20-year data period has been characterized by falling interest rates and low inflation leading to a strong performance for bonds. It is unlikely that the bullish environment for bonds will continue forever, at some point interest rates and inflation are likely to start to increasing.[7] When that happens, an investment in commodities can provide some protection against the impact of inflation on real returns.

Table 2.6
In-sample summary statistics for the daily JPM GBI log-returns.

Mean	SD	Skewness	Kurtosis	ACF(1)	Annual SR
0.00026	0.0018	-0.26	4.7	0.16	2.2

[7]The returns of the JPM GBI exhibit positive autocorrelation at short horizons, but negative autocorrelation at long horizons (the annual first-order autocorrelation is -0.37). The same mean reversion can be found in the rate of inflation (see e.g. Lee and Wu 2001).

2.3 The S&P GSCI

The Standard & Poor's Goldman Sachs Commodity Index[8] quoted in USD is one of the leading measures of general commodity price movements and inflation in the world economy. The total return index has been tradeable since 1991. It contains as many commodities as possible, with rules excluding certain commodities to maintain liquidity and investability in the underlying futures markets. The roll schedule is limited to the most liquid nearby contract months. The total return of the index is significantly different than the return from buying physical commodities.

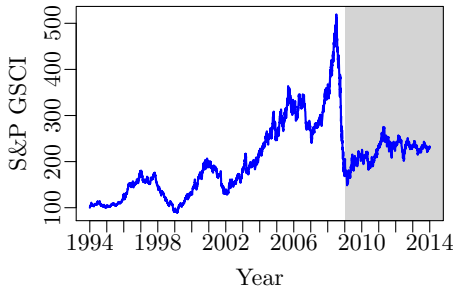


Figure 2.7: The development in the S&P GSCI in-sample and out-of-sample.

At the end of 2013 the index included futures on 24 physical commodities. The index is world production weighted, meaning that the weight assigned to each commodity is in proportion to the amount of that commodity flowing through the economy. As commodity prices are affected by and have an effect on both inflation and real economic growth, commodity exposure is a way of hedging negative shocks to economic growth that might result from higher commodity prices (oil in particular).

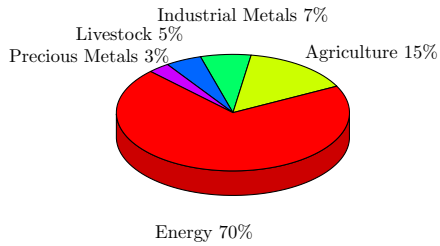


Figure 2.8: The weight of the different sectors in the S&P GSCI in 2013.

Commodities have traditionally been an important component of the basket of goods and services used to measure inflation. Inflation being the annual percentage change in the US Consumer Price Index (CPI) for all urban consumers.[9] Food and energy comprise 25% of the US CPI and they account for an even larger share, about 75%, of its volatility. Since it is unexpected changes in inflation, more than absolute levels, which affect stock and bond prices, this volatility is a major concern for investors (Perrucci and Benaben 2012).

[8]Bloomberg ticker: SPGSCITR Index.
[9]Bloomberg ticker: CPI YOY Index.

Table 2.9
Correlation matrix based on annual returns from 1994 to 2013.

	MSCI ACWI	JPM GBI	S&P GSCI	Gold	ΔInflation
MSCI ACWI	1				
JPM GBI	-0.31	1			
S&P GSCI	0.30	-0.08	1		
Gold	0.00	-0.07	0.20	1	
ΔInflation	0.46	-0.19	0.69	0.24	1

The S&P GSCI has a much higher exposure to energy than the other major commodity indices such as the Dow Jones-UBS Commodity Index. The higher exposure to energy makes the S&P GSCI a better protection against unexpected inflation. Unexpected inflation is notoriously difficult to measure as it requires knowledge of people's average inflation expectations. Erb and Harvey (2006) argued that under the assumption that changes in the rate of inflation are unpredictable, a good proxy for unexpected inflation is the actual change in the rate of inflation. There are certainly some changes that are predictable, but the assumption is very convenient.

Table 2.9 shows the correlations between the three indices, the S&P Gold Index[10], and the annual changes in the rate of inflation. The S&P GSCI has provided some protection against unexpected inflation with a correlation of 0.69 on an annual basis.

A discussion on commodity investing and inflation is not complete without an explicit mention of gold. The empirical evidence in support of gold as an inflation hedge is not very strong with the correlation between the returns of the S&P Gold Index and the changes in inflation being only 0.24. Gold is generally perceived as a defensive asset and there could be diversification benefits from investing in gold as it is almost uncorrelated with both stocks and bonds.

Physical commodities may broadly follow inflationary trends and there appears to be a stronger correlation between commodities and inflation than between traditional asset classes, but commodities are far from perfect or even particularly efficient hedging vehicles for inflation. A far more enticing benefit of commodity investment, even on a passive buy-and-hold basis, is the fact that commodities historically have displayed a low or even negative correlation with traditional investments including stocks and bonds (Perrucci and Benaben 2012).

The in-sample mean return of the S&P GSCI is similar to the mean return of the MSCI ACWI, but the standard deviation is higher which leads to an annual Sharpe ratio of 0.20 compared to 0.29 for the MSCI ACWI. In terms of skewness and kurtosis, the S&P GSCI is closer to the JPM GBI than the MSCI ACWI. Finally, the S&P GSCI is less autocorrelated than the other two indices.

[10]Bloomberg ticker: SPGCGCTR Index.

Table 2.10
In-sample summary statistics for the daily S&P GSCI log-returns.

Mean	SD	Skewness	Kurtosis	ACF(1)	Annual SR
0.00017	0.014	-0.23	6.1	-0.026	0.20

The Commodity Risk Premium

Understanding the size and nature of the commodity risk premium is, as for any other asset class, a necessary precursor to an informed decision about whether to include commodities in a strategic asset allocation. There is, however, strong disagreement about the size of the commodity risk premium, if it at all exists.

The low correlation with stocks and bonds implies that commodities have a low beta (i.e., a low correlation with the market portfolio). Hence, based on the Capital Asset Pricing Model (CAPM) the risk premium is expected to be small. Futures on commodities, however, are not financial claims like stocks and bonds to which the CAPM applies.

Futures on commodities are bets in which risk is transferred from one party to another, but not necessarily increased. This is different from other types of betting, such as betting on sports results, that increase the amount of risk in the world. An investor would expect to be compensated for the transfer of risk undertaken. The insurance premium is the roll return which is the difference in price of the most nearby futures and the most recent futures. The expected roll return is also the expected real return as the price of commodities is expected to follow the general inflation (Hannam and Lejonvarn 2009).

This discussion of the commodity risk premium is not meant to be exhaustive, the purpose is simply to highlight the issue. Based on the data, the characteristics of the S&P GSCI are fairly similar to those of the MSCI ACWI with a slightly lower Sharpe ratio. This indicates that investors have been compensated historically, but past performance is no guarantee of future results.

2.4 Distributional Properties

The in-sample development of the three indices is shown in figure 2.11 together with the pre-tax yield to maturity on generic ten-year on-the-run US government bonds[11]. The MSCI ACWI did well until the outbreak of the GFC at the end of 2007. The index went to about 250 before the burst of the dot-com bubble in 2000. From 2003 until the end of 2007 the index went up from about 125 to 350, but a large part of the gains were lost in 2008 and the index finished below 200 at the end of 2008.

[11] Bloomberg ticker: USGG10YR Index.

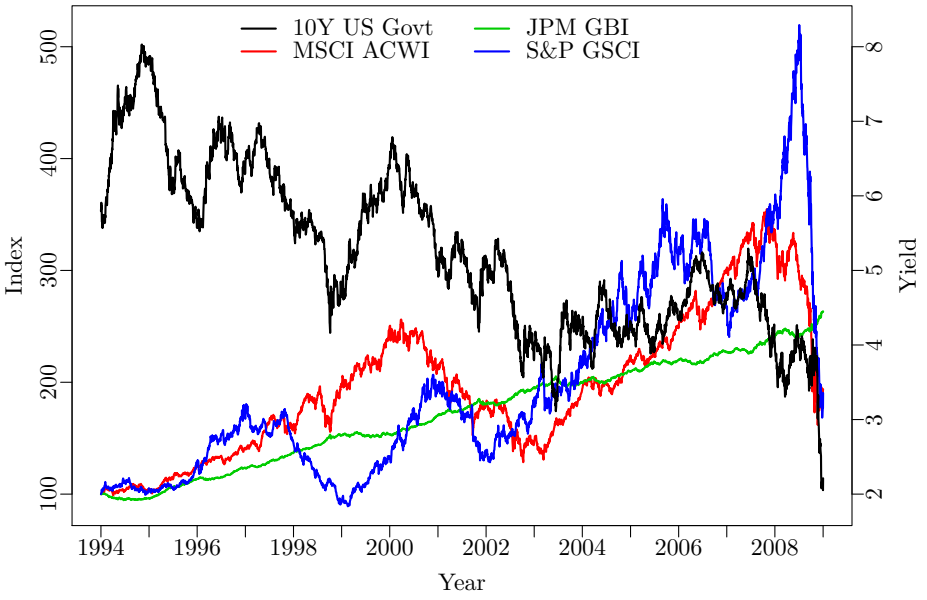


Figure 2.11: The development of the three indices in-sample together with the yield on ten-year US government bonds.

The only points in time where the JPM GBI did not perform well were the first year and the beginning of 1996, where the interest rate increased significantly, and then the index traded horizontally in the end of 1998 and the beginning of 1999 while the interest rate was going up. Overall, the bond index has been moving upward steadily to about 260. The combination of a strong performance and a low volatility naturally leads to the exceptionally high SR of 1.3 that is unlikely to be sustainable in the long run.

The S&P GSCI generally trended in the same direction as the MSCI ACWI, but in some periods the index moved in the opposite direction. For example, in 1998 and 2006, the commodity index incurred large drawdowns while the stock index trended up, or at least did not lose ground. The commodity index experienced large setbacks in both 2001 and 2008 at the same time as the stock index, though the turning points of the two indices are not the exact same. The index was above 500 and much above the stock index before the free fall in 2008 to 175 which was below the stock index.

Commodities have done well in times where there are large interest rate increases in the beginning of the data period, in 1996, and again in 1999, where the bond index underperformed. An investment in commodities appears to be an attractive supplement to investments in bonds.

Table 2.12
In-sample summary statistics and the Jarque–Bera test statistic for the three indices.

	MSCI ACWI	JPM GBI	S&P GSCI
Mean	0.00018	0.00026	0.00017
SD	0.0095	0.0018	0.014
Skewness	-0.40	-0.26	-0.23
Kurtosis	13	4.7	6.1
ACF(1)	0.15	0.16	-0.026
Annual SR	0.29	2.2	0.20
JB-stat	16482	483	1548

Log-Returns

The index series are not stationary as the mean values are growing and there are strong local trends. A transformation is needed to obtain stationarity. A log-transformation narrows the gap between the indices and a difference gets rid of the growing mean values. This leads to the log-return calculated as $r_t = \log(P_t) - \log(P_{t-1})$, where P_t is the closing price of the index on day t and log is the natural logarithm. For returns less than 10%, the log-return is a good approximation to the discrete return, as it is the first order Taylor approximation.[12]

Table 2.12 shows the in-sample summary statistics for the daily log-returns of the three indices together with the first-order autocorrelations and the annual Sharpe ratios. The distributions are left skew and leptokurtic, as already noted. The critical value for the Jarque–Bera test statistic at a 99.9% confidence level is 14.[13] Thus, the Jarque–Bera test strongly rejects the normal distribution for all three indices.

The excess kurtosis relative to the normal distribution is evident from the plot of the kernel density functions in figure 2.13. There is too much mass centered right around the mean and in the tails compared to the normal distribution. The heavy left tail implies that using a normal distribution to model returns underestimates the frequency and magnitude of downside events.

There are 56 observations that deviate more than three standard deviations from the mean for the MSCI ACWI compared to an expectation of 10 if the returns were normally distributed. Out of these, 21 are in the right tail compared to 35 in the left tail. There are 13 observations that are more than five standard

[12] $r_t = \log \frac{P_t}{P_{t-1}} = \log(1 + R_t) = \log(1) + R_t - \frac{R_t^2}{2!} + \frac{R_t^3}{3!} + \mathcal{O}\left(R_t^4\right) \approx R_t$ for discrete returns R_t close to zero.

[13] The Jarque–Bera test statistic is defined as $JB = T\left(\frac{\text{Skewness}^2}{6} + \frac{(\text{Kurtosis}-3)^2}{24}\right)$, where T is the number of observations. If the observations are normally distributed, then the JB test statistic is asymptotically chi-squared distributed with two degrees of freedom.

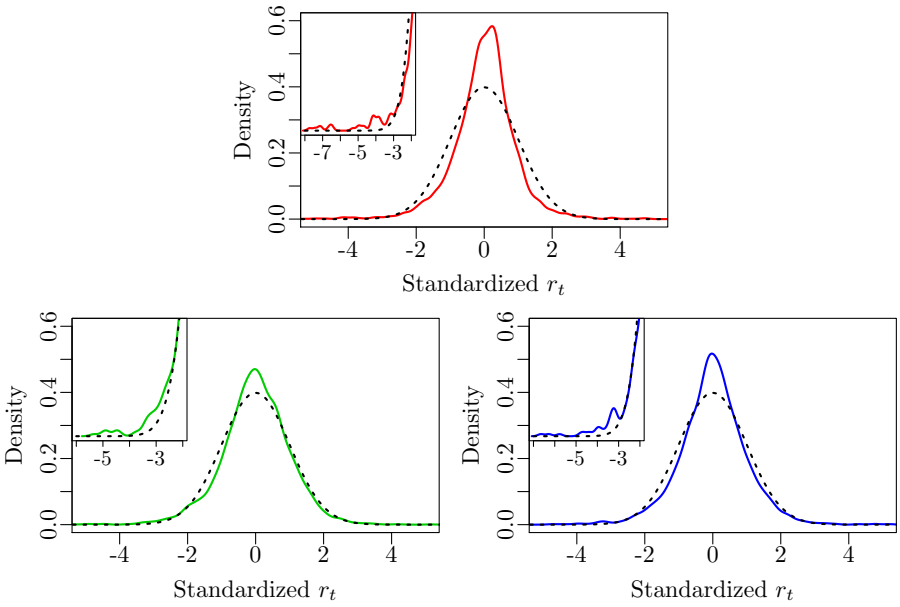


Figure 2.13: Kernel estimates of the densities of the standardized daily log-returns together with the density function for the standard normal distribution.

deviations from the mean—five in the right tail and eight in the left tail—and they are all from the last four months of 2008. The phenomenon, that large drawdowns occur more often than equally large upward movements, is known as gain/loss asymmetry (Cont 2001).

Given the large number of observations ($T_{\text{in-sample}} = 3782$), it takes only a few outliers to reject the normal distribution with a high degree of confidence. It is a general characteristic of financial returns that there are too many extreme values compared to the normal distribution, which results in the very large kurtoses observed.

It is important to distinguish between outliers and extreme observations in this connection; extreme observations deviate considerably from the group mean, but may still represent meaningful conditions, and thus should not be disregarded. Extreme observations should be included in the model estimation in order for the scenarios to be as realistic as possible. After all, the scenarios should reflect the possibility of such extreme events, since they are seen to occur.

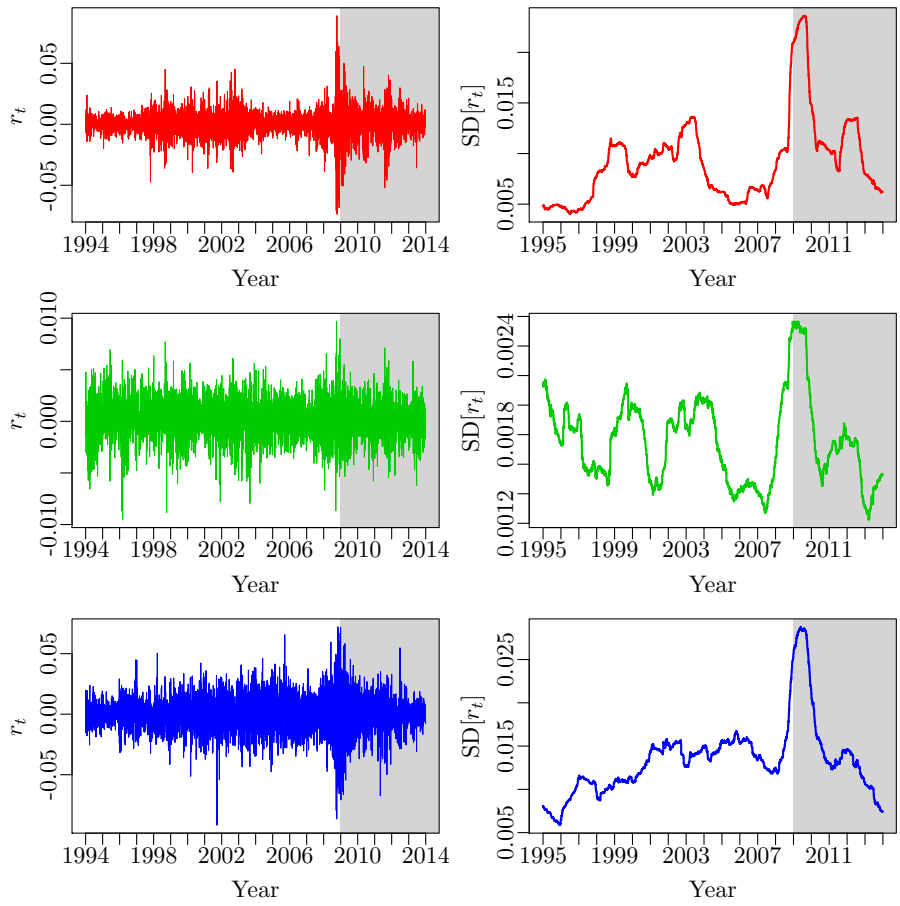


Figure 2.14: The log-returns and their standard deviation estimated using a rolling window of 252 trading days.

2.5 Temporal Properties

The log-returns shown in figure 2.14 are seen to be mean stationary, since they fluctuate around a constant mean level close to zero for all three indices. The log-returns are seen to be much more volatile in some periods than others. This effect, that large price movements tend to be followed by large price movements and vice versa, is referred to as volatility clustering.

The volatility of the commodity index appears to have been increasing through the in-sample period. This is not the case for the JPM GBI and only true to a limited degree for the MSCI ACWI.

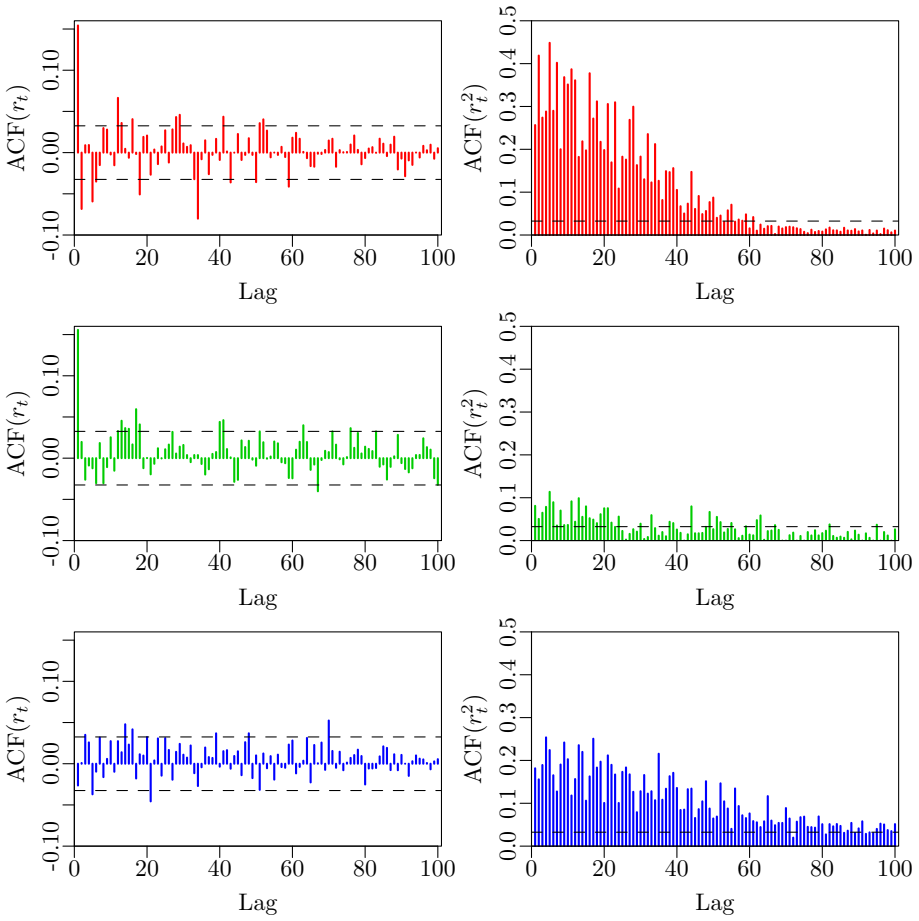


Figure 2.15: The empirical autocorrelation functions for the log-returns and the squared log-returns of the three indices.

The autocorrelation functions (ACFs) for the log-returns and the squared log-returns are shown in figure 2.15. The dashed lines make up approximate 95% confidence intervals under the null hypothesis of independency.[14] The first-order autocorrelation is significant for both the MSCI ACWI and the JPM GBI.

The ACFs for the squared log-returns show significant autocorrelation that persists all the way up to lag 60 for the MSCI ACWI and even further for the S&P

[14]The autocorrelation function for a white noise process is asymptotically normally distributed with mean value zero and variance $1/T$ at all other lags than zero, when T is the number of observations. An approximate 95% confidence interval for the null hypothesis of independency is therefore $\pm 2/\sqrt{T}$ (Madsen 2008).

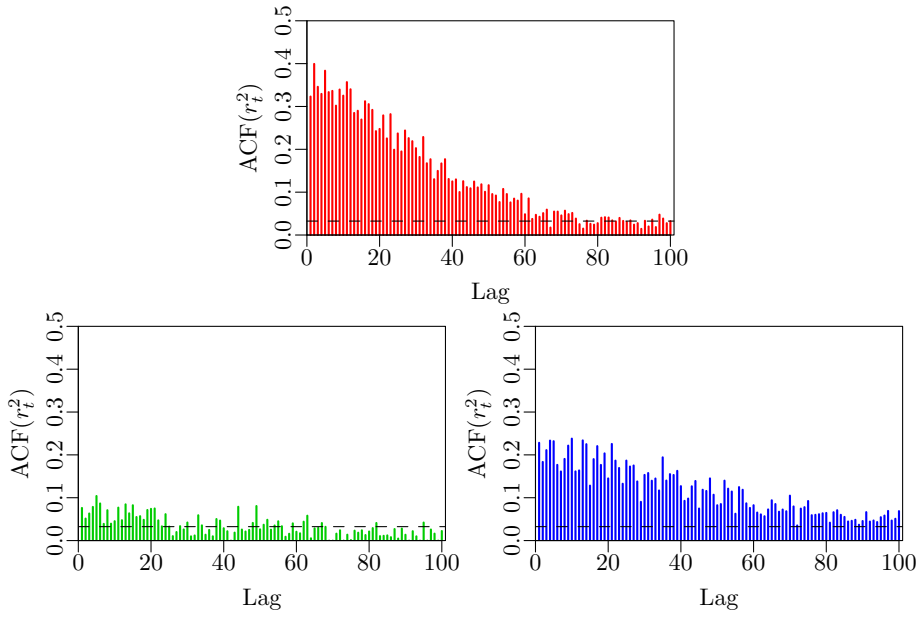


Figure 2.16: The autocorrelation functions for the squared outlier-corrected log-returns of the three indices.

GSCI. The autocorrelation of the squared log-returns of the JPM GBI is only barely significant for lags below 20. The long memory of the squared log-returns is closely related to the volatility clustering noted in relation to figure 2.14.

Figure 2.16 shows the ACFs for the squared outlier-corrected log-returns. Following the approach by Granger and Ding (1995a), values outside the interval $\bar{r_t} \pm 4\widehat{\sigma}$ are set equal to the nearest boundary. Restraining the impact of outliers reduces the amount of noise in the empirical ACFs significantly. There is a weekly variation in the squared log-returns of the MSCI ACWI and the S&P GSCI that could suggest the need for an inhomogeneous model. The size of the weekly variations is negligible compared to the amount of noise, though.

The correlations between the indices are far from constant. This appears from figure 2.17 where the one-year rolling correlations are depicted together with the yield on ten-year US government bonds. The correlation between the MSCI ACWI and the JPM GBI went from 0.5 in 1997 to -0.4 during 1998. Then it went back up to 0.25 in 2000, before it went to a low of almost -0.7 in 2003. From 2004 to 2007 the correlation was close to zero before it fell to -0.6 in 2008.

The correlation between the MSCI ACWI and the S&P GSCI has been close to zero throughout most of the period, but since 2005 it has been increasing steadily to a high of 0.5 at the end of 2008. The correlation between the JPM

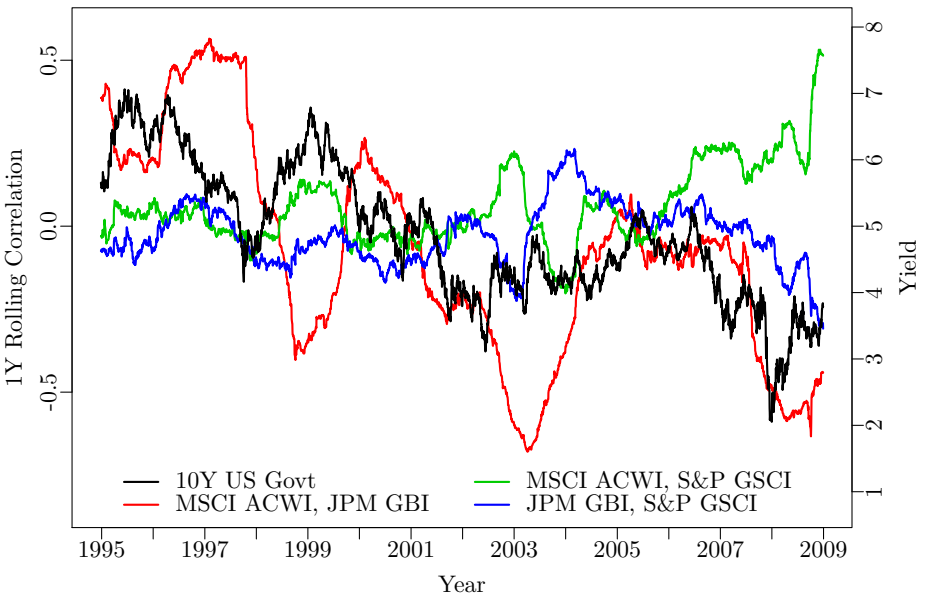


Figure 2.17: One-year rolling correlations between the indices together with the yield on ten-year US government bonds.

