Forecasting Equity Volatility Dynamics with Markov-Switching EGARCH Models

by

Tyson Dennis-Sharma

BASc, University of Toronto, 2022

Submitted to the MIT Sloan School of Management
in partial fulfillment of the requirements for the degree of

MASTER OF FINANCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2024

Authored by:	Tyson Dennis-Sharma MIT Sloan School of Management January 15, 2024
Certified by:	Leonid Kogan Nippon Telegraph and Telephone Professor of Management, Thesis Supervisor
Accepted by:	Urmi Samadar Assistant Dean, MIT Sloan Master of Finance Program MIT Sloan School of Management

Forecasting Equity Volatility Dynamics with Markov-Switching EGARCH Models

by

Tyson Dennis-Sharma

Submitted to the MIT Sloan School of Management
on January 15, 2024 in partial fulfillment of the requirements for the degree of

MASTER OF FINANCE

ABSTRACT

Understanding and anticipating stock market volatility enables better portfolio management. We forecast US equity volatility with a Markov-Switching EGARCH model with one high and one low volatility regime. We show that this model contains similar information about future volatility as the VIX Index. It also outperforms single-regime GARCH and EGARCH models. Moreover, the model's 1-day ahead regime predictions are economically significant: market volatility and kurtosis, equity risk premia, and stock-bond relations shift when the model forecasts a regime change.

Thesis supervisor: Leonid Kogan
Title: Nippon Telegraph and Telephone Professor of Management

Acknowledgments

I firstly and primarily thank Professor Leonid Kogan for granting me his time and providing invaluable support throughout my project. Most of my analyses were pursued in direct response to his suggestions. His expertise in financial econometrics and academic research shone throughout our meetings and allowed me to learn a lot from him. I am especially grateful to Professor Kogan for agreeing to take me as his student even though he was at Harvard last term.

I would also like to thank my many friends in the Master of Finance (MFin) program for their camaraderie and assistance. In particular, the idea for this thesis was born out of a group project I performed with Sarthak Tyagi, Tanya Ratra, Arjun Banerjee and Tanmay Bagree. I am therefore indebted to them. I am also grateful to Da Huo – a fellow thesis student – whom I frequently consulted for advice. His brilliance in time series econometrics enabled him to assist me a lot.

Finally, I'd like to thank my parents and my broader family, including my Boston-based cousin Manisha, for supporting me while I pursued this MFin degree. Moving out of Canada for the first time in my life to pursue graduate studies at MIT was a big life decision. I benefited immensely from unwavering familial support and love.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Equity return volatility is a key determinant of numerous pricing and portfolio allocation decisions. For instance, it is an input to option pricing models. As such, accurate volatility measures and forecasts are invaluable to financial market participants (Andersen and Bollerslev, 1998). GARCH (Generalized Autoregressive Conditional Heteroscedasticity) family models are frequently used to describe and forecast the evolution of volatility. In this study, we implement Nelson (1991)'s EGARCH (Exponential GARCH) model with a Markov switching mechanism, allowing parameters to instantaneously change when a latent Markov process shifts from a high to a low volatility regime (and vice versa). This approach is based on the idea that the market alternates between two states - tranquil and turbulent - with each state influencing volatility dynamics distinctly.

We demonstrate that the volatility forecasting performance of our Markov-Switching EGARCH (MS-EGARCH) model is comparable to that of a regression model which relates volatility to the VIX index. Using a 20-year rolling window, we fit MS-EGARCH via maximum likelihood estimation on daily 1971-2022 S&P500 returns obtained from the CRSP U.S. Stock Database (Center for Research in Security Prices, 2023). We evaluate forecasting performance across 1, 5, 10, and 22 day horizons, and employ mean absolute error (MAE) and root-mean squared error (RMSE) as our evaluation metrics. Our conclusions are similar across metrics and horizons. We also show that MS-EGARCH's forecasts are highly correlated with VIX-based forecasts. Hence, MS-EGARCH forecasts volatility consistently with options markets, making it an especially useful model for indices without liquid options.

We further show that MS-EGARCH obtains superior MAE and RMSE than single-regime EGARCH and GARCH across all (1,5,10, and 22 day) horizons. For this comparison, we fit each model on a daily 1926-2022 NYSE, Nasdaq, and American stock exchange excess returns index (hereon referred to as the 'NYSE/Nasdaq index') , which we also obtain from CRSP (Center for Research in Security Prices, 2023). The improvement in forecasting performance when transitioning from EGARCH to MS-EGARCH is similar to the improvement observed when moving from EGARCH to GARCH. We show that these MS-EGARCH MAE and RMSE gains are statistically significant using Diebold-Mariano tests. Moreover, regression analysis shows that MS-EGARCH's 1-day ahead forecasts effectively predict the discrepancies between realized volatility and fitted, single-regime EGARCH volatility. Admittedly though, similar regressions reveal that MS-EGARCH's multi-step ahead forecasts are less predictive.

Market participants and academics have long known that equity markets undergo regime shifts. In our analysis, we demonstrate that transitions between the high and low volatility regimes detected by MS-EGARCH map to changes in model performance. We define the regime at time t based on MS-EGARCH's filtered time t regime probabilities. Our findings indicate that MS-EGARCH's forecasting out-performance of single regime EGARCH is primarily explained by superior performance in the tranquil state; MS-EGARCH and single-regime EGARCH are statistically indistinguishable within the high volatility state.

We also show that MS-EGARCH's forecasts of market regimes–defined in terms of MS-EGARCH's 1-day ahead predictive regime probabilities–effectively signal shifts in market behavior. Notably, we observe increased volatility and kurtosis in the NYSE/Nasdaq index following a volatile regime prediction. This is coupled with a higher probability of significant market drops (1, 3, 5, and 10%) and an elevated VIX index, which further reinforces MS-EGARCH's allignment with option markets.

Additionally, MS-EGARCH's 1-day ahead regime predictions differentiate the performance of equity risk factors. For example, using data from Kenneth French's website (French, 2023), we discover that the Fama-French size factor yields an annualized return of -1.26% when the low volatility regime is predicted, compared to 3.47% when the high volatility regime is predicted. Similarly, the Fama-French investment factor demonstrates improved performance following a turbulent state prediction.

Our research also reveals that MS-EGARCH's regime predictions anticipate changes in stock-bond relationships. Utilizing data from Bloomberg (Bloomberg L.P., 2023), we find that US government bond yields, as well as investment grade and high-yield bond returns increase when MS-EGARCH predicts the high volatility regime. Furthermore, correlation dynamics shift: both high-yield and investment grade bond returns show a reduced correlation with the NYSE/Nasdaq index in this scenario.

1.1 Relation to the Prior Literature

Our paper contributes to the rich literature on applying Markov-switching models to financial time series. Hamilton (1989) introduced Markov-switching models to the economic mainstream. Subsequently, Hamilton and Susmel (1994) implemented ARCH models with Markov-switching parameters. Gray (1996) was the first to forward a tractable Markov-switching GARCH model. Dueker (1997), Ang and Bekaert (2002), Klaassen (2002), Haas et al. (2004), and Marcucci (2005) extended upon Gray (1996)'s approach.

Klaassen (2002) and Marcucci (2005) specifically showed that regime-switching GARCH models fit on daily data generate superior multi-step ahead realized volatility forecasts than single-regime models. Therefore, our conclusions align with their findings. Moreover, we innovate upon their work in four respects. First, our model allows for EGARCH(1,1) dynamics in each regime, not purely GARCH(1,1) dynamics. This advancement is crucial for accurately modeling the 'leverage effect', a phenomenon whereby stock price declines raise subsequent volatility more than stock price increases of the same magnitude (Tsay, 2010; Nelson, 1991). Henry (2009) also implemented a two-regime Markov-switching EGARCH model on UK equity returns, but he did not examine out-of-sample forecasting.

Second, our out-of-sample forecasting comparison between MS-EGARCH and single regime GARCH and EGARCH models is more extensive. Whereas Marcucci (2005)'s out-of-sample testing period consists of 511 daily observations and Klaassen (2002)'s consists of 2491, our out-of-sample period includes 19,441 daily observations of the NYSE/Nasdaq Index from July 1946 to December 2022. Moreover, both Klaassen (2002) and Marcucci (2005) fit their models once on an in-sample period before testing their models out-of-sample, keeping parameters fixed. We refit our MS-EGARCH and single regime EGARCH and GARCH models daily using a rolling 20-year (5040 day) window. In this manner, we ensure that our models continually re-calibrate in response to new data.

Third, we perform a separate forecasting test using daily S&P500 data from 1990 to 2022 in which we compare MS-EGARCH to a VIX-based regression model. Therefore, our paper serves as a bridge between the literature on Markov-switching GARCH models and the literature on VIX-based realized volatility forecasting. Previous researchers (Christensen and Prabhala, 1998; Blair et al., 2001) have argued that linear models which relate realized volatility to implied volatility perform similarly to standard, single-regime GARCH-type models in realized volatility forecasting. Our study complements these findings by comparing a VIX-based linear model to a Markov-switching GARCH family model.

Fourth, we examine the significance of our MS-EGARCH model's predicted regimes. Thus, our paper connects the literature on Markov-switching for volatility forecasting to the literature on Markov-switching for regime detection and asset allocation. Maheu and McCurdy (2000) introduced a unique Markov-switching model, enhanced with duration-dependent transition probabilities and ARCH effects, to discern bull and bear markets in the U.S. stock market. Their approach prioritized aligning their model's outcomes with historical perceptions of market phases for accurate in-sample fitting. We are more concerned with the out-of-sample, 1-step ahead predicted regimes and their significance; furthermore, we adopt no priors about when each regime should be detected. In that sense, our research aligns more with the literature on out-of-sample regime detection and regime-based asset allocation strategies. This literature includes the works of Ang and Bekaert (2004), Guidolin and Timmermann (2007), and Kritzman et al. (2012). Each of these authors assumed different dynamics under each regime and employed different underlying variables than us. For example, Kritzman et al. (2012) assumed Gaussian dynamics under each regime and fit regimes to inflation, market turbulence and economic growth. Hence, our study uncovers novel insights about the financial significance of MS-EGARCH's predicted US equity regimes.

There are, nevertheless, parallels between our regime-conditional results and prior research. Other researchers have shown that the performance of equity risk factors varies during different market phases. Ahn et al. (2019) argued that the size premium is most present at the bottom of business cycles. This conclusion is congruent with our finding that the size factor performs best after MS-EGARCH predicts the volatile state. Likewise, consistent with our findings, Sheth and Lim (2017) demonstrated that Fama-French's investment factor earns its best returns during recessions and in the early stages following them.

We also add to the body of research characterizing the evolution of the stock-bond relation. Connolly et al. (2005) and Chiang et al. (2014) employed various measures of market uncertainty including implied and conditional equity market volatility as well as stock turnover. They concluded that stock-bond correlations decline and bond returns rise as market uncertainty increases. These conclusions align with our analyses on bond behaviour following a high volatility regime prediction by MS-EGARCH.

The remainder of this paper is organized as follows. Chapter 2 outlines our MS-EGARCH framework and presents in-sample estimation results on the NYSE/Nasdaq index. Subsequently, chapter 3 compares MS-EGARCH's volatility forecasting performance to a VIX-based approach using S&P500 data. Chapter 4 contrasts MS-EGARCH's volatility forecasting performance to that of single-regime EGARCH and GARCH. Chapter 5 discusses the differences in market behaviour we observe conditional upon MS-EGARCH's predicted regime. Chapter 6 concludes and discusses potential extensions to this study.

