UNIVERSITY OF CALIFORNIA

Los Angeles

Covariance Estimation with Markov-Switching Generalized Autoregressive Conditional Heteroskedasticity Models with Applications to Portfolio Management

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Tristan Gardner Wisner

2017

ABSTRACT OF THE THESIS

Covariance Estimation with Markov-Switching
Generalized Autoregressive Conditional Heteroskedasticity Models
with Applications to Portfolio Management

by

Tristan Gardner Wisner
Master of Science in Statistics
University of California, Los Angeles, 2017
Professor Ying Nian Wu, Chair

The objective of this paper is to implement and test the multivariate regime-switching GARCH model as a potential improvement on traditional methods for estimating the covariance matrix for multiple time series. I describe the characteristics and estimation of the primary model of interest, MS-GARCH, and some competitor models. I implement and backtest a portfolio management strategy based on risk minimization using MS-GARCH forecasts and evaluate performance relative to competitors. I find MS-GARCH to be an useful tool in portfolio construction, and to offer some significant advantages over more traditional models in terms of accuracy and interpretability when describing a process.

The thesis of Tristan Gardner Wisner is approved.

Nicolas Christou

Frederic Paik Schoenberg

Ying Nian Wu, Committee Chair

University of California, Los Angeles

2017

Dedicated to my Mom and Dad,

for always encouraging me to pursue my curiosity, and for their unwavering love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to my advisor, Professor Ying Nian Wu, for his expertise and support, and for his brilliance and intuition in teaching. I also thank Professor Nicolas Christou for his dedication to teaching, wry humor, and fascinating applied statistics courses, and Professor Rick Schoenberg for his advice and patience. And thank you to all my UCLA professors, for the knowledge you imparted, and all my fellow graduate students I've met along the way.

I owe a deep debt of gratitude to the many people who have provided encouragement and support along the way. To my Mom, Dad, and Adriane, for encouraging every step of my education and for always believing in me. To Professor Mahtash Esfandiari in particular, one of the warmest and kindest people I know, for the privilege of working with you and for helping me with the transition beyond academia. To Glenda, for always having an open door to drop by, for helping me navigate the procedures and protocols, and giving me the occasional kick in the pants when I needed it. To Jason, Chie, Verghese, and Enrique for making everything run smoothly and keeping the office friendly and calm. To Eric, for giving me a place to vent and helping me stay on track. And most of all to Steph, for putting up with me and being there for me throughout this whole endeavor.

CHAPTER 1

Introduction

Accurate estimates and forecasts of financial time series volatility are crucially important for effective portfolio construction and risk management. The subject has received intense focus in government, industry, and academia in the aftermath of the global financial crisis of 2008-2009. For the last several decades, much work has been done to improve the modeling of time series volatility and provide more accurate forecasts than a simple trailing historical variance. The development of autoregressive models for the variance of a time series marks a significant improvement, but still leaves the issue of persistence in their forecasts. That is, these models will tend to produce elevated forecasts of volatility for an extended period of time in response to a shock such as a stock market crash or an oil price spike. Financial assets tend to have periods of unusually high or unusually low volatility, in response to real market shocks, government actions such as tax cuts or warfare, and shifts in investor sentiment. These periods can change quickly with the "animal spirits" of investor confidence, often leaving volatility forecasting models behind the curve.

One solution is to model a time series as jumping between different states of volatility, rather than smoothly increasing or decreasing. Regime-switching models attempt to describe a time series as being in a certain hidden volatility regime at each point in time. They can be viewed as a form of hidden Markov model, with the unobserved volatility state following a Markov process with unknown transition probabilities on a discrete state space with k possible values. In a seminal paper, economist James Hamilton [Ham89] proposes an ARMA process with regime switching in parameters to capture discrete shifts in the mean growth rate of economic time series. The regime-switching paradigm is soon extended beyond moving average models to volatility models, with a regime-switching ARCH model (SWARCH) [HS94].

This thesis analyzes the accuracy and efficacy of such forecasts from a class of models known as Markov-switching GARCH, for the purposes of optimal portfolio construction within the framework of modern portfolio theory. This model is an evolution of autoregressive time series models and regime-switching models, both of which originated in the econometrics literature. The contribution of this thesis lies in the evaluation of MS-GARCH models as a tool for portfolio management, and comparison with other, more widely used multivariate time series models.

This paper is organized as follows. In Section 2, I review the literature on the origins and evolution of MS-GARCH models, as well as the relevant parts of modern portfolio theory (MPT). In Section 3, I define all models used in the body of the paper, techniques for their estimation, the methodology used to compute optimal portfolios under MPT, and methods for evaluating volatility forecasts and portfolio performance. Section 4 contains a description of the data, results of model fitting, and evaluation of performance for each model. Section 5 concludes.

CHAPTER 2

Literature Review

Much of the pioneering research on time series models comes from the econometrics literature. Engle was the first to propose the autoregressive conditional heteroskedasticity (ARCH) model, in which estimates of the conditional variance σ_t^2 take the form of a linear function of past observed errors [Eng82]. Bollerslev [Bol86] built on this with the introduction of the generalized ARCH (GARCH) model, which adds recursive terms for past values of the conditional variance:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 \tag{2.1}$$

The ubiquitous GARCH model is widely used in finance today, particularly the simplest form with $p = q = 1$. Further research provides us with multivariate generalizations of the GARCH model, including the VECH model of Bollerslev and Engle [BEW88], and the BEKK model of Engle and Kroner [EK95], which model the conditional covariance matrix directly. Correlation models which estimate the conditional variances of each time series and the correlation matrix separately, include the constant conditional correlation model of Bollerslev [Bol90] and the dynamic conditional correlation model of Engle [Eng02]. These latter approaches allow for recovery of the time-varying correlation coefficients, disentangling the contributions of correlation from variance, and providing easier interpretability.

Regime-switching models of volatility were first introduced by Hamilton and Susmel [HS94]. The authors address the problem of persistence for both the unweighted sample variance and more complex ARCH and GARCH models; that is, forecasts of future volatility tend to remain elevated for some time after a shock. To alleviate this problem, they propose modeling volatility as a regime-switching process, which allows their model to adjust in response to shocks while still quickly returning to baseline volatility. Their model also retains ARCH dynamics within each regime, to account for smaller within-regime fluctuations in volatility. They identify the issue of path dependence in the MS-GARCH model, which is not present in Hamilton's SWARCH specification. Specifically, computing the true maximum likelihood estimates of MS-GARCH parameters would require integration over all possible state paths $\tilde{s}_t$, which increases exponentially with the length of the time series. This is because, in the following MS-GARCH specification, the conditional variance at time t depends on the conditional variance at time $t-1$, and so on recursively:

$$\sigma_t^2 = \alpha_{0s_t} + \alpha_{1s_t}\epsilon_{t-1}^2 + \beta_{1s_t}\sigma_{t-1}^2 \tag{2.2}$$

With regime switching, there are k possible states of s_t at each time period. Thus, evaluation of the maximum likelihood requires integrating over k^n states, which quickly becomes computationally intractable. Hamilton addresses this by removing the GARCH terms so that the conditional variance forecast depends only on the current regime; without the β term, there is no prior state path that needs to be accounted for.

Combining these two volatility modeling methods, Gray develops the Markov-switching GARCH model [Gra96]. This model allows the volatility process of the time series to jump between different states, but also allows non-constant volatility within each state due to the GARCH dynamics. He also identifies the issue of regime path dependence that makes maximum likelihood estimation computationally difficult. Gray, and later Klaasen [Kla02] and Haas et al. [HMP04] each propose different methods for avoiding path dependence in their estimation techniques. Briefly, Gray proposes to avoid integrating the regime path in the likelihood function by taking the expectation of the previous conditional variance, so that σ_t^2 depends only on the current regime, s_t:

$$\sigma_t^2 = \alpha_{0s_t} + \alpha_{1s_t}\epsilon_{t-1}^2 + \beta_{1s_t}E_{t-2}[\sigma_{t-1}^2] \tag{2.3}$$

Klaasen's method is similar, but uses additional information by taking the expectation at time $t-1$ rather than $t-2$, and conditioning on the state at time t, s_t:

$$\sigma_t^2 = \alpha_{0s_t} + \alpha_{1s_t}\epsilon_{t-1}^2 + \beta_{1s_t}E_{t-1}[\sigma_{t-1}^2 \mid s_t] \tag{2.4}$$

While Klaasen's and Gray's estimation methods provide a computationally tractable solution, Haas et al. propose an analytically tractable model that allows for the calculation of moment and stationarity conditions, and has a more intuitive meaning behind it. The idea is to model a process for each state, running in parallel, and have the Markov chain switch between GARCH processes rather than switching the parameters within the same process:

$$\sigma^2_{i,t} = \alpha_{0i} + \alpha_{1i}\epsilon^2_{t-1} + \beta_{1i}\sigma^2_{i,t-1}, \quad i = 1, \ldots, k \tag{2.5}$$

and P is a $k * k$ transition matrix with:

$$P(s_t = i \mid s_{t-1} = j) = P_{ij} \tag{2.6}$$

Haas and Mittnik extend their earlier work by proposing a multivariate extension based on the VECH and BEKK models of multivariate GARCH [HM08]. They let an M-dimensional time series $\{\epsilon_t\}$ follow

$$\epsilon_t = H^{1/2}_{i,t}\xi_t$$

where $\xi_t \sim^{iid} N(0_{M*1}, I_M)$, i is the state at time t in a Markov chain with state space $S = \{1, 2, \ldots, k\}$, and the same transition matrix P as in the univariate case. Then, they obtain the natural multivariate extension of the MS-GARCH model by stacking the $M(M+1)/2$ unique lower diagonal values of the covariance matrices H_{jt} columnwise in $h_{jt} = vech(H_{jt})$, and the squared returns in $\eta_t = vech(\epsilon_t\epsilon_t')$. The multivariate GARCH(1,1) version then takes the form:

$$h_{jt} = \alpha_{0j} + \alpha_{1j}\eta_{t-1} + \beta_{1j}h_{j,t-1}, \quad j = 1, \ldots, k \tag{2.7}$$

The issue with this approach is it only allows the estimation of the covariance matrix directly. We cannot see whether a change in covariance is due to a change in correlation between two time series, or if it is instead due to a change in one or both of the time series volatilities. For this reason, I instead follow the work of Chen [Che09], using the Constant Conditional Correlation (CCC) model of MGARCH developed in Bollerslev [Bol90]. The specifics are described in the following section.