GBI and the S&P GSCI has also remained close to zero throughout most of the in-sample period, but from 2007 it started decreasing to a low of -0.3 at the end of 2008. The observation that an investment in commodities seemed to provide protection when the bond index was suffering in figure 2.11 on page 18 is not evident from the one-year correlations.

Figure 2.18 shows the cumulative proportion of variance explained by the first two principal components. Principal component analysis (PCA) uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The first principal component has as high a variance as possible, meaning that it accounts for as much of the variability in the dataset as possible. The succeeding components, in turn, have the highest variance possible under the constraint that each one has to be orthogonal to (uncorrelated with) the preceding components. Higher correlations mean greater redundancy and greater redundancy results in more variation extracted in fewer components.

The solid lines are estimated using a rolling window of 252 trading days. The dashed orange and purple lines show the cumulative proportion of variance that can be explained by one and two principal components, respectively, when looking at the whole in-sample period. The log-returns were standardized prior

to the analysis.

The proportion of variance explained by the first component varies from 0.35 at the lowest to above 0.6 at the highest. The spikes for component one in 1996–1997, in 2003, and again from the middle of 2007 are coinciding with the times where the correlation between the MSCI ACWI and the JPM GBI is significantly strengthened. During the GFC, the correlation between the MSCI ACWI and the S&P GSCI also increased significantly, which is why the first component accounted for a larger proportion of the variance at the expense of component two. Besides the GFC, the second principal component explains an almost constant proportion of the variance of about 0.35.

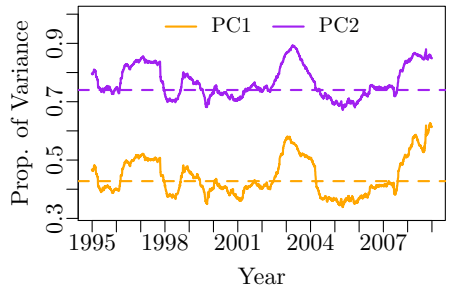


Figure 2.18: The cumulative proportion of variance explained by the first two principal components for the three indices.

The correlations are stronger at the times of high market volatility and stress. Thus, diversification may not materialize precisely when an investor needs it the most. The fact that the correlation between the MSCI ACWI and the JPM GBI gets increasingly negative is not a problem, but the increase in the correlation between the MSCI ACWI and the S&P GSCI, in particular during the GFC, is unfavorable. The impact would be more significant if it was two different stock indices, then the correlations would increase even more around year 2000 and 2008. Although the correlations increased during the GFC, there are definitely diversification benefits between asset classes.

2.6 In-Sample Adjustment

The bond index has clearly outperformed both the stock and commodity index in-sample following the significant decline in the term structure of interest rates with corresponding high returns for long-maturity bonds. For other time periods, stocks and commodities are likely to show a more attractive risk/return profile compared to bonds. Had the in-sample period ended in 2007 rather than 2008, then the Sharpe ratios would have been 0.67 and 0.46 for the MSCI ACWI and the S&P GSCI, respectively.

The end of a major financial crisis, where the level of stress in the markets is at its highest, is a critical time at which to be estimating Sharpe ratios. The SR of 0.29 for the MSCI ACWI is not an unrealistic long-term level. If the SR is set too low, then it will not be profitable to change allocation, and if the SR is

set too high, then it will appear profitable to change allocation very often, as the excess return will quickly cover the transaction costs.

The assumption that will be made when calibrating time series models to the in-sample data to use for scenario generation and asset allocation is that the annual SR of the JPM GBI equals that of the MSCI ACWI. This in attempt to make a neutral assumption although there is no such thing as a neutral assumption in this context. Leverage aversion could be one reason why stocks and bonds should not be expected to yield the same risk-adjusted return in the long run (Asness et al. 2012, Frazzini and Pedersen 2014).

Then there is the dispute about the commodity risk premium. The point of departure will be to use the returns as they are, meaning that commodities will have a lower SR than stocks and bonds in-sample. Hence, it is only the in-sample returns of the JPM GBI that will be adjusted. This is done by subtracting 0.022% from the daily log-returns. In this way, the MSCI ACWI and the JPM GBI will both have an in-sample annual SR of 0.29. It should be emphasized that the data used for out-of-sample testing will remain untransformed.

CHAPTER 3

Markov-Switching Mixtures

The normal distribution is a poor fit to most financial returns. Mixtures of normal distributions provide a much better fit as they are able to reproduce both the skewness and leptokurtosis often observed. An extension to Markov switching mixture models, also referred to as hidden Markov models (HMMs), is frequently applied to capture both the distributional and temporal properties of financial returns.

In an HMM, the distribution that generates an observation depends on the state of an unobserved Markov chain. The transition probabilities of the Markov chain are assumed to be constant implying that the sojourn times are geometrically distributed. The memoryless property of the geometric distribution is not always appropriate. An alternative is the hidden semi-Markov model (HSMM) in which the sojourn time distribution is modeled explicitly for each hidden state.

If the observations are not equidistantly sampled then a continuous-time hidden Markov model (CTHMM) that factors in the sampling times of the observations can be applied. The advantages of a continuous-time formulation include the flexibility to increase the number of states or incorporate inhomogeneity without a dramatic increase in the number of parameters.

The theory of HMMs in discrete time and their estimation is outlined in section 3.1. The HSMM is introduced in section 3.2 and the CTHMM in section 3.3. The models are fitted to the in-sample returns of the MSCI ACWI in section 3.4 and gradient-based methods are considered in section 3.5.

3.1 Hidden Markov Models in Discrete Time

In hidden Markov models, the probability distribution that generates an observation depends on the state of an underlying and unobserved Markov process. HMMs are a particular kind of dependent mixture and are therefore also referred to as Markov-switching mixture models. General references to the subject include Cappé et al. (2005), Frühwirth-Schnatter (2006), and Zucchini and MacDonald (2009).

A sequence of discrete random variables $\{S_t : t \in \mathbb{N}\}$ is said to be a first-order Markov chain if, for all $t \in \mathbb{N}$, it satisfies the Markov property:

$$\Pr\left(S_{t+1} \middle| S_t, \ldots, S_1\right) = \Pr\left(S_{t+1} \middle| S_t\right). \tag{3.1}$$

The conditional probabilities $\Pr\left(S_{u+t} = j \middle| S_u = i\right) = \gamma_{ij}(t)$ are called transition probabilities. The Markov chain is said to be homogeneous if the transition probabilities are independent of u, otherwise inhomogeneous.

A Markov chain with transition probability matrix $\boldsymbol{\Gamma}(t) = \{\gamma_{ij}(t)\}$ has stationary distribution $\boldsymbol{\pi}$ if $\boldsymbol{\pi}\boldsymbol{\Gamma} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1}' = 1$. The Markov chain is termed stationary if $\boldsymbol{\delta} = \boldsymbol{\pi}$ where $\boldsymbol{\delta}$ is the initial distribution, i.e. $\delta_i = \Pr(S_1 = i)$.

If the Markov chain $\{S_t\}$ has m states, then the bivariate stochastic process $\{(S_t, X_t)\}$ is called an m-state HMM. With $\boldsymbol{X}^{(t)}$ and $\boldsymbol{S}^{(t)}$ representing the values from time 1 to time t, the simplest model of this kind can be summarized by

$$\Pr\left(S_t \middle| \boldsymbol{S}^{(t-1)}\right) = \Pr\left(S_t \middle| S_{t-1}\right), \quad t = 2, 3, \ldots, \tag{3.2a}$$

$$\Pr\left(X_t \middle| \boldsymbol{X}^{(t-1)}, \boldsymbol{S}^{(t)}\right) = \Pr\left(X_t \middle| S_t\right), \quad t \in \mathbb{N}. \tag{3.2b}$$

When the current state S_t is known, the distribution of X_t depends only on S_t. This causes the autocorrelation of $\{X_t\}$ to be strongly dependent on the persistence of $\{S_t\}$.

An HMM is a state-space model with finite state space where (3.2a) is the state equation and (3.2b) is the observation equation. A specific observation can usually arise from more than one state as the support of the conditional distributions overlaps. The unobserved state process $\{S_t\}$ is therefore not directly observable through the observation process $\{X_t\}$, but can only be estimated.

As an example, consider the two-state model with Gaussian conditional distributions:

$$X_t = \mu_{S_t} + \varepsilon_{S_t}, \quad \varepsilon_{S_t} \sim N\left(0, \sigma^2_{S_t}\right),$$

where

$$\mu_{S_t} = \begin{cases} \mu_1, & \text{if } S_t = 1, \\ \mu_2, & \text{if } S_t = 2, \end{cases} \quad \sigma^2_{S_t} = \begin{cases} \sigma^2_1, & \text{if } S_t = 1, \\ \sigma^2_2, & \text{if } S_t = 2, \end{cases} \quad \text{and } \boldsymbol{\Gamma} = \begin{bmatrix} 1-\gamma_{12} & \gamma_{12} \\ \gamma_{21} & 1-\gamma_{21} \end{bmatrix}.$$

For this model, the first four central moments are

$$\begin{aligned}
\mathrm{E}\left[X_t|\,\theta\right] &= \pi_1\mu_1 + (1-\pi_1)\,\mu_2,\\
\mathrm{Var}\left[X_t|\,\theta\right] &= \pi_1\sigma_1^2 + (1-\pi_1)\,\sigma_2^2 + \pi_1\,(1-\pi_1)\,(\mu_1-\mu_2)^2,\\
\mathrm{Skewness}\left[X_t|\,\theta\right] &= \pi_1\,(1-\pi_1)\,(\mu_1-\mu_2)\,\frac{(1-2\pi_1)\,(\mu_1-\mu_2)^2 + 3\left(\sigma_1^2-\sigma_2^2\right)}{\sigma^3},\\
\mathrm{Kurtosis}\left[X_t|\,\theta\right] &= \frac{\pi_1\,(1-\pi_1)}{\sigma^4}\Big[3\left(\sigma_1^2-\sigma_2^2\right)^2 + (\mu_1-\mu_2)^4\,(1-6\pi_1\,(1-\pi_1))\\
&\quad + 6\,(2\pi_1-1)\left(\sigma_2^2-\sigma_1^2\right)(\mu_1-\mu_2)^2\Big] + 3.
\end{aligned}$$

θ denotes the model parameters and $\sigma^2 = \mathrm{Var}\left[X_t|\,\theta\right]$ is the unconditional variance (Timmermann 2000, Frühwirth-Schnatter 2006).

The unconditional mean is simply the weighted average of the means. A difference in the means across the states enters both the variance, skewness, and kurtosis. In fact, skewness only arises if there is a difference in the mean values. The unconditional variance is not just the weighted average of the variances; a difference in means also imparts an effect because the switch to a new state contributes to volatility (Ang and Timmermann 2011). Intuitively, the possibility of changing to a new state with a different mean value introduces an extra source of risk. This is similar to a mixed-effects model where the total variance arises from two sources of variability, namely within-group heterogeneity and between-group heterogeneity (see e.g. Madsen and Thyregod 2011).

The value of the autocorrelation function at lag k is

$$\rho_{X_t}\left(k|\,\theta\right) = \frac{\pi_1\,(1-\pi_1)\,(\mu_1-\mu_2)^2}{\sigma^2}\lambda^k$$

and the autocorrelation function for the squared process is

$$\rho_{X_t^2}\left(k|\,\theta\right) = \frac{\pi_1\,(1-\pi_1)\left(\mu_1^2-\mu_2^2+\sigma_1^2-\sigma_2^2\right)^2}{\mathrm{E}\left[X_t^4|\,\theta\right] - \mathrm{E}\left[X_t^2|\,\theta\right]^2}\lambda^k.$$

$\lambda = \gamma_{11} - \gamma_{21}$ is the second largest eigenvalue of $\mathbf{\Gamma}$ (Frühwirth-Schnatter 2006).[15] It is evident from these expressions, as noted by Rydén et al. (1998), that HMMs can only reproduce an exponentially decaying autocorrelation structure.

The ACF of the first-order process becomes zero if the mean values are equal, whereas persistence in the squared process can be induced either by a difference in the means, as for a mixed-effects model, or by a difference in the variances across the states. In both cases, the persistence increases with the combined persistence of the states as measured by λ (Ang and Timmermann 2011).

[15] The other eigenvalue of $\mathbf{\Gamma}$ is $\lambda = 1$.

The sojourn times are implicitly assumed to be geometrically distributed:

$$\Pr\left(\text{'staying } t \text{ time steps in state } i\text{'}\right) = \gamma_{ii}^{t-1}\left(1-\gamma_{ii}\right) \tag{3.3}$$

with the expected duration of state i being

$$\mathrm{E}\left[r_i\right] = \frac{1}{1-\gamma_{ii}}. \tag{3.4}$$

The geometric distribution is memoryless, implying that the time until the next transition out of the current state is independent of the time spent in the state.

The parameters of an HMM are usually estimated using the Maximum Likelihood (ML) method. There are essentially three problems in the context of HMMs:

1. Evaluating the likelihood of the observations $\boldsymbol{x}^{(T)}$ given the parameters.
2. Estimating the model parameters θ that maximize the likelihood of the observed data.
3. Inferring the most likely sequence of states $\boldsymbol{s}^{(T)}$.

The first question is addressed by a dynamic programming technique, the forward–backward algorithm, the second task is carried out using the Baum–Welch algorithm, and the third problem can be solved efficiently using the Viterbi algorithm. The algorithms are outlined in the next three subsections.

The Forward–Backward Algorithm

Under the assumption that successive observations are independent, the likelihood of the observations $\boldsymbol{x}^{(T)}$ given the parameters θ is

$$L_T = \Pr\left(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)} \middle| \theta\right) = \boldsymbol{\delta}\mathbf{D}\left(x_1\right)\boldsymbol{\Gamma}\mathbf{D}\left(x_2\right)\cdots\boldsymbol{\Gamma}\mathbf{D}\left(x_T\right)\mathbf{1}', \tag{3.5}$$

where $\mathbf{D}\left(x\right)$ is a diagonal matrix with the state-dependent conditional densities $d_i\left(x\right) = \Pr\left(X_t = x \middle| S_t = i\right)$, $i \in \{1, 2, \ldots, m\}$, as entries. The conditional distribution of X_t may be either discrete or continuous, univariate or multivariate.

The forward–backward algorithm is used to compute the likelihood and is also employed in the Baum–Welch algorithm. The ith element of the vector $\boldsymbol{\alpha}_t$ is the forward probability

$$\alpha_t\left(i\right) = \Pr\left(\boldsymbol{X}^{(t)} = \boldsymbol{x}^{(t)}, S_t = i \middle| \theta\right) \tag{3.6}$$

and the ith element of $\boldsymbol{\beta}_t$ is the backward probability

$$\beta_t\left(i\right) = \Pr\left(X_{t+1} = x_{t+1}, \ldots, X_T = x_T \middle| S_t = i, \theta\right) \tag{3.7}$$

for $t = 1, 2, \ldots, T$ and $i \in \{1, 2, \ldots, m\}$.

The forward and the backward probabilities can be computed recursively as

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \mathbf{D}\left(x_{t+1}\right) \tag{3.8}$$

$$\boldsymbol{\beta}_t' = \boldsymbol{\Gamma} \mathbf{D}\left(x_{t+1}\right) \boldsymbol{\beta}_{t+1}' \tag{3.9}$$

with initial values $\boldsymbol{\alpha}_1 = \boldsymbol{\delta} \mathbf{D}\left(x_1\right)$ and $\boldsymbol{\beta}_T = \mathbf{1}$.

It follows that

$$\alpha_t\left(i\right) \beta_t\left(i\right) = \Pr\left(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, S_t = i \middle| \theta \right). \tag{3.10}$$

Thus, the likelihood of the observed data $\boldsymbol{x}^{(T)}$ can be evaluated as the sum over all m states:

$$\Pr\left(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)} \middle| \theta \right) = \sum_{i=1}^{m} \alpha_t\left(i\right) \beta_t\left(i\right) = \boldsymbol{\alpha}_t \boldsymbol{\beta}_t'. \tag{3.11}$$

The likelihood is independent of the time point t at which the forward–backward variables are evaluated since

$$\boldsymbol{\alpha}_t \boldsymbol{\beta}_t' = \boldsymbol{\alpha}_{t+1} \boldsymbol{\beta}_{t+1}'.$$

The forward and backward probabilities should be scaled to avoid numerical underflow, as described in Zucchini and MacDonald (2009). The complexity of the forward–backward recursions is $\mathcal{O}\left(m^2 T\right)$ which is feasible in contrast to the straightforward evaluation of (3.5) which has complexity $\mathcal{O}\left(m^T\right)$ (Dittmer 2008).

The Baum–Welch Algorithm

The two most popular approaches to maximizing the likelihood are direct numerical maximization and the Baum–Welch algorithm, a special case of the expectation–maximization (EM) algorithm (Baum et al. 1970, Dempster et al. 1977). The Baum–Welch algorithm is often preferred due to its larger robustness to initial values. The likelihood is guaranteed to increase (or remain the same) for each iteration when using the EM algorithm, but convergence toward the maximum might be slow. Another advantage of using the EM algorithm is that there is no need to transform the parameters.[16]

[16] See Cappé et al. (2005) for a discussion of the relative merits of the EM algorithm and direct maximization of the likelihood of an HMM by gradient-based methods.

The Baum–Welch algorithm maximizes the logarithm of the complete-data likelihood (CDLL), i.e. the log-likelihood of the observations $\boldsymbol{x}^{(T)}$ and the missing data $\boldsymbol{s}^{(T)}$, which can be written as

$$\begin{aligned}\log\left(\Pr\left(\boldsymbol{x}^{(T)},\boldsymbol{s}^{(T)}\right)\right) &= \log\left(\delta_{s_1}\prod_{t=2}^{T}\gamma_{s_{t-1},s_t}\prod_{t=1}^{T}d_{s_t}\left(x_t\right)\right)\\ &= \log\delta_{s_1}+\sum_{t=2}^{T}\log\gamma_{s_{t-1},s_t}+\sum_{t=1}^{T}\log d_{s_t}\left(x_t\right).\end{aligned} \tag{3.12}$$

By introducing the binary random variables

$$v_i\left(t\right)=1 \text{ if and only if } s_t=1$$

and

$$w_{ij}\left(t\right)=1 \text{ if and only if } s_{t-1}=i \text{ and } s_t=j,$$

the CDLL can be written as

$$\begin{aligned}\log\left(\Pr\left(\boldsymbol{x}^{(T)},\boldsymbol{s}^{(T)}\right)\right) &= \sum_{i=1}^{m}v_i\left(1\right)\log\delta_i\\ &\quad+\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\sum_{t=2}^{T}w_{ij}\left(t\right)\right)\log\gamma_{ij}\\ &\quad+\sum_{i=1}^{m}\sum_{t=1}^{T}v_i\left(t\right)\log d_i\left(x_t\right).\end{aligned} \tag{3.13}$$

The idea of the EM algorithm is to replace the quantities $v_i\left(t\right)$ and $w_{ij}\left(t\right)$ by their conditional expectations given the observations $\boldsymbol{x}^{(T)}$ and the current parameter estimates. This is the E-step:

$$\hat{v}_i\left(t\right)=\Pr\left(S_t=i|\,\boldsymbol{x}^{(T)}\right)=\alpha_t\left(i\right)\beta_t\left(i\right)/L_T \tag{3.14}$$

and

$$\hat{w}_{ij}\left(t\right)=\Pr\left(S_{t-1}=i,S_t=j|\,\boldsymbol{x}^{(T)}\right)=\alpha_{t-1}\left(i\right)\gamma_{ij}d_j\left(x_t\right)\beta_t\left(j\right)/L_T. \tag{3.15}$$

Having replaced $v_i\left(t\right)$ and $w_{ij}\left(t\right)$ by their conditional expectations $\hat{v}_i\left(t\right)$ and $\hat{w}_{ij}\left(t\right)$ the CDLL (3.13) is maximized with respect to three sets of parameters: the initial distribution $\boldsymbol{\delta}$, the transition probability matrix $\boldsymbol{\Gamma}$, and the parameters of the state-dependent distributions (e.g., $\mu_1,\ldots,\mu_m$ and $\sigma_1^2,\ldots,\sigma_m^2$ in the case of Gaussian distributions). This is the M-step.

These two steps are repeated until some convergence criterion has been satisfied, e.g. until the resulting change in θ or the log-likelihood is less than some threshold. The resulting value of θ is then a stationary point of the likelihood of the observed data. The stationary point is not necessarily the global maximum, but can also be a local maximum or a saddle point.

The M-step splits into three independent maximizations since the first term in (3.13) depends only on the initial distribution, the second term depends only on the transition probabilities, and the third term depends only on the parameters of the state-dependent distributions.

The solution is to set

$$\delta_i = \frac{\hat{v}_i(1)}{\sum_{i=1}^{m} \hat{v}_i(1)} = \hat{v}_i(1) \tag{3.16}$$

and

$$\gamma_{ij} = \frac{f_{ij}(T)}{\sum_{j=1}^{m} f_{ij}(T)}, \tag{3.17}$$

where $f_{ij}(T) = \sum_{t=2}^{T} \hat{w}_{ij}(t)$ is the total number of transitions from state i to state j.

It does not seem reasonable to try to estimate the initial distribution $\boldsymbol{\delta}$ based on just one observation at time $t = 1$. Another possibility is to maximize the likelihood conditioned on starting in a particular state. Maximizing the conditional likelihood over the m possible starting states is equivalent to maximizing the unconditional likelihood since $\boldsymbol{\delta}$ has to be one of the m unit vectors at a maximum of the likelihood.

If the Markov chain is assumed to be stationary, then $\boldsymbol{\delta}$ is completely determined by the transition probability matrix $\boldsymbol{\Gamma}$ as $\boldsymbol{\delta} = \boldsymbol{\delta}\boldsymbol{\Gamma}$ and $\boldsymbol{\delta}\mathbf{1}' = 1$ and the question of estimating $\boldsymbol{\delta}$ falls away. Then the sum of the first two terms in (3.13) has to be maximized with respect to $\boldsymbol{\Gamma}$ which generally requires a numerical solution (Bulla and Berzel 2008, Zucchini and MacDonald 2009).

The maximization of the third term in (3.13) may be easy or difficult depending on the nature of the state-dependent distributions. In some cases numerical maximization will be necessary, but in the case of conditional univariate normal distributions the maximizing values of the state-dependent parameters are simply

$$\hat{\mu}_i = \frac{\sum_{t=1}^{T} \hat{v}_i(t)\, x_t}{\sum_{t=1}^{T} \hat{v}_i(t)},$$

$$\hat{\sigma}_i^2 = \frac{\sum_{t=1}^{T} \hat{v}_i(t)\, (x_t - \hat{\mu}_i)^2}{\sum_{t=1}^{T} \hat{v}_i(t)}.$$

For conditional multivariate normal distributions the M-step reads

$$\hat{\boldsymbol{\mu}}_i = \frac{\sum_{t=1}^{T} \hat{v}_i(t)\, \boldsymbol{x}_t}{\sum_{t=1}^{T} \hat{v}_i(t)},$$
$$\hat{\boldsymbol{\Sigma}}_i = \frac{\sum_{t=1}^{T} \hat{v}_i(t)\, (\boldsymbol{x}_t - \hat{\boldsymbol{\mu}}_i)\, (\boldsymbol{x}_t - \hat{\boldsymbol{\mu}}_i)'}{\sum_{t=1}^{T} \hat{v}_i(t)}.$$

In the case of conditional t-distributions[17] the maximizing values of the state-dependent parameters at iteration $k+1$ are

$$\mu_i^{(k+1)} = \frac{\sum_{t=1}^{T} \hat{v}_i(t)\, u_{it}^{(k)} x_t}{\sum_{t=1}^{T} \hat{v}_i(t)\, u_{it}^{(k)}},$$
$$\sigma_i^{2^{(k+1)}} = \frac{\sum_{t=1}^{T} \hat{v}_i(t)\, u_{it}^{(k)} \left(x_t - \mu_i^{(k+1)}\right)^2}{\sum_{t=1}^{T} \hat{v}_i(t)},$$

where

$$u_{it}^{(k)} = \sigma_i^{2^{(k)}} \frac{\nu_i^{(k)} + 1}{\sigma_i^{2^{(k)}} \nu_i^{(k)} + \left(x_t - \mu_i^{(k)}\right)^2}.$$

The estimator $\nu_i^{(k+1)}$ is the unique solution to the equation

$$1 - \psi\left(\frac{1}{2}\nu_i^{(k)}\right) + \log\left(\frac{1}{2}\nu_i^{(k)}\right) + \psi\left(\frac{\nu_i^{(k)}+1}{2}\right) - \log\left(\frac{\nu_i^{(k)}+1}{2}\right)$$
$$+ \frac{1}{\sum_{t=1}^{T} \hat{v}_i(t)} \left[\sum_{t=1}^{T} \hat{v}_i(t) \left(\log u_{it}^{(k)} - u_{it}^{(k)}\right)\right] = 0,$$

where $\psi(z) = \{\partial \Gamma(z)/\partial z\}/\Gamma(z)$ is the Digamma function. The solution can be determined without relevant complications, e.g. using a bisection algorithm or quasi-Newton methods, because the function on the left-hand side is monotonically increasing in $\nu_i^{(k)}$ (Bulla 2011).

In order to increase the speed of convergence a hybrid algorithm can be applied that starts with the EM algorithm and switches to a Newton-type algorithm when a certain stopping criterion is fulfilled as outlined by Bulla and Berzel (2008). The resulting algorithm combines the large circle of convergence from the EM algorithm with the superlinear convergence of the Newton-type algorithm in the neighborhood of the maximum.

[17] The univariate Student t-distribution, $X \sim t_\nu(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$, $\sigma > 0$, and $\nu > 0$, is defined for $x \in \mathbb{R}$. The density is $d_{t_\nu}(x) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi\sigma^2}} \left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}$.

The likelihood function of an HMM is in general a complicated function of the parameters with several local maxima. Depending on the initial values, it can easily happen that the algorithm identifies a local rather than the global maximum. A sensible strategy is therefore to try a range of (randomly generated) initial values for the maximization.

In the case of continuous state-dependent distributions the likelihood can be unbounded in the vicinity of certain parameter combinations. For example, if the conditional distribution is Gaussian, then the likelihood can be arbitrarily large if the mean is equal to one of the observations and the conditional variance tends to zero (Frühwirth-Schnatter 2006). Podgórski and Wallin (2013) showed how this can be handled in an elegant way by leaving out the observation that makes the largest contribution to the likelihood at each step of the maximization.[18]

Alternatively, the likelihood function can be maximized using a Bayesian approach based on Markov chain Monte Carlo (MCMC) sampling; by including prior information on the ratio between the conditional variances the likelihood becomes bounded (Frühwirth-Schnatter 2006). Rydén (2008) compared the EM and MCMC approaches and found that MCMC can be advantageous for interval estimation and inferential problems, whereas EM is a simpler and faster way of obtaining point estimates.