Chapter 2

Theoretical Framework and In-Sample Estimation of MS-EGARCH

In this chapter, we outline the mathematical theory underlying our MS-EGARCH model and present in-sample results. Section 2.1 introduces single-regime GARCH and EGARCH models. Section 2.2 explores our Markov-Switching GARCH-EGARCH framework. Section 2.3 describes our estimation procedure, and justifies why we choose MS-EGARCH in particular. Section 2.3 also presents our in-sample MS-EGARCH estimates and discusses them.

2.1 GARCH and EGARCH Models

Consider a stock market index and its level at time t, p_t. This index's log rate of return in percentage points is defined as

$$r_t = 100[\ln(p_t) - \ln(p_{t-1})] \tag{1}$$

Bollerslev (1986)'s GARCH(1,1) model can be written as

$$r_t = \mu + a_t = \mu + \sigma_t \varepsilon_t \tag{2}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. (independent and identically distributed) random variables with zero mean and unit variance, and $\alpha_0 > 0, \alpha_1 \geqslant 0$, and $\beta_1 \geqslant 0$ are required to ensure that the conditional variance is always positive (Tsay, 2010). Equation 2 can be modified to allow for more complex modelling of the mean of r_t, but we ignore this to instead focus on volatility dynamics. Furthermore, while it is possible to generalize GARCH(1,1) to a GARCH(m,s) model, we focus on GARCH(1,1) for simplicity.

GARCH(1,1) models the conditional variance of log returns, σ_t^2, as an affine function of its prior value, σ_{t-1}^2, and the most recent shock, a_{t-1}^2. This means that large shocks tend to be followed by more large shocks, producing the well-known phenomenon of volatility clustering. However, one downside of GARCH(1,1) is that it treats positive and negative shocks symmetrically, failing to account for the leverage effect (Tsay, 2010). In order to overcome this challenge, Nelson (1991) proposes the EGARCH(1,1) model[1], in which equation 3 is modified as follows (Ardia et al., 2019):

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1\Big(\frac{|a_{t-1}|}{\sigma_{t-1}} - \mathbb{E}\big[\frac{|a_{t-1}|}{\sigma_{t-1}}\big]\Big) + \alpha_2\frac{a_{t-1}}{\sigma_{t-1}} + \beta_1\ln(\sigma_{t-1}^2) \tag{4}$$

Notice that if $\alpha_2 < 0$ then this model takes into account the leverage effect as negative shocks raise conditional variance more than positive shocks. Consequently, we expect to estimate $\alpha_2 < 0$.

One convenient choice for the distribution of the $\{\varepsilon_t\}$ is the standard normal distribution. To better account for the leptokurtic nature of financial returns, another possibility is to assume that the ε_t follow a standardized student-t distribution. Under this specification, the probability density function of the ε_t is

$$f(\varepsilon_t;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{(\nu-2)\pi}\Gamma(\frac{\nu}{2})}\big(1+\frac{\varepsilon_t^2}{(\nu-2)}\big)^{\frac{-(\nu+1)}{2}} \tag{5}$$

where $\Gamma(\cdot)$ is the Gamma function, and ν is a shape parameter that must be greater than 2 (Ardia et al., 2019). The kurtosis of this distribution is elevated for smaller ν[2].

2.2 Markov-Switching GARCH and EGARCH Models

Regime-switching models allow parameters to switch across different regimes according to a Markov process, which is described by a latent state variable, s_t. This state variable is assumed to progress according to a Markov chain with transition probability function

$$\mathbb{P}(s_t = j | s_{t-1} = i) = p_{ij} \tag{6}$$

that specifies the probability of switching to state j at time t after being in state i at time $t-1$ (Marcucci, 2005). One can group these probabilities into the ergodic transition probability matrix:

$$P = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} = \begin{pmatrix} p_{11} & p_{21} \\ (1-p_{11}) & (1-p_{21}) \end{pmatrix} \tag{7}$$

where, following Marcucci (2005), we limit ourselves to two-regime models in this study.

Consider an index log return series, r_t. We will again use equation 2 as our mean equation. Our state variable s_t will affect the parameters of the conditional variance equation. Let $\alpha_0(s_t)$, $\alpha_1(s_t)$, and $\beta_1(s_t)$ denote the GARCH(1,1) parameters as a function of the current state, s_t. If we proceed naively with allowing Markov-switching in these parameters and change nothing else about the GARCH(1,1) model, then our conditional variance equation will be

$$\sigma_t^2 = \alpha_0(s_t) + \alpha_1(s_t)a_{t-1}^2 + \beta_1(s_t)\sigma_{t-1}^2 \tag{8}$$

---

[1] Once again, one can generalize to an EGARCH(m,s) model, but we will stick to EGARCH(1,1) for simplicity.

[2] As ν rises to ∞, the standardized student-t distribution approaches the standard normal distribution (Ardia et al., 2019).

where, as a reminder, $a_t = r_t - \mu$. Given some initial σ_0^2, recursive substitution into equation 8 allows us to express conditional volatility as

$$\sigma_t^2 = \sum_{i=0}^{t-1} \Big(\big[\alpha_0(s_{t-i}) + \alpha_1(s_{t-i}) a_{t-1-i}^2\big] \prod_{j=0}^{i-1} \beta_1(s_{t-j}) \Big) + \sigma_0^2 \prod_{i=0}^{t-1} \beta_1(s_{t-i}) \tag{9}$$

which elucidates that σ_t^2 depends on the entire (unknown) regime path. Consequently, computing the likelihood for a sample of size T involves summing over 2^T possible regime paths (Haas et al., 2004). This is highly impractical. Haas et al. (2004) circumvent this path dependence problem by suggesting that there be one independent conditional variance path for each regime. Specifically, let $k \in \{1, 2\}$ denote the regimes with 1 corresponding to the tranquil state and 2 to the turbulent state. Adopt the notation that $\alpha_{1,k}$,$\beta_{1,k}$ and $\alpha_{0,k}$ are the regime k-specific GARCH parameters. Haas et al. (2004)'s Markov-switching GARCH(1,1) model can be expressed as

$$\sigma_{t,k}^2 = \alpha_{0,k} + \alpha_{1,k} a_{t-1}^2 + \beta_{1,k} \sigma_{t-1,k}^2, \ \forall k \in \{1, 2\} \tag{10}$$

where $\{\sigma_{t,k}^2\}$ is the conditional variance process of regime k. As Haas et al. (2004) argue, an advantage of this Markov-Switching GARCH model is that it can account for instantaneous shifts in variance dynamics. For instance, Dueker (1997) observes that volatility often rises substantially in a short period of time at the start of a turbulent period, a fact equation 10 can recognize via it's turbulent regime-specific parameter $\alpha_{0,2}$.

We can apply similar logic to define a regime-switching EGARCH(1,1) model as follows:

$$\ln(\sigma_{t,k}^2) = \alpha_{0,k} + \alpha_{1,k} \Big(\frac{|a_{t-1}|}{\sigma_{t-1,k}} - \mathbb{E}\big[\frac{|a_{t-1}|}{\sigma_{t-1,k}}\big] \Big) + \alpha_{2,k} \frac{a_{t-1}}{\sigma_{t-1,k}} + \beta_{1,k} \ln(\sigma_{t-1,k}^2), \ \forall k \in \{1, 2\} \tag{11}$$

We can even allow for a mixed Markov-Switching GARCH-EGARCH model in which the dynamics under one regime are as specified by equation 10 while the dynamics in the other regime are as specified by equation 11. Moreover, just like with single regime models, we can assume the $\{\varepsilon_t\}$ follow a standard normal or a standardized student-t distribution. In fact, we will assume that the ε_t distribution is regime dependent, which necessitates a regime-specific shape parameter, ν_k, in the standardized student-t case. Regime-specific ε_t distributions account for regime-specific kurtosis.

2.3 Estimation Procedure and In-Sample Estimation Results

2.3.1 Data and Estimation Methodology

We employ the NYSE, American, and Nasdaq value-weighted returns (including distributions) series from the CRSP U.S. Stock Database (Center for Research in Security Prices, 2023). We use the daily risk-free rate series from Kenneth French's website (French, 2023). We subtract the log percentage risk-free rate from the log percentage returns of the NYSE, American, and Nasdaq exchanges. This defines our NYSE/Nasdaq index series, which spans

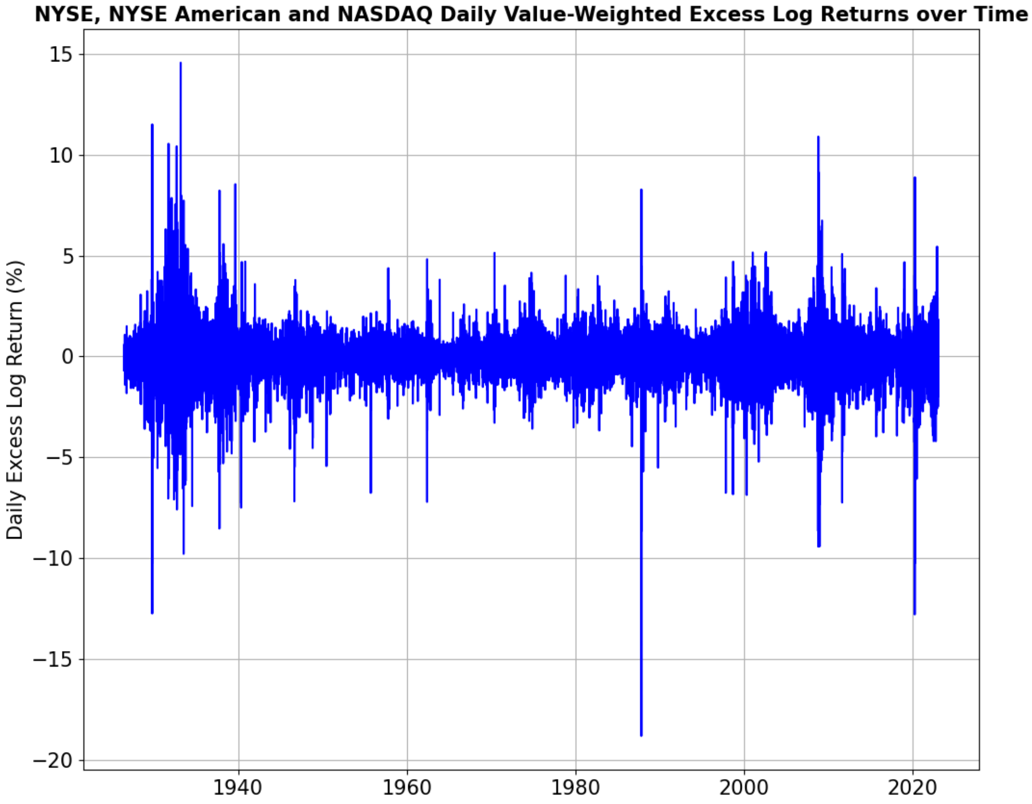

Figure 1: Plot of Full NYSE/Nasdaq Index Series

from July 1st, 1926 to December 30th, 2022. A plot of this full time series is contained in figure 1.