Finally, to examine the efficacy of these models I test the performance of portfolios allocated according to Modern Portfolio Theory, first outlined in Markowitz (1952) and since expanded on by many other authors. In Markowitz's formulation, an *efficient portfolio* is one with the highest possible expected return given a level of expected volatility. The appropriate portfolio weights can be calculated precisely given estimates for expected returns, expected variance and covariance, and a risk-free rate. I use the period-ahead forecasts obtained from multivariate MS-GARCH models to calculate the optimal portfolios and rebalance as needed. Performance is compared to that attained by a simple close-to-close sample variance and by multivariate non-switching GARCH models. The exact specifications for computing minimum variance portfolios are given in Section 3.

CHAPTER 3

Methodology

This section describes the various GARCH models used in the comparative analysis of the next section, as well as outlining the procedures for computing optimal, minimum-variance portfolios under the Modern Portfolio Theory framework, and the calculation of Value-at-Risk to assess the empirical performance of the volatility estimates.

3.1 Multivariate GARCH Models

Various methods have been proposed for extending the GARCH model to multiple variables, as discussed in section 2. These include the early VEC and BEKK parameterizations, linear combinations of individual GARCH models such as OGARCH and GOGARCH, and the nonlinear combination models discussed here. An excellent overview can be found in Bauwens et al. [BLR06]. To compare the performance of the multivariate MS-GARCH model, I use the constant conditional correlation (CCC) model and the dynamic conditional correlation (DCC) model. Both of these models allow the user to recover the correlation estimate directly, contain fewer parameters, and are far easier to estimate than earlier specifications. For brevity and to simplify the comparison of these models, I restrict my overview here to the bivariate case, for time series a_1 and a_2. Letting the log return series $\boldsymbol{r_t} \equiv [r_{a_1,t}, r_{a_2,t}]'$, with $\boldsymbol{\mu}_t \equiv [\mu_{a_1,t}, \mu_{a_2,t}]'$, $\boldsymbol{\epsilon}_t \equiv [\epsilon^1_{a_1,t} \quad \epsilon^2_{a_2,t}]$, $\boldsymbol{h}_t \equiv [h^1_{a_1,t} \quad h^2_{a_2,t}]$, we characterize the series and its volatility process as:

$$\boldsymbol{r_t} = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t$$
$$\boldsymbol{\epsilon}_t \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{h}_t), \quad or \quad \boldsymbol{\epsilon}_t = \boldsymbol{H}_t^{1/2} \boldsymbol{z}_t, \quad where \quad \boldsymbol{z}_t \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{1})$$

For both of the conditional correlation models, the conditional variance $\boldsymbol{h}_t$ follows a univariate GARCH process for each component time series,

$$h_{j,t} = \omega_j + \alpha_j {\epsilon_{j,t-1}}^2 + \beta_j h_{j,t-1}, \quad j = a_1, a_2$$

or in matrix notation:

$$\boldsymbol{h}_t = \begin{bmatrix} \omega_{a_1} \\ \omega_{a_2} \end{bmatrix} + \begin{bmatrix} \alpha_{a_1} & 0 \\ 0 & \alpha_{a_2} \end{bmatrix} \begin{bmatrix} \epsilon^2_{a_1,t-1} \\ \epsilon^2_{a_2,t-1} \end{bmatrix} + \begin{bmatrix} \beta_{a_1} & 0 \\ 0 & \beta_{a_2} \end{bmatrix} \begin{bmatrix} h_{a_1,t-1} \\ h_{a_2,t-1} \end{bmatrix} \tag{3.1}$$

We can then decompose the conditional covariance matrix $\boldsymbol{H}_t$ as follows to extract the conditional correlation matrix. Letting $\boldsymbol{D}_t = diag(\sqrt{h_t^a}, \ldots, \sqrt{h_t^N})$, we write:

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{P}_t \boldsymbol{D}_t, \tag{3.2}$$

where $\boldsymbol{P}_t$ is a positive definite $N * N$ matrix with $p_{ii} = 1$, and the elements of $\boldsymbol{H}_t$ are $[\boldsymbol{H}_t]_{ij} = \rho_{ij}\sqrt{h_t^i h_t^j}$. In the CCC-GARCH model, we simply have a time-invariant matrix $\boldsymbol{P}_t = \boldsymbol{P}$. In the bivariate case I focus on in this paper, we have just one parameter to estimate.

$$\boldsymbol{P} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

Because of the relative paucity of variables to estimate (3 for each GARCH process and one correlation), the CCC-GARCH is computationally efficient and easy to interpret. It can be estimated numerically by minimizing the log-likelihood:

$$\begin{aligned} \sum_{t=1}^{T} l_t(\boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{T} nlog(2\pi) - \frac{1}{2}\sum_{t=1}^{T} log|\boldsymbol{H}_t| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{r}_t' \boldsymbol{H}_t^{-1} \boldsymbol{r}_t \\ &= -\frac{1}{2}\sum_{t=1}^{T} nlog(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{2} log|h_t^j| - \frac{1}{2}\sum_{t=1}^{T} log|\boldsymbol{P}| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{r}_t' \boldsymbol{D}_t^{-1} \boldsymbol{P}^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{r}_t \end{aligned} \tag{3.3}$$

We can recover the GARCH parameters for each series as well as an estimate of the correlation between the two (or more) series, and it is easily extensible into larger multivariate models. For some applications, where the correlation between the series of interest is expected to remain relatively constant, this simplicity can result in a valuable and effective model. However, particularly with financial return series, correlation is almost certainly not constant. For example, the correlation between stock and bond indices may be weak and fluctuating during calm or growing economic periods, but become sharply negatively correlated in financial crises, as investors dump their stock portfolios and drive up the prices of bonds in search of relative safety. The DCC parameterization partially addresses this by modeling correlation itself as a time series, allowing it to fluctuate with each time period.

In CCC-GARCH, we had a constant $\boldsymbol{P}_t = \boldsymbol{P}$. DCC-GARCH extends the CCC parameterization by giving the correlation matrix a dynamic process of its own, similar to a GARCH process where the coefficients must sum to 1. It starts with:

$$\boldsymbol{Q}_t = (1 - \alpha - \beta)\boldsymbol{S} + a z_{t-1} z'_{t-1} + b\boldsymbol{Q}_{t-1}, \quad a, b > 0, \quad a + b < 1 \tag{3.4}$$

where $\boldsymbol{S}$ is an unconditional correlation matrix, and $\boldsymbol{Q}_0$ is positive definite. To ensure that the resulting matrix conforms to the required values for a correlation matrix ($\boldsymbol{P}_{ii} = 1,\ 0 < \boldsymbol{P}_{ij} < 1$), Engle applies the following scaling to $\boldsymbol{Q}_t$ to give the final result for the correlation matrix:

$$\boldsymbol{P}_t = (\boldsymbol{Q}_t \odot \boldsymbol{I}_N)^{-1/2} \boldsymbol{Q}_t (\boldsymbol{Q}_t \odot \boldsymbol{I}_N)^{-1/2} \tag{3.5}$$

where $\odot$ indicates element-wise multiplication, i.e. extracting the diagonal of $\boldsymbol{Q}_t$. Then the log-likelihood is as follows:

$$\begin{aligned}
\sum_{t=1}^{T} l_t(\boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{T} nlog(2\pi) - \frac{1}{2}\sum_{t=1}^{T} log|\boldsymbol{H}_t| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{r}'_t \boldsymbol{H}_t^{-1} \boldsymbol{r}_t \\
&= -\frac{1}{2}\sum_{t=1}^{T} (nlog(2\pi) + log|\boldsymbol{D}_t \boldsymbol{P}_t \boldsymbol{D}_t| + \boldsymbol{r}'_t \boldsymbol{D}_t^{-1} \boldsymbol{P}^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{r}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T} (nlog(2\pi) + 2log|\boldsymbol{D}_t| + \boldsymbol{r}'_t \boldsymbol{D}_t^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{r}_t - \epsilon'_t \epsilon_t + log|\boldsymbol{P}_t| + \epsilon'_t \boldsymbol{P}_t^{-1} \epsilon_t)
\end{aligned} \tag{3.6}$$

The log-likelihood is straightforward to define and compute numerically, though the dynamic process of the correlation matrix means that the estimation process entails significantly more matrix operations. This parameterization is also fairly frugal, with just one additional parameter compared with CCC-GARCH. With time series whose correlations likely vary over time, a DCC-GARCH approach may represent an improvement over the constant correlation assumption. However, we can see that due to the GARCH-like nature of the matrix process, changes in the estimated correlation parameters are slow and dependent on trailing information. Revisiting the situation described above, where a financial crisis or other shock rapidly changes the relationship between a pair of financial instruments, we can imagine that a model that allows for more sudden changes in correlation would be desirable. By applying a regime-switching process analogous to a hidden Markov Model, in conjunction with a GARCH process, we can describe a time series in a way that allows rapid "jumps" in parameter estimates in response to evidence of shocks.

3.2 The Univariate Markov-Switching GARCH Model

Before describing the multivariate model, I begin with an overview of the univariate Markov-switching GARCH model proposed in [HMP04]. They employ a set of independent GARCH processes running in parallel over the whole time series. These are then coupled with a latent Markov state variable $s(t)$, which determines which volatility process is followed by the return series at each time t. The return series and its K state volatility processes are as follows:

$$\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\epsilon_t &\sim N(0, V_{s(t),t}) \\
V_{i,t} &= \omega_i + \alpha_i \epsilon_{t-1}^2 + \beta_i V_{i,t-1}, \quad i = 1, \dots, K \\
P_{ij} &= \mathbb{P}(s(t) = i | s(t-1) = j), \quad i = 1, \dots, K, \ j = 1, \dots, K
\end{aligned} \tag{3.7}$$

Like the basic GARCH model, this formulation allows for smoothly time varying estimates of the variance within a given state. In addition to these dynamics, we can also model abrupt changes between states, which allows for more accurate models of processes which don't always move smoothly. For instance, financial asset prices may be relatively stable for a period of time before a new piece of asset-specific information, change in sentiment, or a geopolitical or systemic event precipitates a change in the volatility dynamics. Because the parameterization of univariate MS-GARCH is straightforward, with just 3 GARCH parameters for each of the K states and a transition matrix, the interpretation of different states is straightforward. For instance, a greater ω indicates a higher baseline volatility, a greater α indicates that a shock has a greater effect on the volatility going forward, and a greater β indicates a greater persistence of volatility.