The uncertainties of the parameter estimates can also be obtained through bootstrapping as outlined by Zucchini and MacDonald (2009) or from the inverse of the Hessian of the log-likelihood function at the optimum (see e.g. Madsen and Thyregod 2011). Lystig and Hughes (2002) described an algorithm for exact computation of the score vector and the observed information matrix in HMMs that can be performed in a single pass through the data. The algorithm was derived from the forward–backward algorithm. If some of the parameters are on or near the boundary of their parameter space, which is often the case in HMMs, the use of the Hessian to compute standard errors is unreliable. Moreover, the conditions for asymptotic normality of the ML estimator are often violated, thus making approximate confidence intervals based on the computed standard errors unreliable.

Decoding

Decoding is the process of determining the states of the Markov chain that are most likely (under the fitted model) to have given rise to the observation sequence. This can be done either locally by determining the most likely state at each time t or globally by determining the most likely sequence of states.

[18] Seo and Kim (2012) proposed a similar strategy for avoiding spurious roots where one or more components are overfitting a small random localized pattern in the data rather than any underlying group structure.

Local decoding is the maximization of the conditional distribution of S_t for each $t = 1, 2, \ldots, T$ given the observations $\boldsymbol{x}^{(T)}$ and model parameters θ:

$$\underset{i=1,\ldots,m}{\arg\max} \Pr\left(S_t = i | \, \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, \theta\right) \tag{3.18}$$

where the smoothing probabilities can be computed using

$$\Pr\left(S_t = i \,\middle|\, \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, \theta\right) = \frac{\alpha_t(i)\,\beta_t(i)}{L_T}. \tag{3.19}$$

Local decoding can lead to impossible state sequences as it does not take the transition probabilities into account.

Global decoding is the maximization of the conditional probability[19]

$$\underset{\boldsymbol{s}^{(T)}}{\arg\max} \Pr\left(\boldsymbol{S}^{(T)} = \boldsymbol{s}^{(T)} \,\middle|\, \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, \theta\right). \tag{3.20}$$

Maximizing (3.20) over all possible state sequences by brute force involves m^T function evaluations. The Viterbi algorithm (Viterbi 1967) can be used to compute the most likely sequence of states in an efficient manner. This is done by defining the highest probability along a single path for the first t observations ending in state i at time t:

$$\xi_{ti} = \underset{\boldsymbol{s}^{(t-1)}}{\arg\max} \Pr\left(\boldsymbol{S}^{(t-1)} = \boldsymbol{s}^{(t-1)}, S_t = i, \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)} \,\middle|\, \theta\right). \tag{3.21}$$

The probabilities ξ_{tj} can be computed recursively for $t = 2, 3, \ldots, T$ and $i \in \{1, 2, \ldots, m\}$ using

$$\xi_{tj} = \left(\underset{i=1,\ldots,m}{\arg\max}\, \xi_{t-1,i}\gamma_{ij}\right) d_j(x_t) \tag{3.22}$$

with the initial condition $\xi_{1i} = \Pr(S_1 = i, X_1 = x_1) = \delta_i d_i(x_1)$.

The dynamic programming technique is the same as in the forward–backward algorithm, the only difference is the substitution of the sum with a maximization. The complexity remains $\mathcal{O}\left(m^2 T\right)$ (Dittmer 2008).

3.2 Hidden Semi-Markov Models

If the assumption of geometrically distributed sojourn times is unsuitable, then hidden semi-Markov models can be applied. HMMs and HSMMs differ only

[19] This is equivalent to maximizing the joint probability $\Pr\left(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, \boldsymbol{S}^{(T)} = \boldsymbol{s}^{(T)} \,\middle|\, \theta\right)$.

in the way that the state process is defined. In HSMMs, the sojourn time distribution is modeled explicitly for each state i:

$$r_i(u) = \Pr(S_{t+u+1} \neq i,\, S_{t+u-v} = i,\, v = 0, \ldots, u-2 \,|\, S_{t+1} = i,\, S_t \neq i) \quad (3.23)$$

and the transition probabilities are defined as

$$\gamma_{ij} = \Pr(S_{t+1} = j \,|\, S_{t+1} \neq i,\, S_t = i) \quad (3.24)$$

for each $j \neq i$ with $\gamma_{ii} = 0$ and $\sum_j \gamma_{ij} = 1$.

The conditional independence assumption for the observation process is similar to a simple HMM, but the semi-Markov chain associated with HSMMs does not have the Markov property at each time t. This property is transferred to the imbedded first-order Markov chain, that is the sequence of visited states (Bulla and Bulla 2006). In other words, the future states are only conditionally independent of the past states when the process changes state.

The HMM with geometrically distributed sojourn times is just a special case of the HSMM and the EM algorithm as outlined in the previous section can easily be extended to cover other sojourn time distributions (see Bulla and Bulla 2006).

The CDLL of an HSMM includes the sojourn times $u_1, u_2, \ldots, u_N$ and the sequence of visited states $\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_N$, where N is the number of visited states, in addition to the observations $x_1, x_2, \ldots, x_T$ and the hidden states $s_1, s_2, \ldots, s_T$:

$$\begin{aligned}
\log\left(\Pr\left(\boldsymbol{x}^{(T)}, \boldsymbol{s}^{(T)}\right)\right) &= \log\left(\delta_{\tilde{s}_1} d_{\tilde{s}_1}(u_1) \prod_{n=2}^{N} \gamma_{\tilde{s}_{n-1},\tilde{s}_n} r_{\tilde{s}_n}(u_n) \prod_{t=1}^{T} d_{s_t}(x_t)\right) \\
&= \log \delta_{\tilde{s}_1} + \sum_{n=1}^{N} \log d_{\tilde{s}_n}(u_n) + \sum_{n=2}^{N} \log \gamma_{\tilde{s}_{n-1},\tilde{s}_n} \\
&\quad + \sum_{t=1}^{T} \log d_{s_t}(x_t)\,.
\end{aligned} \quad (3.25)$$

The four terms in (3.25) can be maximized independently as the first term depends only on the initial distribution, the second term depends only on the parameters of the sojourn time distributions, the third term depends only on the transition probabilities, and the fourth term depends only on the parameters of the state-dependent distributions.

HSMMs provide a considerable flexibility compared to HMMs without complicating the estimation significantly, though the maximization of the term that depends on the parameters of the sojourn time distributions in (3.25) typically requires a numerical solution. The selection of the most appropriate sojourn time distribution can, however, be a complicated problem (Bulla et al. 2010).

3.3 Hidden Markov Models in Continuous Time

In continuous-time Markov chains, transitions may occur at all times rather than at discrete and equidistant time points. There is no smallest time step and the quantities of interest are the transition probabilities

$$p_{ij}(\Delta t) = \Pr(S(t+\Delta t) = j \,|\, S(t) = i) \tag{3.26}$$

as $\Delta t \to 0$. Clearly, $p_{ij}(0) = 0$ for different states i and j, and it can be shown that under certain regularity conditions

$$\lim_{t \to 0} \mathbf{P}(t) = \mathbf{I}. \tag{3.27}$$

Assuming that $p_{ij}(\Delta t)$ is differentiable at 0, the transition rates are defined as

$$\begin{aligned} p'_{ij}(0) &= \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t) - p_{ij}(0)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\Pr(S(t+\Delta t) = j \,|\, S(t) = i)}{\Delta t} \\ &= q_{ij} \end{aligned} \tag{3.28}$$

with the additional definition $q_{ii} = q_i = -\sum_{j \neq i} q_{ij}$. The transition intensity matrix $\mathbf{Q} = \{q_{ij}\}$ has non-negative off-diagonal elements q_{ij}, non-positive diagonal entries q_i, and all rows sum to zero.

The stationary distribution $\boldsymbol{\pi}$, if it exists, is found by solving the system of equations

$$\begin{cases} \boldsymbol{\pi}\mathbf{Q} = \mathbf{0} \\ \boldsymbol{\pi}\mathbf{1}' = 1. \end{cases} \tag{3.29}$$

If it has a strictly positive solution (all elements in $\boldsymbol{\pi}$ are strictly positive), then the stationary distribution exists and is independent of the initial distribution.

The matrix of transition probabilities $\mathbf{P}(t) = \{p_{ij}(t)\}$ can be found as the solution to Kolmogorov's differential equation

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} = \mathbf{P}(t)\,\mathbf{Q} \tag{3.30}$$

with the initial condition $\mathbf{P}(0) = \mathbf{I}$. The solution being

$$\mathbf{P}(t) = e^{\mathbf{Q}t}\mathbf{P}(0) = e^{\mathbf{Q}t}. \tag{3.31}$$

It follows that the transition intensity matrix $\mathbf{Q}$ is the matrix-logarithm of the one-step transition probability matrix

$$\mathbf{Q} = \log \mathbf{P}(1). \tag{3.32}$$

When the process enters state i, it remains there according to an exponential distribution with parameter $-q_i > 0$ before it instantly jumps to another state $j \neq i$ with probability $-q_{ij}/q_i$. A continuous-time Markov chain is fully characterized by its initial distribution $\boldsymbol{\delta}$ and the transition intensity matrix $\mathbf{Q}$.

The exponential distribution is memoryless just like its discrete analogue, the geometric distribution. By mapping multiple latent states to the same output distribution, it is possible to allow for non-exponentially distributed sojourn times. The distribution of sojourn times will then be a mixture of exponential distributions, which is a phase-type distribution, and the Markov property will be transferred to the imbedded Markov chain as for the HSMM. Phase-type distributions can be used to approximate any positive-valued distribution with arbitrary precision.

It is often reasonable to assume that in a short time interval Δt, the only possible transitions are to the neighboring states:

$$\left.\begin{aligned} p_{ij} &= o\left(\Delta t\right), \;\; |i-j| \geq 2, \\ p_{ii}\left(\Delta t\right) &= 1 - q_i \Delta t + o\left(\Delta t\right), \\ p_{i,i-1}\left(\Delta t\right) &= w_i q_i \Delta t + o\left(\Delta t\right), \\ p_{i,i+1}\left(\Delta t\right) &= \left(1 - w_i\right) q_i \Delta t + o\left(\Delta t\right), \\ i &\in \left\{1, 2, \ldots, m\right\}, \end{aligned}\right\} \tag{3.33}$$

where $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$. The notation includes transitions from state 1 to m and reverse with the definition that state 0 = state m and state $(m+1)$ = state 1.

The Markov chain $\{S\left(t\right)\}$ is then a birth-and-death process in the sense that it can change its state index by at most one in each step. Although the process cannot go straight from state i to state $i+2$ there is no limit to how fast the transition can occur. The matrix of transition intensities has the structure

$$\mathbf{Q} = \begin{bmatrix} -q_1 & \left(1-w_1\right) q_1 & 0 & \cdots & 0 & w_1 q_1 \\ w_2 q_2 & -q_2 & \left(1-w_2\right) q_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \left(1-w_m\right) q_m & 0 & 0 & \cdots & w_m q_m & q_m \end{bmatrix}. \tag{3.34}$$

The number of parameters increases linearly with the number of states. Thus, a continuous-time Markov chain yields a parameter reduction over its discrete-time analogue if the number of states exceeds three. The higher the number of states, the larger the reduction. In addition, it is possible to incorporate inhomogeneity without a dramatic increase in the number of parameters using splines, harmonic functions, or similar.[20]

[20] See Iversen et al. (2013) for an example of the use of splines to reduce the number of parameters in an inhomogeneous Markov model.

Another advantage of a continuous-time formulation is the flexibility to use data with any sampling interval as the data is not assumed to be equidistantly sampled. In a discrete-time model, the trading days are aggregated meaning that weekends and bank holidays are ignored so that Friday is followed by Monday in a normal week. A continuous-time model is able to recognize the longer time span between Friday and Monday, which could lead to a different behavior (see e.g. Rogalski 1984, Asai and McAleer 2007). The time span should not necessarily be three days, the point is that the sampling times are modeled.

A Markov process $\{S(t)\}$ observed at equidistant time points $t_1 = \varsigma, \ldots, t_T = \varsigma T$ defines a discrete Markov chain $Y_1 = S(\varsigma), \ldots, Y_T = S(\varsigma T)$ with transition matrix $\boldsymbol{\Gamma} = e^{\mathbf{Q}\varsigma}$. It is not every discrete Markov chain that belongs to a continuous Markov process. In other words, not every discrete Markov chain can be imbedded into a continuous-time Markov process. This is called the imbedding problem (Dittmer 2008).

A Continuous-Time Version of the Baum–Welch Algorithm

The CDLL of a CTHMM, i.e. the log-likelihood of the observations $x_1, \ldots, x_T$ and the unobserved states $s_1, \ldots, s_T$ at the discrete time points $t_1, \ldots, t_T$, is

$$\begin{aligned}\log\left(\Pr\left(\boldsymbol{x}^{(T)}, \boldsymbol{s}^{(T)}\right)\right) &= \log\left(\delta_{s_1} \prod_{i=2}^{T} p_{s_{i-1},s_i}(t_i - t_{i-1}) \prod_{i=1}^{T} d_{s_i}(x_i)\right) \\ &= \log \delta_{s_1} + \sum_{i=2}^{T} \log p_{s_{i-1},s_i}(t_i - t_{i-1}) \\ &\quad + \sum_{i=1}^{T} \log d_{s_i}(x_i).\end{aligned} \tag{3.35}$$

The probability of the transitions can be written as the probability of a sojourn of a specific length multiplied by the probability of an instant transition to another state as it was done for the HSMM. With $f_{ij}(\tau)$ denoting the total number of transitions from state i to j and $R_i(\tau)$ the total duration spent in state i the CDLL can be written as

$$\begin{aligned}\log\left(\Pr\left(\boldsymbol{x}^{(T)}, \boldsymbol{s}^{(T)}\right)\right) &= \log \delta_{s_1} + \sum_{i=1}^{m} \sum_{j \neq i} f_{ij}(\tau) \log q_{ij} \\ &\quad + \sum_{i=1}^{m} q_i R_i(\tau) + \sum_{i=1}^{T} \log d_{s_i}(x_i).\end{aligned} \tag{3.36}$$

In this way the CDLL is expressed in terms of the parameters. The first and the last term can be dealt with in the same way as for the discrete-time models, but the middle terms require more attention.

The maximum likelihood estimator of the transition rates is

$$\hat{q}_{ij} = \frac{f_{ij}(\tau)}{R_i(\tau)} \tag{3.37}$$

where the diagonal entries follow from the generator constraint $\hat{q}_i = -\sum_{j \neq i} \hat{q}_{ij}$. The problem is that neither f nor R are available and it is not trivial to evaluate their expectations.

Hobolth and Jensen (2011) considered the problem of estimating the summary statistics needed in the EM algorithm. The expected time spent in state i conditional on starting in state A at time $t_1 = 0$ and ending in state B at time $t_T = \tau$ is

$$\mathrm{E}\left[R_i(\tau) \middle| S(0) = A, S(\tau) = B\right] = \frac{1}{p_{AB}(\tau)} \int_0^\tau p_{Ai}(t)\, p_{iB}(\tau - t)\, \mathrm{d}t. \tag{3.38}$$

The expected number of transitions between state i and j conditional on starting in state A at time $t_1 = 0$ and ending in state B at time $t_T = \tau$ is

$$\mathrm{E}\left[f_{ij}(\tau) \middle| S(0) = A, S(\tau) = B\right] = \frac{q_{ij}}{p_{AB}(\tau)} \int_0^\tau p_{Ai}(t)\, p_{jB}(\tau - t)\, \mathrm{d}t. \tag{3.39}$$

The joint expectation integral can be evaluated using eigenvalue decomposition under the assumption that $\mathbf{Q}$ has no repeated eigenvalues. If $\mathbf{Q}$ is diagonalizable with real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$, then $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is a diagonal matrix with the eigenvalues on the diagonal and the columns of $\mathbf{U}$ are the eigenvectors. The transition probability matrix $\mathbf{P}(t) = e^{\mathbf{Q}t} = \mathbf{U}e^{\mathbf{\Lambda}t}\mathbf{U}^{-1}$ and the integral can be evaluated as

$$\begin{aligned}\int_0^\tau p_{A\alpha}(t)\, p_{\beta B}(\tau - t)\, \mathrm{d}t &= \int_0^\tau (\mathbf{U}e^{\mathbf{\Lambda}t}\mathbf{U}^{-1})_{Ai} (\mathbf{U}e^{\mathbf{\Lambda}(\tau - t)}\mathbf{U}^{-1})_{jB} \mathrm{d}t \\ &= \sum_{i=1}^m \mathbf{U}_{Ai}\mathbf{U}^{-1}_{i\alpha} \sum_{j=1}^m \mathbf{U}_{\beta j}\mathbf{U}^{-1}_{jB} \int_0^\tau e^{t\lambda_i + (\tau - t)\lambda_j} \mathrm{d}t.\end{aligned}$$

The last integral is easy to evaluate in the case of real eigenvalues.

Let $\Psi(0, \tau)$ be either of the complete-data sufficient statistics $R_i(\tau)$ or $f_{ij}(\tau)$. $\Psi(0, \tau)$ is then an additive statistic, that is

$$\Psi(0, \tau) = \sum_{l=2}^T \mathrm{E}\left[\Psi(t_{l-1}, t_l) \middle| S_{l-1} = s_{l-1}, S_l = s_l\right].$$

It follows from the Markov property that

$$\mathrm{E}\left[\Psi(0, \tau) \middle| \boldsymbol{s}^{(T)}\right] = \sum_{l=2}^T \mathrm{E}\left[\Psi(t_{l-1}, t_l) \middle| S_{l-1} = s_{l-1}, S_l = s_l\right]. \tag{3.40}$$

In the E-step, the expectation that is needed is

$$\begin{aligned} \mathrm{E}\left[\Psi\left(0,\tau\right)\middle|\boldsymbol{x}^{(T)}\right] = &\sum_{l=2}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}\mathrm{E}\left[\Psi\left(t_{l-1},t_l\right)|S_{l-1}=i,S_l=j\right] \\ &\cdot\Pr\left(S_{l-1}=i,S_l=j\middle|\boldsymbol{x}^{(T)}\right), \end{aligned} \tag{3.41}$$

where the conditional probabilities

$$\Pr\left(S_{l-1}=i,S_l=j\middle|\boldsymbol{x}^{(T)}\right)=\frac{\alpha_{t_{l-1}}\left(i\right)p_{ij}\left(t_l-t_{l-1}\right)d_j\left(x_l\right)\beta_{t_l}\left(j\right)}{L_T} \tag{3.42}$$

are outputs from the forward–backward algorithm.

Each step of the forward–backward recursions (3.8) and (3.9) requires the calculation of the transition probability matrix $\mathbf{P}\left(t\right)=e^{\mathbf{Q}t}$. This can be done in an efficient manner using the eigenvalue decomposition of $\mathbf{Q}$. The relatively small size of the state space in financial applications does not prohibit numerical estimation of the matrix exponential. It is, however, still desirable to avoid this computationally intensive calculation whenever possible.

The scaling by $1/p_{AB}\left(\tau\right)$ in (3.38) and (3.39) cancels with the transition probability in (3.42). The evaluation of the expectations of the summary statistics f and R does therefore not require further evaluations of the matrix exponential. The really time-consuming task is the sum over all possible sequences of states in (3.41).

Lange and Minin (2013) reported good results using the R package `SQUAREM` due to Varadhan (2012) to accelerate the convergence of the EM algorithm. They found that the accelerated EM algorithm outperformed other optimization approaches including direct maximization of the observed data likelihood by numerical methods as implemented in the R package `msm` due to Jackson (2011) and the EM algorithm of Bureau et al. (2000) that uses numerical maximization to update the parameter estimates in the M-step. Alternatively, a hybrid algorithm that switches to a Newton-type algorithm in the neighborhood of the maximum can be applied to increase the speed of convergence as outlined in section 3.1. As to the present application, there is no need to consider means of reducing the time to convergence.

3.4 Model Estimation and Selection

The focus will be on modeling the stock returns as stocks are typically the largest risk contributor in a portfolio. Another reason to focus on the stock index is that stock markets generally lead the economy. Thus, a stock index will often bottom (and head higher) before the economy begins to pick up, and top out before the economy begins to slow down.

The focus will be on univariate models, but a multivariate model will be tested in chapter 5. The models that will be estimated are both discrete and continuous-time HMMs with conditional normal distributions following the approach by Rydén et al. (1998) and Nystrup et al. (2014), respectively; discrete-time HMMs with conditional normal distributions and a t-distribution in the state with the highest volatility following the approach by Bulla (2011); discrete-time HMMs with conditional t-distributions in all the states; and discrete-time HSMMs with negative binomial distributed sojourn times and conditional normal or t-distributions following the approach by Bulla and Bulla (2006).

The estimated models will be compared in terms of how they fit the empirical moments and the autocorrelation function of the squared log-returns. Finally, the likelihood of the models will be compared to the number of parameters using model selection criteria.

The discrete-time models are estimated using the R package `hsmm` due to Bulla et al. (2010) that implements the EM algorithm for various sojourn time distributions including the geometric and the negative binomial distribution. The only exception is the HMM with conditional normal distributions and a t-component in the state with the highest variance. The `hsmm` package cannot handle different conditional distributions across the states so those models are estimated using the implementation of the Baum–Welch algorithm that can be found in appendix A.1 on page 93.

The CTHMM is estimated using the implementation of the continuous-time version of the Baum–Welch algorithm that can be found in appendix A.2 on page 96. This algorithm definitely seemed superior to the `msm` package due to Jackson (2011) that is based on direct numerical maximization of the likelihood function. The Baum–Welch algorithm was much more robust towards initial values without a notable decrease in the speed of convergence.

Exploring the Estimated Models

This subsection will focus on the normal HMMs, but the estimated models all have the same structure. The estimated parameters for the fitted two, three, and four-state HMMs with conditional normal distributions are shown in table 3.1. Approximate standard errors of the estimates based on the observed information are reported in parentheses.[21] The stationary distributions are derived from the estimated transition probability matrices. Parameter estimates for all the fitted models can be found in appendix B on page 103.

The two-state model has a low-variance state with a positive mean return and a high-variance state with a negative mean return implying that turbulence is associated with lower returns, on average. The two states are mostly identified

[21] See section 3.1 for comments on the use of the Hessian to compute standard errors.

Table 3.1
Parameter estimates for the fitted m-state HMMs with conditional normal distributions.

m	$\boldsymbol{\Gamma}$				$\mu \times 10^4$	$\sigma^2 \times 10^4$	δ	π
2	0.987	0.013			6.8	0.32	1	0.69
		(0.003)			(1.2)	(0.01)		
	0.029	0.971			−10.0	2.26	0	0.31
	(0.008)				(4.6)	(0.14)		
3	0.988	0.012	0		8.8	0.24	1	0.48
		(0.004)			(1.3)	(0.01)		
	0.012	0.981	0.006		−1.0	0.88	0	0.45
	(0.004)		(0.002)		(2.5)	(0.04)		
	0	0.043	0.957		−32.5	6.05	0	0.07
		(0.016)			(16.3)	(0.65)		
4	0.983	0.017	0	0	8.9	0.47	1	0.49
		(0.004)			(1.2)	(0.01)		
	0	0.651	0.331	0.018	−69.1	0.71	0	0.19
			(0.042)	(0.012)	(7.7)	(0.05)		
	0.033	0.240	0.727	0	50.7	0.74	0	0.25
	(0.010)	(0.036)			(6.9)	(0.05)		
	0	0	0.053	0.947	−27.9	2.48	0	0.07
			(0.014)		(16.8)	(0.67)		

by the variance as the difference in the mean values is of a similar magnitude as the variances. Nevertheless, the difference in the mean values is a prerequisite for the model's ability to reproduce the observed first-order autocorrelation and skewness. The two states are both very persistent.

The main difference compared to the two-state models with conditional t-distributions is the longer tails of the t-distribution that lead to a lower variance and a higher persistence in the high-variance state. Consequently, the stationary probabilities of the two states are more even in the case of conditional t-distributions.

When adding a third state to the model, an interesting structure emerges where the two outer states only communicate through the middle state. The probabilities of switching between state one and three are fixed to zero in order to save two parameters as they were effectively zero anyway. The third state has a high variance, a low mean return, and is less persistent than the two other states. The third state is sometimes referred to as an outlier state due to its low unconditional probability and high variance (Rydén et al. 1998, Bulla et al. 2011).

The four-state model has two states that are less persistent with a similar variance. Six of the estimated transition probabilities were effectively zero. The

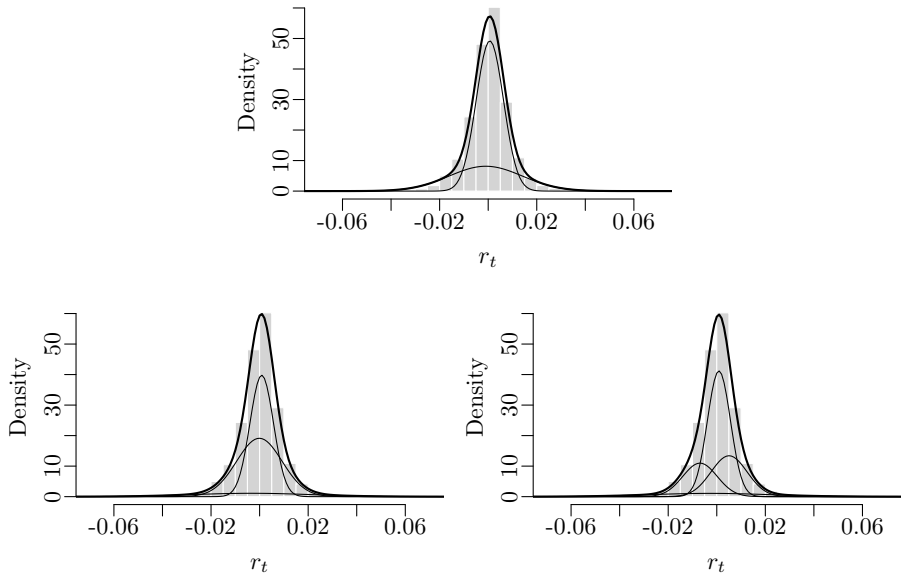


Figure 3.2: Density histograms of the MSCI ACWI log-returns together with density functions for the state-dependent normal distributions scaled by the stationary distribution of the underlying Markov chain and the resulting unconditional distribution for the estimated two, three, and four-state models.

high-variance state again has a low unconditional probability. The simple structure of the transition probability matrix offsets the advantage of a continuous-time formulation when the Markov chain is assumed to be homogeneous and the observations equidistantly sampled.

The conditional densities scaled by the stationary probabilities are shown in figure 3.2 together with the resulting unconditional density and a density histogram of the log-returns of the MSCI ACWI. The unconditional densities appear to be a reasonable fit to the empirical distribution for all three models although the two-state model does not fully capture the kurtosis of the empirical distribution.