In chapter 4, we perform a forecasting test on this series, using a 20-year rolling window. Hence, we define the first 20-years of our series (i.e., from July 1, 1926 to June 30, 1946) as our in-sample period. Figure 2 plots the auto-correlation function ('ACF') of our squared NYSE/Nasdaq index series based on data from this in-sample period[3]. It is apparent in this figure that strong ARCH effects are present, which justifies our use of GARCH family models.

We estimate Markov-switching GARCH-EGARCH models via maximum likelihood estimation (MLE). Assume we first demean our log returns series, $\{r_t\}$, to obtain the series $\{a_t\}$ (this is equivalent to estimating the mean equation 2). Let $\boldsymbol{\theta}$ represent the vector of parameters for our Markov-switching GARCH-EGARCH model – including GARCH-EGARCH parameters, regime-specific shape parameters, and entries in the transition probability ma-

---

[3] The shaded blue region in figure 2 corresponds to 95% confidence intervals under the null that the auto-correlations are zero. Auto-correlations outside this region are statistically significant.

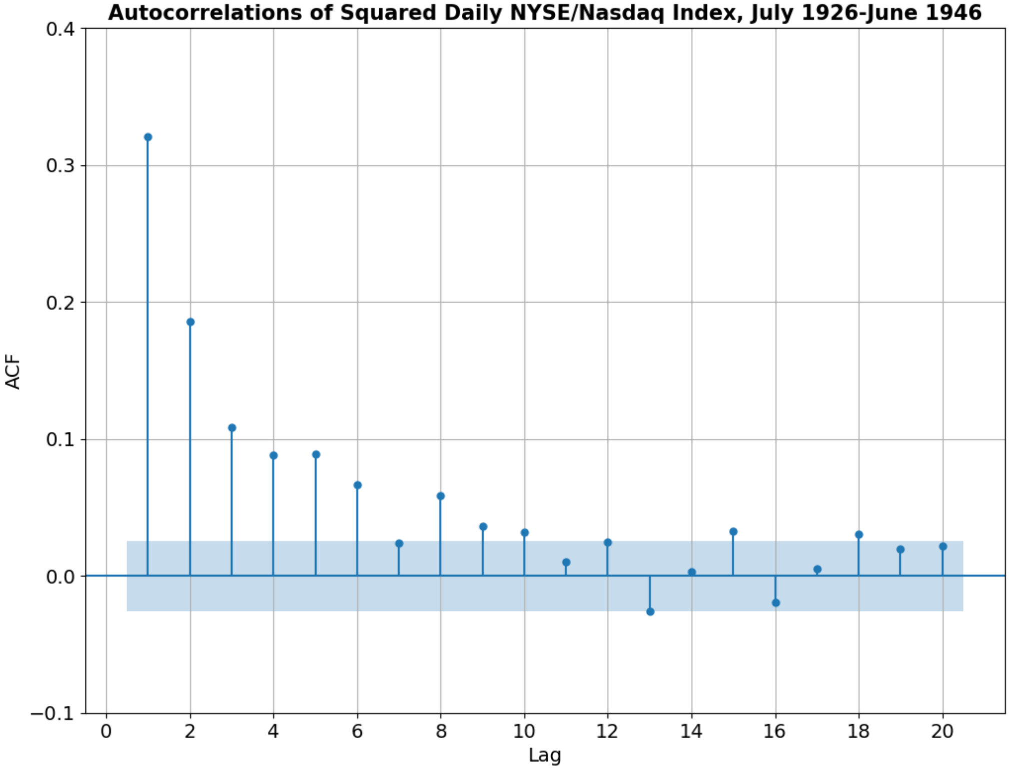

Figure 2: Sample ACF of Squared In-Sample NYSE/Nasdaq Index Series

trix P (see equation 7). Then, the likelihood function

$$\mathcal{L}(\boldsymbol{\theta}|I_T) = \prod_{t=1}^{T} f(a_t|\boldsymbol{\theta}, I_{t-1}) \tag{12}$$

where I_t denotes the set of all measurable variables available at time t. As noted by Ardia et al. (2019), for Markov-switching GARCH models, the conditional density of a_t can be expressed as

$$f(a_t|\boldsymbol{\theta}, I_{t-1}) = \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij} z_{i,t-1} f_{\mathcal{D}}(a_t|s_t = j, \boldsymbol{\theta}, I_{t-1}) \tag{13}$$

with p_{ij} is defined in equation 6, $z_{i,t-1} = \mathbb{P}(s_{t-1} = i|\boldsymbol{\theta}, I_{t-1})$ is the filtered probability of being in regime i at time $t-1$, and $f_{\mathcal{D}}(a_t|s_t = j, \boldsymbol{\theta}, I_{t-1})$ is the density of a_t conditional upon being in regime j with parameter vector $\boldsymbol{\theta}$. Note that $z_{i,t-1}$ can be obtained via Hamilton's filter; see Hamilton (1989) or Hamilton (1994) for further details. Maximizing the log of this likelihood yields the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$. In practise, we use the 'MSGARCH' package in R[4] (Ardia et al., 2019) to perform estimation of Markov-Switching GARCH-EGARCH models.

---

[4] This package is accessible at https://cran.r-project.org/web/packages/MSGARCH/index.html.

2.3.2 In-Sample Markov-Switching GARCH-EGARCH Model Selection

Based on the discussion above, there are multiple possible Markov-Switching EGARCH-GARCH specifications to choose from. Table 1 summarizes the log-likelihood and Akaike Information Criterion (AIC) values we obtain from fitting all possible specifications on our in-sample NYSE/Nadaq index series. 'σ_t^2' refers to the variance dynamics. '$\boldsymbol{\varepsilon_t}$' refers to the error distribution. 'R1' denotes regime 1, while 'R2' denotes regime 2. Mathematically,

Table 1: Comparison of Markov-Switching Fits on 1926-1946 NYSE/Nasdaq Index Series

σ_t^2 – **R1**	σ_t^2– **R2**	$\boldsymbol{\varepsilon_t}$ – **R1**	$\boldsymbol{\varepsilon_t}$ – **R2**	**Log-Lik.**	**AIC**
EGARCH(1,1)	EGARCH(1,1)	Std. Student-t	Std. Student-t	-8276	16575
GARCH(1,1)	EGARCH(1,1)	Std. Student-t	Std. Student-t	-8279	16579
GARCH(1,1)	EGARCH(1,1)	Normal	Std. Student-t	-8287	16593
EGARCH(1,1)	EGARCH(1,1)	Std. Student-t	Normal	-8287	16595
EGARCH(1,1)	GARCH(1,1)	Normal	Std. Student-t	-8295	16609
GARCH(1,1)	EGARCH(1,1)	Std. Student-t	Normal	-8298	16616
EGARCH(1,1)	GARCH(1,1)	Std. Student-t	Std. Student-t	-8300	16622
EGARCH(1,1)	GARCH(1,1)	Std. Student-t	Normal	-8313	16645
EGARCH(1,1)	EGARCH(1,1)	Normal	Normal	-8326	16672
GARCH(1,1)	GARCH(1,1)	Normal	Std. Student-t	-8331	16681
EGARCH(1,1)	EGARCH(1,1)	Normal	Std. Student-t	-8330	16682
GARCH(1,1)	EGARCH(1,1)	Normal	Normal	-8334	16687
GARCH(1,1)	GARCH(1,1)	Std. Student-t	Std. Student-t	-8339	16698
GARCH(1,1)	GARCH(1,1)	Std. Student-t	Normal	-8355	16728
EGARCH(1,1)	GARCH(1,1)	Normal	Normal	-8391	16800
GARCH(1,1)	GARCH(1,1)	Normal	Normal	-8428	16871

$$AIC = 2k - 2\ln\left(\mathcal{L}(\boldsymbol{\theta}|I_T)\right) \tag{14}$$

where k is the number of estimated parameters in the model and $\mathcal{L}(\boldsymbol{\theta}|I_T)$ is the likelihood (refer back to equation 12). As Vrieze (2012) explains, asymptotically, choosing the model with the lowest AIC minimizes prediction error. Therefore, we will choose the specification in table 1 which minimizes AIC: the one with EGARCH(1,1) dynamics and standardized student-t distributed errors in both regimes. This is our MS-EGARCH model.

2.3.3 MS-EGARCH Estimation Results

MS-EGARCH is fully specified by the EGARCH parameters in equation 11, the transition matrix parameters in equation 7, as well as two regime-specific ν_k. Table 2 presents estimates, and standard errors[5] for each of these parameters based upon our in-sample NYSE/Nasdaq index series.

---

[5] The standard errors are computed in MSGARCH based on standard formulas for the asymptotic standard errors of maximum likelihood estimators; see Douc et al. (2004) for details and rigorous proofs.

Table 2: MS-EGARCH Parameter Estimates based on 1926-1946 NYSE/Nasdaq Series

Parameter	**Estimate**	**Standard Error**
$\alpha_{0,1}$	-0.001	0.003
$\alpha_{1,1}$	0.133	0.149
$\alpha_{2,1}$	-0.038	0.044
$\beta_{1,1}$	0.997	0.011
ν_1	4.327	0.655
$\alpha_{0,2}$	0.007	0.012
$\alpha_{1,2}$	0.172	0.077
$\alpha_{2,2}$	-0.244	0.063
$\beta_{1,2}$	0.910	0.059
ν_2	7.291	4.020
p_{11}	0.997	0.012
p_{21}	0.005	0.004

For comparison, we also employ MATLAB's 'Econometrics Toolbox' (LeSage, 1998) to estimate a single regime EGARCH(1,1) model (refer back to equation 4) with standardized student-t errors. Table 3 presents these parameter estimates and standard errors.

Table 3: EGARCH(1,1) Parameter Estimates based on 1926-1946 NYSE/Nasdaq Series

Parameter	**Estimate**	**Standard Error**
α_0	0.003	0.002
α_1	0.208	0.017
α_2	-0.079	0.010
β_1	0.984	0.003
ν	4.701	0.300

Comparing the results in tables 2 and 3, we make the following observations:

i) $\widehat{\upsilon}_1 < \widehat{\nu}_2$, which implies that regime 1 has fatter tails compared to regime 2. We will later show that this fact also holds out-of-sample.

ii) The leverage effect coefficient, α_2 is negative in all cases as expected. Interestingly though, $\widehat{\alpha}_{2,2} < \widehat{\alpha}_{1,2}$ which suggests that the leverage effect is more pronounced in the volatile regime. This agrees with Henry (2009)'s findings.

iii) The estimated transition probabilities p_{11} and p_{21} imply that the regimes are highly persistent. Moreover, based on the estimated transition probability matrix, the stationary distribution row vector[6] is $\begin{bmatrix}0.641 & 0.359\end{bmatrix}$. In other words, our MS-EGARCH estimates imply that 64.1% of trading days are spent in the tranquil state while the remaining 35.9% are spent in the volatile state.

---

[6] This is the row vector $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}\widehat{P} = \boldsymbol{\pi}$ and $\iota'\boldsymbol{\pi} = 1$, where ι is a vector of ones, and $\widehat{P}$ is the estimated transition probability matrix.