Haas and Mittnick also show that a MS-GARCH model need not be stationary within each state, so long as the whole process is stationary. That is, $\alpha_i + \beta_i > 1$ can be possible without destabilizing the whole process. The necessary conditions for this are outlined in their 2004 paper [HMP04]. Thus, we can have a"super-persistent" volatility process, in which the effects of shocks *increase* over time, rather than gradually returning to baseline like in a stationary process with $\alpha_i + \beta_i < 1$. This may be useful in describing the underlying process during a financial crisis, where panic can self-perpetuate into a stampede for the exits.

As discussed in the section 2, this formulation of the MS-GARCH addresses the computational issues faced by Hamilton and Susmel's model due to the recursive nature of the GARCH formula. This allows for a simpler likelihood function which doesn't require integrating over the whole historical path of the process:

$$\begin{aligned} l(\theta) = log\ f_{R_T,\ldots,R_1}(\theta) &= \sum_{t=1}^{T} log\ f_{R_t|R_{t-1},\ldots,R_1}(\theta) \\ &= \sum_{t=1}^{T}\sum_{k=1}^{K} log\ f_{R_t|R_{t-1}}(\theta_i|s(t)=i) * P(s(t)=i|\theta_i, R_{t-1}) \end{aligned} \tag{3.8}$$

The likelihood can be maximized numerically to find the optimal parameters $\theta_i = \{\omega_i, \alpha_i, \beta_i, p_{ij}\}$. I follow a two-stage estimation process, first estimating the univariate MS-GARCH models for each time series, then using those estimates to find the remaining parameters.

3.3 The Multivariate Markov-Switching GARCH Model

Building on the univariate MS-GARCH model described in equation 3.7, we can generalize to a multivariate version. Haas and Mittnick [HM08] propose a version based on the BEKK formulation of multivariate GARCH. I instead follow the model proposed by Chen [Che09], which adapts the CCC-GARCH model described in equations 3.1 and 3.2 to a regime-switching model. The two-state bivariate version consists of two return series, r_{a_1} and r_{a_2}, each with a univariate MS-GARCH parameterization with high and low volatility states. The correlation between the two series is also a two-state variable, representing high and low correlations. Thus in the full parameterization, we have $2^3 = 8$ possible combinations of these state parameters. This can be represented as a single eight-state variable, S_t. By breaking down the correlations in this manner, we can richly describe the behavior of the bivariate series in any of these 8 states using the state-dependent GARCH parameters for each of the return series, as well as the state-dependent correlation between the two.

$$\boldsymbol{R}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \quad where \;\; \boldsymbol{r}_t = [r_{a_1,t}, r_{a_2,t}]', \;\; \boldsymbol{\mu} = [\mu_{a_1,t}, \mu_{a_2,t}]'$$

$$\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{V}_{s(t),t}), \quad where \;\; \boldsymbol{\epsilon}_t = [\epsilon_{a_1,t}, \epsilon_{a_2,t}]', \mathbf{0} = [0,0]' \;\; and$$

$$\boldsymbol{V}_{s(t),t} = \begin{bmatrix} V_{a_1,s(t),t} & C_{s(t),t}\sqrt{V_{a_1,s(t),t}V_{a_2,s(t),t}} \\ C_{s(t),t}\sqrt{V_{a_1,s(t),t}V_{a_2,s(t),t}} & V_{a_2,s(t),t} \end{bmatrix}$$

$$V_{a_k,i,t} = \omega_{a_k,i} + \alpha_{a_k,i}\epsilon^2_{a_k,t-1} + \beta_{a_k,i}V_{a_k,i,t-1}, \quad i = 1,\ldots,8, \quad k = 1,2$$

$$C_{i,t} = C_i, \quad i = 1,\ldots,8$$

$$P_{ij} = \mathbb{P}(s(t) = i|s(t-1) = j), \quad i = 1,\ldots,8, \; j = 1,\ldots,8$$

The maximum likelihood does not have a straightforward analytical solution, and is estimated numerically. The following description of the estimation procedure is largely reproduced from [Che09], with some additional details on computational implementation. One difference is that I used the estimates for the GARCH parameters from the first stage as the final estimates, rather than just as initial estimates in the second stage. The reasons are twofold. First, if using the first stage estimates as initial values in the second stage, there is little difference between the final estimates obtained by just estimating the correlation and transition matrix in the second stage versus estimating all 76 parameters in the second stage. Second, the computation time of the estimating the full parameter set greatly exceeds that of estimating just C and P. We can simplify the likelihood function as follows:

$$
\begin{aligned}
p_t(r_{a_1,t}, r_{a_2,t}|C_i, P_{i,j}, i = 1,\ldots,8, j = 1,\ldots,8; r_{a_1,t-1},\ldots,r_{a_1,0}, r_{a_2,t-1},\ldots,r_{a_2,0})& \\
= p_t(\boldsymbol{r}_t|\boldsymbol{\theta}; \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)& \\
= \sum_{i=1}^{8} p_t(\boldsymbol{r}_t, s(t) = i|\boldsymbol{\theta}; \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)& \\
= \sum_{i=1}^{8} p_t(\boldsymbol{r}_t|\boldsymbol{\theta}; s(t) = i, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0) \boldsymbol{\cdot} \mathbb{P}(s(t) = i|\boldsymbol{\theta}, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)& \\
= \sum_{i=1}^{8} p_t(\boldsymbol{r}_t|\boldsymbol{\theta}; s(t) = i, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)& \\
\boldsymbol{\cdot} \left[\sum_{j=1}^{8} \mathbb{P}(s(t) = i|\boldsymbol{\theta}, s(t-1) = j) \cdot \mathbb{P}(s(t-1) = j|\boldsymbol{\theta}, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)\right]& \\
= \sum_{i=1}^{8} \left[\eta_t(i) \cdot \sum_{j=1}^{8} [p_{i,j} \cdot \xi_{t-1,t-1}(j)]\right]& \qquad (3.9)
\end{aligned}
$$

where: (3.10)

$$\eta_t(i) = p_t(\boldsymbol{r}_t|\boldsymbol{\theta}; s(t) = i, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0) \qquad (3.11)$$

$$p_{i,j} = \mathbb{P}(s(t) = i|\boldsymbol{\theta}, s(t-1) = j), \qquad (3.12)$$

$$\xi_{t-1,t-1}(j) = \mathbb{P}(s(t-1) = j|\boldsymbol{\theta}; \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0) \qquad (3.13)$$

We can then compute the t and $t-1$ conditional state probabilities; that is, the probability the process is in state j at time t given all information through time t, $\xi_{t|t}(j)$, and the probability the process is in state j at time t given all information through time $t-1$, $\xi_{t|t-1}(j)$. The latter is used for forecasting the period ahead.

$$
\begin{aligned}
\xi_{t|t}(j) &= \mathbb{P}(s(t) = j|\boldsymbol{\theta}; \boldsymbol{r}_t,\ldots,\boldsymbol{r}_0) \\
&= \frac{\mathbb{P}(s(t) = j, \boldsymbol{r}_t|\boldsymbol{\theta}, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)}{\mathbb{P}(\boldsymbol{r}_t|\boldsymbol{\theta}, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)} \\
&= \frac{\mathbb{P}(\boldsymbol{r}_t|\boldsymbol{\theta}, s(t) = j, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)) \cdot \mathbb{P}(s(t) = j|\boldsymbol{\theta}, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0))}{\sum_{j=1}^{8} \mathbb{P}(\boldsymbol{r}_t|\boldsymbol{\theta}, s(t) = j, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0)) \cdot \mathbb{P}(s(t) = j|\boldsymbol{\theta}, \boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_0))} \\
&= \frac{\eta_t(j) \cdot \xi_{t|t-1}(j)}{\sum_{j=1}^{8} \eta_t j \cdot \xi_{t|t-1}(j)}\ , \qquad (3.14)
\end{aligned}
$$

$$\xi_{t|t-1}(j) = P \cdot \xi_{t-1|t-1}(j) \quad \text{(by the Markov property)} \qquad (3.15)$$

So, we can get the conditional state probabilities by iterating from t=0, by setting $\xi(0|0) = \boldsymbol{\pi}_0$, where $\boldsymbol{\pi}_0$ is the stationary distribution of the transition matrix P. $\eta_t(i)$ is just the likelihood conditioned on one state; that is, a typical CCC-GARCH likelihood that can be estimated as in section 3.1. So the MLE parameters can be obtained by maximizing the log-likelihood over the whole timespan:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} \sum_{t=1}^{T} \log p_t(\boldsymbol{r}_t|\boldsymbol{\theta}, \boldsymbol{r}_{t-1}, \ldots, \boldsymbol{r}_0) \tag{3.16}$$

3.4 The Markowitz/MPT Framework

After estimating expected returns and the covariance matrix using the previous models, we need some mechanism to translate these estimates into an optimal portfolio allocation. The idea is to combine multiple assets into a portfolio in a way that maximizes the portfolio expected return for a given level of risk (volatility), or minimizes variance for a required expected return. For a collection of assets with non-perfect correlations, you can construct a minimum variance portfolio (MVP) with lower volatility than any one asset on its own, as shown in the seminal paper of Markowitz [Mar52]. The global MVP is calculated as follows:

Given j assets with expected return $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_j)$ and covariance matrix $\boldsymbol{\Sigma}_{j\times j}$, find allocation $\boldsymbol{x} = (x_1, \ldots, x_j)$ with minimum portfolio variance σ_p^{2*}:

$$\min_{x_1,\ldots,x_j} \sigma_p^2 = \boldsymbol{x'\sigma x} \tag{3.17}$$

$$s.t. \quad \boldsymbol{x'1} = 1$$

which can be solved for the optimal allocation $\boldsymbol{x}^* = (\mu_1^*, \ldots, \mu_j^*)$ using the method of Lagrange multipliers.