Looking at the conditional densities for the three and four-state models it is evident that the high-variance state, which is the density lying very close to the first axis due to its low unconditional probability, captures the tails on both sides of the distribution. It illustrates that turbulent periods are not characterized only by low or negative returns in agreement with the definition by Chow et al. (1999).

Table 3.3
The first four moments of the MSCI ACWI daily log-returns and the fitted models.

Model	Mean	SD	Skewness	Kurtosis
r_t	0.00018	0.0095	-0.40	13
HMM_N (2)	0.00017	0.0097	-0.21	5.9
HMM_{Nt} (2)	0.00026	0.0095	-0.03	17
HMM_t (2)	0.00028	0.0095	-0.18	15
HSMM_N (2)	0.00013	0.0096	-0.31	6.6
HSMM_t (2)	0.00027	0.0095	-0.20	12
HMM_N (3)	0.00011	0.0097	-0.48	10
HMM_{Nt} (3)	0.00015	0.0097	-0.46	19
HMM_t (3)	0.00018	0.0098	-0.44	18
HSMM_N (3)	0.00015	0.0096	-0.52	10
HSMM_t (3)	0.00021	0.0096	-0.55	15
HMM_N (4)	0.00014	0.0096	-0.42	10
HMM_{Nt} (4)	0.00016	0.0096	-0.39	15
HSMM_N (4)	0.00017	0.0094	-0.50	10
CTHMM_N (4)	0.00021	0.0095	-0.39	10

Matching the Moments

Table 3.3 shows the first four empirical moments of the in-sample log-returns of the MSCI ACWI together with the moments of the fitted models based on 250,000 Monte Carlo simulations. The moments and the autocorrelation functions could have been computed numerically, but it would mainly affect the mean values; the simulated mean value is relatively far from the empirical mean value for some of the models as a result of the large standard deviation. This is not the case for the theoretical values.

The two-state models with conditional normal distributions capture some of the observed skewness, but only half of the observed kurtosis. The two-state models with one or more t-components are able to reproduce the large kurtosis, but not the skewness. In fact, the two-state HMM with one t-component does not capture the observed skewness at all.

A much better fit to the empirical moments is obtained by increasing the number of states from two to three as concluded by Nystrup et al. (2014). The HMMs are the best at capturing the skewness. The HMMs with one or more t-components overestimate the kurtosis, but the three-state models are all able to reproduce the large kurtosis. For the HMM and the HSMM with conditional t-distributions the number of degrees of freedom in the middle state is above 100, meaning that the conditional distribution is effectively normal.

Increasing the number of states from three to four improves the fit to the empirical moments although the improvement is less significant compared to going from two to three states. It was not possible to estimate models with condi-

tional t-distributions in all four states as the number of degrees of freedom went towards infinity for some of the states. The four-state models provide more or less an equally good fit to the moments.

Reproducing the Long Memory

The empirical autocorrelation function for the squared log-returns and the squared outlier-corrected log-returns of the MSCI ACWI and the fitted models are shown in figure 3.4. Of the two-state models, the two models with conditional normal distributions do the best job at reproducing the shape of the ACF. The fluctuations at lag one to ten for the two-state HSMM with conditional normal distributions is a result of the short expected sojourns of this model.

The models with one or more t-components are very persistent, but at too low of a level. As pointed out by Bulla (2011), the increased persistence is most likely a result of the excess kurtosis of the t-distributed component(s) that allows for less frequent state changes. Looking at the outlier-corrected squared log-returns in the right-hand column, the models with conditional normal distributions are still better at capturing the decay.

Increasing the number of states to three leads to a better fit to the empirical autocorrelation function of the squared log-returns. The differences between the models are smaller, but the two models with conditional normal distributions are still the best fit. Looking at the outlier-corrected squared log-returns, the three-state models all provide a similar fit.

A further increase in the number of states does not lead to a better fit. The HMM with one t-component is the worst fit when looking at the squared log-returns, while the differences in performance are small when reducing the impact of outliers.

Model Selection

A generalized likelihood ratio test (GLRT) cannot be applied to choose between models with different types of conditional distributions as they are not hierarchically nested. Models with the same type of conditional distribution and different numbers of states are hierarchically nested, but the asymptotic distribution of the likelihood ratio statistic is not the usual χ^2 as there is a continuum of models with $m + 1$ states that are equivalent to a model with m states. The GLRT can be bootstrapped, but it is very time consuming (Rydén 2008).

Instead, penalized likelihood criteria can be used to select the model that is estimated to be closest to the "true" model, as suggested by Zucchini and MacDonald (2009). The disadvantage is that model selection criteria provide no information about the confidence in the selected model relative to others.

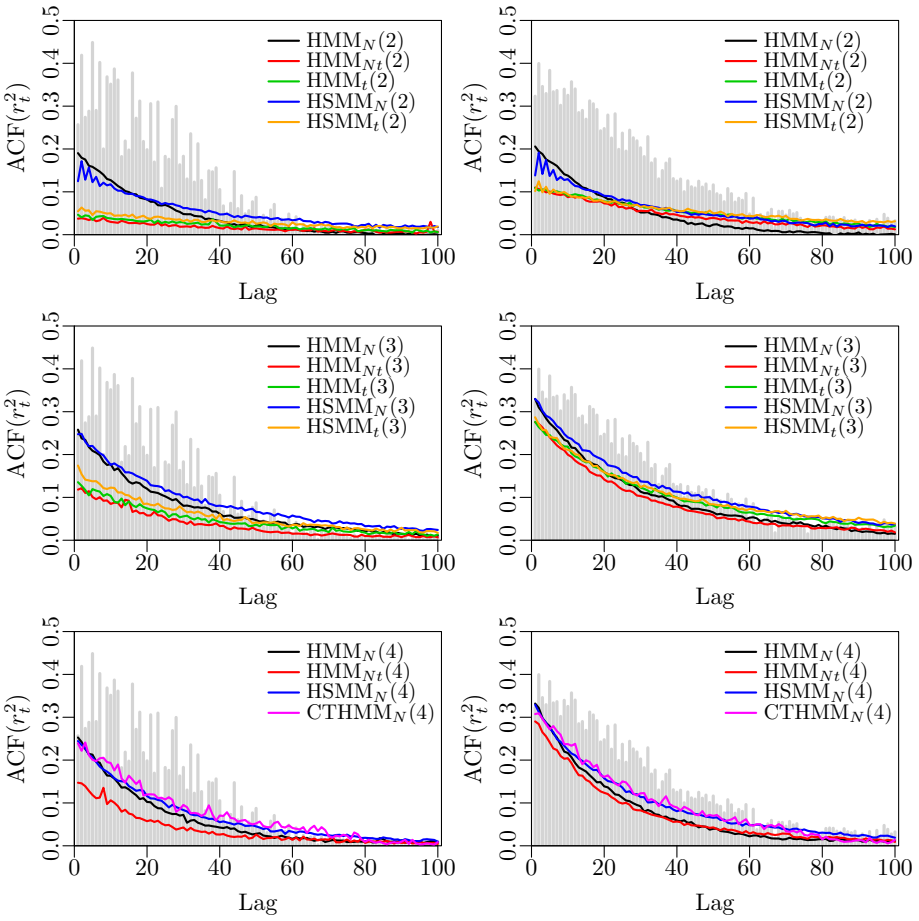


Figure 3.4: The empirical autocorrelation function for the squared log-returns (left column) and the squared outlier-corrected log-returns (right column) of the MSCI ACWI and the fitted models.

Table 3.5
Model selection based on the Akaike and the Bayesian information criterion.

Model	No. of parameters	Log-lik	AIC	BIC
HMM_N (2)	7	12949	-25884	-25840
HMM_{Nt} (2)	8	13032	-26049	-25999
HMM_t (2)	9	13037	-26056	-26000
HSMM_N (2)	9	12991	-25964	-25908
HSMM_t (2)	11	13048	-26075	-26006
HMM_N (3)	12	13135	-26246	-26172
HMM_{Nt} (3)	13	13140	-26253	-26172
HMM_t (3)	15	13143	-26256	-26162
HSMM_N (3)	15	13140	-26251	-26157
HSMM_t (3)	18	13148	-26260	-26147
HMM_N (4)	17	13174	-26315	-26209
HMM_{Nt} (4)	18	13178	-26320	-26208
HSMM_N (4)	21	13198	-26353	-26222
CTHMM_N (4)	17	13170	-26307	-26201

Table 3.5 shows the number of parameters, the log-likelihood, and the value of the Akaike[22] and the Bayesian[23] information criterion for the estimated models. The four-state models are preferred by the two information criteria. Most emphasis is put on the BIC as various simulation studies have shown that the AIC tends to select models with too many states as it puts less weight on the number of parameters (Bacci et al. 2012). Rydén (2008) remarked that BIC is based on approximating the distribution of the ML estimator by a normal and may be unreliable for data of small or moderate size, though.

Looking at the two-state models, the models with conditional t-distributions are preferred by a large margin. The differences between the three and four-state models are a lot smaller. The increase in the number of parameters with the number of states for the discrete-time models is not quadratic as the estimated transition probability matrices have a simple structure where it is not all transitions that are possible. As a consequence, the four-state CTHMM has the same number of parameters as the four-state HMM and provides a similar fit.

The four-state HSMM with conditional normal distributions is preferred by the BIC although the model is not a better fit to the empirical moments nor the long memory of the squared process than the three-state HMM with conditional normal distributions that has 12 instead of 21 parameters. As argued by Dacco and Satchell (1999), the performance of the models should be evaluated by

[22]The Akaike information criterion is defined as $\text{AIC} = -2\log L + 2p$, where p is the number of parameters.

[23]The Bayesian information criterion is defined as $\text{BIC} = -2\log L + p\log T$, where T is the number of observations.

methods appropriate for the intended application rather than in-sample fit to the data. The topic of model selection will, therefore, be revisited in chapter 5.

Parameter Stationarity

Following the approach by Bulla et al. (2011) the parameters of the two-state HMM with conditional normal distributions are estimated using a rolling window of 2000 trading days corresponding to about eight years. The result is shown in figure 3.6 where the dashed lines are the in-sample ML estimates. It is evident that the parameters are far from constant throughout the in-sample period. The size of the variations seems consistent with the approximate standard errors reported in table 3.1 on page 44. For all parameters, the movements at the end of 2008 stand out.

Figure 3.7 shows the parameters of a two-state HMM with a conditional t-distribution in the high-variance state estimated using a rolling window of 2000 trading days. The parameters are still far from constant, but the variations are smaller especially towards the end of 2008. It is only the degrees of freedom of the t-distribution that change dramatically at the end of 2008. The degrees of freedom vary between 8 and 14 throughout most of the in-sample period before dropping below the ML estimate of 4.5 in 2008.

The length of the rolling window affects the parameter estimates. Bulla et al. (2011) chose 2000 days based on the average length of an economic cycle. If the window length is reduced to 1000 days, then the degrees of freedom of the t-distribution exceed 100 throughout most of the in-sample period, meaning that the distribution in the high-variance state is effectively normal. It might suggest that the t-distribution is simply a compensation for inadequacies of the model. The shorter the rolling window, the larger the variations in the parameters, but the models change character if the window is too short. The parameters of the two-state HMM with conditional normal distributions, when estimated using a rolling window of 1000 trading days, are shown in figure 3.8. Compared to figure 3.6, the impact of the GFC on the variance parameters is seen to die out before the end of the sample period due to the shorter window.

To summarize, it is evident that the parameters cannot be assumed to be stationary. This will be particularly important in the out-of-sample testing. As a consequence of the non-constant transition probabilities, the sojourn time distribution becomes a mixture of geometric distributions that does not possess the memoryless property. Accounting for the non-constant transition probabilities is, therefore, likely to offset the advantage of an HSMM.

3.5 Gradient-Based Methods

As pointed out by Cappé et al. (2005) it is possible to evaluate derivatives of the likelihood function with respect to the parameters for any model that

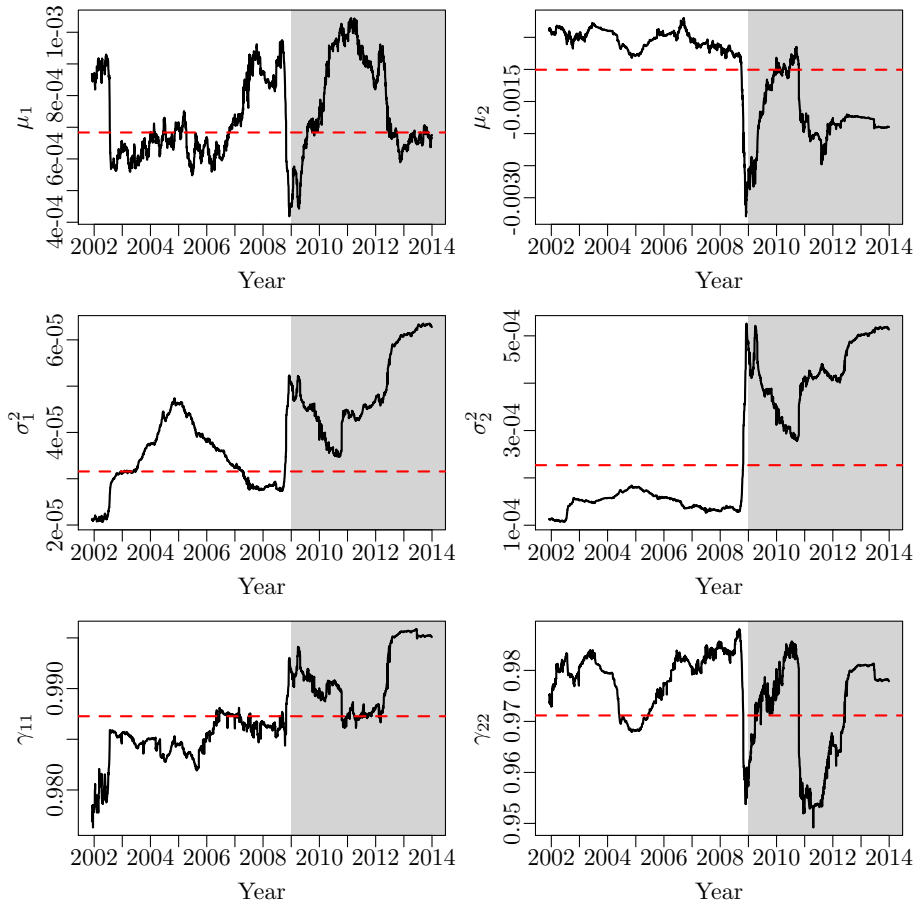


Figure 3.6: The parameters of a two-state Gaussian HMM estimated using a rolling window of 2000 trading days. The dashed lines are the in-sample ML estimates.

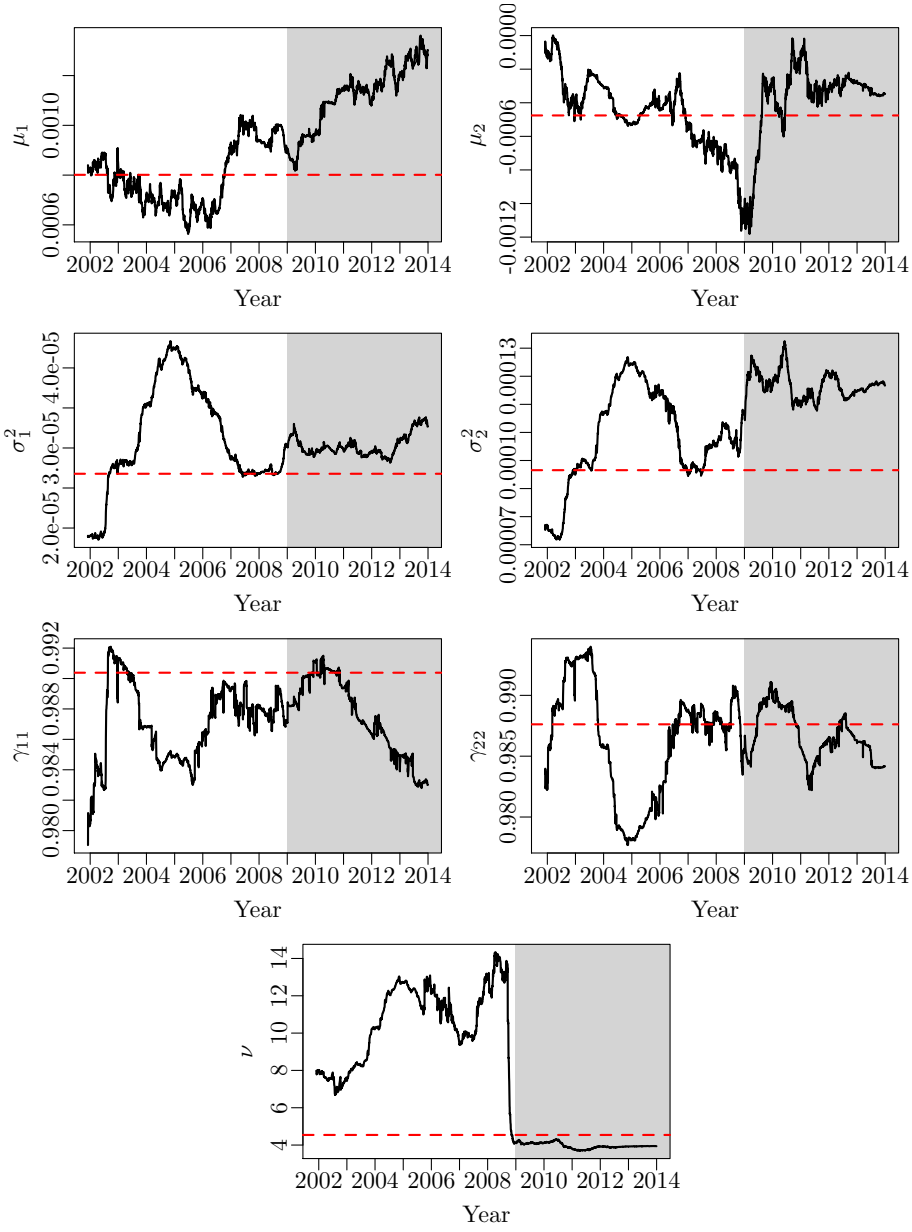


Figure 3.7: The parameters of a two-state HMM with a conditional t-distribution in the high-variance state estimated using a rolling window of 2000 trading days. The dashed lines are the in-sample ML estimates.

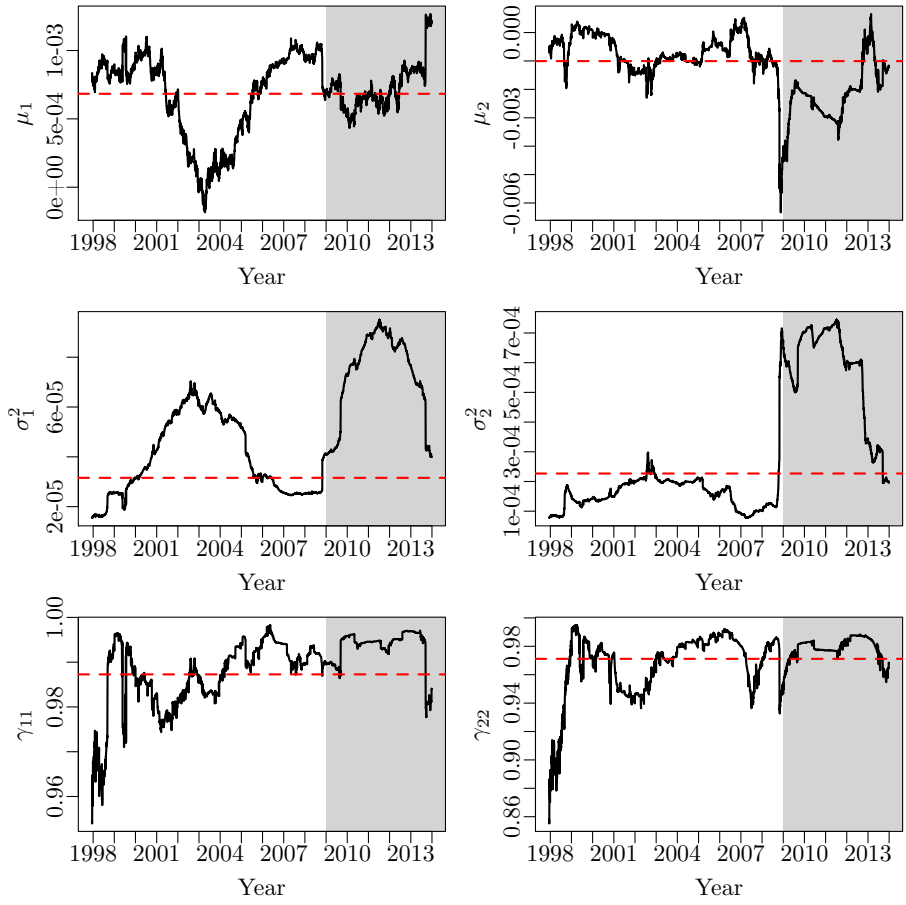


Figure 3.8: The parameters of a two-state Gaussian HMM estimated using a rolling window of 1000 trading days. The dashed lines are the in-sample ML estimates.

the EM algorithm can be applied to. This is obvious because the maximizing quantities in the M-step are derived based on the derivatives of the likelihood function. As a consequence, instead of resorting to a specific algorithm such as the EM algorithm, the likelihood can maximized using gradient-based optimization methods.

As already argued, the EM algorithm is preferred to direct numerical maximization of the likelihood function due to its larger robustness to initial values. The reason for exploring gradient-based methods is the flexibility to make the estimator recursive and subsequently adaptive. Using the EM algorithm, every observation is assumed to be of equal importance no matter how long the sample period is. This approach works well when the sample period is short and the underlying process is not time-sensitive. The time-varying behavior of the parameters uncovered in the previous section calls for an adaptive approach that assigns more weight to the most recent observations while keeping in mind the past patterns at a reduced confidence.

Recursive Estimation

The estimation of the parameters through a maximization of the conditional log-likelihood function can be done recursively using the estimator

$$\hat{\theta}_t = \arg\max_{\theta} \sum_{n=1}^{t} \log \Pr\left(X_n \left| \boldsymbol{X}^{(n-1)}, \theta \right.\right) = \arg\max_{\theta} \ell_t(\theta). \tag{3.43}$$

A second order Taylor expansion of $\ell_t(\theta)$ around $\hat{\theta}_{t-1}$ gives

$$\begin{aligned} \ell_t(\theta) = \ell_t\left(\hat{\theta}_{t-1}\right) &+ \nabla_\theta \ell_t\left(\hat{\theta}_{t-1}\right)\left(\theta - \hat{\theta}_{t-1}\right) \\ &+ \frac{1}{2}\left(\theta - \hat{\theta}_{t-1}\right)' \nabla_{\theta\theta} \ell_t\left(\hat{\theta}_{t-1}\right)\left(\theta - \hat{\theta}_{t-1}\right) + R_t. \end{aligned} \tag{3.44}$$

This expression is maximized with respect to θ assuming that $R_t \simeq 0$:

$$\nabla_\theta \ell_t(\theta) = \nabla_\theta \ell_t\left(\hat{\theta}_{t-1}\right) + \nabla_{\theta\theta} \ell_t\left(\hat{\theta}_{t-1}\right)\left(\theta - \hat{\theta}_{t-1}\right) = 0. \tag{3.45}$$

The solution is defined as the estimator

$$\hat{\theta}_t = \hat{\theta}_{t-1} - \left[\nabla_{\theta\theta} \ell_t\left(\hat{\theta}_{t-1}\right)\right]^{-1} \nabla_\theta \ell_t\left(\hat{\theta}_{t-1}\right). \tag{3.46}$$

It is typically assumed that near an optimum, the score function is approximately equal to the score function of the latest observation

$$\begin{aligned} \nabla_\theta \ell_t\left(\hat{\theta}_{t-1}\right) &= \nabla_\theta\left(\ell_{t-1}\left(\hat{\theta}_{t-1}\right) + \log \Pr\left(X_t \left| \boldsymbol{X}^{(t-1)}, \hat{\theta}_{t-1}\right.\right)\right) \\ &\approx \nabla_\theta \log \Pr\left(X_t \left| \boldsymbol{X}^{(t-1)}, \hat{\theta}_{t-1}\right.\right). \end{aligned} \tag{3.47}$$

The Hessian can be replaced by

$$\begin{aligned}\nabla_{\theta\theta}\ell_t\left(\hat{\theta}_{t-1}\right) &= \sum_{n=1}^{t}\nabla_{\theta\theta}\ell_n\left(\hat{\theta}_{n-1}\right)\\ &\simeq t\cdot \mathrm{E}\left[\nabla_{\theta\theta}\ell_t\left(\hat{\theta}_{t-1}\right)\right]\\ &= -t\cdot I_t\left(\hat{\theta}_{t-1}\right),\end{aligned} \tag{3.48}$$

where $I_t(\theta)$ is the Fisher information, leading to the Fisher scoring algorithm

$$\hat{\theta}_t \approx \hat{\theta}_{t-1} + \frac{1}{t}\left[I_t\left(\hat{\theta}_{t-1}\right)\right]^{-1}\nabla_\theta \log \Pr\left(X_t\left|\boldsymbol{X}^{(t-1)},\hat{\theta}_{t-1}\right.\right). \tag{3.49}$$

The approximation of the score function with the score function of the latest observation is not accurate in this particular case. The algorithm of Lystig and Hughes (2002), therefore, has to be run for each iteration, which increases the computational complexity significantly.

The Fisher information can be updated using the Fisher information identity $\mathrm{E}\left[\nabla_{\theta\theta}\ell_t\right] = \mathrm{E}\left[\nabla_\theta\ell_t\nabla_\theta\ell_t'\right]$:

$$\begin{aligned}I_t\left(\hat{\theta}\right) &= -\frac{1}{t}\sum_{n=1}^{t}\nabla_\theta\ell_n\left(\hat{\theta}\right)\nabla_\theta\ell_n\left(\hat{\theta}\right)'\\ &= -\frac{t-1}{t}\frac{1}{t-1}\left\{\sum_{n=1}^{t-1}\nabla_\theta\ell_n\left(\hat{\theta}\right)\nabla_\theta\ell_n\left(\hat{\theta}\right)'\right.\\ &\quad\left. +\ \nabla_\theta\log\Pr\left(X_t\left|\boldsymbol{X}^{(t-1)},\hat{\theta}\right.\right)\nabla_\theta\log\Pr\left(X_t\left|\boldsymbol{X}^{(t-1)},\hat{\theta}\right.\right)'\right\}\\ &= I_{t-1}\left(\hat{\theta}\right) + \frac{1}{t}\left[\nabla_\theta\log\Pr\left(X_t\left|\boldsymbol{X}^{(t-1)},\hat{\theta}\right.\right)\right.\\ &\quad\left.\cdot\nabla_\theta\log\Pr\left(X_t\left|\boldsymbol{X}^{(t-1)},\hat{\theta}\right.\right)' - I_{t-1}\left(\hat{\theta}\right)\right].\end{aligned} \tag{3.50}$$

This is simply calculating a mean recursively. The Fisher information can be updated using the matrix inversion lemma, since the estimator only makes use of the inverse of the Fisher information.[24] The diagonal elements of the inverse of the Fisher information provide uncertainties of the parameter estimates as a by-product of the algorithm.[25]

[24]The matrix inversion lemma is $(A+BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1}+DA^{-1}B\right)^{-1}DA^{-1}$, where A is an n-by-n matrix, B is n-by-k, C is k-by-k, and D is k-by-n.