Furthermore, the in-sample log-likelihood and AIC values of the fitted single EGARCH model were -8314, and 16638 respectively; MS-EGARCH's AIC value of 16575 is lower and its log-likelihood of -8276 is higher, which implies that it fits the in-sample data better. Further evidence of the MS-EGARCH's strong in-sample fit comes from examining $\tilde{z}_{i,t} = \mathbb{P}(s_t = i|\widehat{\boldsymbol{\theta}}, I_T)$, MS-EGARCH's smoothed regime probabilities, which are obtained via Hamilton's smoother; consult Hamilton (1989) or Hamilton (1994) for further insights. Figure 3 presents the smoothed regime 2 probabilities. Additionally, figure 3 annotates Black Tuesday (October 1929), the introduction of FDR's Social Security Act (August 1935), the German occupation of Czechoslovakia (March 1939), and the end of World War 2 (September 1945). It is evident in figure 3 that these historical events align with spikes in the smoothed turbulent state probability. Thus, MS-EGARCH appears to be detecting economically meaningful regimes in-sample.

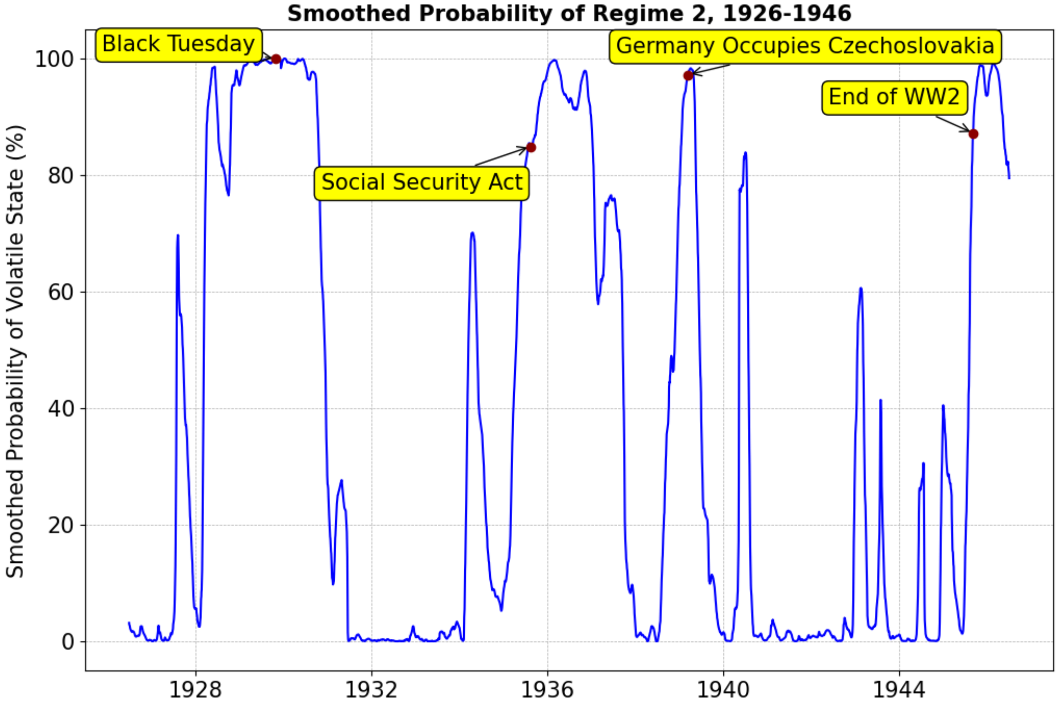

Figure 3: Smoothed MS-EGARCH In-Sample Regime 2 Probabilities

Chapter 3

Relation between MS-EGARCH Volatility Forecasts and VIX

This chapter focuses on relating MS-EGARCH's out-of-sample predictive capabilities to the CBOE (Chicago Board Options Exchange) VIX Index. Section 3.1 details our methodology for MS-EGARCH forecasting. Section 3.2 describes how we use the VIX to produce realized volatility forecasts. Finally, section 3.3 discusses our quantitative results.

3.1 MS-EGARCH Realized Volatility Forecasting

In this chapter we switch from the NYSE/Nasdaq index to the daily S&P500 log percentage returns series, which we also derive from CRSP (Center for Research in Security Prices, 2023) data. For our study, this series extends from 1971 to December 2022. However, the observations and model selection detailed in the previous chapter (section 2.3) are still relevant. This is because the NYSE/Nasdaq index is value-weighted, so its performance mirrors that of the S&P 500 index. In fact, our calculations reveal a 98.9% correlation between the log daily returns of these two indices from 1971 to 2022.

For single-regime GARCH(1,1) and EGARCH(1,1) models, volatility forecasting is a straightforward, recursive procedure. For instance, Tsay (2010) notes that the 1-step ahead volatility forecast of a GARCH(1,1) model (refer back to equation 3) at time (origin) t is given by

$$\sigma^2_{t,GAR}(1) = \alpha_0 + \alpha_1 a_t^2 + \beta_1 \sigma^2_{t,GAR} \tag{15}$$

where $a_t = r_t - u_t$ and $\sigma_{t,GAR}$ (the fitted conditional variance) are known at time t. For $h > 1$, the h-step ahead GARCH(1,1) conditional volatility forecast is given by the following recursive formula:

$$\sigma^2_{t,GAR}(h) = \alpha_0 + (\alpha_1 + \beta_1)\sigma^2_{t,GAR}(h-1),\ h > 1 \tag{16}$$

A similar recursive formula can be derived for an EGARCH(1,1) model, though the exact form in that case does depend on the ε_t distribution assumed (see section 3.8.4 of Tsay (2010) for further details).

With Markov-Switching GARCH-EGARCH models, the forecasting procedure is rendered more challenging by the underlying state process s_t, and the existence of a separate conditional variance process in each regime. We continue to employ the MSGARCH package for forecasting which employs a closed-form equation for 1-step ahead MS-EGARCH forecasts; see Ardia et al. (2019) and A.3 of Haas et al. (2004) for further details. For multi-step ahead forecasts (i..e, $h > 1$), the MSGARCH package utilizes simulations of the demeaned return process, $a_t = r_t - u_t$. The package draws a new observation from the one-step ahead predictive distribution[1], updates the 2 conditional variance processes accordingly, and then redraws another new observation from the one-step ahead predictive distribution, iterating for as many steps as needed based on the value of h. This iterative procedure produces a simulated value of a^s_{t+h} for each simulation s. MSGARCH uses the variance of a^s_{t+h} across all simulations as the h-step ahead conditional variance forecast at origin t. 10,000 simulations are performed.

In our analysis, we wish to work with daily log return series to forecast future volatility 1, 5, 10, and 22 days ahead. However, for multi-step forecasts, such as the 22-day horizon, our interest is not in the instantaneous daily conditional volatility on day 22. Rather, we want to predict the volatility in the entire 22-day period following time t. As such, we need to aggregate our daily volatility forecasts. Mathematically, let H denote the horizon we are interested in, and let $\sigma^2_{t,MSEGAR}(h)$ denote the h-step ahead daily variance forecast produced by the MSGARCH package via the process described above. Our horizon H cumulative volatility forecast at time (origin) t can be expressed as follows:

$$\sigma_{t,MSEGAR}(H) = \sqrt{\sum_{h=1}^{H} \sigma^2_{t,MSEGAR}(h)} \tag{17}$$

In the special case of $H = 1$, this formula simplifies to $\sqrt{\sigma^2_{t,MSEGAR}(1)}$.

Our out-of-sample period spans from 1991 to 2022. Therefore, for each trading day during this period, we re-fit an MS-EGARCH model on data from the prior 20 years (5040 trading days) and then use the procedure described above to forecast S&P500 volatility 1, 5, 10, and 22 days ahead. The rolling window length of 20-years (5040 observations) was chosen a priori, with the goal allowing the model to observe a sufficient number of economic cycles; it is also similar to the in-sample length of 4095 observations that Marcucci (2005) utilized. Finally, note that when fitting our MS-EGARCH model, we use the values in table 2 as initial guesses.

3.2 VIX-Based Realized Volatility Forecasting

Recall that $\{r_t\}$ is the daily log S&P500 return series. Since $\mathbb{E}(r_t) \approx 0$, a consistent, albeit quite noisy measure of daily volatility is $\sqrt{r_t^2}$ (Tsay, 2010). We shall proceed with this measure[2]. In addition to daily S&P500 log returns, we also obtain daily VIX data, spanning

---

[1] The one-step ahead predictive distribution was also derived by Haas et al. (2004).

[2] Andersen and Bollerslev (1998) detail volatility measures that can be derived from intra-day data; we avoid these to ensure our findings are applicable to indices and periods lacking quality intra-day data.

from 1990 to the end 2022, from Bloomberg (Bloomberg L.P., 2023). The VIX index represents the (options) market's expectation for realized S&P500 volatility in the next 30 days. However, is is well-known that the VIX is typically higher than future volatility because of an embedded risk premium, often termed the 'variance premium' (Bekaert and Hoerova, 2014).

To adjust for scaling and embedded risk premiums, Blair et al. (2001) and Christensen and Prabhala (1998) propose fitting a linear model between realized volatility and implied volatility. We employ such a linear model (hereon referred to as 'Linear VIX'), which can be written as follows

$$\sqrt{r_t^2} = \alpha + \beta_{VIX} \cdot VIX_t + u_t \tag{18}$$

where VIX_t denotes the daily value of the VIX index on day t and the u_t are assumed to be exogenous errors.

For each trading day from 1991 to the end of December 2022, we estimate α and β_{VIX} via OLS (Ordinary Least Squares) using a rolling window of the previous 252 trading day's log S&P500 returns. We would ideally have liked to use a rolling window of 5040 days to be consistent with our MS-EGARCH procedure. However, this would have afforded us less out-of-sample data. Using our rolling $\widehat{\alpha}$ and $\widehat{\beta}_{VIX}$, our horizon H forecast of future realized S&P500 volatility at time (origin) t is given by the following equation:

$$\sqrt{H} \cdot \left[\widehat{\alpha} + \widehat{\beta}_{VIX} \cdot VIX_t\right] \tag{19}$$

3.3 Quantitative Results

3.3.1 Volatility Forecasting Metrics

For the 1-step ahead horizon, we define

$$h_{t+1} = \sqrt{r_{t+1}^2} \tag{20}$$

as the true realized 1-step ahead volatility. As discussed above, this is a consistent, but noisy measure. For the multi-step ahead horizons (5, 10, and 22-days ahead), letting $\bar{r}_t = \frac{1}{H}\sum_{h=1}^{H} r_{t+h}$, we define

$$h_{t+H} = \sqrt{\frac{H}{H-1} \cdot \sum_{h=1}^{H} (r_{t+h} - \bar{r}_t)^2} \tag{21}$$

as the true realized H-step ahead volatility. We can compare these true volatilities to the MS-EGARCH and Linear VIX forecasts at each time step t within our out-of-sample period. To evaluate the performance of our forecasting models, we will consider the mean-absolute error (MAE) metric, which is defined as

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |\widehat{h}_{t+H} - h_{t+H}| \tag{22}$$

where T is the length of the out-of-sample period (in this case, 8557), $\widehat{h}_{t+H}$ is the horizon H volatility prediction made at time t and h_{t+H} is the true horizon H realized volatility. Another forecasting error metric is root-mean-squared error (RMSE), which is defined as

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\widehat{h}_{t+H} - h_{t+H})^2} \tag{23}$$

with all terms defined the same as in equation 22. These two forecasting evaluation metrics are extremely common. As Chai and Draxler (2014) discuss, neither metric is, in general, superior. The two metrics have different interpretations: MAE weighs all errors equally whereas RMSE penalizes larger errors more. We will consider both metrics.