I further restrict the portfolio to prohibit short sales by adding an additional constraint, $x_i > 0, \;\; i = (1, \ldots, j)$. We can then solve for the MVP as a quadratic programming problem,

$$\min_{\boldsymbol{x}} \frac{1}{2}\boldsymbol{x'Dx} - \boldsymbol{d'x}, \qquad \boldsymbol{D} = \boldsymbol{D}_{j\times j}, \quad \boldsymbol{d} = \boldsymbol{d}_{j\times 1} \tag{3.18}$$

$$\boldsymbol{A}'_{in}\boldsymbol{x} \geq \boldsymbol{b}_{in}, \qquad \boldsymbol{A}_{in} = \boldsymbol{A}_{inj\times j}, \quad \boldsymbol{b}_{in} = \boldsymbol{b}_{inj\times 1}$$

$$\boldsymbol{A}'_{eq}\boldsymbol{x} = \boldsymbol{b}_{eq}, \qquad \boldsymbol{A}_{eq} = \boldsymbol{A}_{eq2\times j}, \quad \boldsymbol{b}_{eq} = \boldsymbol{b}_{eq2\times 1}$$

where:

$$\boldsymbol{D} = 2\boldsymbol{\Sigma}, \quad \boldsymbol{d} = (0, \ldots, 0)'$$
$$\rightarrow \frac{1}{2}\boldsymbol{x}'\boldsymbol{D}\boldsymbol{x} - \boldsymbol{d}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{\sigma}\boldsymbol{x},$$

$$\boldsymbol{A}_{eq}'\boldsymbol{x} = \begin{bmatrix} \boldsymbol{\mu}' \\ \boldsymbol{1}' \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{\mu}'\boldsymbol{x} \\ \boldsymbol{1}'\boldsymbol{x} \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{\mu}_p \\ \boldsymbol{1}' \end{bmatrix} \boldsymbol{x} = \boldsymbol{b}_{eq},$$

$$\boldsymbol{A}_{in}'\boldsymbol{x} = \boldsymbol{I}_j\boldsymbol{x} = \boldsymbol{x} \geq \boldsymbol{0}_j = \boldsymbol{b}_{in}$$

which can be easily solved with any quadratic programming solver, such as `solve.QP()` in the `R` package `quadprog`. So, given an estimated covariance matrix and expected return vector, the global MVP with no short sales allowed can be computed for the time period ahead.

3.5 Value-at-Risk Calculations

To assess the effectiveness of each model in accurately modeling and minimizing market risk, I examine Value-at-Risk (VaR). As the name implies, VaR is a measure of potential losses under normal market conditions. The basic formulation comes out of the perspective of modern portfolio theory, looking at risk in a portfolio as a function of expected returns and covariance, with the first published metric recognizable as VaR published in 1945 by Leavens (1945), and expanded on by Markowitz (1952), Roy (1952), Tobin (1958), and Sharpe (1963), among others, though the phrase "value-at-risk" had yet to enter the lexicon. The adoption of VaR as a widely used metric by financial institutions and their regulators evolved from the 1970's onward as financial markets grew more liberalized, complex, and leveraged [Hol02]. The 70's and 80's were marked by a number of financial innovations, including the standardization of "vanilla" derivatives such as options, futures, and currency forwards, the beginning of securitization, and ever more complicated "bespoke" derivatives in the over-the-counter markets. Leverage grew with the introduction and mainstreaming of securities leasing, repo markets, swaps, and the new high-yield market, as banks moved beyond their deposit bases and senior debt issuance to fund ever more speculative activity.

With the increasing leverage, complexity, and volatility of their books, financial institutions moved to quantitative risk measures out of necessity. The history of legislative and regulatory activity relating to capital requirements and risk management spans decades, borders, and shifts in philosophy, and while fascinating, is far too extensive to detail here. The term "value-at-risk" came into common usage out of JP Morgan's RiskMetrics service in 1994, which provided a standardized methodology and updated correlations for a variety of asset classes. VaR became the gold standard metric of risk, with the SEC virtually requiring its disclosure in 1997 and the Basel II Accords adopting it as a preferred measure of risk in 1999.

I use a 95% one-tailed weekly VaR measure to assess the empirical performance of the portfolios obtained by each model. This is the largest amount we would expect to lose in 19 out of 20 trading weeks.

$$\begin{aligned} VaR_{t,\alpha} &= q_{\alpha}^{\boldsymbol{R}_t} P_t, \\ q_{\alpha}^{\boldsymbol{R}_t} &= \mu_t + V_{s(t),t} q_{\alpha}^{z} \end{aligned} \tag{3.19}$$

where P_t is the value of the portfolio at time t, q_{α}^{z} is the α% quantile of the normal distribution, and $\alpha = 1 - 0.95 = 0.05$, or whichever confidence level is selected.

The converse, of course, is that in 1 out of 20 weeks, we expect to see a loss exceeding VaR - a VaR "break". Rather than focusing on the size of the potential loss, as is the typical usage of VaR in risk management, I am interested in comparing the expected number of breaks with the observed. This helps determine the accuracy of the expected return, and forecasted volatility. A perfect estimate of those quantities used in portfolio construction should result in VaR breaks occurring, on average, in 1 in 20 weeks over a long period time. We can test this using a *coverage test* such as the Kupiec proportion-of-failures test.

$$LR_{PoF} = -2 \log \frac{(1-p)^{N-x} p^{x}}{(1-\frac{x}{N})^{N-x} (\frac{x}{N})^{x}} \tag{3.20}$$

This test statistic follows a χ_1^2 distribution. Moreover, the breaks should happen uniformly over time. We can examine plots of VaR breaks or test the randomness of those sequences with a runs test to see if this assumption is satisfied.

CHAPTER 4

Results

To test the performance of the MS-GARCH model as a tool for portfolio construction and risk management, I backtest a hypothetical portfolio over time. At each time t, an MS-GARCH model is estimated using 200 weeks of trailing data for the assets in the portfolio, and used to forecast the volatility at time $t + 1$. With this rolling volatility forecast, I compute the global minimum variance portfolio and rebalance the allocation of the portfolio components from the previous period. I compare the performance of the MS-GARCH model to four competing models, with portfolios constructed in the same manner, by examining their realized returns, volatility, and VaR breaks.

4.1 Description of Data

I examine simple, 2-asset portfolios consisting of two popular exchange traded funds (ETFs), selected as proxies for the US stock market and US bond market. I use ETFs rather than historical indices or bond yields to better reflect a typical individual retail portfolio, which are often categorized based on their allocation of stocks vs. bonds[1]. I use the iShares Core U.S. Aggregate Bond ETF for the bond portion, and the SPDR S&P 500 ETF for the stock portion. These funds were selected for their size, liquidity, age, and adherence to underlying indices[2] commonly used as benchmarks for broad US mutual funds and ETFs.

[1] http://www.aaii.com/asset-allocation

[2] Bloomberg Barclays US Aggregate Bond Index and Standard & Poor's 500

For both ETFs, weekly closing price and dividend data was obtained from Yahoo! Finance for the time period from October 13, 2003 to May 29, 2017. I then compute a total return series, showing the true return over time of the asset with dividends reinvested as they are issued. Because each model must be parameterized using trailing data, the time period over which the portfolios are computed spans the 10-year period from August 13, 2007 to May 29, 2017. Descriptive statistics for each total return series are shown in Table 4.1, and the prices over time are charted in Figure A.1. We see that the stock fund had an average weekly return approximately twice as large as the bond fund. It also had much more variability, with the largest one-week loss of over 22% coming in the depths of the financial crisis of 2008. The bond fund saw a similar large drawdown around the same time, as seen in the squared returns in Figure A.2. As we will see later, these unusually large returns can have a significant impact on the estimation of model parameters. Also in Figure A.1, we can see how the trailing correlation between the two ETFs fluctuates over the whole time period, potentially indicating the importance of accounting for these shifts when trying to craft a properly balanced portfolio.

Table 4.1: Descriptive Statistics for Return Series

	AGG	SPY
Mean	0.0677	0.1343
Std. Dev.	0.7841	2.6315
Skewness	-4.1277	-1.1285
Minimum	-10.58	-22.06
Maximum	6.238	12.48
Annualized Return	3.5228	7.2701

Weekly returns in % for iShares Core U.S. Aggregate Bond ETF (ticker symbol AGG) and SPDR S&P 500 ETF (ticker symbol SPY), 8/13/2007 - 5/29/2017.

4.2 Results of Model Parameter Estimation

The maximum likelihood parameter estimates for the MS-GARCH model are computed numerically. Because of the large search space, approximating a global optimum rather than a local one can be computationally and time intensive. After experimentation with several optimization techniques and software implementations, I settled on simulated annealing using the R package GenSA. This offers a good balance of accuracy and speed, and avoids the rapid convergence to a local optimum that often occurs when using gradient descent or related methods. Standard errors are also estimated numerically using block bootstrapping, where the time series is resampled in random length blocks with a mean length of 20.

To illustrate the MS-GARCH model parameterization and get a sense of how the state and volatility dynamics occur in these assets, I begin by fitting the model to the entire time series, from 2003 to 2017. The estimates are given in tables A.1 and A.2 in the appendix, along with estimates for the CCC-GARCH and DCC-GARCH models fitted to the same data. We can see that the CCC-GARCH and DCC-GARCH parameters are very similar, with the only notable difference coming from the correlation estimates. For the DCC model, β_{dcc} is the dominant coefficient, meaning $\boldsymbol{P}_t$ is the main component of $\boldsymbol{P}_{t+1}$. Yet this small difference has a large impact on the model's log-likelihood, indicating the importance of correlation to model fit.