[25]See section 3.1 for comments on the use of the Hessian to compute standard errors.

The Fisher scoring algorithm is a variant of the Newton-Raphson method. The algorithm can be very sensitive to the initial conditions, especially if the Fisher information is poorly estimated. The problem is that the algorithm takes very large steps initially, when t is small due to the $\frac{1}{t}$ term in (3.49). There are different ways to make sure that the initial steps are small enough. One possibility is to replace $\frac{1}{t}$ by $\frac{a}{A+t}$, where $0 < a < 1$ and/or A is some number, typically corresponding to 10% of the size of the total data set. Another possibility is to begin the recursion at some $t > 2$. Furthermore, it is necessary to apply a transformation to all constrained parameters for the estimator to converge.

Figure 3.9 shows the parameters of the two-state HMM with conditional normal distributions estimated using the recursive estimator. The estimation was started at $t = 500$ in order to avoid very large initial steps. The ML estimate based on the first 1000 observations was used as initial value with the Fisher information being initialized as one fifth of the observed information matrix. The burn-in period is very long as a result of the high persistence of the states. The dynamics of the model are very different when based upon less than 1000 observations as evidenced by the low values of γ_{11} and γ_{22} in the beginning. The impact of the GFC on the estimated parameters is illustrated in that the recursive estimates of the variance parameters do not converge to the ML estimate until the GFC.

Adaptive Estimation

The recursive estimator (3.43) can be made adaptive by introducing a weighting:

$$\hat{\theta}_t = \arg\max_{\theta} \sum_{n=1}^{t} w_n \log \Pr\left(X_n \left| \boldsymbol{X}^{(n-1)}, \theta \right.\right) = \arg\max_{\theta} \tilde{\ell}_t(\theta). \tag{3.51}$$

A popular choice is to use exponential weights $w_n = \lambda^{t-n}$, where the forgetting factor $0 < \lambda < 1$ (see e.g. Madsen 2008).

$\tilde{\ell}_t(\theta)$ can be Taylor expanded around $\hat{\theta}_{t-1}$ similarly to (3.44). Maximizing the second order Taylor expansion with respect to θ under the assumption that $R_t \simeq 0$ and defining the solution as the estimator $\hat{\theta}_t$ leads to

$$\hat{\theta}_t = \hat{\theta}_{t-1} - \left[\nabla_{\theta\theta} \tilde{\ell}_t\left(\hat{\theta}_{t-1}\right)\right]^{-1} \nabla_{\theta} \tilde{\ell}_t\left(\hat{\theta}_{t-1}\right). \tag{3.52}$$

It is typically assumed that near an optimum, the score function is approximately equal to the score function of the latest observation

$$\begin{aligned} \nabla_{\theta} \ell_t\left(\hat{\theta}_{t-1}\right) &= \nabla_{\theta}\left(\ell_{t-1}\left(\hat{\theta}_{t-1}\right) + \log \Pr\left(X_t \left| \boldsymbol{X}^{(t-1)}, \hat{\theta}_{t-1} \right.\right)\right) \\ &\approx \nabla_{\theta} \log \Pr\left(X_t \left| \boldsymbol{X}^{(t-1)}, \hat{\theta}_{t-1} \right.\right). \end{aligned} \tag{3.53}$$

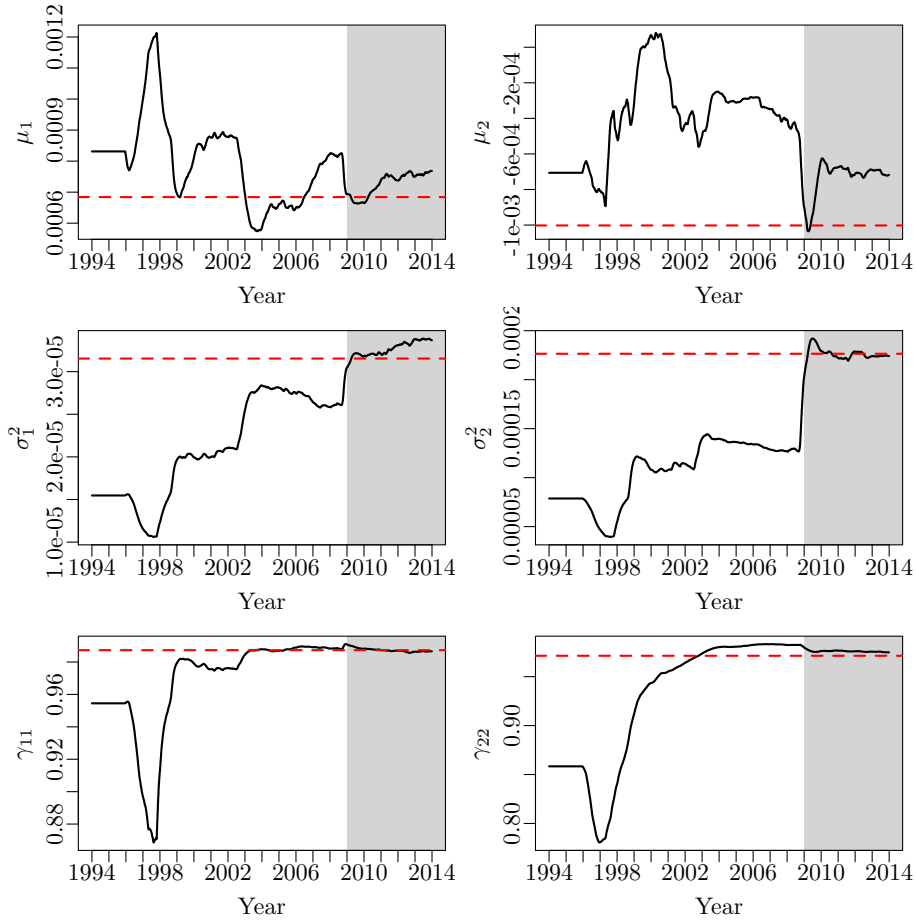


Figure 3.9: The parameters of a two-state Gaussian HMM estimated recursively. The dashed lines are the in-sample ML estimates.

The typical assumption, that near an optimum the score function is approximately equal to the score function of the latest observation

$$\begin{aligned}\nabla_\theta \tilde{\ell}_t\left(\hat{\theta}_{t-1}\right) &= \nabla_\theta\left(\lambda \nabla_\theta \tilde{\ell}_{t-1}\left(\hat{\theta}_{t-1}\right) + \lambda^0 \log \Pr\left(X_t \,\middle|\, \boldsymbol{X}^{(t-1)}, \hat{\theta}_{t-1}\right)\right) \\ &\approx \log \Pr\left(X_t \,\middle|\, \boldsymbol{X}^{(t-1)}, \hat{\theta}_{t-1}\right),\end{aligned} \tag{3.54}$$

is not accurate in this case. In order to compute the weighted score function, the algorithm of Lystig and Hughes (2002) has to be run for each iteration and the contribution of each observation has to be weighted.

The Hessian can be approximated by

$$\begin{aligned}\nabla_{\theta\theta} \tilde{\ell}_t\left(\hat{\theta}_{t-1}\right) &= \nabla_{\theta\theta} \sum_{n=1}^{t} \lambda^{t-n} \log \Pr\left(X_n \,\middle|\, \boldsymbol{X}^{(n-1)}, \hat{\theta}_{t-1}\right) \\ &= \sum_{n=1}^{t} \lambda^{t-n} \nabla_{\theta\theta} \log \Pr\left(X_n \,\middle|\, \boldsymbol{X}^{(n-1)}, \hat{\theta}_{t-1}\right) \\ &\approx \sum_{n=1}^{t} \lambda^{t-n}\left(-I_t\left(\hat{\theta}_{t-1}\right)\right).\end{aligned} \tag{3.55}$$

This leads to the recursive, adaptive estimator

$$\hat{\theta}_t \approx \hat{\theta}_{t-1} + \frac{1-\lambda}{1-\lambda^t}\left[I_t\left(\hat{\theta}_{t-1}\right)\right]^{-1} \nabla_\theta \tilde{\ell}_t\left(\hat{\theta}_{t-1}\right), \tag{3.56}$$

where the Fisher information can be updated recursively using (3.50).[26] The fraction $\frac{1-\lambda}{1-\lambda^t}$ can be replaced by $\frac{1}{\min(t,t_0)}$, where t_0 is a constant, in order to improve the clarity. The two fractions share the property that they decrease towards a constant when t increases. A forgetting factor of $\lambda = 0.998$, for example, corresponds to an effective window length of $t_0 = \frac{1}{1-0.998} = 500$.

Figure 3.10 shows the parameters of the two-state HMM with conditional normal distributions estimated using the adaptive estimator (3.56) with an effective window length of $t_0 = 500$. The dashed lines show the in-sample ML estimates. The initialization is similar to the recursive estimation. The adaptivity is most evident through the estimated variance parameters as the impact of the GFC is seen to die out through the out-of-sample period compared to the recursive estimates in figure 3.9 on page 57. $\lambda = 0.998$ is the lowest value of the forgetting factor that leads to reasonable estimates.

Using exponential forgetting the effective window length can be reduced compared to using fixed-length forgetting thereby allowing a faster adjustment to

[26] $\sum_{n=1}^{t} \lambda^{t-n} = \frac{1-\lambda^t}{1-\lambda}$.

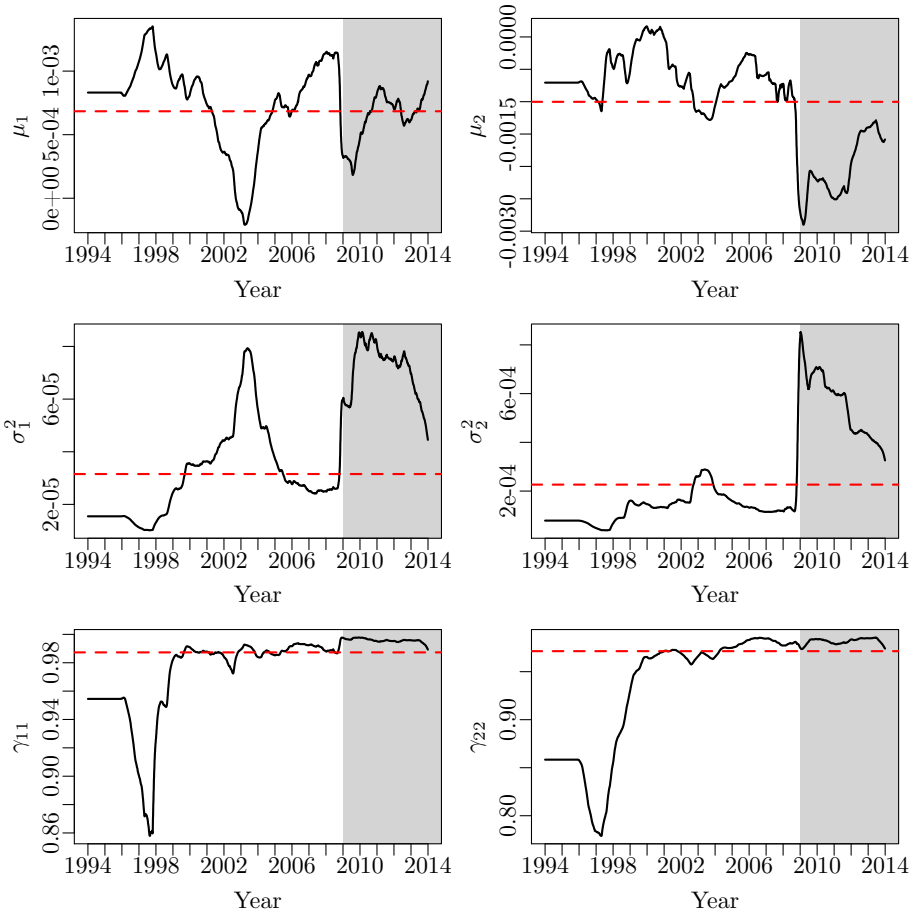


Figure 3.10: The parameters of the two-state HMM with conditional normal distributions estimated adaptively using a forgetting factor of $\lambda = 0.998$. The dashed lines are the in-sample ML estimates.

changes and a better reproduction of the current parameter values. Exponential forgetting is more meaningful as an observation is not just excluded from the estimation from one day to the next. In principle, all past observations are included in the estimation, but some are assigned an infinitesimally small weight.

CHAPTER 4

Strategic Asset Allocation

Strategic asset allocation is long-term in nature and based on long-term views of asset class performance. Dahlquist and Harvey (2001) distinguished between conditional and unconditional allocation based on how information is used to determine weight changes—unconditional implying no knowledge of the current regime. This chapter considers SAA in an unconditional framework to clearly distinguish it from the regime-based asset allocation that is the topic of chapter 5.

Based on theoretical arguments, Merton (1973) showed that the optimal allocation is affected by the possibility of uncertain changes in future investment opportunities—such as regime changes—even if the current regime cannot be identified. A risk averse investor will, to some degree, want to hedge against changes to the investment opportunity set. A better description of the behavior of financial markets, e.g. using a regime-switching model, will therefore be valuable, also for SAA.

With an empirical approach based on returns for eight asset classes including US and non-US stocks and bonds, high-yield bonds, EM equities, commodities, and cash equivalents, Chow et al. (1999) found that portfolios optimized based on the full-sample covariance matrix could be significantly suboptimal in periods of financial stress. Kritzman and Li (2010) later showed, as an extension of the work of Chow et al. (1999), that by considering the conditional behavior of assets it is possible to construct portfolios that are conditioned to better

withstand turbulent events and, at the same time, perform relatively well in all market conditions.

The performance of the optimized portfolios serve as a benchmark for the performance of the dynamic strategies that will be tested in the next chapter. The portfolios are optimized based on scenarios. The scenario generation is discussed in section 4.1 and the portfolios are optimized in section 4.2. Finally, the performance of the portfolios in sample and out of sample is examined in section 4.3.

4.1 Scenario Generation

A general way to describe risk is by using scenarios. A scenario being a realization of the future value of all parameters that influence the portfolio. A collection of scenarios should capture the range of variations that is likely to occur in these parameters including the impact of the shocks that are likely to come. These representations of uncertainty are the cornerstone of risk management. The purpose of generating scenarios is not to forecast what will happen.

There are three overall approaches to generating scenarios that should be mentioned. The first is to generate scenarios by sampling historical data through bootstrapping. The second approach is to generate scenarios through random sampling and then accept each scenario if its statistical moments match those of the observed data (see e.g. Høyland and Wallace 2001). The third approach, which is the approach that will be emphasized in this chapter, is to develop a theoretical model with parameters calibrated to historical data and then simulate the model to generate scenarios. Simple bootstrapping and moment matching are well suited to capture the empirical moments including the mean, covariance, skewness, and kurtosis, but they are unable to reproduce the autocorrelation. Autocorrelation can reduce risk estimates from a time series by inappropriately smoothing the volatility.

The frequency of the analyzed data has important implications for the measured risk throughout an investment horizon. Risk is typically measured as the probability of a given loss or the amount that can be lost with a given probability at the end of the investment horizon. Kritzman and Rich (2002) argued that the exposure to loss throughout an investment horizon, not only at its conclusion, is important to investors, as it is substantially greater than investors normally assume. Scenarios based on daily rather than monthly data lead to more reliable estimates of the within-horizon exposure to loss, as lower-frequency data smoothens the estimated risk as discussed in section 1.2.

It is a common belief that time diversification reduces risk and as a consequence long-term investors should have a higher proportion of risky assets in their portfolio than short-term investors (see e.g. Siegel 2007). It depends crucially on how the market behaves. Under a random walk, short and long-term investors

should have the same exposure to risk as there is no risk-reduction by staying in the market for an extended period of time, whereas mean reversion generates a higher proportion of risky assets for long-term investors. The probability of losses in the long run is never zero, regardless of the market behavior, thus it must also depend on the level of risk aversion. As noted by Kritzman and Rich (2002), the within-horizon probability of loss rises as the investment horizon expands even if the end-of-horizon probability of loss diminishes with time.

The time horizon has a significant impact on the optimal allocation in a conditional framework as shown by Guidolin and Timmermann (2007). In an unconditional framework, the time horizon is less important, as the initial distribution is assumed to be the stationary distribution. The scenarios generated in this chapter will have a one-year horizon. The one-year horizon reflects that the long-term views of asset class performance that SAA is based upon are typically updated once a year as noted by Dahlquist and Harvey (2001).

Regime-switching models are well suited to capture the time-varying behavior of risk premiums, variances, and correlation patterns. A multivariate model is chosen to secure a proper representation of the correlation patterns in financial returns. As noted by Sheikh and Qiao (2010), the correlations between asset classes tend to strengthen during periods of high market volatility and stress meaning that diversification might not materialize when it is needed the most. Inappropriately assuming linearity of correlations can lead to a significant underestimation of joint negative returns during a market downturn.

The parameter estimates for the fitted three-state multivariate Gaussian HMM are shown in table 4.1 together with approximate standard errors based on bootstrapping. The multivariate models were estimated using the R package `RHmm` due to Taramasco and Bauer (2013) that offers a faster but less comprehensive implementation of the EM algorithm. The number of states is selected based on the results in chapter 3. Based on model selection criteria it would be optimal with five states to capture the correlation patterns, but there is a strong preference for a less comprehensive model.

With three states the number of parameters is 35. The structure of the model is similar to the univariate three-state models estimated in chapter 3; there are two almost equally probable states and one recession state with a low unconditional probability. The correlation between stocks and bonds is significantly positive in the bull state and significantly negative in the two other states. The commodity index has a low correlation with both the stock and the bond index in the bull and bear state, but the size of the correlations increases significantly in the recession state. Three states seem to give a reasonable representation of the time-varying behavior of the mean values, variances, and correlations. It will remain a possibility for future work to examine the impact of the number of regimes on the SAA performance.

Table 4.1
Parameter estimates for the fitted three-state multivariate Gaussian HMM.

$\boldsymbol{\Gamma}$			$\boldsymbol{\mu} \times 10^4$			$\boldsymbol{\sigma}^2 \times 10^4$		
0.984	0.015	0.001	7.6	−0.4	7.0	0.26	0.03	1.20
	(0.004)	(0.001)	(1.3)	(0.4)	(2.6)	(0.01)	(0.00)	(0.07)
0.016	0.969	0.015	−2.5	1.2	3.8	0.93	0.03	1.90
(0.004)		(0.005)	(2.5)	(0.5)	(3.5)	(0.05)	(0.00)	(0.09)
0.008	0.143	0.850	−20.5	−0.4	−75.4	7.60	0.11	8.91
(0.014)	(0.039)		(21.8)	(2.7)	(24.0)	(0.93)	(0.01)	(1.13)

δ	π	$\rho_{\text{ACWI, GBI}}$	$\rho_{\text{ACWI, GSCI}}$	$\rho_{\text{GBI, GSCI}}$
1	0.49	0.31	0.03	−0.01
		(0.03)	(0.02)	(0.02)
0	0.46	−0.37	0.07	−0.06
		(0.03)	(0.03)	(0.03)
0	0.05	−0.42	0.43	−0.19
		(0.09)	(0.09)	(0.08)

Figure 4.2 shows ten of the 10,000 simulated scenarios (gray) together with the 5%, 25%, 50%, 75%, and 95%-quantile (black), and the maximum drawdown (MDD) scenario (red) for each of the indices. The MDD is the largest relative decline from a historical peak in the index value. It provides a measure of the within-horizon exposure to loss. It is not necessarily the same scenario that contains the MDD for all three indices.

With a multivariate model of the conditional behavior of the three indices it is not necessary to generate scenarios as the unconditional distribution is known. The optimal asset allocation could be inferred directly from the unconditional distribution. The reason for exploring scenario generation is the ease at which it can be implemented and generalized to more complex settings where the unconditional distribution is unknown.

4.2 The Mean–CVaR Model

In an asset allocation framework that incorporates non-normality, the variance is ineffective as the primary quantifier of portfolio risk because it equally penalizes desirable upside and undesirable downside movements. Non-normality can impact asset allocation as the downside risk associated with different asset classes is very different (see e.g. Sheikh and Qiao 2010).

The Mean–CVaR model is an alternative to the widely used mean–variance optimization introduced by Markowitz (1952). The Conditional Value-at-Risk at level $1-\alpha$ ($\text{CVaR}_{(1-\alpha)}$) is defined as the average loss in the $100\,(1-\alpha)\,\%$ worst scenarios. The literature on CVaR has its roots in the calculation of

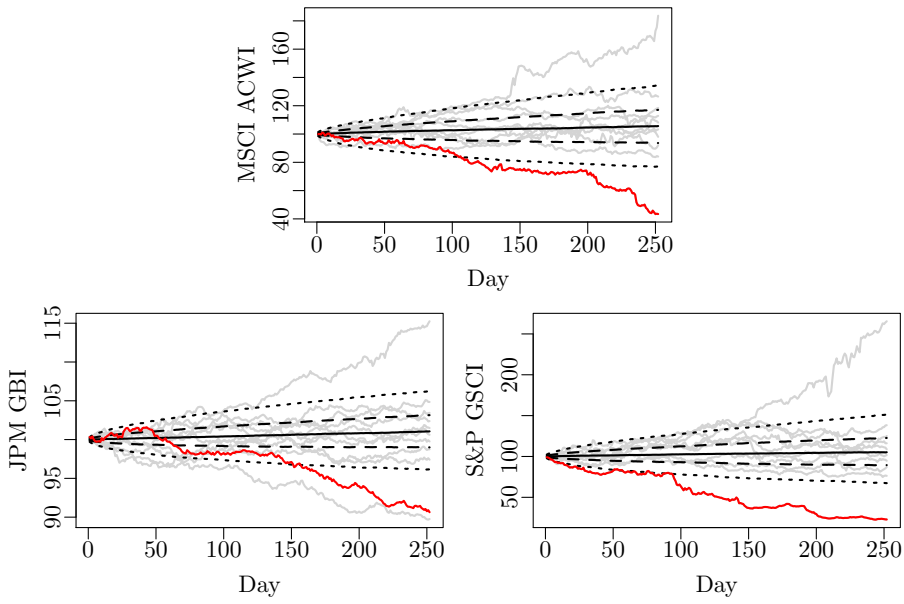


Figure 4.2: Ten of the 10,000 simulated scenarios (gray) together with the 5%, 25%, 50%, 75%, and 95%-quantile (black), and the maximum drawdown scenario (red).

expected shortfall in insurance where losses are the focus and, therefore, by definition positive. The α-quantile of the loss distribution is called the Value-at-Risk ($\text{VaR}_{(1-\alpha)}$). The CVaR, unlike the VaR, is a coherent risk measure (Artzner et al. 1999).

The objective of the Mean–CVaR model is to maximize the expected return (ER) and at the same time minimize the CVaR of the portfolio. The objective function is a weighted combination of risk and return. Each value of the weight λ in the interval $[0; 1]$ yields a portfolio that has the maximum ER for a given level of CVaR or, equivalently, the minimum CVaR for a given level of ER. The weight λ is a utility parameter measuring the risk aversion. The value $\lambda = 1$ yields the portfolio that has the lowest possible CVaR and $\lambda = 0$ yields the portfolio that has the highest possible ER. In practice, it is advisable to use $\lambda = 0.001$ (or some small number greater than zero) to still have some control of the CVaR when maximizing the ER and similarly $\lambda = 0.999$ instead of $\lambda = 1$ when minimizing the risk.

Model 4.1 outlines the optimization. Here $\overline{r}_i$ is the average return of asset i, r_i^l is the return of asset i in scenario l, x_i is the weight of asset i in the optimal portfolio, VaR is the Value-at-Risk, p^l is the probability of scenario l, and $y_+^l = \max\left[0, -\sum_i r_i^l x_i - VaR\right]$ is the maximum of zero and the loss exceeding the

Model 4.1: Mean–CVaR efficient portfolios

$$\text{Maximize} \quad (1-\lambda)\, ER - \lambda \cdot CVaR \tag{4.1}$$

$$\text{subject to} \quad ER = \sum_i \overline{r}_i x_i, \tag{4.2}$$

$$CVaR = VaR + \frac{\sum_l p^l y_+^l}{(1-\alpha)}, \tag{4.3}$$

$$y_+^l + \sum_i r_i^l x_i + VaR \geq 0, \quad \forall l, \tag{4.4}$$

$$y_+^l \geq 0, \quad \forall l, \tag{4.5}$$

$$\sum_i x_i = 1, \tag{4.6}$$

$$x_i \geq 0, \quad \forall i, \tag{4.7}$$

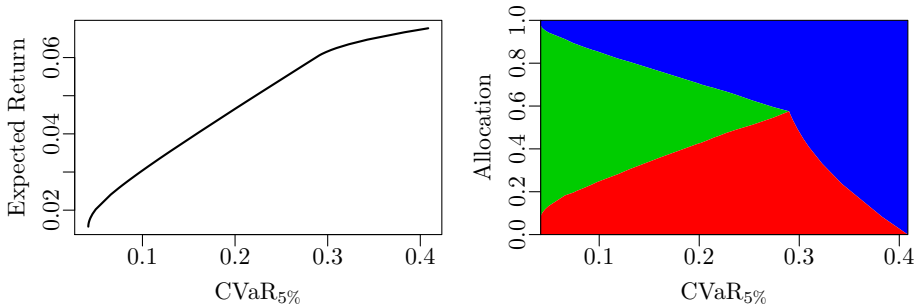
$$ER,\, CVaR,\, VaR \in \mathbb{R}. \tag{4.8}$$

VaR. The CVaR is formulated as the VaR plus the average of the losses exceeding the VaR in (4.3) in order to make the problem linear. An implementation of the Mean–CVaR model using the R package `lpSolve` (Berkelaar and others 2014) can be found in appendix A.3 on page 101.

Bertsimas et al. (2004) showed that mean–CVaR optimization might be preferable to the standard mean–variance optimization, even if the distribution of the asset returns is in fact normal or elliptical, because in this case it leads to an efficient and stable computation of the same optimal weights and does not require the often problematic estimation of large covariance matrices necessary under the mean–variance approach.

The efficient frontier and the allocations along it based on the 10,000 scenarios generated in section 4.1 are shown in figure 4.3. The portfolios on the efficient frontier are efficient in the sense that they have the highest possible ER for their level of risk. The steeper the slope of the frontier the better the ratio of ER to CVaR.