3.3.2 Volatility Forecasting Results

Using MAE and RMSE, we evaluate both our MS-EGARCH model and the Linear VIX model across the entire out-of-sample period (1991-2022). We will also test a naive model (hereon referred to as 'Naive') that forecasts volatility across horizon H using the the standard deviation of daily returns in the prior 252 days, scaled by $\sqrt{H}$. Table 4 presents our results. Inspecting the numbers in table 4, we make the following two comments:

Table 4: S&P500 Volatility Forecasting Results, 1991-2022

	1-step		**5-step**		**10-step**		**22-step**	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MS-EGARCH	0.569	0.771	0.744	**1.021**	0.883	**1.303**	1.263	**1.976**
Linear VIX	**0.540**	**0.761**	**0.691**	1.062	**0.866**	1.441	**1.185**	2.072
Naive	0.660	0.896	1.001	1.440	1.252	1.895	1.726	2.678

i) Both MS-EGARCH and Linear VIX outperform Naive across all horizons and both metrics. This suggests that there is value in either adopting a sophisticated model to forecast realized volatility as a function of past volatility (MS-EGARCH) or in using market information to forecast volatility (Linear VIX).

ii) Generally, MS-EGARCH and Linear VIX obtain similar performance. For all horizons except 1-step ahead, MS-EGARCH earns moderately better (lower) RMSE; conversely, Linear VIX earns slightly better (lower) MAE across all horizons.

Figure 4 plots the 10-step ahead forecast errors of Linear VIX and MS-EGARCH across the entire out-of-sample period. For brevity, we shall not display plots for the other horizons as those plots are generally similar. Figure 4 also annotates the dot-com bubble pop (April 2000), the financial crisis (October 2008), and the COVID-19 market crash (March 2020).

Inspecting Figure 4, it appears as though Linear VIX has a greater tendency to under-predict volatility. In contrast, MS-EGARCH has a greater tendency to over-predict volatility, particularly immediately following large shocks (e.g., following the Covid-19 market crash). Overall though, the two models' forecasts appear similar. As further evidence of this, table 5 contains the models' forecast correlations for each of the horizons, all of which are above 90%.

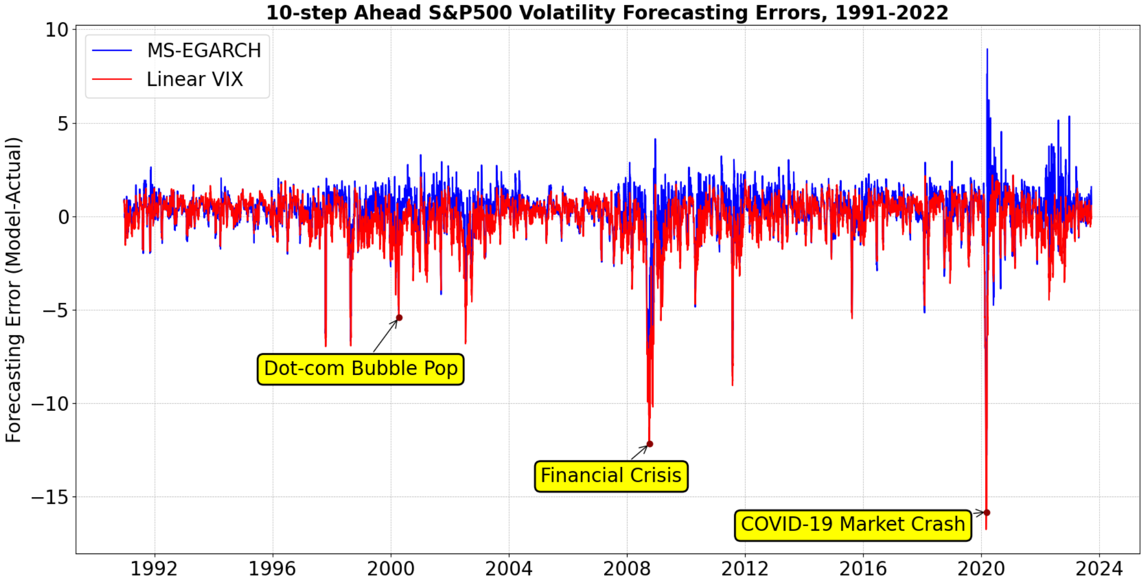

Figure 4: MS-EGARCH and Linear-VIX Out-of-Sample Forecasting Errors

Table 5: MS-EGARCH and Linear VIX Forecast Correlations, 1991-2022

Forecasting Horizon	**Sample Correlation**
1-step	91.0%
5-step	92.3%
10-step	92.1%
22-step	90.8%

3.3.3 Diebold-Mariano Tests of Superior Predictive Ability

For rigor, we examine whether the results in table 4 are statistically significant by performing Diebold-Mariano Tests. The Diebold-Mariano test, proposed by Diebold and Mariano (1995), examines whether two competing models, i and j, possess equal predictive ability. Let $\{e_{i,t}\}$ and $\{e_{j,t}\}$ be the series of forecasting errors of the two models. Assume we can define the loss differential series between the two competing models as $d_t = [g(e_{i,t}) - g(e_{j,t})]$ where $g(\cdot)$ is the loss function for a single forecast error. For MAE, $g(x) = |x|$, while for mean-squared error (MSE)[3], $g(x) = x^2$. Diebold and Mariano (1995) show that, assuming the series $\{d_t\}$ is weakly stationary and has decaying dependence, $\sqrt{T}(\bar{d} - \mu) \Rightarrow \mathcal{N}(0, V)$, where T is the sample size, and V is the long-run variance of d_t. We shall employ the

---

[3] We use MSE here because it is not possible to formulate RMSE in terms of individual forecasts errors since the square root is applied after the mean.

following estimator of V (Newey and West, 1987):

$$\widehat{V} = \widehat{\gamma}_0 + 2\sum_{k=1}^{S_T}(1 - \frac{k}{S_T})\widehat{\gamma}_k \tag{24}$$

where $\widehat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-K}(d_t - \bar{d})(d_{t+k} - \bar{d})$. For the bandwidth parameter, S_T, we follow Stock and Watson (2003)'s rule of thumb and set it to $0.75T^{\frac{1}{3}}$; in this case, that means $S_T = 15$. The null is that the two models have equal predictive ability and it is formulated mathematically as $H_0 : \mu = 0$. To test this null, we examine the Diebold-Mariano (DM) test-statistic:

$$DM = \frac{\bar{d}}{\sqrt{\frac{\widehat{V}}{T}}} \overset{H_0}{\Rightarrow} \mathcal{N}(0,1) \tag{25}$$

MAE Diebold-Mariano Tests

Let's first test the null that both MS-EGARCH and Linear-VIX have equivalent MAE to the Naive model. Table 6 presents the relevant DM test-statistics and p-values[4].

Table 6: S&P500 MAE Diebold-Mariano Tests, MS-EGARCH/Linear-VIX vs. Naive

	1-step		**5-step**		**10-step**		**22-step**	
Model	DM	p-value	DM	p-value	DM	p-value	DM	p-value
MS-EGARCH	8.88	**0.00**	9.47	**0.00**	9.75	**0.00**	8.32	**0.00**
Linear VIX	12.22	**0.00**	11.66	**0.00**	9.95	**0.00**	8.89	**0.00**

At a 5% significance level, we easily reject the null in all cases. MS-EGARCH and Linear VIX's MAE out-performances of the Naive model across all horizons are highly statistically significant. Table 7 tests whether MS-EGARCH and Linear VIX earn equivalent MAE.

Table 7: S&P500 MAE Diebold-Mariano Tests, Linear-VIX vs. MS-EGARCH

1-step		**5-step**		**10-step**		**22-step**	
DM	p-value	DM	p-value	DM	p-value	DM	p-value
7.42	**0.00**	4.10	**0.00**	0.83	0.40	2.41	**0.02**

Analyzing the results in table 7, Linear VIX's MAE out-performances of MS-EGARCH are statistically significant at all horizons except $H = 10$.

MSE Diebold-Mariano Tests

Table 8 displays the relevant Diebold-Mariano test statistics and p-values for assessing whether MS-EGARCH and Linear-VIX earn superior MSE than the Naive model. Based on the results, we reject the null that MS-EGARCH and Linear VIX are equivalent in MSE to the Naive model. Next, in table 9, we assess whether MS-EGARCH's out-performances in

Table 8: S&P500 MSE Diebold-Mariano Tests, MS-EGARCH/Linear-VIX vs. Naive

	1-step		**5-step**		**10-step**		**22-step**	
Model	DM	p-value	DM	p-value	DM	p-value	DM	p-value
MS-EGARCH	5.80	**0.00**	5.96	**0.00**	5.93	**0.00**	6.21	**0.00**
Linear VIX	8.48	**0.00**	7.40	**0.00**	6.28	**0.00**	5.99	**0.00**

Table 9: S&P500 MSE Diebold-Mariano Tests, MS-EGARCH vs. Linear VIX

1-step		**5-step**		**10-step**		**22-step**	
DM	p-value	DM	p-value	DM	p-value	DM	p-value
-0.79	0.43	0.98	0.33	2.27	**0.02**	1.49	0.14

MSE over Linear VIX are statistically significant. At a significance level of 5%, we fail to reject the null that MS-EGARCH and Linear VIX are equivalent in terms of MSE for all horizons except $H = 10$. Therefore, MS-EGARCH's MSE out-performance of Linear VIX is only statistically significant for 10-day ahead volatility forecasts.

Commentary

Diebold-Mariano tests validate that MS-EGARCH and Linear VIX outperform the Naive model, which uses a trailing standard deviation of log S&P500 returns. The tests also confirm that Linear VIX's MAE out-performances of MS-EGARCH are statistically significant for all horizons except 10-days ahead. Conversely, from an MSE standpoint, the tests suggest that MS-EGARCH is a superior volatility forecaster for the 10-day ahead horizon. Ultimately though, particularly given how correlated the models' forecasts are, our main takeaway is that, despite being quite different models, MS-EGARCH and Linear VIX perform comparably. This suggests that MS-EGARCH and the VIX Index contain similar information about future volatility. Hence, MS-EGARCH is especially useful for indices without liquid and efficient option markets.

---

[4] When labelling tables, we hereon follow the convention that 'model 1 vs. model 2' means that significant, positive values imply out-performance by model 1.

Chapter 4

Comparison of MS-EGARCH Forecasts to Single-Regime EGARCH/GARCH

In this chapter, we prove MS-EGARCH's superiority over single-regime models in out-of-sample forecasting. Section 4.1 briefly summarizes our data and methodology, borrowing from earlier sections. Section 4.2 presents our main NYSE/Nasdaq index forecasting results. Subsequently, section 4.3 details the results of more Diebold-Mariano tests. Section 4.4 examines how forecasting performance varies by regime. Finally, section 4.5 examines whether MS-EGARCH can predict the errors of single-regime EGARCH.

4.1 Data and Methodology

For this forecasting comparison, we revert to the NYSE/Nasdaq index series, whose construction is detailed in Section 2.3.1. We prefer this index because it is not purely a large-cap index, it factors returns from distributions, and we were able to obtain a more extensive history for it. We only employ the S&P500 in the previous chapter because doing so is necessary to compare to a VIX-based model.