The GARCH parts of MS-GARCH reveal an interesting contrast between the low and high volatility states. For both series, in the low volatility state, β is very low, and ω and α are fairly high. The volatility persistence is very low, with the dynamic primarily driven by the baseline volatility ω, and the observed variance in the most recent time period. In the high volatility state, all three parameters are elevated far beyond what you would expect in a typical GARCH model. Were the series to stay in this high volatility state for long, it would continue to grow in an unstable manner. The state transition matrix offers some clues as to why this is viable. The $t+1$ transition probabilities are fairly uniform in each of the states with at least one asset in the high volatility state. In the "low-low" states 1 and 2, nearly all transition probability is on remaining in 1 or 2. So, barring a major shock, the state probability is likely to remain in 1 and 2, and to revert there quickly if in an elevated state. Figure A.5 shows the $t-$ and $(t+1)-$ conditional transition probabilities for each week in the time period. Elevated probability of being in a high volatility state only registers in response to major market gyrations, such as the 2008 financial crisis, the 2011 and 2013 debt ceiling debates, and the Euro crisis.

Turning attention to the 200 week rolling window used in portfolio construction, figures

A.6 and A.7 show the conditional state probabilities from models fit over two timespans, one containing the 2008 crash and the other without it. There is a stark difference in the estimated parameters depending on whether or not major shocks are included. In more normal times, the state probabilities fluctuate more, and the GARCH parameters are closer to typical models. In periods containing major shocks, the high volatility state is essentially reserved for those dramatic events, while the "low-low" states cover the majority of the timespan; in other words, like CCC-GARCH with a 2-state correlation parameter, with occasional spikes.

4.3 Portfolio Construction and Performance

To examine the effectiveness of MS-GARCH as a tool for portfolio management, I compare the performance of the two-asset portfolio based on MS-GARCH estimates with similar portfolios constructed from other volatility estimates. The estimated covariance matrix for the MS-GARCH portfolios is an average of the covariance matrices for each state, weighted by the $t+1$ conditional state probability. The expected returns for each of the two assets are the same for each portfolio at each time t. I use an arithmetic mean of the trailing 200 weeks of return values. This keeps the expected return standard across all portfolios, and makes asset allocation dependent only on the covariance matrices for each. The estimates used for comparison come from an unweighted 200 day covariance, CCC-GARCH, DCC-GARCH, and an exponentially weighted moving average (EWMA) with $\lambda = 0.94$. Figure 4.1 shows the comparative performance of each of these portfolios.

Overall, there is not a substantial difference in performance. This is not particularly surprising, given that all portfolios are constructed from the same two assets and optimized to minimize variance. We can see that the CCC-GARCH and DCC-GARCH portfolios underperform the others, which achieve approximately the same annualized returns over the 10 year time span. As shown in Table 4.2, all five portfolios have similar volatility, while DCC-GARCH and MS-GARCH do best at reducing skewness, as they experience smaller maximum losses in any given week in the 10 year time frame. The simple trailing and

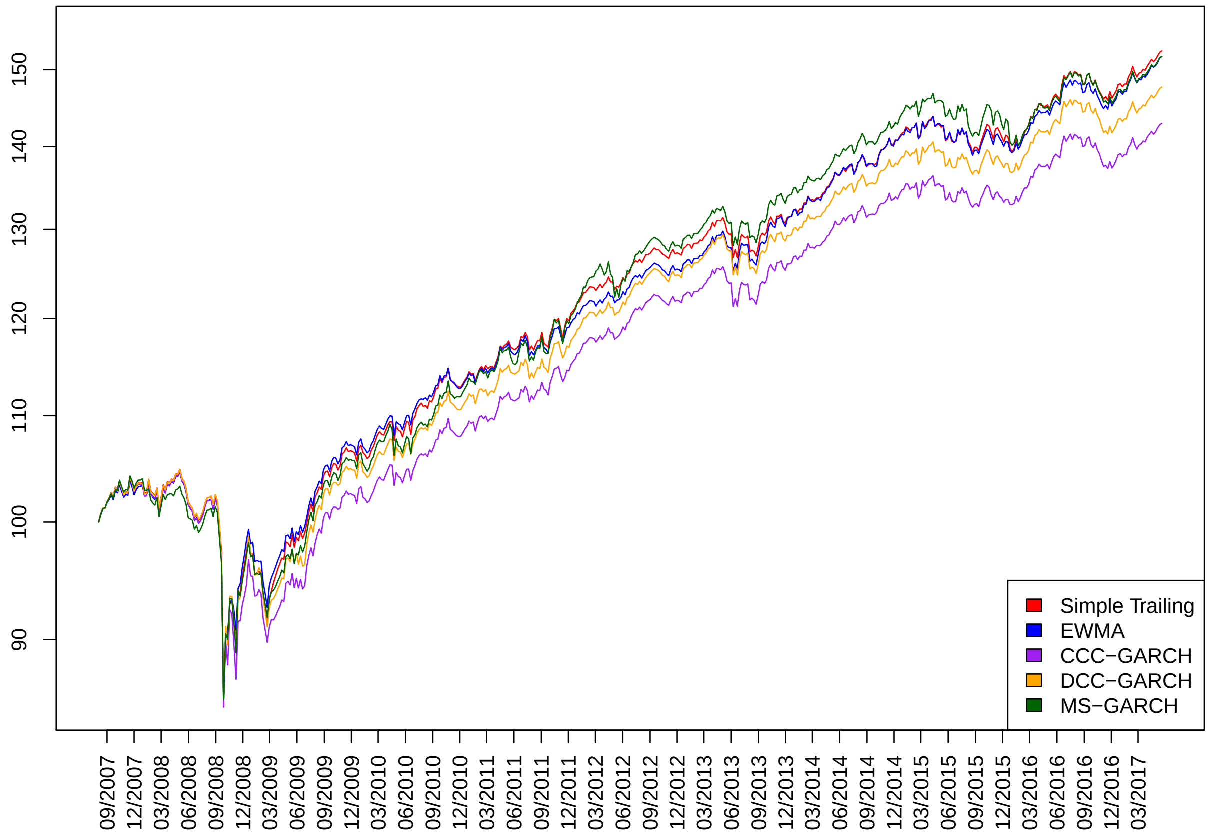
Table 4.2: Descriptive Statistics for Portfolios

	Simple Trailing	EWMA	CCC-GARCH	DCC-GARCH	MS-GARCH
Mean	0.0826	0.0816	0.0699	0.0763	0.0817
Std. Dev.	0.8976	0.8730	0.9374	0.8994	0.9207
Skewness	-5.6755	-5.7491	-4.9656	-4.2233	-3.9768
Minimum	-13.1945	-12.8720	-13.3446	-12.1970	-12.2802
Maximum	6.0141	6.1669	5.8540	6.1743	5.9459
Annualized Return	4.4031	4.3486	3.7143	4.0588	4.3514

EWMA portfolios slightly outperform MS-GARCH until about 200 weeks after the declines of late 2008 and early 2009. This seems to be due to the larger and steadier allocation to AGG, while the allocations in GARCH type portfolios, particularly DCC and CCC, fluctuate far more and seem to miss periods of rapid gains in SPY, as seen in Figure A.4. Essentially, with the "memory" of the financial crisis still in their rolling window, they overreact to subsequent, smaller market declines, decrease the allocation to SPY, and miss out on subsequent rebounds. This "whipsaw" effect seems to be an inherent problem with this type of model in volatility forecasting. The MS-GARCH model fares slightly better, as spells of slightly increased volatility will not always register as being in high volatility states.

Around the end of 2011, MS-GARCH begins to outperform until stock market volatility of 2015 and early 2016 enters the covariance estimation windows. The MS-GARCH model is at its best in times of relative calm and steady returns like this period. Here, the state probabilities are not pulled to extremes as they are in time periods including the financial crisis, as shown in A.6. In periods of calm, traditionally riskier allocations can be made, and in periods of increasing volatility, they can be quickly made less risky. Figure A.4 shows the evolution of allocations over time for each portfolio. The simple trailing portfolio is steadiest, consisting of about 75-80% AGG over the whole time period. EWMA has a wider range of allocations, but moves gradually over time, as older observations decrease in the weighting for covariance calculation. CCC-GARCH and DCC-GARCH allocations respond to every

Figure 4.1: Comparing Portfolio Performance



twist and turn in the market varying from 60% to 90% allocation to AGG, sometimes within the same year. They become steadier after the volatility of the financial crisis leaves the 200 week sample window. MS-GARCH has the highest range, from nearly 100% AGG to just above 50%. Allocations move in quick jumps followed by periods of relatively steady trending, rather than the weekly up and down of the other GARCH type models.

4.4 Model Accuracy: Value at Risk

Finally, I evaluate the performance of each portfolio based on VaR. For each time t in the covered timespan, we compute VaR for the following period based on our volatility forecast and expected return. The next period, we check if a VaR break occurred. Figure A.3 shows log returns, the VaR threshold, and all VaR breaks for each of the five portfolios. First, we can see that all five models are reasonably close to our expected 25.55 VaR breaks; none of them wildly over- or under-estimates portfolio riskiness at any given time. The CCC-GARCH portfolio may overestimate risk a little, and the EWMA portfolio may underestimate it, but neither are significantly off under the Kupiec coverage test. Similarly, no model exhibits a strong lack of randomness in the Wald-Wolfowitz runs test, though visually there is some cause for concern.

Table 4.3: 95% Value-at-Risk Breaks

	VaR Breaks	LR_{Kupiec_PoF}	p-value	Runs Test	p-value
Simple Trailing	26	0.0083	0.927	-0.626	0.531
EWMA	31	1.1491	0.284	-1.658	0.0973
CCC-GARCH	18	2.6073	0.106	-1.8025	0.07147
DCC-GARCH	24	0.1009	0.751	-1.8696	0.06153
MS-GARCH	28	0.2401	0.624	-1.2627	0.2067

Value-at-risk breaks and Kupiec proportion of failure coverage test results for each model. For the 511 observations observed, we expect $511 \times 0.05 = 25.55$ VaR breaks. The test statistic $LR_{Kupiec_{PoF}}$ follows the χ^2 distribution with 1 degree of freedom, with critical value $\chi^2_{1,0.05} = 3.841$. The second test statistic is the Wald-Wolfowitz runs test statistic, and follows an approximately normal distribution.