A portfolio consisting only of the S&P GSCI has the maximum $\text{CVaR}_{5\%}$ of 41% with a one-year horizon. The allocation to the MSCI ACWI is then gradually increased until it reaches about 58% and the portfolio has a $\text{CVaR}_{5\%}$ of 29%. From that point the allocation to the JPM GBI increases while the allocation to the two other indices decreases. The inclusion of the JPM GBI increases the steepness of the frontier as the JPM GBI had a better in-sample ratio of ER to CVaR even after the adjustment to equal Sharpe ratios described in section 2.6. The minimum $\text{CVaR}_{5\%}$ of 4% is not obtained by holding a portfolio



(a) Mean–CVaR frontier (b) Allocations

Figure 4.3: The efficient frontier and the allocations along it based on 10,000 scenarios with a one-year horizon.

consisting only of the JPM GBI as an 8% allocation to the MSCI ACWI and a 2% allocation to the S&P GSCI is beneficial for diversification purposes.

4.3 Results

In Sample

Table 4.4 shows the in-sample realized return (RR), standard deviation, maximum drawdown, and Sharpe ratio starting from three portfolios on or near the efficient frontier. The three portfolios realized almost identical annualized SRs in sample. Using the SR to summarize the performance is ambivalent after commenting on the variance being ineffective as a risk measure for non-elliptical distributions. The RR should, therefore, also be compared to the MDD, although the MDD cannot be annualized in the same way. It is expected that the MDD for the 15-year in-sample period exceeds the one-year $\text{CVaR}_{5\%}$ as losses can accumulate over time but the difference is substantial.

Table 4.4
The performance of a buy-and-hold strategy over the 15-year in-sample period starting from three different portfolios.

w_{ACWI}	w_{GBI}	w_{GSCI}	ER	$\text{CVaR}_{5\%}$	RR	SD	MDD	SR
0.25	0.6	0.15	0.031	0.10	0.025	0.08	0.37	0.33
0.5	0.25	0.25	0.047	0.21	0.037	0.11	0.48	0.32
0.45	0	0.55	0.062	0.30	0.044	0.15	0.58	0.30

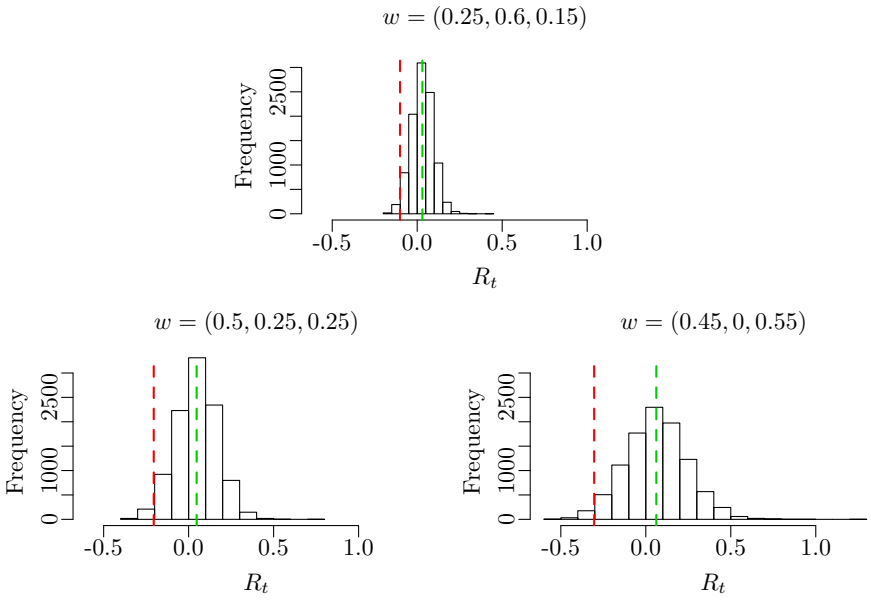


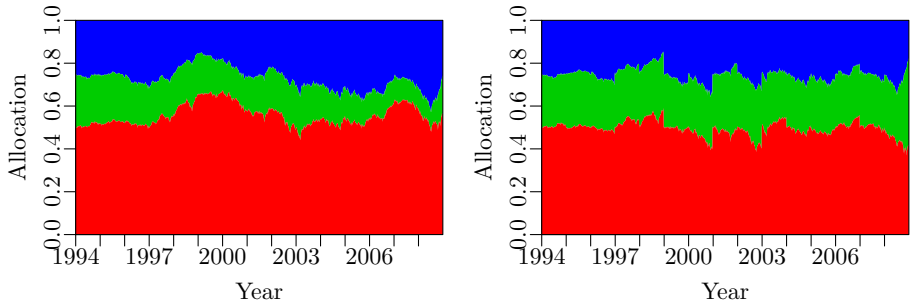
Figure 4.5: Frequency histograms of the portfolio returns based on the 10,000 scenarios. The dashed green line is the ER and the dashed red line is the $\text{CVaR}_{5\%}$ with opposite sign.

Frequency histograms of the portfolio returns for the three portfolios are shown in figure 4.5. The dashed green line shows the ER and the dashed red line is the $\text{CVaR}_{5\%}$ with opposite sign. The realized MDD exceeds the worst-case one-year scenario for the two portfolios that include the JPM GBI.

Rebalancing

The results presented thus far ignore the possibility of rebalancing. Almost immediately upon implementation, however, the portfolio weights become sub-optimal as changes in asset prices cause the portfolio to drift away from the optimal targets. The development in the relative weights starting from $w = (0.5, 0.25, 0.25)$ is depicted in figure 4.6a. The divergence from the initial weights is significant.

Rebalancing is the most basic and fundamental long-run investment strategy, and is naturally counter-cyclical. Rebalancing to constant weights is an example of a strategy that leans against the wind. It ensures that the risk level is kept constant when the current regime is assumed to be unknown. Crockett (2000) suggested that it may be helpful to think of risk as increasing during upswings, as financial imbalances build up, and materializing in recessions. If risk increases in



(a) No rebalancing (b) Annual rebalancing

Figure 4.6: The development in the portfolio weights in-sample with and without rebalancing.

upswings and materializes in recessions it stands to reason that defenses should be built up in upswings so as to be relied upon when the rough times arrive. By leaning against the wind, it could reduce the amplitude of the financial cycle.

In an idealized world without transaction costs investors would rebalance continually to the optimal weights. In the presence of transaction costs investors must trade off the cost of sub-optimality with the cost of restoring the optimal weights. Most investors employ heuristics that rebalance the portfolio as a function of the passage of time or the size of the misallocation (see e.g. Sun et al. 2006, Kritzman et al. 2009). Only the effect of periodic rebalancing on the optimal asset allocation will be considered here.

The development in the relative weights when starting from $w = (0.5, 0.25, 0.25)$ and rebalancing annually is shown in figure 4.6b. The divergence from the optimal weights within a calendar year can be significant. The result of monthly rebalancing is shown in table 4.7 after accounting for transactions costs of 0.1%. Similar results for annual rebalancing are shown in table 4.8.

Rebalancing is seen to improve the RR and, at the same time, reduce the MDD. The impact on the realized SD is smaller, but the overall improvement of the

Table 4.7
The performance of the three portfolios over the 15-year in-sample period after subtracting 0.1% transaction costs when rebalancing monthly.

w_{ACWI}	w_{GBI}	w_{GSCI}	ER	$\text{CVaR}_{5\%}$	RR	SD	MDD	SR
0.25	0.6	0.15	0.031	0.10	0.028	0.05	0.24	0.53
0.5	0.25	0.25	0.047	0.21	0.042	0.10	0.42	0.43
0.45	0	0.55	0.062	0.30	0.051	0.15	0.56	0.35

Table 4.8
The performance of the three portfolios over the 15-year in-sample period after subtracting 0.1% transaction costs when rebalancing annually.

w_{ACWI}	w_{GBI}	w_{GSCI}	ER	$\text{CVaR}_{5\%}$	RR	SD	MDD	SR
0.25	0.6	0.15	0.031	0.10	0.033	0.05	0.22	0.63
0.5	0.25	0.25	0.047	0.21	0.048	0.10	0.40	0.50
0.45	0	0.55	0.062	0.30	0.056	0.15	0.57	0.37

Table 4.9
The performance of the three portfolios over the five-year out-of-sample period after subtracting 0.1% transaction costs when rebalancing annually.

w_{ACWI}	w_{GBI}	w_{GSCI}	ER	$\text{CVaR}_{5\%}$	RR	SD	MDD	SR
0.25	0.6	0.15	0.031	0.10	0.063	0.06	0.10	0.99
0.5	0.25	0.25	0.047	0.21	0.094	0.13	0.19	0.75
0.45	0	0.55	0.062	0.30	0.090	0.18	0.25	0.49

SR is significant. The improvement is most significant for the portfolio that has a 60% weight on the JPM GBI and this portfolio now outperforms the two other portfolios in terms of SR. Monthly rebalancing does not appear to reduce the risk beyond what annual rebalancing does, but the higher transaction costs lead to a lower RR and, consequently, a lower SR. Based on these results annual rebalancing is preferred to monthly rebalancing. This would be the case even if transaction costs were only 0.01%.

The importance of allowing for rebalancing is evident from the significantly improved SR. Given the improvement in performance that can be achieved by rebalancing, a regime-based strategy would be expected to outperform a buy-and-hold strategy as rebalancing is a natural part of a dynamic strategy. The regime-based strategies should be compared to an SAA strategy with rebalancing in order to distinguish the contribution from being regime-based from rebalancing.

Out of Sample

Table 4.9 shows the out-of-sample performance for the three portfolios when rebalancing annually. The five-year out-of-sample period has been favorable for a strategy based on static weights, as it does not include any major setbacks. The RR by far exceeds the ER and, at the same time, the MDD is below the one-year $\text{CVaR}_{5\%}$. The out-of-sample SR exceeds the in-sample SR for all three portfolios by a good margin. Obviously, the in-sample SRs would have been higher for two of the portfolios without the adjustment of the JPM GBI discussed in section 2.6.

CHAPTER 5

Regime-Based Asset Allocation

If economic conditions are persistent and strongly linked to asset class performance, then a DAA strategy should add value over static weights. The purpose of a dynamic strategy is to take advantage of favorable economic regimes, as well as withstand adverse economic regimes and reduce potential drawdowns. A regime-based approach has the flexibility to adapt to changing economic conditions within a benchmark-based investment policy. It straddles a middle ground between strategic and tactical asset allocation (Sheikh and Sun 2012).

Previous studies have established a volatility reduction by dynamically shifting into cash in the most turbulent periods, with no adverse and sometimes even a positive effect on the realized return. All in all leading to significantly improved Sharpe ratios. Not all studies considered out-of-sample testing and the importance of transaction costs. This chapter examines whether the volatility reduction found in previous studies on dynamic asset allocation can be achieved when there is no risk-free asset, but possibilities for diversification by holding a portfolio of assets which may include short positions.

The asset classes considered are stocks, bonds, and commodities to keep it simple, yet complex enough for diversification possibilities to arise. The focus on modeling stock returns continues as portfolio risk is typically dominated by stock market risk. The implementation builds on the analysis in chapter 3 that showed a need for an adaptive approach. The performance of the regime-based strategies is compared to the rebalancing strategies considered in chapter 4.

5.1 Decoding the Hidden States

As discussed in section 3.1, there are two different ways of inferring the most likely sequence of hidden states; it can be done locally by determining the most likely state at each time t or globally by determining the most likely sequence of states using the Viterbi algorithm (Viterbi 1967). Intuitively, there is a preference for global decoding as local decoding can lead to impossible state sequences since it does not take the transition probabilities into account, but at the same time local decoding reduces the probability of misclassification.

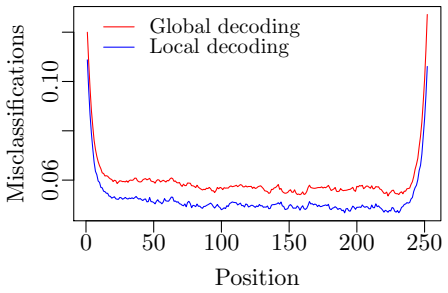


Figure 5.1: The proportion of misclassifications at each position.

Figure 5.1 shows the proportion of misclassifications at each position based on 50,000 simulated series of 252 observations from a three-state Gaussian HMM with parameters estimated from the MSCI ACWI log-returns. Local decoding gives the lowest number of misclassifications, but both techniques perform reasonably well for the larger part of the observations with average errors of 5.9% and 5.2%, respectively. The probability of misclassification increases with the number of states as Bulla et al. (2011) reported average errors of 3.5% and 3.2%, respectively, based on a similar study of a two-state Gaussian HMM.

The proportion of misclassifications increases strongly at the beginning and at the end of the series to more than 10%. This is problematic as the last position plays a central role for the state prediction at time $T + 1$ in an out-of-sample setting. The difference between local and global decoding is almost 2%-points at position 252.

An important feature in relation to the intended application is that global decoding reduces the number of transitions compared to smoothing. Table 5.2 shows the inferred number of in-sample transitions for the estimated models based on global and local decoding, respectively. Looking at the two-state models, it appears that conditional t-distributions lead to a significant reduction in the number of transitions, whereas the semi-Markov models yield too many transitions. The differences in the number of transitions are smaller between the three-state models. The number of transitions increases dramatically when going from three to four states for all the univariate models. The multivariate normal HMM leads to a large number of transitions compared to the univariate two and three-state models, but the increase when expanding the model to four states is smaller.

Table 5.2
The inferred number of transitions during the 15-year in-sample period ending in 2008 using global and local decoding, respectively.

Model	Global Decoding	Local Decoding
HMM_N (2)	33	61
HMM_{Nt} (2)	19	31
HMM_t (2)	17	25
HSMM_N (2)	221	559
HSMM_t (2)	67	315
MHMM_N (2)	53	85
HMM_N (3)	33	55
HMM_{Nt} (3)	35	53
HMM_t (3)	31	41
HSMM_N (3)	35	55
HSMM_t (3)	34	51
MHMM_N (3)	64	90
HMM_N (4)	361	525
HMM_{Nt} (4)	359	535
HSMM_N (4)	581	719
MHMM_N (4)	69	85
CTHMM_N (4)	344	528

There is no doubt that the Viterbi path is preferred over the smoothed path in a setting with perfect foresight with regard to the future returns as it leads to significantly fewer transitions. Based on the inferred number of transitions the four-state models and the multivariate models are ruled out as possible candidates for a successful strategy.

Figure 5.3 shows the decoded states in sample using the Viterbi algorithm for six of the models. Similar plots for all the estimated models can be found in figures C.1 and C.2 on page 108. There is a noticeable difference between the decoded states of the two-state models in the first row. The longer tails of the t-distribution increase the persistence of the bear state and lead to fewer transitions. The decoded states of the HMM with conditional t-distributions in both states are very similar to those shown for the HMM with one conditional t-distribution. The decoded states of the HSMMs do not look that different because there is a number of sojourns that are too short-lived to be visible.

The decoded states of the three-state models in the second row of figure 5.3 are more alike. The decoded states of the three-state HSMMs are similar to those shown for the HMMs. The multivariate models in the last row are seen to lead to more frequent state changes as a consequence of the lower persistence of the bear and recession state in these models.

It will be worth testing a two-state normal HMM and a two-state model with a conditional t-distribution in the high-variance state, as there is a considerable

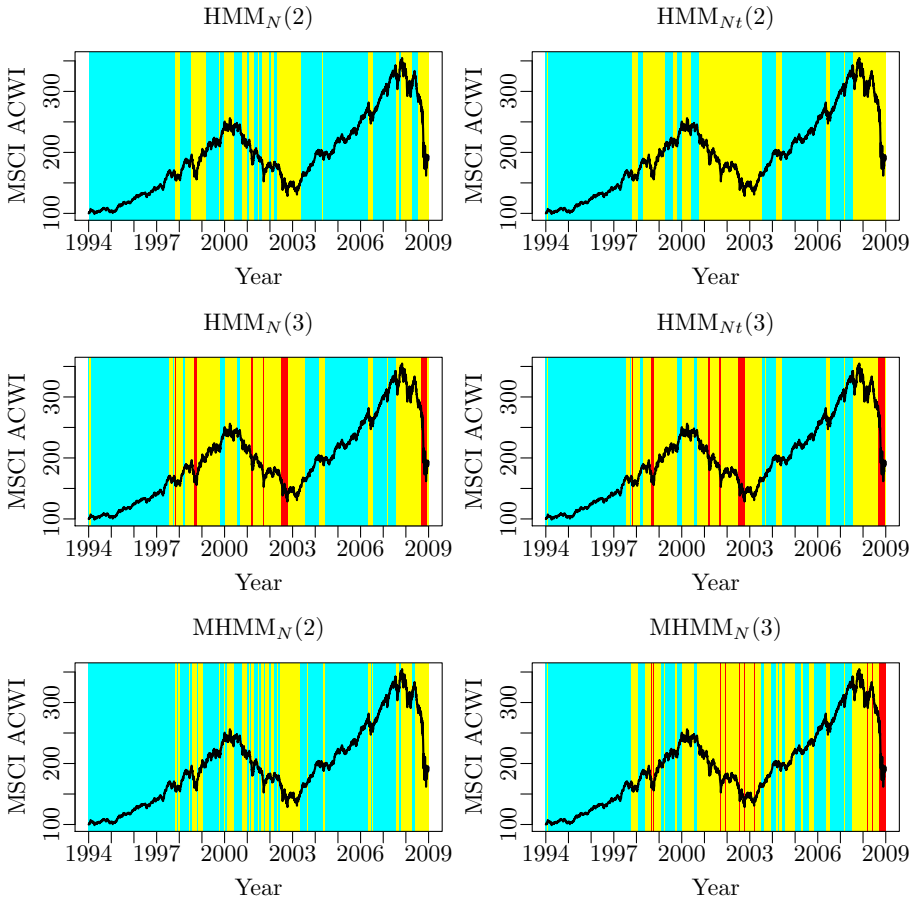


Figure 5.3: The MSCI ACWI and the decoded states using the Viterbi algorithm. Cyan is the bull state, yellow is the bear state, and red is the recession state.

difference between the decoded states of these two models. Since the decoded states of the three-state models are very similar, it will only be the three-state normal HMM that will be tested. Based on the results of the decoding, it is not worth testing the HSMMs, the four-state models, and the multivariate models.

The work of Bulla et al. (2011) relied on the Viterbi algorithm for decoding both in and out of sample. They applied a median filter in their out-of-sample test to reduce the number of unnecessary state shifts. That is, the predicted state at time $t + 1$ was given by the median of the last six one-day-ahead state predictions. The filtering procedure reduced the number of state changes by 50–65%. A future study should consider alternative filtering procedures such as total variation regularization (Rudin et al. 1992), a generalization of the median filter.

In an in-sample setting with perfect foresight the Viterbi path is the obvious choice as it is the most likely sequence of states with fewer transitions than the smoothed path. This is, however, not the case in an out-of-sample setting based on one-step predictions. The arguments for using the Viterbi algorithm rely on the knowledge of future returns. Out of sample, where only past returns are known, it makes more sense to consider the smoothing probabilities in order to obtain the best possible prediction (cf. figure 5.1 on page 72). This facilitates a filtering procedure based on the confidence in the predicted state rather than the number of times the state has been predicted with no minimum delay in responding to regime changes.

The confidence threshold in the probability filter plays the same role as the time lag in the median filter; a higher threshold value reduces the number of transitions at the same time as increasing the risk of delaying the reaction to regime changes. Figure 5.4 shows the predicted states of the MSCI ACWI based on the adaptively estimated two-state normal HMM analyzed in section 3.5. The dashed lines indicate the 92%-threshold used. If the state of day $t + 1$ is predicted to be different from the predicted state for day t and the confidence in the prediction exceeds the threshold, then the state is predicted to change, otherwise the state remains the same.

The predicted states are fairly similar to the Viterbi path shown in figure 5.3. With fewer short-lived sojourns and five less transitions the predicted states are almost more similar to the Viterbi path for the two-state model that has a conditional t-distribution in the high-variance state. The result is very sensitive to the selected threshold value as a lower threshold would increase the number of unnecessary transitions, whereas a higher threshold value would make it difficult to identity any regime changes.

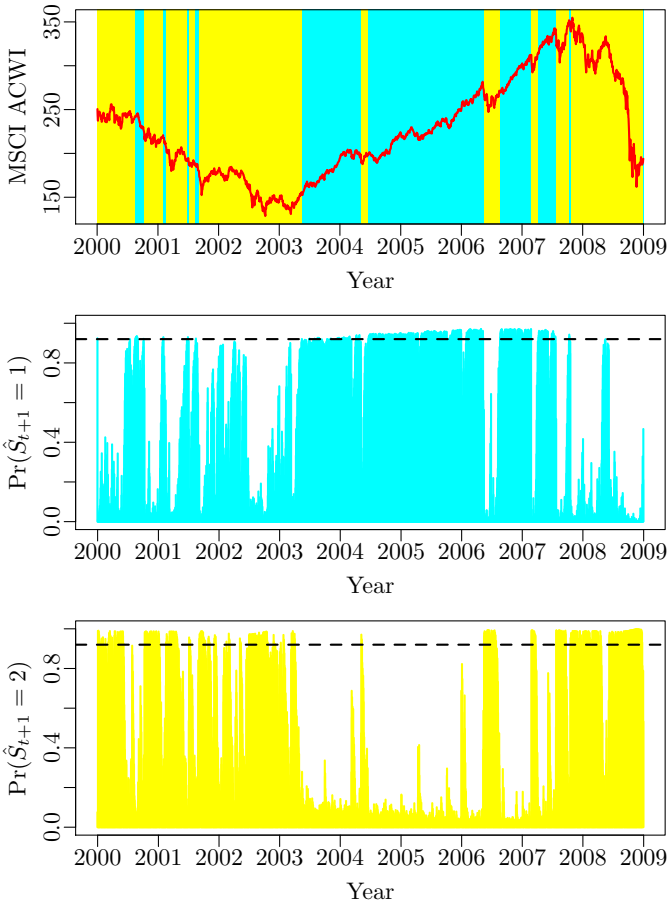


Figure 5.4: The MSCI ACWI and the predicted states based on an adaptively estimated two-state normal HMM and a probability filter with a 92%-threshold. Cyan is the bull state and yellow is the bear state.

Table 5.5
The performance of DAA strategies over the 15-year in-sample period after subtracting 0.1% transaction costs with no short positions allowed.

Model	w_{Bull}	w_{Bear}	$w_{\text{Recession}}$	RR	SD	MDD	SR
$\text{HMM}_N\,(2)$	$(1,0,0)$	$(0,1,0)$		0.127	0.08	0.15	1.67
$\text{HMM}_{Nt}\,(2)$	$(1,0,0)$	$(0,1,0)$		0.114	0.06	0.09	1.75
$\text{HMM}_N\,(3)$	$(1,0,0)$	$(0,0,1)$	$(0,1,0)$	0.138	0.16	0.46	0.87

Table 5.6
The annualized in-sample performance of the three indices.

Index	RR	SD	MDD	SR
MSCI ACWI	0.045	0.15	0.54	0.30
JPM GBI	0.009	0.03	0.12	0.29
S&P GSCI	0.044	0.22	0.68	0.20

5.2 In-Sample Testing

The in-sample testing will assume perfect foresight with regard to the future returns and be based on the Viterbi path. This assumption is unrealistic for practical purposes, nevertheless the performance of the strategies should be evaluated in this setting to facilitate a comparison to previous studies. A strategy that is not profitable under perfect foresight is not worth testing out of sample. The performance under perfect foresight provides an upper boundary for the potential of a strategy subject to the considered time period.

Table 5.5 shows the in-sample performance of a dynamic strategy with no short positions allowed based on three different models with transaction costs fixed at 0.1% per one-way trade. The three strategies clearly outperform the strategies based on rebalancing to static weights reported in table 4.8 on page 70. They also outperform the three indices as summarized in table 5.6.[27] The three-state strategy realizes a lower SR than the two-state strategies due to the risk associated with holding the S&P GSCI in the bear state.

The optimal allocation in each state was found by partitioning the in-sample returns based on the Viterbi path. For each state 5,000 one-year scenarios were bootstrapped from the returns belonging to that particular state. Simple bootstrapping is applicable as the returns are assumed to be independent within each state. Applying the Mean–CVaR model described in section 4.2 to the generated scenarios with a risk aversion parameter $\lambda = 0.15$, corresponding to the least risky of the three portfolios highlighted in chapter 4, led to state-

[27] The reported in-sample performance is based on geometric average returns, whereas the numbers reported in chapter 2 were based on arithmetic averages of the log-returns.

Table 5.7
The in-sample performance of DAA strategies after subtracting 0.1% transaction costs when a short position in the stock index is allowed.

Model	w_{Bull}	w_{Bear}	$w_{\text{Recession}}$	RR	SD	MDD	SR
$\text{HMM}_N\,(2)$	$(1, 0, 0)$	$(-1, 0, 0)$		0.197	0.15	0.24	1.30
$\text{HMM}_{Nt}\,(2)$	$(1, 0, 0)$	$(-1, 0, 0)$		0.182	0.15	0.24	1.20
$\text{HMM}_N\,(3)$	$(1, 0, 0)$	$(-0.32, 0, 0.68)$	$(-1, 0, 0)$	0.193	0.16	0.36	1.24

Table 5.8
The in-sample performance of DAA strategies after subtracting 0.1% transaction costs when the deviations from the benchmark allocation are constrained.

Model	w_{Bull}	w_{Bear}	$w_{\text{Recession}}$	RR	SD	MDD	SR
$\text{HMM}_N\,(2)$	$(0.6, 0.15, 0.25)$	$(0.4, 0.35, 0.25)$		0.069	0.09	0.35	0.77
$\text{HMM}_{Nt}\,(2)$	$(0.6, 0.15, 0.25)$	$(0.4, 0.35, 0.25)$		0.061	0.09	0.37	0.70
$\text{HMM}_N\,(3)$	$(0.6, 0.15, 0.25)$	$(0.4, 0.25, 0.35)$	$(0.4, 0.35, 0.25)$	0.065	0.10	0.37	0.68
Rebalancing	$(0.5, 0.25, 0.25)$	$(0.5, 0.25, 0.25)$	$(0.5, 0.25, 0.25)$	0.048	0.10	0.40	0.50

dependent allocations that only include one of the indices at a time with the exception of the three-state strategy, when a short position can be held.

Table 5.7 shows the in-sample results when a short position in the MSCI ACWI is allowed. A short position does not release any capital for other positions as it should be thought of as implemented using an inverse ETF. With a risk aversion parameter of 0.15, the optimal allocation in the bear/recession state is a short position in the MSCI ACWI rather than a long position in the JPM GBI, despite the strong in-sample performance of the JPM GBI. It is possible to achieve a higher SR by holding the JPM GBI, but a short position in the MSCI ACWI leads to a significant excess return. The three strategies realize similar SRs, but the three-state strategy has a higher MDD than the two-state strategies.