In this chapter, our focus is on generating 1,5,10, and 22 day ahead volatility forecasts for every day in the out-of-sample period, which commences in July, 1946 and terminates in December, 2022 (19,441 observations). Our MS-EGARCH forecasting procedure is unchanged from the prior chapter (refer back to section 3.1). We also now estimate GARCH(1,1) and EGARCH(1,1) models for every trading day in the out-of-sample period, using a rolling 20-year (5040 day) window. These models are estimated using MATLAB's 'Econometrics Toolbox' (LeSage, 1998). EGARCH(1,1) and GARCH(1,1) forecasting was also described in section 3.1 and is easy to implement in MATLAB.

We can define the true volatility of the NYSE/Nasdaq index in the same manner as we did for the S&P500 index, and then compute MAE and RMSE values for each of our models. For this forecasting comparison, we will again include a 'Naive' baseline model, which forecasts volatility across horizon H using the standard deviation of daily log returns in the prior 252 days, scaled by $\sqrt{H}$.

4.2 Forecasting Results

Table 10 presents the 1, 5, 10, and 22-day ahead MAEs and RMSEs of all our models across the entire out-of-sample period. Examining the results, all the GARCH-type models outperform the Naive model in terms of MAE and RMSE across all horizons. The performance of the three GARCH family models is similar, which makes sense because they are, ultimately, similar models. Nevertheless, unlike in the prior section when we were comparing two different models, our goal in this section is to distinguish amongst these similar models. With this objective in mind, MS-EGARCH performs the best, followed by EGARCH, and then GARCH. This ranking of the three GARCH models is the same across all metrics and horizons.

Table 10: NYSE/Nasdaq Index Volatility Forecasting Results, 1946-2022

	1-step		**5-step**		**10-step**		**22-step**	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MS-EGARCH	**0.471**	**0.652**	**0.652**	**0.915**	**0.769**	**1.189**	**1.096**	**1.765**
EGARCH	0.475	0.653	0.667	0.943	0.798	1.237	1.143	1.821
GARCH	0.489	0.674	0.701	0.982	0.836	1.252	1.170	1.827
Naive	0.529	0.738	0.824	1.193	1.013	1.559	1.382	2.184

EGARCH's out-performance of GARCH is unsurprising. EGARCH improves upon GARCH by factoring in the asymmetric effect negative returns have upon volatility (i.e., the leverage effect). Promisingly, MS-EGARCH's degree of out-performance over EGARCH appears comparable to EGARCH's out-performance over GARCH. This suggests that incorporating Markov-Switching is as useful an innovation as accounting for the leverage effect.

Figure 5 contrasts the 5-day ahead forecasting errors of MS-EGARCH and EGARCH throughout the entire out-of-sample period. This plot also highlights four significant historical events: the World War II recovery in the early 1950s, Black Monday in October 1987, the 2008 financial crisis, and the March 2020 COVID-19 crash. Ultimately, in the plot, both models show similar performance, with a notable exception during the World War II recovery period, where EGARCH tends to over-predict 5-day volatility significantly. This pattern is also observed in the 1, 10, and 22 day ahead forecast plots, which are excluded from this report for brevity.

4.3 Diebold-Mariano Tests of Superior Predictive Ability

We now test for the statistical significance of the results in table 10 using Diebold-Mariano tests. Refer back to section 3.3.3 for a reminder of the theory behind these tests. In this case, the band-width parameter $S_T = 0.75T^{\frac{1}{3}} \approx 20$.

4.3.1 MAE Diebold-Mariano Tests

We first test the null that each of GARCH-type models have equivalent MAE to the naive model. Table 11 depicts the relevant Diebold-Mariano (DM) test-statistics and p-values for

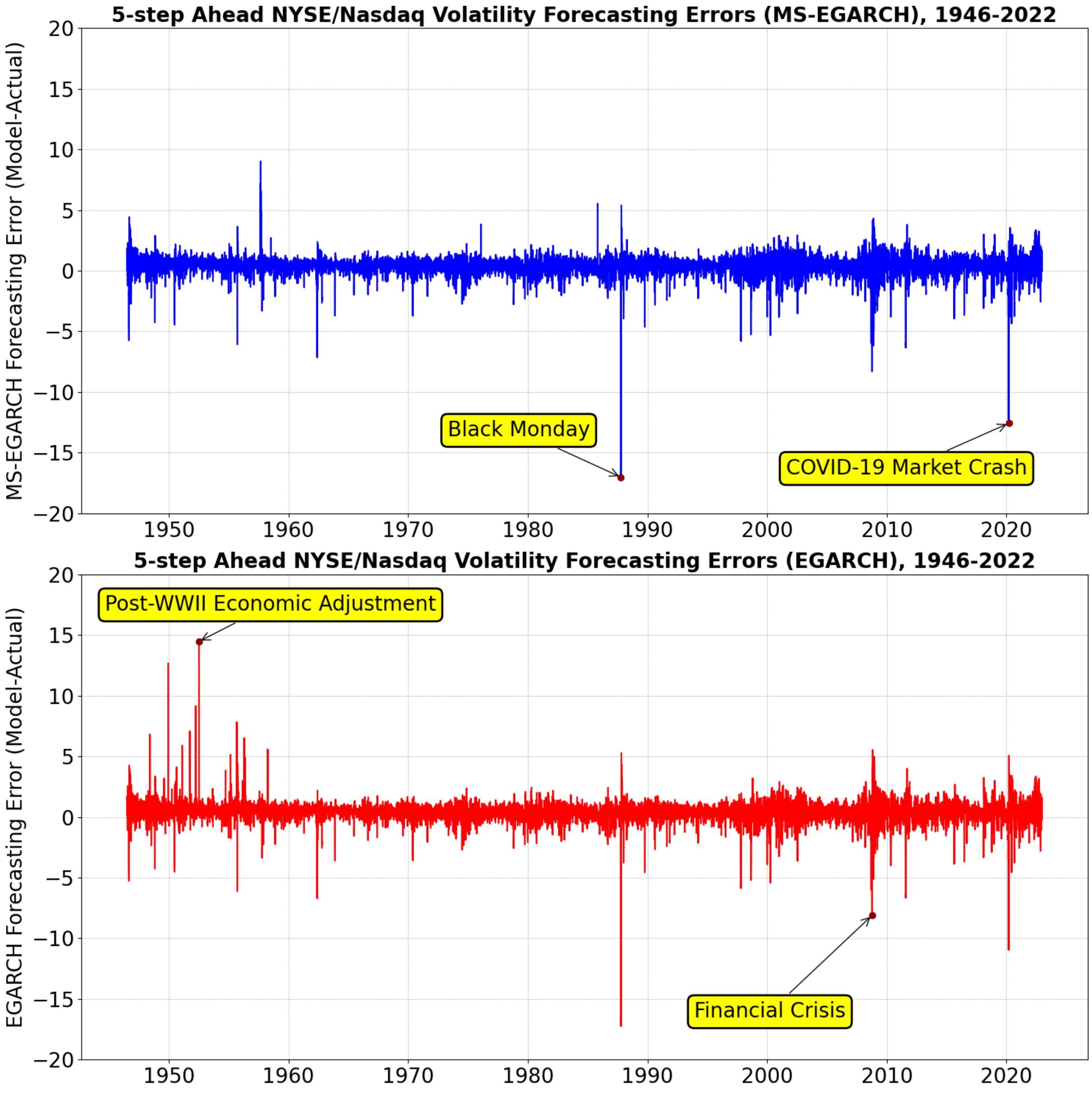

Figure 5: MS-EGARCH and EGARCH NYSE/Nasdaq Out-of-Sample Forecasting Errors

these tests. Based on these test-statistics and p-values, we reject all the nulls. All three GARCH-type models' MAE out-performances of Naive across all four horizons are highly statistically significant.

Table 11: NYSE/Nasdaq Index MAE Diebold-Mariano Tests, GARCH models vs. Naive

	1-step		**5-step**		**10-step**		**22-step**	
Model	DM	p-value	DM	p-value	DM	p-value	DM	p-value
MS-EGARCH	9.53	**0.00**	10.49	**0.00**	10.63	**0.00**	8.19	**0.00**
EGARCH	8.72	**0.00**	9.05	**0.00**	8.63	**0.00**	6.20	**0.00**
GARCH	6.48	**0.00**	7.25	**0.00**	7.09	**0.00**	5.33	**0.00**

Next, we test whether MS-EGARCH's MAE out-performances of EGARCH, the best single-regime model, across the four horizons, are statistically significant. Based on the relevant p-values of these tests in table 12, we reject the nulls that MS-EGARCH and single-regime EGARCH obtain equivalent MAEs. MS-EGARCH's MAE out-performances of single-regime EGARCH are statistically significant for 1,5,10, and 22 day ahead NYSE/-Nasdaq index volatility forecasts.

Table 12: NYSE/Nasdaq Index MAE Diebold-Mariano Tests, MS-EGARCH vs. EGARCH

1-step		**5-step**		**10-step**		**22-step**	
DM	p-value	DM	p-value	DM	p-value	DM	p-value
2.61	**0.01**	3.39	**0.00**	3.66	**0.00**	3.37	**0.00**

4.3.2 MSE Diebold-Mariano Tests

We commence by testing the null that each of the GARCH family models have equivalent MSE[1] to the Naive model. Table 13 demonstrates the test-statistics and p-values for these tests. Based on these p-values, we reject the null that each GARCH model has equivalent MSE to the Naive model for all four horizons. In other words, the GARCH family models' MSE out-performances of the Naive model are statistically significant.

Table 13: NYSE/Nasdaq Index MSE Diebold-Mariano Tests, GARCH models vs. Naive

	1-step		**5-step**		**10-step**		**22-step**	
Model	DM	p-value	DM	p-value	DM	p-value	DM	p-value
MS-EGARCH	6.82	**0.00**	6.96	**0.00**	6.67	**0.00**	5.76	**0.00**
EGARCH	6.56	**0.00**	5.77	**0.00**	5.02	**0.00**	4.52	**0.00**
GARCH	5.15	**0.00**	5.17	**0.00**	4.91	**0.00**	4.10	**0.00**

Next, table 14 shows the DM test-statistics and corresponding p-values for tests of the null that MS-EGARCH and EGARCH have equivalent MSE. Inspecting these values, at

---

[1] Recall that RMSE is incompatible with Diebold-Mariano tests so we use MSE, which is directly proportional to RMSE.

a 5% significance level, we fail to reject the null that MS-EGARCH and EGARCH have equivalent MSE for the 1-step and 22-step ahead volatilility forecasts. However, we do reject the null that they have equivalent MSE for the 5-step and 10-step ahead volatility forecasts; for these horizons, we conclude that MS-EGARCH is superior from an MSE standpoint. MS-EGARCH's MSE out-performance over EGARCH is less pronounced than its MAE out-performance. Nevertheless, overall, we still assess it to be the superior model according to both metrics.