For example, the simple trailing model shows a much lower VaR threshold for an extended period following the 2008 crash during which no VaR breaks occur. This suggests some over-estimation of risk during this time that the test statistics in Table 4.3 don't register due to the small sample period and low power of the tests. Still, they provide some reassurance that none of the models is completely off base. The GARCH class models do appear to produce an intuitively more realistic VaR, based on the closeness with which the thresholds follow the ebbs and flows of volatility in the portfolios. The MS-GARCH model in particular seems to react quickly to changes in market dynamics, due to its regime-switching property.

CHAPTER 5

Conclusions

"The major difference between a thing that might go wrong and a thing that cannot possibly go wrong is that when a thing that cannot possibly go wrong goes wrong it usually turns out to be impossible to get at or repair."

-Douglas Adams

Douglas Adams' observation on the fallibility of technology extends to many aspects of the human condition, not least the financial markets and prediction of risk. This thesis takes a small step in addressing the need for more accurate forecasts of volatility. Markov-switching GARCH models show promise as a tool for portfolio construction, with minimum variance portfolios based on MS-GARCH estimates of covariance performing at least as well over the sample period as traditional covariance estimation, and better then competing multivariate GARCH models without regime switching. MS-GARCH also proves useful as a descriptive method, to aid in understanding the volatility dynamics in different states of markets. It is also extensible to larger numbers of assets and numbers of states.

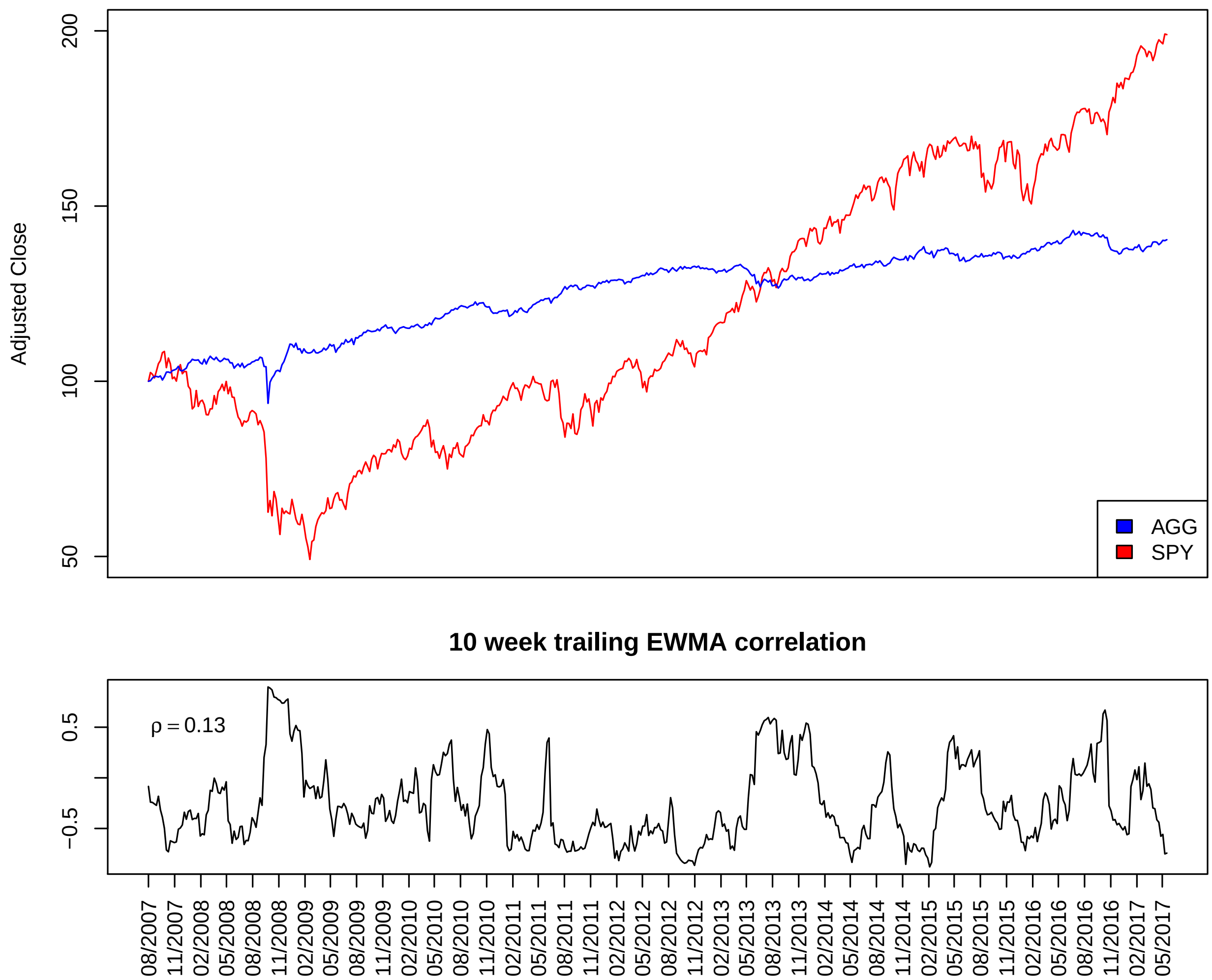
There are many opportunities for further exploration and innovation on this topic. Faster computational techniques and an easy to use software implementation would allow this method to be accessible to more users. The `MSGARCH` package was recently made available on CRAN, but as of this writing there are no `R` packages that implement multivariate MS-GARCH available. Another fruitful avenue of research would be the extension of multivariate MS-GARCH beyond the bivariate case. With the current framework of a state-switching variable for each return series and each pairwise correlation, the total number of states quickly becomes unwieldy, both computationally and in terms of interpretability. One possible solution is to allow different assets and correlations to share state variables based on some criterion, limiting the number of state combinations that need to be estimated.

Though it represents an improvement in the modeling of periods of very high volatility, MS-GARCH still faces the difficulty of anticipating sudden spikes in volatility. It may adjust faster than other models after the fact, but does no better at predicting that first spike, as we see in our hypothetical portfolio in late 2008. Users would do well to remember the words of another of Adams' characters, Vroomfondel, who absurdly declared "We demand rigidly defined areas of doubt and uncertainty!" More complex models may improve portfolio outcomes in periods of normal market behavior, but practitioners should not overconfidently place too much faith in any one model and be lulled into a false sense of invulnerability regarding unforeseeable tail risks.

Appendix A

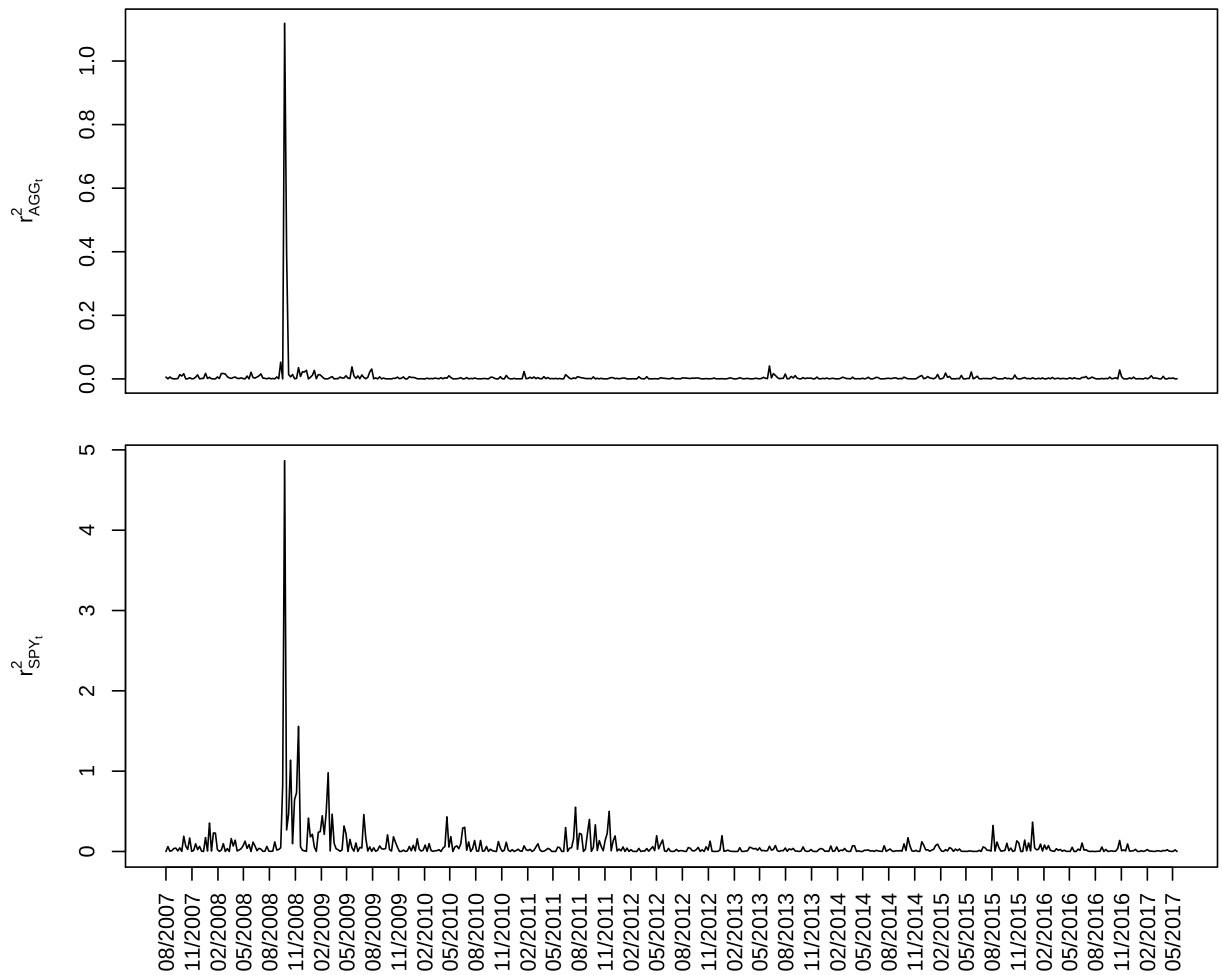
Tables and Figures

Figure A.1: Original Return Series



Price performance over the relevant sample period for SPY and AGG, the constituent ETFs of the portfolios examined in Section 4. The plot beneath shows the correlation between the two assets over the preceeding 10 weeks, with the most recent weeks weighted more heavily (EWMA parameter of $\lambda = 0.94$).