As noted by Kritzman et al. (2012), strategic investors that invest with a long time horizon typically face constraints on the size of possible deviations from their benchmark allocation. It is therefore relevant to establish the potential of a dynamic strategy when the possible allocations are less concentrated. Table 5.8 reports the results when the possible deviations from the benchmark allocation $w = (0.5, 0.25, 0.25)$ are limited to ± 0.1. Kritzman et al. (2012) would refer to this as a regime-dependent rather than regime-based strategy as opposed to the stocks–bonds and the long–short strategy.

The dynamic tilts lead to a higher realized return and a slightly lower SD and MDD compared to rebalancing to the benchmark allocation. The dynamic strategy based on the two-state Gaussian HMM has a higher SR than the rebalancing strategy as long as transaction costs are below 1.53% per one-way transaction (rebalancing also becomes more expensive).

Table 5.10
The performance of the DAA strategies over the five-year out-of-sample period after subtracting 0.1% transaction costs with a one-day delay in the portfolio adjustment.

Model	w_{Bull}	w_{Bear}	RR	SD	MDD	SR
HMM_N (2)	$(1, 0, 0)$	$(0, 1, 0)$	0.116	0.12	0.14	0.99
HMM_N (2)	$(1, 0, 0)$	$(-1, 0, 0)$	0.082	0.18	0.35	0.47
HMM_N (2)	$(0.6, 0.15, 0.25)$	$(0.4, 0.35, 0.25)$	0.098	0.13	0.17	0.78
Rebalancing	$(0.5, 0.25, 0.25)$	$(0.5, 0.25, 0.25)$	0.094	0.13	0.19	0.75

Table 5.11
The annualized out-of-sample performance of the three indices.

Index	RR	SD	MDD	SR
MSCI ACWI	0.15	0.17	0.26	0.85
JPM GBI	0.03	0.03	0.04	1.12
S&P GSCI	0.04	0.22	0.28	0.17

To summarize, the in-sample testing has shown that the strategies based on the two-state models outperform the three-state strategies. As discussed in section 5.1, an approach based on two rather than three states should reduce the probability of misclassification, as it is easier to distinguish between two than three states. It cannot be ruled out that other more profitable three-state strategies exist, but it will remain a possibility for future research to examine more advanced strategies. The out-of-sample testing will employ an adaptive estimation approach, which should offset the advantage of a conditional t-distribution, as discussed in section 3.5. The out-of-sample testing will, therefore, focus on the two-state Gaussian HMM.

5.3 Out-of-Sample Testing

Figure 5.9 shows the predicted states out of sample based on the adaptively estimated two-state Gaussian HMM when applying a probability filter with a 92%-threshold. The effective window length used in the estimation was 500 trading days. The result of the decoding looks very robust as the identified regimes do not appear to be very sensitive to the chosen threshold. It might have been optimal with a slightly lower value out of sample as the threshold influences how fast the regime changes are detected.

Table 5.10 reports the out-of-sample performance of the three different strategies based on the decoded states of the adaptively estimated two-state Gaussian HMM. Switching between the MSCI ACWI and the JPM GBI led to a higher return and SR than rebalancing to static weights (see table 4.9 on page 70). It also led to a higher SR and lower drawdown than the MSCI ACWI as shown in table 5.11. Switching between a long and a short position in the MSCI ACWI

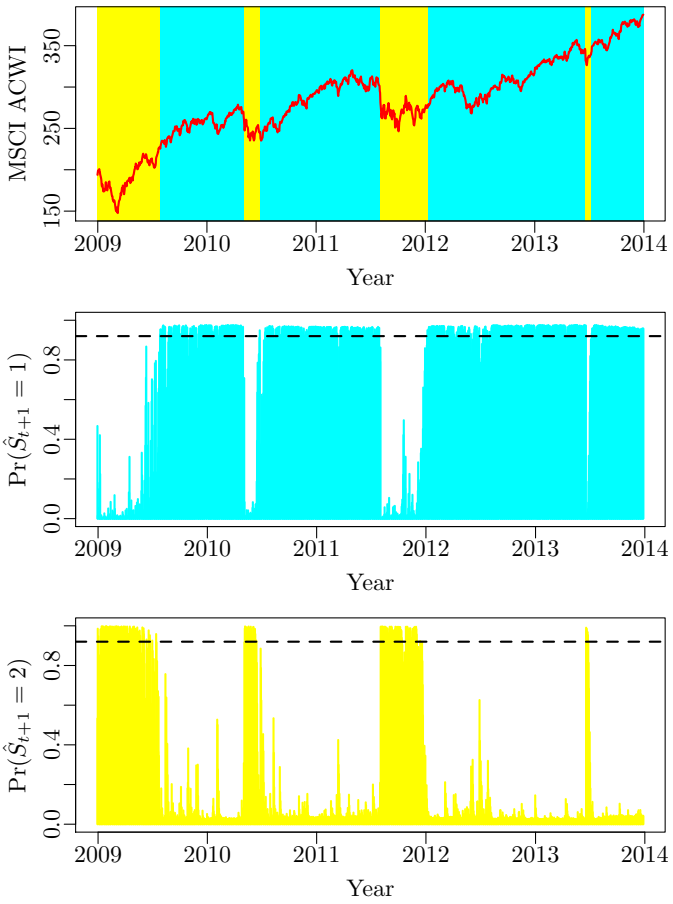


Figure 5.9: The MSCI ACWI and the predicted states based on the adaptively estimated two-state Gaussian HMM and a probability filter with a 92%-threshold. Cyan is the bull state and yellow is the bear state.

Table 5.12
The performance of the DAA strategies over the five-year out-of-sample period after subtracting 0.1% transaction costs with no delay in the portfolio adjustment.

Model	w_{Bull}	w_{Bear}	RR	SD	MDD	SR
HMM_N (2)	$(1, 0, 0)$	$(0, 1, 0)$	0.138	0.11	0.13	1.21
HMM_N (2)	$(1, 0, 0)$	$(-1, 0, 0)$	0.130	0.18	0.35	0.74
HMM_N (2)	$(0.6, 0.15, 0.25)$	$(0.4, 0.35, 0.25)$	0.103	0.12	0.17	0.83

Table 5.13
The performance of the DAA strategies over the five-year out-of-sample period after subtracting 0.1% transaction costs with a two-day delay in the portfolio adjustment.

Model	w_{Bull}	w_{Bear}	RR	SD	MDD	SR
HMM_N (2)	$(1, 0, 0)$	$(0, 1, 0)$	0.129	0.12	0.15	1.10
HMM_N (2)	$(1, 0, 0)$	$(-1, 0, 0)$	0.104	0.17	0.34	0.59
HMM_N (2)	$(0.6, 0.15, 0.25)$	$(0.4, 0.35, 0.25)$	0.101	0.13	0.17	0.80

has not been the best strategy in the five-year out-of-sample period, as there have been no major setbacks. The long–short strategy realized a lower return than the two most risky rebalancing strategies and at the same time a larger MDD.

The strategy with constrained deviations from the benchmark allocation realized a higher return with a slightly lower SD and MDD than the benchmark allocation in a period that did not favor a dynamic strategy. The dynamic strategy realized a higher SR and a lower MDD compared to the rebalancing strategy as long as the transaction costs did not exceed 0.6% per one-way trade.

The reported results are based on a one-day delay between the prediction of a state change and the portfolio adjustment. That is, if the predicted state of day $t + 1$ based on the closing price at day t is different from the state that the allocation at day t is based on and the confidence in the prediction is above the 92%-threshold, then the allocation is changed at the closing of day $t + 1$. If the reallocation could be implemented at the closing of day t, i.e. with no delay, then the RR would have been a lot higher, as shown in table 5.12, with the SD and MDD remaining largely unchanged.

It would also increase the RR if the reallocations were implemented with a two rather than one-day delay as shown in table 5.13. The relatively large differences in the RR depending on the implementation delay might be a coincidence, but they can also be a result of the significantly positive first-order autocorrelation that both the MSCI ACWI and the JPM GBI exhibit and the significantly negative second-order autocorrelation of the MSCI ACWI. A one-day delay causes the portfolio to miss out on the positive lag-one momentum, at the same time as being hurt by the day-two correction, which is avoided with a two-day delay.

Table 5.14
The performance of the DAA strategies over the five-year out-of-sample period after subtracting 0.1% transaction costs and with a one-day delay in the portfolio adjustment when the model is estimated using a rolling window of 1000 trading days.

Model	w_{Bull}	w_{Bear}	RR	SD	MDD	SR
HMM_N (2)	$(1, 0, 0)$	$(0, 1, 0)$	0.104	0.12	0.17	0.87
HMM_N (2)	$(1, 0, 0)$	$(-1, 0, 0)$	0.057	0.18	0.48	0.33
HMM_N (2)	$(0.6, 0.15, 0.25)$	$(0.4, 0.35, 0.25)$	0.098	0.13	0.17	0.77

The same strategies would have realized lower returns and higher MDDs if based on a two-state Gaussian HMM estimated using a rolling window of 1000 days. The results, when using a rolling window for the estimation, are shown in table 5.14 based on a one-day delay in the implementation. The performance of the strategy with dynamic tilts is almost the same. The rolling window estimation would have led to 13 allocation changes instead of 7 due to some very short-lived sojourns. Hence, more intelligent adaptivity does add value in particular in terms of reducing the tail-risk.

CHAPTER 6

Summary and Conclusion

The thesis' point of departure was that different economic regimes require different asset allocations. In the presence of time-varying investment opportunities, portfolio weights should be adjusted as new information arrives. Regime-switching models are well suited to capture the sudden changes of behavior and the phenomenon that the new dynamics of asset prices persist for several periods after a change. The asset classes considered were stocks, bonds, and commodities. Unlike the majority of previous studies on regime-switching strategies, there was no risk-free asset.

The data analysis confirmed well-known stylized facts of financial returns including skewness, leptokurtosis, volatility persistence, and time-varying correlations. The stylized behavior was most distinct for the stock index, while the sudden changes in the behavior of the government bond index seemed to be following interest rate changes. There was a focus on modeling the stock returns as portfolio risk is typically dominated by stock market risk.

The estimated models turned out to have a simple structure as only certain transitions were seen to occur. This offset the advantage of a continuous-time formulation in relation to the intended use. Even so, the implementation of a continuous-time version of the Baum–Welch algorithm was an important contribution of the project that will be useful for future work on continuous-time hidden Markov models. Three states seemed to be sufficient to reproduce the stylized facts with the most exceptional observations being allocated in a sepa-

rate state. A fourth state did not seem to improve the fit significantly compared to the added complexity and two states were not enough to adequately capture the stylized behavior. Although a conditional t-distribution in the high-variance state improved the ability of a two-state model to reproduce the excess kurtosis, it did not lead to a satisfactory fit to the long memory of the squared returns.

The estimation identified a call for an adaptive approach as the parameters of the estimated models were far from constant throughout the sample period. Time-varying parameters are likely part of the reason why many studies that employ Markov-switching models to financial returns obtain significantly better results in sample compared to out of sample. An adaptive estimation method based on exponential forgetting compensated for the need for a third state or a conditional t-distribution in the high-variance state to capture the most exceptional observations. Allowing for non-constant transition probabilities implies that the sojourn time distribution can take any shape.

A three-state multivariate Gaussian hidden Markov model was estimated and used to generate scenarios useful for a strategic asset allocation decision. Three efficient portfolios with different levels of risk and expected return were found based on mean–CVaR optimization. Rebalancing was shown to improve the return and at the same time reduce the risk—both the variance and the maximum drawdown—compared to buy-and-hold strategies in sample with an annual frequency being optimal. The three portfolios realized significantly higher returns and Sharpe ratios out of sample compared to in sample.

Dynamic asset allocation strategies were shown to add value over strategies based on rebalancing to static weights. The considered two-state models were found to outperform a three-state model in sample as foundation for dynamic asset allocation strategies. One strategy was based on switching between the stock and the bond index, another strategy was either long or short the stock index, and a third strategy included dynamic tilts of the benchmark allocation to stocks and bonds based on the prevailing regime while keeping a fixed allocation to commodities.

The outperformance in sample was very large, among other things, due to the major setbacks at the end of the period that favored a dynamic strategy. In an out-of-sample period that did not favor a dynamic strategy, the tested strategies based on an adaptively estimated two-state Gaussian hidden Markov model still outperformed a rebalancing strategy after accounting for transaction costs, assuming no knowledge of future returns, and with a realistic delay between the identification of a regime change and the portfolio adjustment.

6.1 Discussion

The thesis extends the work of Bulla et al. (2011) and addresses most of their suggestions for future research. The consideration of a continuous-time approach;

the model selection and testing of models with more than two states, different univariate and multivariate state-dependent distributions, and non-geometrical sojourn time distributions; the adaptive estimation method; and the asset allocation decision between three asset classes allowing for short positions in the stock index are new to the majority of the literature on applications of Markov-switching models in empirical finance as reviewed by Guidolin (2011).

Previous studies on regime-based asset allocation have primarily found that it is possible to obtain the same or a slightly higher return and at the same time reduce the portfolio risk by switching to a risk-free asset in the most turbulent periods. Holding a risk-free asset obviously yields a risk reduction, but leverage would be needed to significantly increase the return. This is consistent with the finding that safer assets offer higher risk-adjusted returns than riskier assets because of leverage aversion (Asness et al. 2012, Frazzini and Pedersen 2014).

This study has shown that it is possible to obtain a significant excess return with the same risk in terms of portfolio variance but a lower drawdown risk by taking advantage of the time-varying investment opportunities rather than rebalancing to static weights. The documented in-sample improvements of the Sharpe ratio exceed those of Bulla et al. (2011) by a good margin while the out-of-sample improvements are of a similar magnitude.

The need for an adaptive estimation approach to capture the time-varying behavior of the model parameters is consistent with previous studies that have found hidden semi-Markov models with non-geometric sojourn time distributions to be better at reproducing some of the stylized facts (Bulla and Bulla 2006) and that conditional t-distributions increased the persistence of the visited states and improved the ability to reproduce the excess kurtosis (Bulla 2011).

6.2 Future Work

Adaptive methods are a relatively undiscovered topic within empirical finance despite the strong evidence of non-stationary data-generating processes. The work done in this thesis has focused on capturing the time-varying behavior of the model parameters by assigning more weight to the most recent observations which is an extension of the fixed-length forgetting Bulla et al. (2011) applied in their study. A better description of the time-varying behavior of the parameters is an open route for future research that can be pursued in various ways. One would be to allow different forgetting factors for each parameter or consider more advanced state-dependent or time-dependent forgetting. Another way would be to formulate a model for the parameter changes.

The hierarchical regime-switching model studied by Ang and Bekaert (2004), where the parameters of the regime-switching process depended on an exogenous variable (a short-term interest rate) is an interesting extension of the frequently used Markov-switching models with seemingly great potential for future

work. Testing other possible predictors such as credit spreads or implied volatility is another possibility for future work. It is worth establishing whether an approach based on modeling regimes in assets returns outperforms approaches based on forecasting regimes in drivers of asset returns. It would also be relevant to do further research on whether an approach based on regime-switching models outperforms more simple approaches based on, for instance, the Mahalanobis distance, implied volatility, or simple moving average rules as suggested by Zakamulin (2014).

Inclusion of other major asset classes will remain a possibility for future work. This study already extends previous studies that only considered stocks and cash by including both a bond and a commodity index. Stocks, bonds, and commodities are likely to be three of the main common risk factors across a higher number of asset classes (see e.g. Ilmanen 2001, Bender et al. 2010). Rather than examining a large number of asset classes it would be useful to have a model describing the behavior of a smaller number of common risk factors.

A final suggestion for future work is to consider more smooth switching procedures for the portfolio adjustment for instance based on the confidence in the predicted state similar to the idea behind the Black–Litterman model (Black and Litterman 1992). Rather than doing a separate optimization of the model, the forgetting factor, the probability threshold, and the state-dependent allocations it could all be combined into one large optimization problem, although the magnitude of the problem is likely to make it very difficult to solve.

References

Ahlgren, P., Dahl, H., April 2010. Hidden Markov models—detecting regimes in financial time series. Nykredit Asset Management, 1–9.

Ammann, M., Verhofen, M., 2006. The effect of market regimes on style allocation. Financial Markets and Portfolio Management 20 (3), 309–337.

Ang, A., Bekaert, G., 2002. International asset allocation with regime shifts. Review of Financial Studies 15 (4), 1137–1188.

Ang, A., Bekaert, G., 2004. How regimes affect asset allocation. Financial Analysts Journal 60 (2), 86–99.

Ang, A., Timmermann, A., June 2011. Regime changes and financial markets. Working Paper 17182, National Bureau of Economic Research.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9 (3), 203–228.

Asai, M., McAleer, M., 2007. Non-trading day effects in asymmetric conditional and stochastic volatility models. Econometrics Journal 10 (1), 113–123.

Asness, C., Frazzini, A., Pedersen, L. H., 2012. Leverage aversion and risk parity. Financial Analysts Journal 68 (1), 47–59.

Bacci, S., Pandolfi, S., Pennoni, F., 2012. A comparison of some criteria for states selection in the latent markov model for longitudinal data. ArXiv e-prints.

Bauer, R., Haerden, R., Molenaar, R., 2004. Asset allocation in stable and unstable times. J. Investing 13 (3), 72–80.

Baum, L. E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. The Annals of Mathematical Statistics 41 (1), 164–171.

Bender, J., Briand, R., Nielsen, F., Stefek, D., 2010. Portfolio of risk premia: A new approach to diversification. J. Portfolio Management 36 (2), 17–25.

Berkelaar, M., others, 2014. lpSolve: Interface to Lp_solve v. 5.5 to solve linear/integer programs. R package version 5.6.8.

Bertsimas, D., Lauprete, G. J., Samarov, A., 2004. Shortfall as a risk measure: properties, optimization and applications. J. Economic Dynamics & Control 28 (7), 1353–1381.

Black, F., Litterman, R., 1992. Global portfolio optimization. Financial Analysts Journal 48 (5), 28–43.

Brinson, G. P., Hood, L. R., Beebower, G. L., 1986. Determinants of portfolio performance. Financial Analysts Journal 42 (4), 39–44.

Bulla, J., 2011. Hidden Markov models with t components. Increased persistence and other aspects. Quantitative Finance 11 (3), 459–475.

Bulla, J., Berzel, A., 2008. Computational issues in parameter estimation for stationary hidden Markov models. Computational Statistics 23 (1), 1–18.

Bulla, J., Bulla, I., 2006. Stylized facts of financial time series and hidden semi-Markov models. Computational Statistics and Data Analysis 51, 2192–2209.

Bulla, J., Bulla, I., Nenadic, O., 2010. hsmm—an R package for analyzing hidden semi-Markov models. Computational Statistics and Data Analysis 54, 611–619.

Bulla, J., Mergner, S., Bulla, I., Sesboüé, A., Chesneau, C., 2011. Markov-switching asset allocation: Do profitable strategies exist? J. Asset Management 12 (5), 310–321.

Bureau, A., Hughes, J. P., Shiboski, S. C., 2000. An S-Plus implementation of hidden Markov models in continuous time. J. Computational and Graphical Statistics 9 (4), 621–632.

Burns, A. F., Mitchell, W. C., 1946. Measuring Business Cycles. National Bureau of Economic Research.

Campbell, J. Y., 1998. Asset prices, consumption, and the business cycle. Working Paper 6485, National Bureau of Economic Research.

Campbell, J. Y., Lo, A. W., Mackinlay, A. C., 1997. The Econometrics of Financial Markets. Princeton University Press.

Cappé, O., Moulines, E., Rydén, T., 2005. Inference in hidden Markov models. Springer.

Carhart, M. M., 1997. On persistence in mutual fund performance. J. Finance 52 (1), 57–82.

Chow, G., Jacquier, E., Kritzman, M., Lowry, K., 1999. Optimal portfolios in good times and bad. Financial Analysts Journal 55 (3), 65–73.

Cochrane, J., 2005. Financial markets and the real economy. Working Paper 11193, National Bureau of Economic Research.

Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1, 223–236.

Crockett, A. D., September 2000. Marry the micro- and macro-prudential dimensions of financial stability. Speech at the Eleventh International Conference of Banking Supervisors.

Dacco, R., Satchell, S., 1999. Why do regime-switching models forecast so badly? J. Forecasting 18 (1), 1–16.

Dahlquist, M., Harvey, C. R., 2001. Global tactical asset allocation. Emerging Markets Quarterly 5 (1), 6–14.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statistical Society 39 (1), 1–38.

Dittmer, E., November 2008. Hidden Markov models with time-continuous output behavior. Ph.D. thesis, Freie University of Berlin.

Edwards, P., Manjrekar, C., 2014. Asset allocation survey. Mercer.

Erb, C. B., Harvey, C. R., 2006. The strategic and tactical value of commodity futures. Financial Analysts Journal 62 (2), 69–97.

Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. J. Finance 25 (2), 383–417.

Frazzini, A., Pedersen, L. H., 2014. Betting against beta. J. Financial Economics 111 (1), 1–25.

Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer.

Granger, C. W. J., Ding, Z., 1995a. Some properties of absolute return: An alternative measure of risk. Annales D'Economie Et Statistique (40), 67–92.

Granger, C. W. J., Ding, Z., 1995b. Stylized facts on the temporal and distributional properties of daily data from speculative markets. Unpublished paper, Department of Economics, University of California, San Diego.

Guidolin, M., 2011. Markov switching models in empirical finance. Working Paper 415, IGIER, Bocconi University.

Guidolin, M., Timmermann, A., 2007. Asset allocation under multivariate regime switching. J. Economic Dynamics & Control 31 (11), 3503–3544.

Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.

Hamilton, J. D., 1990. Analysis of time series subject to changes in regime. J. Econometrics 45 (1–2), 39–70.

Hannam, M., Lejonvarn, J., 2009. The commodity risk premium.
URL `http://www.markhannam.com/essays/essay2a.htm`

Hobolth, A., Jensen, J. L., 2011. Summary statistics for endpoint-conditioned continuous-time Markov chains. J. Applied Probability 48 (4), 901–1204.

Høyland, K., Wallace, S. W., February 2001. Generating scenario trees for multistage decision problems. Management Science 47 (2), 295–307.

Ibbotson, R. G., Kaplan, P. D., 2000. Does asset allocation policy explain 40, 90, or 100 percent of performance? Financial Analysts Journal 56 (1), 26–33.

Ilmanen, A., 2001. Expected Returns. Wiley.

Iversen, E. B., Møller, J. K., Morales, J. M., Madsen, H., 2013. Inhomogeneous Markov models for describing driving patterns. Technical report, Technical University of Denmark.

Jackson, C. H., 2011. Multi-state models for panel data: The msm package for R. J. Statistical Software 38 (8), 1–29.

Kritzman, M., Li, Y., 2010. Skulls, financial turbulence, and risk management. Financial Analysts Journal 66 (5), 30–41.

Kritzman, M., Myrgren, S., Page, S., 2009. Optimal rebalancing: A scalable solution. J. Investment Management 7 (1), 9–19.

Kritzman, M., Page, S., Turkington, D., 2012. Regime shifts: Implications for dynamic strategies. Financial Analysts Journal 68 (3), 22–39.

Kritzman, M., Rich, D., 2002. The mismeasurement of risk. Financial Analysts Journal 58 (3), 91–99.

Krstic, M., Kanellakopoulos, I., Kokotovic, P. V., 1995. Nonlinear and Adaptive Control Design. Wiley.

Lange, J. M., Minin, V. N., 2013. Fitting and interpreting continuous-time latent Markov models for panel data. Statistics in Medicine 32 (26), 4581–4595.

Lee, H.-Y., Wu, J.-L., 2001. Mean reversion of inflation rates: Evidence from 13 OECD countries. J. Macroeconomics 23 (3), 477–487.

Levich, R. M., 2001. International Financial Markets: Prices and Policies, 2nd Edition. McGraw–Hill.

Lystig, T. C., Hughes, J. P., 2002. Exact computation of the observed information matrix for hidden Markov models. J. Computational and Graphical Statistics 11 (3), 678–689.

Madsen, H., 2008. Time Series Analysis. Chapman & Hall.

Madsen, H., Thyregod, P., 2011. Introduction to General and Generalized Linear Models, 1st Edition. Chapman & Hall.

Mahalanobis, P. C., 1936. On the generalised distance in statistics. Proceedings of the National Institute of Sciences of India 2 (1), 49–55.

Markowitz, H. M., 1952. Portfolio selection. J. Finance 7 (1), 77–91.

Merton, R. C., 1973. An intertemporal capital asset pricing model. Econometrica 41 (5), 867–887.

Nystrup, P., Madsen, H., Lindström, E., 2014. Stylized facts of financial time series and hidden Markov models in continuous time. Submitted to Quantitative Finance.

Perrucci, S., Benaben, B., 2012. Inflation Sensitive Assets: Instruments and Strategies. Risk Books.

Pinson, P., Madsen, H., 2012. Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models. J. Forecasting 31 (4), 281–313.

Podgórski, K., Wallin, J., 2013. Maximizing leave-one-out likelihood for the location parameter of unbounded densities. Annals of the Institute of Statistical Mathematics, 1–20.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org/`

Rogalski, R. J., 1984. New findings regarding day-of-the-week returns over trading and non-trading periods. J. Finance 39 (5), 1603–1614.

Rudin, L. I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. Physica D 60 (1–4), 259–268.

Rydén, T., 2008. EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. Bayesian Analysis 3 (4), 659–688.

Rydén, T., Teräsvirta, T., Åsbrink, S., 1998. Stylized facts of daily return series and the hidden Markov model. J. Applied Econometrics 13 (3), 217–244.

Seo, B., Kim, D., 2012. Root selection in normal mixture models. Computational Statistics and Data Analysis 56 (8), 2454–2470.

Sheikh, A. Z., Qiao, H., 2010. Non-normality of market returns: A framework for asset allocation decision making. J. Alternative Investments 12 (3), 8–35.

Sheikh, A. Z., Sun, J., 2012. Regime change: Implications of macro shifts on asset class and portfolio performance. J. Investing 21 (3), 36–54.

Siegel, J. J., 2007. Stocks for the Long Run, 4th Edition. McGraw–Hill.

Silvestrov, D., Stenberg, F., 2004. A pricing process with stochastic volatility controlled by a semi-markov process. Communications in Statistics - Theory and Methods 33 (3), 591–608.

Sun, W., Fan, A., Chen, L.-W., Schouwenaars, T., Albota, M. A. ., 2006. Optimal rebalancing for institutional portfolios. J. Portfolio Management 32 (2), 33–43.

Taramasco, O., Bauer, S., 2013. RHmm: Hidden Markov Models simulations and estimations. R package version 2.0.3.

Timmermann, A., 2000. Moments of Markov switching models. J. Econometrics 96 (1), 75–111.

Varadhan, R., 2012. SQUAREM: Squared extrapolation methods for accelerating fixed-point iterations. R package version 2012.7–1.