Table 14: NYSE/Nasdaq Index MSE Diebold-Mariano Tests, MS-EGARCH vs. EGARCH

1-step		**5-step**		**10-step**		**22-step**	
DM	p-value	DM	p-value	DM	p-value	DM	p-value
0.49	0.63	2.11	**0.04**	2.10	**0.04**	1.62	0.11

4.4 Regime-Conditional Results

4.4.1 Raw Results

Table 15: Regime 1 NYSE/Nasdaq Index Volatility Forecasting Results

	1-step		**5-step**		**10-step**		**22-step**	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MS-EGARCH	**0.388**	**0.518**	**0.596**	**0.827**	**0.724**	**1.136**	**1.049**	**1.728**
EGARCH	0.402	0.529	0.647	0.898	0.809	1.240	1.208	1.874
GARCH	0.408	0.536	0.649	0.864	0.801	1.143	1.162	1.734

Table 16: Regime 2 NYSE/Nasdaq Index Volatility Forecasting Results

	1-step		**5-step**		**10-step**		**22-step**	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MS-EGARCH	0.539	0.743	0.698	0.981	0.806	**1.231**	1.133	1.794
EGARCH	**0.534**	**0.739**	**0.683**	**0.979**	**0.789**	1.233	**1.090**	**1.776**
GARCH	0.555	0.768	0.743	1.070	0.865	1.334	1.177	1.900

In this section, we examine how the forecasting performance of the GARCH family models varies across the two MS-EGARCH regimes[2]. To define the regime on day t, we utilize MS-EGARCH's filtered regime probabilities, $z_{i,t} = \mathbb{P}(s_t = i|\widehat{\boldsymbol{\theta}}, I_t)$, which are the probabilities that the regime (s_t) at time t is i, based on the information set I_t, and the MS-EGARCH parameter vector $\widehat{\boldsymbol{\theta}}.\widehat{\boldsymbol{\theta}}$ is obtained from fitting MS-EGARCH to the 5040 most recent daily NYSE/Nasdaq returns. The filtered probabilities are obtained via Hamilton's filter; consult

---

[2] We ignore the Naive model since its relative performance in each regime is not interesting.

Hamilton (1989) or Hamilton (1994) for more information. We classify time t as regime 2 if $z_{2,t} > 0.5$ and as regime 1 otherwise. With these classifications, we compute regime-specific MAE and RMSE statistics for each of the models.

Table 15 presents the forecasting results in regime 1 (the low volatility regime), whereas table 16 presents the forecasting results in regime 2 (the high volatility regime). We make the following observations:

Regime 1: In this regime, MS-EGARCH's out-performance of the other two models is quite pronounced. Recall that in 2.3.3, we found that in regime 1 and relative to EGARCH, MS-EGARCH had a smaller (in magnitude) leverage effect coefficient, in-sample. We hypothesize that this smaller coefficient helps MS-EGARCH out-perform EGARCH within this regime. As further evidence of this, regular GARCH out-performs EGARCH in most of the columns of table 15.

Regime 2: In this regime, MS-EGARCH still outperforms regular GARCH. However, in all columns of table 16 but one, it under-performs single-regime EGARCH. The out-performance of EGARCH over MS-EGARCH is less substantial in this regime than MS-EGARCH's out-performance of EGARCH in regime 1, which is why, overall, MS-EGARCH performs best. Nevertheless, the regime 2 under-performance of MS-EGARCH is puzzling as, in theory, MS-EGARCH should be able to adapt best to both regimes. We investigate the statistical significance of this result below.

4.4.2 Assessing Significance

Block Bootstrapping Framework

We are interested in the statistical significance of the differences between the models' performances within each regime. Here, Diebold-Mariano tests will not work as regime 2 and regime 1 are not contiguous in time, meaning we can not easily estimate the auto-correlation structure of the regime-specific loss differential series. Our solution is to perform a block bootstrap. Specifically, we sample 252 day blocks with replacement from our original NYSE/Nasdaq series to construct 100 artificial, bootstrapped samples. We refit our models on each artificial sample, produce forecasts on each of them along with 'true' volatility measures, and then use the $z_{i,t}$ to classify each sample's observations into the two regimes. We then use the standard deviations of the regime 1 and regime 2 differences in MAE and MSE[3] between our models across these 100 bootstrap samples to compute bootstrapped standard errors.

Ideally, we would like to construct bootstrap samples of the same length as our original series. However, this is impractical as we need to re-fit our GARCH family models on each of the bootstrapped samples. With $T = 19,441$ observations, our code takes over 20 hours to run. Another wrinkle is that the regime 1 and regime 2 MAE and MSE differences between models depend not on the overall out-of-sample period length, but instead on the number of out-of-sample observations classified in regime 1 and regime 2. These numbers depend on the MS-EGARCH model outputs (i.e., its filtered probabilities) and will, therefore, vary in each

---

[3] We are again using MSE because it has more straightforward asymptotics.

bootstrapped sample. Our solution to both of these problems is to perform an m-out-of-n bootstrap. Specifically, let the regime-i-specific estimator of interest be D_i (in this case D_i is the regime-specific difference in MSE and MAE between models). Let $D_{b,i}$ denote the regime-i-estimator value in bootstrapped sample b. Assume that our original sample has n_i occurrences of regime i and that each bootstrapped sample has m_i occurrences of regime i[4]. Our bootstrapped estimate of the sample size n_i standard error of D_i,

$$\widehat{S} = \sqrt{\frac{m_i}{n_i} \cdot \frac{1}{100} \sum_{b=1}^{100} (D_{b,i} - \bar{D}_i)^2} \tag{26}$$

where $\bar{D}_i = \frac{1}{100} \sum_{b=1}^{100} D_{b,i}$. Each of our bootstrapped samples consists of 2520 (out-of-sample) observations[5], which, results in $m_1 = 927$ and $m_2 = 882$. Furthermore, in our original sample, $n_1 = 10,719$ while $n_2 = 8722$. Admittedly, the $\frac{m_i}{n_i}$ scaling applied in equation 26 is imperfect for time-series data. Future researchers, with additional computational resources, could perform full n-out-n bootstraps.

MAE Signifiicance Results

Applying the block-bootstrapping procedure described above, we obtain bootstrap standard errors for the differences in MAE within both regimes between MS-EGARCH and the best singe-regime model, EGARCH. We start with the MAE differences. Table 17 contains the regime 1 results. The 'T-statistic' is the MAE difference divided by the standard error since we are testing the null that the true difference in MAE between the two models is zero for each horizon. To facilitate inference, we will assume that these MAE differences are asymptotically normal (when scaled by the $\sqrt{n_i}$, which is scaling we already accounted for in equation 26). Based on Diebold and Mariano (1995)'s findings, this is equivalent to assuming that the non-contiguous sequence of absolute error differentials between the two models (i.e., the loss differentials) in regime 1 is covariance-stationary and has short-term memory. Under this assumption, since all the T-statistics are greater than 1.96, we reject the nulls that MS-EGARCH and GARCH obtain equivalent MAE in regime 1. We conclude that MS-EGARCH's out-performances of single-regime EGARCH in regime 1 across all horizons are statistically significant.

Table 17: Regime 1 MAE Difference Significance, MS-EGARCH vs. EGARCH

Horizon	**MAE Difference**	**Bootstrap Standard Error**	**T-statistic**
1-step	0.014	0.003	4.798
5-step	0.052	0.010	4.957
10-step	0.085	0.012	6.880
22-step	0.158	0.023	6.734

---

[4] We ensure that m_i is constant by setting m_i to be the minimum number observations of regime i across all bootstrapped samples, and then only computing $D_{b,i}$ using the first m_i observations for all samples.

[5] To be exact, each bootstrapped sample consists of 7560 observations, but the first 5040 observations are used purely for the first model fits.

Next, we investigate whether single-regime EGARCH's MAE out-performances of MS-EGARCH in regime 2 are statistically significant. Table 18 contains the relevant statistics. Investigating these results, it is apparent that the bootstrapped standard errors for the MAE differences at all four horizons are larger than they were in regime 1. This is unsurprising; we have less observations of regime 2 and it is, by definition, the turbulent state. Ultimately, if we again assume asymptotic normality, then, at a 5% significance level, we fail to reject the null hypotheses that MS-EGARCH and EGARCH obtain equivalent 1, 5, 10, and 22 step ahead MAE in regime 2.

Table 18: Regime 2 MAE Difference Significance, EGARCH vs. MS-EGARCH

Horizon	**MAE Difference**	**Bootstrap Standard Error**	**T-statistic**
1-step	0.005	0.005	1.001
5-step	0.014	0.012	1.184
10-step	0.018	0.014	1.305
22-step	0.043	0.025	1.714

MSE Significance Results

We can also apply the block bootstrapping procedure to obtain MSE significance results. Table 19 contains the relevant statistics for the regime 1 MSE differences between MS-EGARCH and EGARCH. Inspecting the results and making the asymptotic normality assumption again, we reject the null in all instances and conclude that MS-EGARCH's MSE out-performances of EGARCH in regime 1 are statistically significant.

Table 19: Regime 1 MSE Difference Significance, MS-EGARCH vs. EGARCH

Horizon	**MSE Difference**	**Bootstrap Standard Error**	**T-statistic**
1-step	0.011	0.004	2.720
5-step	0.122	0.029	4.218
10-step	0.248	0.051	4.861
22-step	0.527	0.125	4.215

Next, we investigate whether EGARCH's MSE out-performances of MS-EGARCH in regime 2 are statistically significant. Analyzing table 20, we conclude that none of the regime 2 MSE differences are statistically significant.

Table 20: Regime 2 MSE Difference Significance, EGARCH vs. MS-EGARCH

Horizon	**MSE Difference**	**Bootstrap Standard Error**	**T-statistic**
1-step	0.006	0.012	0.491
5-step	0.005	0.031	0.152
10-step	-0.006	0.061	-0.102
22-step	0.065	0.127	0.510

Commentary

In terms of both MSE (RMSE) and MAE, MS-EGARCH's regime 1 out-performance of the best alternative model, single-regime EGARCH, is considerable and statistically significant. While single-regime EGARCH generally out-performs in terms of MSE (RMSE) and MAE in the turbulent regime 2, this out-performance is not statistically significant. Therefore, we conclude that MS-EGARCH's overall out-performance of single-regime EGARCH is primarily explained by superior performance in the tranquil regime (regime 1).

4.5 Testing MS-EGARCH's Ability to Forecast EGARCH misspecifications

In addition to comparing the RMSE and MAE of MS-EGARCH and the best alternative model, single-regime EGARCH, we also examine whether MS-EGARCH can forecast the misspecifications of single-regime EGARCH. Let h_t denote the true realized daily volatility at time t (computed according to equation 20), and let $\sigma_{t,EGAR}$ denote the fitted conditional volatility of single-regime EGARCH at time t (obtained during estimation). The misspecification of single-regime EGARCH is $h_t - \sigma_{t,EGAR}$. Assume $\sigma_{t-1,MSEGAR}(1)$ is the 1-step prior prediction of the time t volatility made by MS-EGARCH. Then, we are interested in the following regression:

$$(h_t - \sigma_{t,EGAR}) = \alpha + \beta_{MSEGAR} \cdot \sigma_{t-1,MSEGAR}(1) + u_t \tag{27}$$

We estimate the coefficients of this regression on the entire 1946-2022 out-of-sample period via OLS (Ordinary Least Squares) with Newey and West (1987) heteroscedasticity and autocorrelation consistent (HAC) standard errors[6]. The parameter estimates, standard errors, and p-values are contained in table 21.