Figure A.2: Weekly Return Series



Squared weekly returns for SPY and AGG, showing the magnitude of the spike during the financial crisis for squared values.

Table A.1: Parameter Estimates

	CCC-GARCH		DCC-GARCH		MS-GARCH	
	Estimate	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
ω_1^b	7.26×10^{-6}	1.16×10^{-6}	6.52×10^{-6}	3.38×10^{-6}	0.017	0.042
α_1^b	0.351	1.45×10^{-3}	0.314	0.062	0.360	0.127
β_1^b	0.545	0.247	0.571	0.101	0.036	0.043
ω_1^s	4.43×10^{-5}	0.053	4.06×10^{-5}	0.296	0.154	0.149
α_1^s	0.239	1.26×10^{-5}	0.229	1.55×10^{-5}	0.435	0.127
β_1^s	0.689	0.035	0.698	0.097	0.076	0.088
ω_2^b					0.538	0.304
α_2^b					0.893	0.217
β_2^b					0.981	0.255
ω_2^s					0.416	0.344
α_2^s					0.866	0.111
β_2^s					0.891	0.153
α_{dcc}			0.073	0.413		
β_{dcc}			0.822	0.571		
ρ_1	-0.139	0.037			-0.431	0.192
ρ_2					0.018	0.259
ρ_3					-0.099	0.379
ρ_4					0.161	0.325
ρ_5					0.166	0.427
ρ_6					0.723	0.238
ρ_7					0.179	0.398
ρ_8					0.822	0.362
LL	5841.946		4454.436		1986.477	
AIC	-11669.89		-8892.872		-3820.954	

Standard error estimates for CCC and DCC parameters obtained using the inverse Hessian method, from the `ccgarch` package. Standard error estimates for MS-GARCH obtained using block bootstrapping, with the original time series resampled 5000 times using random blocks with mean length 20 to obtain a series the same length as the original, using the `tsboot` package.

Table A.2: Transition Matrix Estimates

state	1	2	3	4	5	6	7	8
1	0.4881	0.4789	0.0002	0.0317	0.0001	0.0003	0.0003	0.0005
	0.2173	*0.1479*	*0.1268*	*0.1003*	*0.0808*	*0.0564*	*0.0409*	*0.0298*
2	0.6328	0.3570	0.0004	0.0064	0.0004	0.0002	0.0000	0.0029
	0.1500	*0.1895*	*0.1130*	*0.0940*	*0.1024*	*0.0777*	*0.0480*	*0.0349*
3	0.1166	0.0808	0.1611	0.1500	0.0848	0.0889	0.1595	0.1583
	0.1218	*0.1147*	*0.1430*	*0.0977*	*0.0619*	*0.0525*	*0.0573*	*0.0714*
4	0.3652	0.3528	0.0076	0.1936	0.0092	0.0187	0.0090	0.0439
	0.0932	*0.0930*	*0.0812*	*0.0952*	*0.0638*	*0.0644*	*0.0659*	*0.0736*
5	0.1483	0.1466	0.1414	0.1470	0.1020	0.1017	0.1463	0.0667
	0.0597	*0.0598*	*0.0549*	*0.0629*	*0.0725*	*0.0858*	*0.0560*	*0.0627*
6	0.1764	0.1775	0.1837	0.1769	0.1038	0.0329	0.1163	0.0325
	0.0656	*0.0717*	*0.0631*	*0.0675*	*0.0736*	*0.0911*	*0.0592*	*0.0565*
7	0.1433	0.1412	0.0907	0.0944	0.1029	0.1414	0.1408	0.1454
	0.0444	*0.0461*	*0.0437*	*0.0417*	*0.0474*	*0.0484*	*0.0503*	*0.0486*
8	0.1665	0.1648	0.1631	0.0890	0.0630	0.0653	0.1520	0.1363
	0.0485	*0.0454*	*0.0464*	*0.0454*	*0.0493*	*0.0515*	*0.0491*	*0.0653*

Estimate of transition matrix probabilities with standard errors shown beneath estimates, italicized. Also obtained via block bootstrapping, as in Table A.1.