Viterbi, A. J., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory 13 (2), 260–269.

Zakamulin, V., 2014. Dynamic asset allocation strategies based on unexpected volatility. J. Alternative Investments 16 (4), 37–50.

Zucchini, W., MacDonald, I. L., 2009. Hidden Markov Models for Time Series: An Introduction Using R, 2nd Edition. Chapman & Hall.

APPENDIX A

R-code

A.1 Discrete-time Baum–Welch Algorithm

```
EMNt = function(x, mu, sigma, Gamma, delta, nu, maxiter = 500,
                tol = 1e-8, print = FALSE){
  n = length(x)
  m = length(mu)
  mu.next = mu
  sigma.next = sigma
  Gamma.next = Gamma
  delta.next = delta
  nu.next = nu
  u = numeric(n)
  llk.prev = 0

  for (iter in 1:maxiter){
    lallprobs = matrix(NA, n, m)
    for (i in 1:n){
      for (j in 1:m){
        if (j < m)
          lallprobs[i, j] = dnorm(x[i], mean = mu.next[j],
                                  sd = sigma.next[j], log = T)
        else
          lallprobs[i, j] = dtmod(x[i], mu.next[j], sigma.next[j],
                                  nu.next, log = T)
      }
```

```
}
fb = lalphabeta(x, mu.next, sigma.next, Gamma.next,
                delta.next, nu.next)
la = fb$la
lb = fb$lb
c = max(la[, n])
llk = c + log(sum(exp(la[, n] - c)))
for (j in 1:m){
  for (k in 1:m)
    Gamma.next[j, k] = Gamma[j, k]*sum(exp(la[j, 1:(n-1)] +
                        lallprobs[2:n, k] + lb[k, 2:n] - llk))
  if (j < m){
    mu.next[j] = exp(la[j, ] + lb[j, ] - llk)%*%x/
      sum(exp(la[j, ] + lb[j, ] - llk))
    sigma.next[j] = sqrt((exp(la[j, ] + lb[j, ] - llk)*
                          (x - mu.next[j]))%*%(x - mu.next[j])/
                            sum(exp(la[j, ] + lb[j, ] - llk)))
  }
  else{
    u = (nu.next + 1)/(nu.next + (x - mu.next[j])*
                         (x - mu.next[j])/sigma.next[j]^2)
    w = (exp(la[j, ] + lb[j, ] - llk)*u)%*%x
    z = exp(la[j, ] + lb[j, ] - llk)%*%u
    mu.next[j] = w/z
    w = (exp(la[j, ] + lb[j, ] - llk)*u*(x - mu.next[j]))%*%
      (x - mu.next[j])
    sigma.next[j] = sqrt(w/z)
    w = exp(la[j, ] + lb[j, ] - llk)%*%(log(u) - u)
    z = sum(exp(la[j, ] + lb[j, ] - llk))
    estimator = function(nu, frac){
      -digamma(exp(nu)/2)+log(exp(nu)/2) + 1 + frac +
        digamma((exp(nu) + 1)/2) - log((exp(nu) + 1)/2)
    }
    nu.next = exp(uniroot(estimator, frac = w/z, lower =
                            log(0.01), upper = log(500))$root)
  }
}
Gamma.next = Gamma.next/apply(Gamma.next, 1, sum)
delta.next = exp(la[, 1] + lb[, 1] - llk)
delta.next = delta.next/sum(delta.next)
crit = llk - llk.prev
if(crit < tol){
  np = (m + 2)*m
  AIC = -2*(llk - np)
```

```
      BIC = -2*llk + np*log(n)
      return(list(mu = mu, sigma = sigma, Gamma = Gamma,
                  delta = delta, nu = nu, iterations = iter,
                  mllk = llk, AIC = AIC, BIC = BIC))
    }
    mu = mu.next
    sigma = sigma.next
    Gamma = Gamma.next
    delta = delta.next
    nu = nu.next
    llk.prev = llk
    if(print == T)
      print(paste('Iteration', iter, 'LogLik', round(llk, 4)))
  }

  print(paste('No convergence after', maxiter, 'iterations'))
  return(list(mu = mu, sigma = sigma, Gamma = Gamma,
              delta = delta, nu = nu, iterations = iter, mllk = llk))
}

lalphabeta = function(x, mu, sigma, Gamma, delta, nu){
  n = length(x)
  m = length(mu)
  lalpha = lbeta = matrix(NA, m, n)
  P = rep(NA, m)
  for (j in 1:m){
    if (j < m)
      P[j] = dnorm(x[1], mean = mu[j], sd = sigma[j])
    else
      P[j] = dtmod(x[1], mu[j], sigma[j], nu)
  }
  foo = delta*P
  sumfoo = sum(foo)
  lscale = log(sumfoo)
  foo = foo/sumfoo
  lalpha[, 1] = log(foo) + lscale
  for (i in 2:n){
    for (j in 1:m){
      if (j < m)
        P[j] = dnorm(x[i], mean = mu[j], sd = sigma[j])
      else
        P[j] = dtmod(x[i], mu[j], sigma[j], nu)
    }
    foo = foo%*%Gamma*P
```

```
    sumfoo = sum(foo)
    lscale = lscale + log(sumfoo)
    foo = foo/sumfoo
    lalpha[, i] = log(foo) + lscale
  }
  lbeta[, n]  = rep(0, m)
  foo = rep(1/m, m)
  lscale = log(m)
  for (i in (n - 1):1){
    for (j in 1:m){
      if (j < m)
        P[j] = dnorm(x[i + 1], mean = mu[j], sd = sigma[j])
      else
        P[j] = dtmod(x[i+1], mu[j], sigma[j], nu)
    }
    foo = Gamma%*%(P*foo)
    lbeta[, i] = log(foo) + lscale
    sumfoo = sum(foo)
    foo = foo/sumfoo
    lscale = lscale + log(sumfoo)
  }
  list(la = lalpha, lb = lbeta)
}

dtmod = function(x, mu = 0, sigma = 1, nu = 30, log = FALSE){
  den1 = try(sigma*beta(1/2, nu/2))
  num1 = try(nu^(-1/2))
  den2 = try(nu*sigma^2)
  num2 = try((x - mu)^2)
  dtmod = try(num1/den1*(1 + num2/den2)^(-1*(nu + 1)/2))
  if (log == TRUE)
    dtmod = log(dtmod)
  return (dtmod)
}
```

A.2 Continuous-time Baum–Welch Algorithm

```
CTEM = function(x, t, mu, sigma, Q, delta, m,
                maxiter = 100, tol = 1e-8){
  n              = length(x)
  mu.next        = mu
  sigma.next     = sigma
  Q.next         = Q
```

```
delta.next     = delta
llk.prev       = 0

for (iter in 1:maxiter){
  fb = lalphabeta(x, t, mu.next, sigma.next, Q.next,
                  delta.next, m)
  la = fb$la
  lb = fb$lb
  c = max(la[, n])
  llk = c + log(sum(exp(la[, n] - c)))
  R = numeric(m)
  N = matrix(0, m, m)
  tmp = eigen(Q.next)
  U = tmp$vectors
  Uinv = solve(U)
  lambda = tmp$values
  maxtau = max(diff(t))
  TiS = array(NA, c(maxtau, m, m, m))
  Trans = array(NA, c(maxtau, m, m, m, m))
  for (tau in 1:maxtau){
    for (a in 1:m){
      for (b in 1:m){
        TiS[tau, a, b, ] = TimeInStates(a, b, tau, U, Uinv,
                                        lambda)
        Trans[tau, a, b, , ] = Transitions(a, b, tau, U, Uinv,
                                           lambda, Q.next)
      }
    }
  }
  for (l in 2:n){
    tau =  t[l] - t[l - 1]
    for (a in 1:m){
      for (b in 1:m){
       Pab = dnorm(x[l], mean = mu.next[b], sd = sigma.next[b])*
          exp(la[a, l - 1] + lb[b, l] - llk)
        R = R + TiS[tau, a, b, ]*Pab
        N = N + Trans[tau, a, b, , ]*Pab
      }
    }
  }
  Q.next = N/R
  diag(Q.next) = -apply(Q.next, 1, sum)
  for (j in 1:m){
    mu.next[j] = exp(la[j, ] + lb[j, ] - llk)%*%x/
```

```
        sum(exp(la[j, ] + lb[j, ] - llk))
      sigma.next[j] = sqrt((exp(la[j, ] + lb[j, ] - llk)*
                            (x - mu.next[j]))%*%(x - mu.next[j])/
                             sum(exp(la[j, ] + lb[j, ] - llk)))
    }
    delta.next = exp(la[, 1] + lb[, 1] - llk)
    delta.next = delta.next/sum(delta.next)
    crit = llk - llk.prev
    if(crit < tol){
      np = 5*m - 1
      AIC = -2*(llk - np)
      BIC = -2*llk + np*log(n)
      return(list(mu = mu, sigma = sigma, Q = Q, delta = delta,
                  iterations = iter, mllk = llk, AIC = AIC, BIC = BIC))
    }
    mu = mu.next
    sigma = sigma.next
    Q = Q.next
    delta = delta.next
    llk.prev = llk
    print(paste('Iteration', iter, 'LogLik', round(llk, 4)))
  }
  print(paste('No convergence after', maxiter, 'iterations'))
  return(list(mu = mu, sigma = sigma, Q = Q, delta = delta,
              iterations = iter, mllk = llk))
}

lalphabeta = function(x, t, mu, sigma, Q, delta, m){
  n = length(x)
  lalpha = lbeta = matrix(NA, m, n)
  P = rep(NA, m)
  for (j in 1:m)
    P[j] = dnorm(x[1], mean = mu[j], sd = sigma[j])
  foo = delta * P
  sumfoo = sum(foo)
  lscale = log(sumfoo)
  foo = foo/sumfoo
  lalpha[, 1] = log(foo)+lscale
  tmp = eigen(Q)
  U = tmp$vectors
  Uinv = solve(U)
  lambda = tmp$values
  for (i in 2:n){
    for (j in 1:m)
```

```
      P[j] = dnorm(x[i], mean = mu[j], sd = sigma[j])
    tau = t[i] - t[i - 1]
    Gamma = U%*%diag(exp(tau*lambda))%*%Uinv
    foo = foo%*%Gamma*P
    sumfoo = sum(foo)
    lscale = lscale+log(sumfoo)
    foo = foo/sumfoo
    lalpha[, i] = log(foo)+lscale
  }
  lbeta[, n] = rep(0, m)
  foo = rep(1/m, m)
  lscale = log(m)
  for (i in (n - 1):1){
    for (j in 1:m)
      P[j] = dnorm(x[i + 1], mean = mu[j], sd = sigma[j])
    tau = t[i + 1] - t[i]
    Gamma = U%*%diag(exp(tau*lambda))%*%Uinv
    foo = Gamma%*%(P*foo)
    lbeta[, i] = log(foo)+lscale
    sumfoo = sum(foo)
    foo = foo/sumfoo
    lscale = lscale+log(sumfoo)
  }
  list(la = lalpha, lb = lbeta)
}

TimeInStates = function(a, b, tau, U, Uinv, lambda,
                        epsilon = 1e-7){
  m = nrow(U)
  J = matrix(0, m, m)
  for (i in 1:m){
    for (j in 1:m){
      if (abs(lambda[i] - lambda[j]) <= epsilon)
        J[i, j] = tau*exp(tau*lambda[i])
      else
        J[i, j] = (exp(tau*lambda[i]) - exp(tau*lambda[j]))/
          (lambda[i] - lambda[j])
    }
  }
  Times = numeric(m)
  for (k in 1:m){
    second_layer = 0
    for (i in 1:m){
      third_layer = 0
```

```
      for (j in 1:m)
        third_layer = third_layer + U[k, j]*Uinv[j, b]*J[i,j]
      second_layer = second_layer + U[a, i]*Uinv[i, k]*third_layer
    }
    Times[k] = second_layer
  }
  return(Times)
}

Transitions = function(a, b, tau, U, Uinv, lambda, Q,
                       epsilon = 1e-7){
  m = nrow(Q)
  J = matrix(0, m, m)
  for (i in 1:m){
    for (j in 1:m){
      if (abs(lambda[i] - lambda[j]) <= epsilon)
        J[i, j] = tau*exp(tau*lambda[i])
      else
        J[i, j] = (exp(tau*lambda[i]) - exp(tau*lambda[j]))/
          (lambda[i] - lambda[j])
    }
  }
  Transitions = matrix(0, m, m)
  for (k in 1:m){
    for (l in 1:m){
      second_layer = 0
      if (k != l){
        for (i in 1:m){
          third_layer = 0
          for (j in 1:m)
            third_layer = third_layer + U[l, j]*Uinv[j, b]*
              J[i, j]*Q[k, l]
          second_layer = second_layer + U[a, i]*Uinv[i, k]*
            third_layer
        }
        Transitions[k, l] = second_layer
      }
    }
  }
  return(Transitions)
}
```

A.3 Mean–CVaR optimization

```
MeanCVaR = function(scenarios, lambda = 0.5, alpha = 0.05,
                    print = FALSE){
  require('lpSolve')
  nassets = ncol(scenarios)
  nscenarios = nrow(scenarios)
  nlambda = length(lambda)
  allocation = matrix(NA, nlambda, nassets)

  ER = CVaR = numeric(nlambda)
  rhat = as.vector(apply(scenarios, 2, mean))
  A = rbind(c(rep(1, nassets), numeric(nscenarios + 1)),
            cbind(scenarios, 1, diag(nscenarios)))
  b = c(1, numeric(nscenarios))

  for (i in 1:nlambda){
    c = c((1 - lambda[i])*rhat, -lambda[i],
          -rep(lambda[i]/(alpha*nscenarios), nscenarios))
    sol = lp('max', c, A, c('=', rep('>=', nscenarios)), b)
    allocation[i, ] = sol$solution[1:nassets]
    ER[i] = allocation[i, ]%*%rhat
    CVaR[i] = ((1 - lambda[i])*ER[i] - sol$objval)/lambda[i]
    if(print)
      print(i)
  }

  list(lambda = lambda, CVaR = CVaR, ER = ER, Allocation = allocation)
}
```

APPENDIX B

Parameter Estimates

Table B.1
Parameter estimates for m-state HMMs with conditional normal distributions fitted to the daily in-sample log-returns of the MSCI ACWI.

m	$\mathbf{\Gamma}$				$\mu \times 10^4$	$\sigma \times 10^2$	δ	π
2	0.987	0.013			6.8	0.56	1	0.69
	0.029	0.971			−10.0	1.50	0	0.31
3	0.988	0.012	0		8.8	0.48	1	0.48
	0.012	0.981	0.006		−1.0	0.94	0	0.45
	0	0.043	0.957		−32.5	2.46	0	0.07
4	0.983	0.017	0	0	8.9	0.47	1	0.49
	0	0.651	0.331	0.018	−69.1	0.71	0	0.19
	0.033	0.240	0.727	0	50.7	0.74	0	0.25
	0	0	0.053	0.947	−27.9	2.48	0	0.07

Table B.2
Parameter estimates for m-state HMMs with conditional normal and t-distributions fitted to the daily in-sample log-returns of the MSCI ACWI.

m	$\mathbf{\Gamma}$				$\mu \times 10^4$	$\sigma \times 10^2$	ν	δ	π
2	0.990	0.010			8.0	0.52		1	0.56
	0.012	0.988			−4.8	0.98	4.5	0	0.44
3	0.988	0.012	0		8.7	0.48		1	0.48
	0.012	0.981	0.007		−0.8	0.93		0	0.45
	0	0.039	0.961		−29.3	1.97	6.3	0	0.07
4	0.983	0.017	0	0	9.0	0.47		1	0.49
	0	0.651	0.331	0.018	−68.3	0.70		0	0.19
	0.033	0.241	0.726	0	50.6	0.73		0	0.25
	0	0	0.046	0.954	−24.0	2.00	6.6	0	0.07

Table B.3
Parameter estimates for m-state HMMs with conditional t-distributions fitted to the daily in-sample log-returns of the MSCI ACWI.

m	$\mathbf{\Gamma}$			$\mu \times 10^4$	$\sigma \times 10^2$	ν	δ	π
2	0.992	0.008		8.4	0.49	14.0	1	0.56
	0.010	0.990		−4.8	0.98	4.5	0	0.44
3	0.991	0.009	0	8.9	0.46	15.5	1	0.48
	0.009	0.984	0.006	−0.8	0.92	119.8	0	0.44
	0	0.038	0.962	−29.9	1.98	6.4	0	0.08

Table B.4
Parameter estimates for m-state HSMMs with conditional normal distributions fitted to the daily in-sample log-returns of the MSCI ACWI. p and r are the parameters of the negative binomial sojourn time distribution.

m	$\mathbf{\Gamma}$	$1-p$	$r \times 10$	$\mu \times 10^4$	$\sigma \times 10^2$	δ
2	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	0.995	0.37	6.7	0.55	1
		0.964	0.81	−12.1	1.60	0
3	$\begin{pmatrix} 0 & 1 & 0 \\ 0.817 & 0 & 0.183 \\ 0 & 1 & 0 \end{pmatrix}$	0.997	1.21	8.6	0.49	1
		0.941	14.59	−1.3	0.95	0
		0.972	5.59	−32.8	2.47	0
4	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0.934 & 0.066 \\ 0.630 & 0.370 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	0.995	0.64	8.9	0.48	1
		0.350	39.74	−68.1	0.72	0
		0.577	16.71	59.7	0.74	0
		0.974	4.07	−28.3	2.47	0

Table B.5
Parameter estimates for m-state HSMMs with conditional t-distributions fitted to the daily in-sample log-returns of the MSCI ACWI. p and r are the parameters of the negative binomial sojourn time distribution and ν is the degrees of freedom for the conditional t-distributions.

m	$\mathbf{\Gamma}$	$1-p$	$r \times 10$	$\mu \times 10^4$	$\sigma \times 10^2$	ν	δ
2	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	0.996	0.31	8.0	0.50	21.7	1
		0.985	0.75	−6.7	1.08	5.1	0
3	$\begin{pmatrix} 0 & 1 & 0 \\ 0.79 & 0 & 0.21 \\ 0 & 1 & 0 \end{pmatrix}$	0.997	1.25	8.9	0.46	17.5	1
		0.940	16.93	−1.0	0.94	1382.0	0
		0.974	5.98	−29.7	1.99	6.4	0

Table B.6
Parameter estimates for a four-state CTHMM with conditional normal distributions fitted to the daily in-sample log-returns of the MSCI ACWI.

$\mathbf{Q}$				$\mu \times 10^4$	$\sigma \times 10^2$	δ	π
−0.016	0.016	0	0	8.8	0.47	1	0.48
0	−0.492	0.474	0.018	−67.5	0.71	0	0.20
0.030	0.353	−0.383	0	49.7	0.74	0	0.25
0	0	0.053	−0.053	−27.8	2.47	0	0.07

Table B.7
Parameter estimates for m-state HMMs with conditional multivariate normal distributions fitted to the daily in-sample log-returns of the three indices.

m	$\boldsymbol{\Gamma}$				$\boldsymbol{\mu} \times 10^4$			$\boldsymbol{\sigma} \times 10^2$		
2	0.981	0.019			7.2	−0.2	6.0	0.57	0.17	1.16
	0.045	0.955			−11.6	1.7	−9.0	1.53	0.21	1.82

δ	π	$\rho_{\text{ACWI, GBI}}$	$\rho_{\text{ACWI, GSCI}}$	$\rho_{\text{GBI, GSCI}}$
1	0.71	0.16	0.03	−0.01
0	0.29	−0.45	0.26	−0.12

m	$\boldsymbol{\Gamma}$				$\boldsymbol{\mu} \times 10^4$			$\boldsymbol{\sigma} \times 10^2$		
3	0.984	0.015	0.001		7.6	−0.4	7.0	0.51	0.17	1.10
	0.016	0.969	0.015		−2.5	1.2	3.8	0.96	0.17	1.38
	0.008	0.143	0.850		−20.5	−0.4	−75.4	2.76	0.33	2.98

δ	π	$\rho_{\text{ACWI, GBI}}$	$\rho_{\text{ACWI, GSCI}}$	$\rho_{\text{GBI, GSCI}}$
1	0.49	0.31	0.03	−0.01
0	0.46	−0.37	0.07	−0.06
0	0.05	−0.42	0.43	−0.19

m	$\boldsymbol{\Gamma}$				$\boldsymbol{\mu} \times 10^4$			$\boldsymbol{\sigma} \times 10^2$		
4	0.988	0.004	0.009	0	9.2	−0.3	3.6	0.52	0.14	1.46
	0.003	0.988	0.008	0.001	5.1	−0.3	7.5	0.54	0.20	0.84
	0.007	0.005	0.973	0.015	−4.3	1.4	5.0	1.04	0.18	1.31
	0	0.008	0.135	0.857	−17.2	−1.1	−82.3	2.84	0.34	3.10

δ	π	$\rho_{\text{ACWI, GBI}}$	$\rho_{\text{ACWI, GSCI}}$	$\rho_{\text{GBI, GSCI}}$
0	0.29	0.02	0.04	0.02
1	0.27	0.39	−0.00	−0.03
0	0.39	−0.39	0.10	−0.09
0	0.04	−0.42	0.44	−0.19

APPENDIX C

Additional Figures and Tables

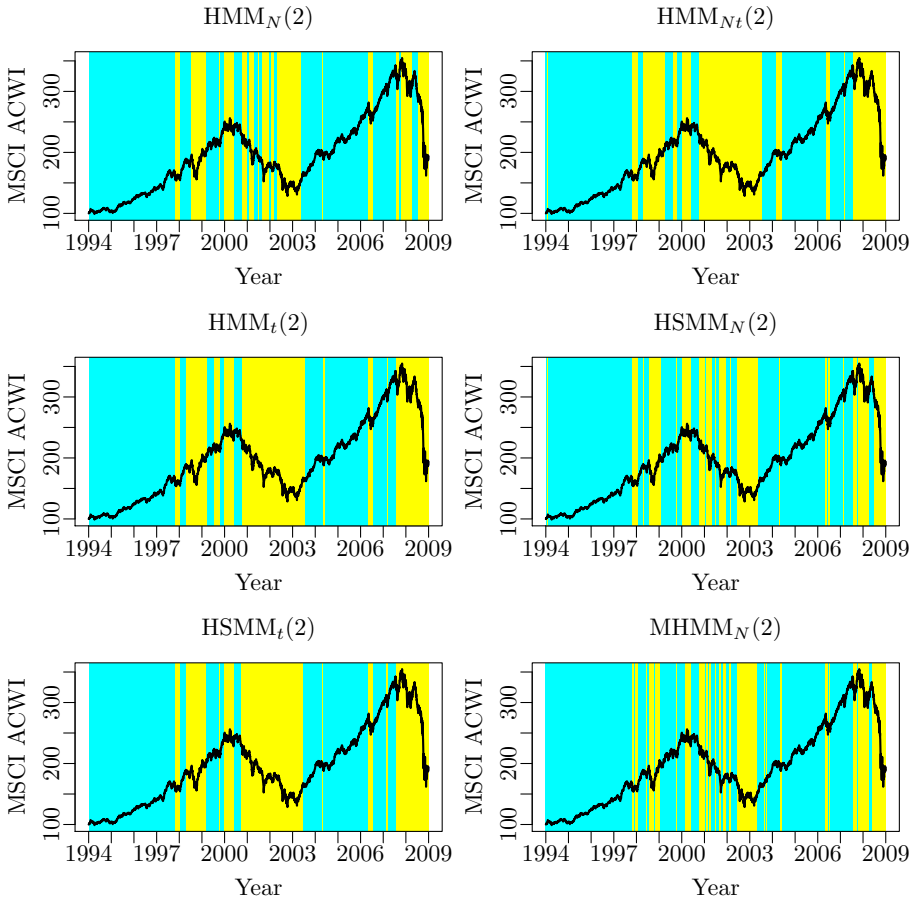


Figure C.1: The MSCI ACWI and the decoded states of the two-state models using the Viterbi algorithm. Cyan is the bull state and yellow is the bear state.

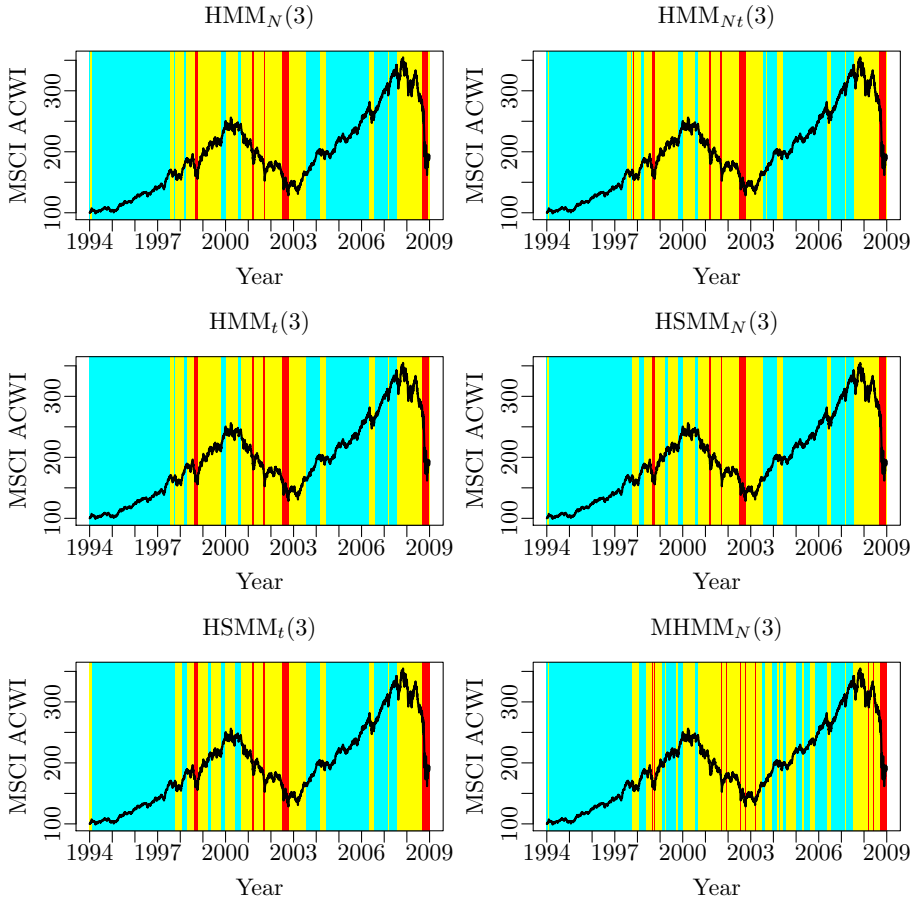


Figure C.2: The MSCI ACWI and the decoded states of the three-state models using the Viterbi algorithm. Cyan is the bull state, yellow is the bear state, and red is the recession state.

{2.7182818284}ϑφερτυθιοπσδφγηξκλ