Figure A.3: 95% Value-at-Risk Breaks

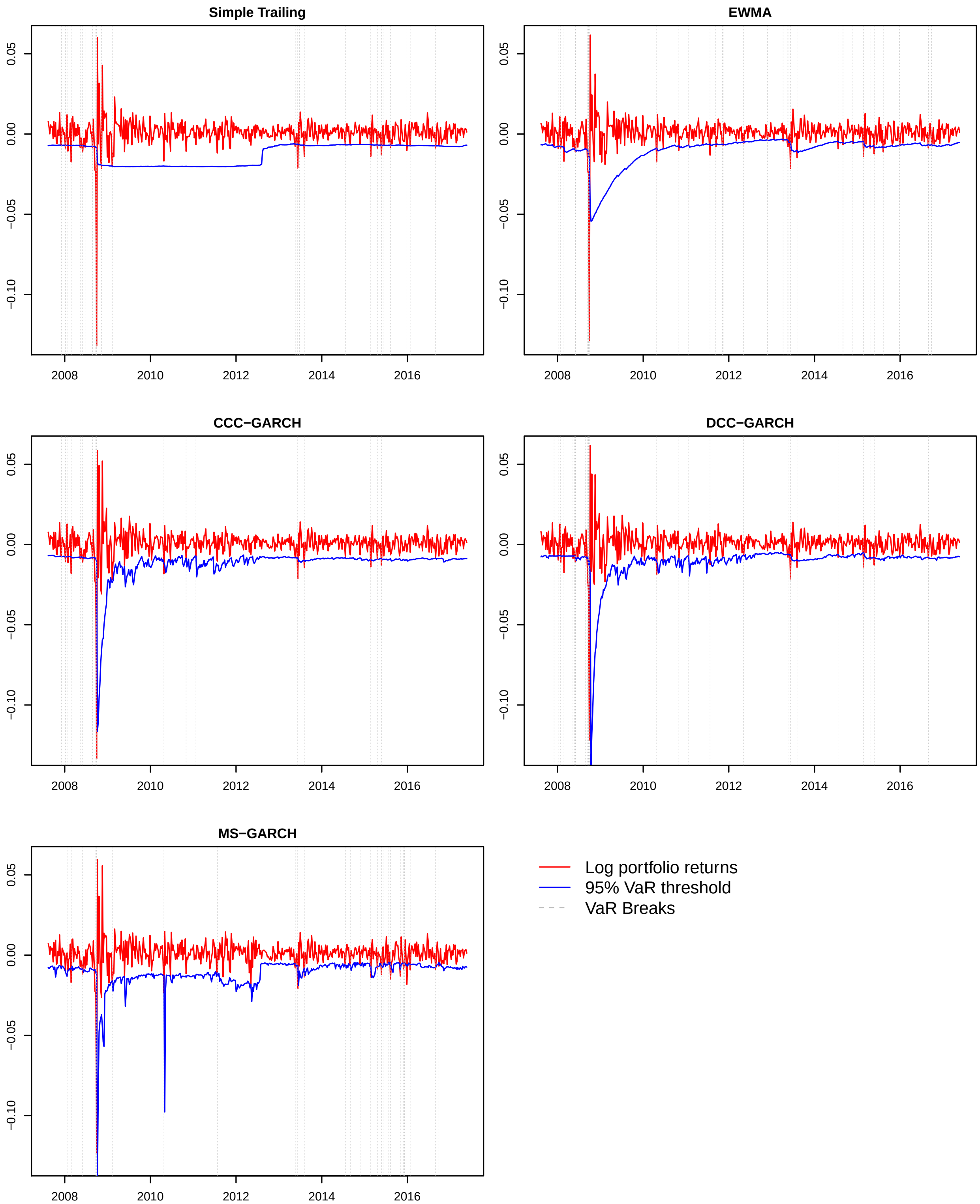


Figure A.4: Portfolio Allocation Over Time

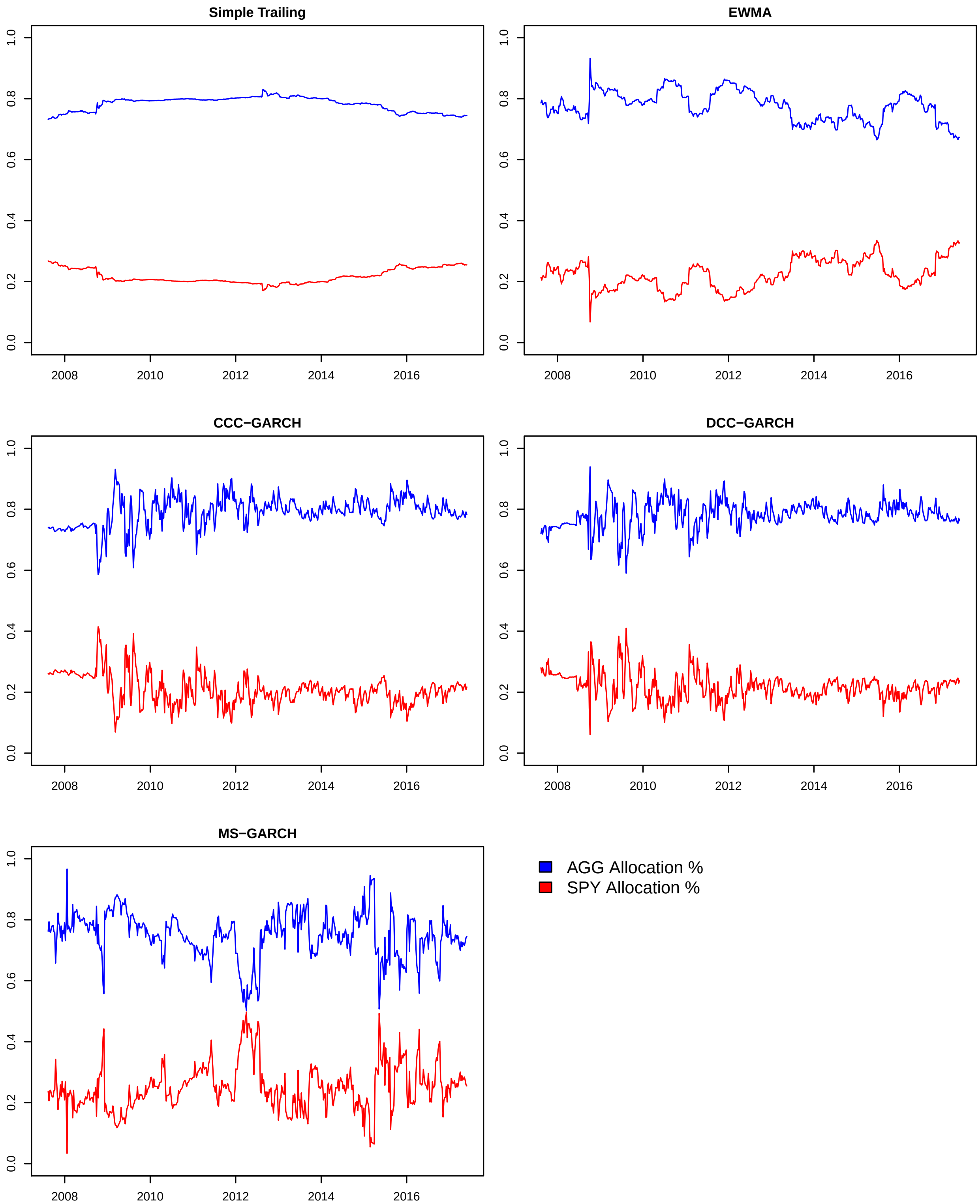


Figure A.5: Conditional State Probabilities, Whole Series

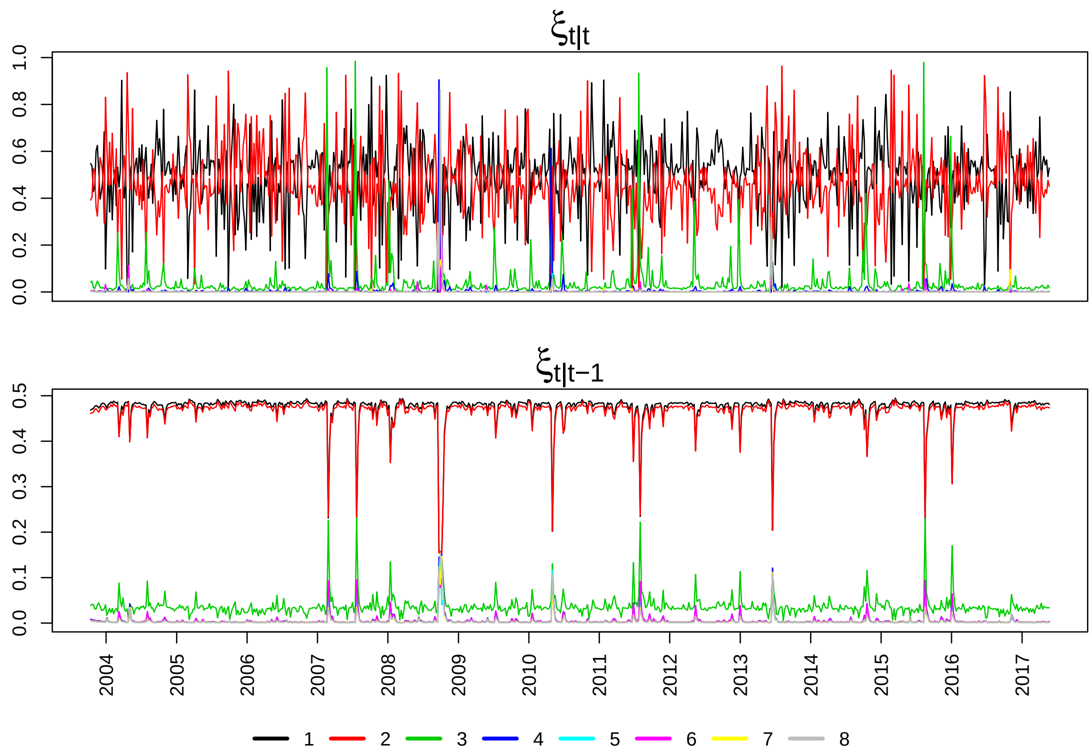


Figure A.6: Conditional State Probabilities, 200-week Period Including Crash

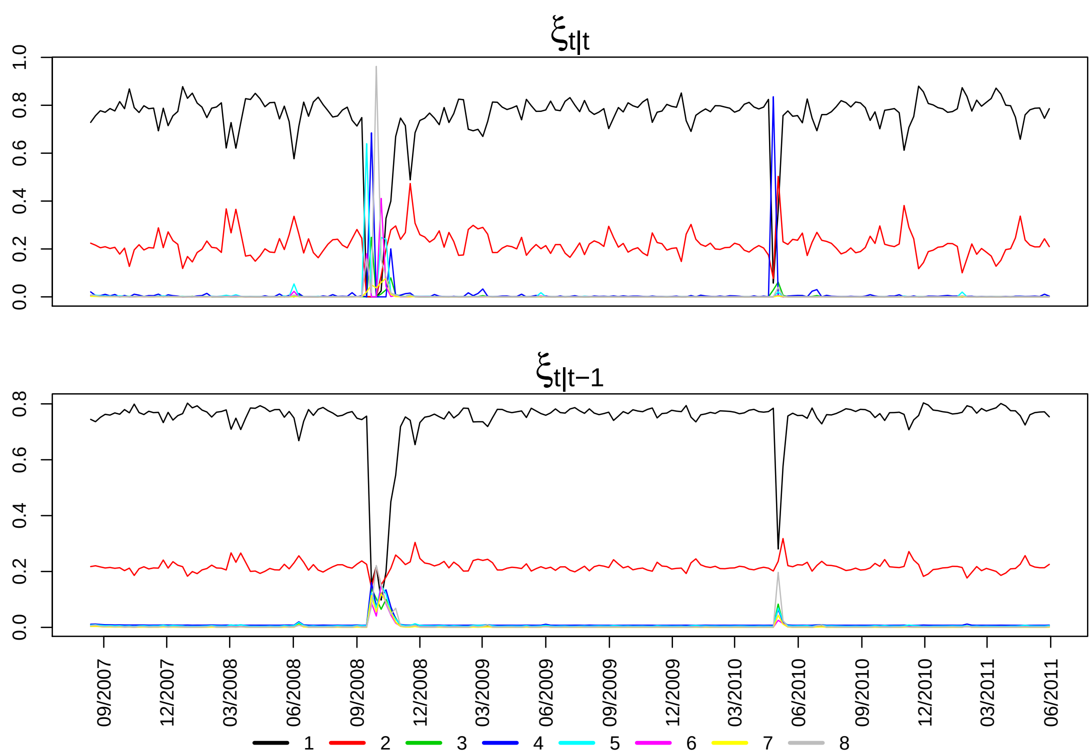


Figure A.7: Conditional State Probabilities, 200-week Period Not Including Crash

REFERENCES

[BEW88] Tim Bollerslev, Robert Engle, and Jeffrey Wooldridge. “A Capital Asset Pricing Model with Time-Varying Covariances.” *Journal of Political Economy*, **96**(1):116–31, 1988.

[BLR06] Luc Bauwens, Sébastien Laurent, and Jeroen V. K. Rombouts. “Multivariate GARCH models: a survey.” *Journal of Applied Econometrics*, **21**(1):79–109, 2006.

[Bol86] Tim Bollerslev. “Generalized Autoregressive Conditional Heteroskedasticity.” *Journal of Econometrics*, **31**:307–327, February 1986.

[Bol90] Tim Bollerslev. “Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model.” *The Review of Economics and Statistics*, **72**(3):498–505, 1990.

[Che09] Runquan Chen. “Regime Switching in Volatilities and Correlation between Stock and Bond Markets.” Discussion paper 640, Financial Markets Group, London School of Economics, 2009.

[EK95] Robert Engle and Kenneth F. Kroner. “Multivariate Simultaneous Generalized ARCH.” *Econometric Theory*, **11**(01):122–150, 1995.

[Eng82] Robert F. Engle. “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.” *Econometrica*, **50**(4):987–1007, 1982.

[Eng02] Robert Engle. “Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models.” *Journal of Business & Economic Statistics*, **20**(3):339–350, 2002.

[Gra96] Stephen F Gray. “Modeling the conditional distribution of interest rates as a regime-switching process.” *Journal of Financial Economics*, **42**(1):27–62, 1996.

[GT07] Massimo Guidolin and Allan Timmermann. “Asset allocation under multivariate regime switching.” *Journal of Economic Dynamics and Control*, **31**(11):3503–3544, 2007.

[Ham89] James D. Hamilton. “A New Approach to the Economic Analysis of Nonstationary Times Series and the Business Cycle.” *Econometrica*, **57**(2):357–384, March 1989.

[HM08] Markus Haas and Stefan Mittnik. “Multivariate Regime Switching GARCH with an Application to International Stock Markets.” CFS Working Paper Series 2008/08, Center for Financial Studies (CFS), 2008.

[HMP04] Markus Haas, Stefan Mittnik, and Marc S. Paolella. “A New Approach to Markov-Switching GARCH Models.” *Journal of Financial Econometrics*, **2**(4):493–530, July 2004.

[Hol02] Glyn A Holton. "History of Value-at-Risk.", 2002.

[HS94] James D. Hamilton and Raul Susmel. "Autoregressive Conditional Heteroskedasticity and Changes in Regime." *Journal of Econometrics*, **64**:307–333, September 1994.

[Hul09] J. Hull. *Options, Futures and Other Derivatives*. Options, Futures and Other Derivatives. Pearson/Prentice Hall, 2009.

[Kla02] Franc Klaassen. "Improving GARCH volatility forecasts with regime-switching GARCH." *Empirical Economics*, **27**(2):363–394, Mar 2002.

[Mar52] Harry Markowitz. "Portfolio Selection." *The Journal of Finance*, **7**(1):77–91, 1952.

[ST09] Annastiina Silvennoinen and Timo Teräsvirta. "Multivariate GARCH models." *Handbook of financial time series*, pp. 201–229, 2009.