Table 21: Equation 27 Parameter Estimates, 1946-2022 Data

Parameter	**Estimate**	**Standard Error**	**P-value**
α	-0.1173	0.027	**0.00**
β_{MSEGAR}	0.0962	0.036	**0.01**

The magnitude, positive sign, and statistical significance of this β_{MSEGAR} suggests that MS-EGARCH is able to predict the 1-day misspecifications of single-regime EGARCH. A higher 1-step ahead MS-EGARCH forecast implies that single-regime EGARCH will underfit volatility. Naturally, we can also reverse this regression to test whether the 1-step ahead forecasts of single-regime EGARCH anticipate the misspecifications of MS-EGARCH. Let $\sigma_{t,MSEGAR}$ denote the time t fitted conditional volatility of MS-EGARCH, and let $\sigma_{t-1,EGAR}(1)$ denote the 1-step prior prediction of time t volatility made by single-regime EGARCH. We can estimate the following regression via OLS with HAC standard errors:

$$(h_t - \sigma_{t,MSEGAR}) = \alpha + \beta_{EGAR} \cdot \sigma_{t-1,EGAR}(1) + u_t \tag{28}$$

---

[6] We again set the band-width parameter to $S_T = 0.75T^{\frac{1}{3}} \approxeq 20$.

The results of this are contained in table 22. β_{EGAR} is smaller than β_{MSEGAR} in magnitude, and it is not statistically significant. Therefore, we deduce that MS-EGARCH is able to anticipate the 1-day ahead misspecifications of single-regime EGARCH, but that EGARCH can not do the same for MS-EGARCH misspecifications. This implies that MS-EGARCH contains information about 1-day ahead volatility dynamics that single-regime EGARCH lacks.

Table 22: Equation 28 Parameter Estimates, 1946-2022 Data

Parameter	**Estimate**	**Standard Error**	**P-value**
α	-0.1449	0.033	**0.00**
β_{EGAR}	0.0533	0.043	0.21

We can generalize equation 27 to multi-step ahead misspecifications and forecasts. Let h_{t+H} denote the horizon H realized volatility between time t and $t+H$, and let $\sigma_{t-1,MSEGAR}(H)$ denote the horizon H volatility forecast made by MS-EGARCH at time $t-1$. We can estimate the following regression via OLS with HAC standard-errors:

$$(h_{t+H} - \sqrt{H} \cdot \sigma_{t,EGAR}) = \alpha + \beta_{MSEGARCH} \cdot \sigma_{t-1,MSEGAR}(H) + u_t \tag{29}$$

where $(h_{t+H} - \sqrt{H} \cdot \sigma_{t,EGAR})$ is the multi-step misspecification of single-regime EGARCH. Unfortunately, performing these regressions with $H \in \{5, 10, 22\}$ on the out-of-sample period does not yield promising results. Analyzing table 23, the beta estimates are not significant.

Table 23: Equation 29 $\beta_{MSEGARCH}$ Estimates across Horizons, 1946-2022 Data

Horizon	**Beta Estimate**	**Standard Error**	**P-value**
5	0.068	0.041	0.10
10	0.062	0.046	0.18
22	0.109	0.078	0.08

These results may seem contradictory with our earlier forecasting results, which suggested that MS-EGARCH forecasts realized volatility better than single-regime EGARCH at multi-step ahead horizons. However, predicting the misspecification of single-regime EGARCH is a distinct task from obtaining superior MAE and RMSE. We have found in this exercise that MS-EGARCH is capable of predicting the 1-step ahead misspecifications of single-regime EGARCH, but that its ability to anticipate multi-step ahead misspecifications is limited.

Chapter 5

Analysis of Regime-Conditional Market Behaviour

In this chapter, we show that MS-EGARCH's predicted regime changes are economically significant. Section 5.1 discusses our general methodology. Section 5.2 discusses how the behaviour of the NYSE/Nasdaq series used to generate the regime predictions varies based on them. Section 5.3 shows that the performance of certain equity risk factors varies after MS-EGARCH anticipates a regime change. Section 5.4 highlights changes to cross-asset behaviour following predicted regime shifts.

5.1 General Methodology

Our general approach in this section is to classify the regime on day t based on $z^*_{i,t} = \mathbb{P}(s_t = i|\widehat{\boldsymbol{\theta}}, I_{t-1})$, MS-EGARCH's estimate of the probability that the regime at time t, s_t, is i based on the information set I_{t-1} (i.e., excluding the day t return) and the parameter vector $\widehat{\boldsymbol{\theta}}$ obtained from fitting MS-EGARCH on the 5040 NYSE/Nasdaq index returns prior to (but not including) day t. This is MS-EGARCH's 1-step prior predicted probability of regime i on day t, obtained via Hamilton's filter. We classify day t as regime 2 if $z^*_{2,t} > 0.5$ and as regime 1 otherwise. This allows us to study characteristics on regime 1 days and on regime 2 days.

As with our regime-conditional forecasting performance analysis, we are not just interested in the raw estimates of the regime 1 and regime 2 characteristics, but also in the standard errors of these estimates. To determine these standard errors, we again employ an m-out-of-n block bootstrapping procedure (refer back to section 4.4.2). More precisely, we construct 100 bootstrapped samples of size 2520 by sampling 252-day blocks with replacement, and then utilize equation 26 to estimate the sample-size adjusted regime i characteristic D_i's standard error. In some of the tests below, we estimate regime-conditional characteristics that are not directly computed using MS-EGARCH outputs nor the NYSE/Nasdaq index, but rather computed using additional variables. Therefore, we sample blocks of these additional time series in parallel with our sampling of NYSE/Nasdaq index return blocks. That is, if in a particular bootstrap sample b, we include the block of NYSE/Nasdaq returns from day t to day $t + 252$, then we simultaneously add the day t to day $t + 252$ additional characteristic series block to b. In this manner, the additional variables become additional series within each bootstrapped sample.

Moreover, as in section 4.4.2 we will generally assume our regime-specific estimates are asymptotically normal without proof. Many of our characteristics are means of some sort, so, in those cases, this assumption amounts to assuming the conditions of the (generalized) Central Limit Theorem.

5.2 NYSE/Nasdaq Index Regime-Conditional Behaviour

We first compute the NYSE/Nasdaq log returns series mean, volatility, and excess (i.e., Fisher) kurtosis within each regime as well as the associated bootstrapped regime-specific standard errors for these estimates. We employ these bootstrapped standard errors in conjunction with our asymptotic normality assumption to produce 95% confidence intervals. These values are presented in tables 24 and 25.

Table 24: Regime 1-NYSE/Nasdaq Daily Returns Series Moments

Characteristic	**Regime 1 Value**	**Regime 1 SE**	**Regime 1 CI**
Mean	2.62×10^{-4}	5.23×10^{-6}	$[2.51 \times 10^{-4}, 2.72 \times 10^{-4}]$
Volatility	7.64×10^{-3}	4.24×10^{-5}	$[7.55 \times 10^{-3}, 7.72 \times 10^{-3}]$
Kurtosis	25.75	0.94	[23.90,27.59]

Table 25: Regime 2-NYSE/Nasdaq Daily Returns Series Moments

Characteristic	**Regime 2 Value**	**Regime 2 SE**	**Regime 2 CI**
Mean	2.33×10^{-4}	1.22×10^{-5}	$[2.09 \times 10^{-4}, 2.57 \times 10^{-4}]$
Volatility	1.09×10^{-2}	6.44×10^{-5}	$[1.08 \times 10^{-2}, 1.10 \times 10^{-2}]$
Kurtosis	15.62	1.13	[13.40,17.83]

Examining the results, the daily means correspond to an annualized mean return of 6.60% in regime 1 and 5.87% in regime 2. The fact that the lower volatility regime has a higher mean return aligns with Moreira and Muir (2017)'s finding that reducing risk exposure when volatility is higher generates superior returns. Granted, the overlap between the regime 1 and regime 2 confidence intervals suggests that this difference in means is unlikely to be statistically significant. For volatility and kurtosis on the other hand, the confidence intervals do not overlap. This suggests[1] that predicted shifts to regime 2 map to increases in volatility and declines in kurtosis. In fact, these daily volatility figures translate to an annualized volatility of 17.28% in regime 2 and 12.12% in regime 1. This is a reasonable discrepancy in regime-conditional volatility across a wide sample. The gap between the excess kurtosis values of the two regimes is also considerable. By construction, MS-EGARCH fits regime 1

---

[1] To be rigorous, a non-overlap of confidence intervals is insufficient to prove statistical significance because it is likely wrong to assume that the values of characteristics are independent across regimes. Since we do not know the relation between the regimes, we can not perform fully rigorous statistical significance tests.

to be the lower volatility period in-sample. We also saw in section 2.3.3 that MS-EGARCH fit higher kurtosis in regime 1 in-sample. Our results confirm that these regime features also hold out-of-sample.

Next, we analyze the frequency of large drops within each regime. Specifically, for each day t in regime i, we measure how frequently a cumulative NYSE/Nasdaq drop of X% or more occurs within the following 10 days. We assess this for $X \in \{1, 3, 5, 10\}$. While we will stick to a look-ahead of 10 days in this paper for brevity, our results were not that sensitive to this value and look-aheads of 5 and 22 days yielded comparable results. The sample estimates, bootstrap standard errors and corresponding 95% confidence intervals for these regime-conditional drop probabilities are contained in tables 26 and 27.

Table 26: Regime 1-NYSE/Nasdaq Index Drop Probabilities

Characteristic	**Regime 1 Value**	**Regime 1 SE**	**Regime 1 CI**
P(1% Drop)	**23.31%**	1.43%	[20.5%, 26.12%]
P(3% Drop)	**8.31%**	1.11%	[6.1%,10.49%]
P(5% Drop)	**2.73%**	0.89%	[1.0%, 4.48%]
P(10% Drop)	**0.38%**	0.30%	[0.00%, 0.96%]

Table 27: Regime 2-NYSE/Nasdaq Index Drop Probabilities

Characteristic	**Regime 2 Value**	**Regime 2 SE**	**Regime 2 CI**
P(1% Drop)	**31.68%**	1.40%	[28.94%,34.42%]
P(3% Drop)	**14.72%**	1.09%	[12.58%,16.87%]
P(5% Drop)	**6.31%**	0.73%	[4.87%,7.74%]
P(10% Drop)	**0.96%**	0.36%	[0.25%,1.67%]

Analyzing the data, the point estimates and confidence intervals suggest a higher frequency of drops of various sizes in the turbulent regime. This is in line with our expectations. Overall, MS-EGARCH appears able to anticipate periods of increased downside risk in the US stock market.

5.3 Equity Risk Premia Regime-Conditional Performance

Using daily Fama-French factor return data from Kenneth French's website (French, 2023), we next investigate whether equity risk-factors earn different mean log returns within each regime. To be more precise, we examine the arithmetic mean daily log returns of the Fama-French size ('SMB'), value ('HML'), investment ('CMA'), and profitability ('RMW') factors as well as the momentum factor ('MOM'), conditional upon the predicted regime. Tables 28 and 29 contain our results. Inspecting the numbers, various factors earn different mean returns depending upon MS-EGARCH's regime predictions. The two factors whose outperformance in the turbulent regime appears to be the most significant are the size and investment factors. The size factor earns negative annualized returns of -1.26% in regime 1, compared to positive annualized returns of 3.47% in regime 2. In other words, the size

Table 28: Regime 1-Fama-French Factor Mean Daily Log Returns

Characteristic	**Regime 1 Mean**	**Regime 1 SE**	**Regime 1 CI**
MOM	3.3×10^{-4}	5.4×10^{-5}	$[2.3 \times 10^{-4}, 4.4 \times 10^{-4}]$
SMB	-5.0×10^{-5}	4.4×10^{-5}	$[-1.4 \times 10^{-4}, 3.7 \times 10^{-5}]$
HML	2.9×10^{-5}	5.2×10^{-5}	$[-7.3 \times 10^{-5}, 1.3 \times 10^{-4}]$
RMW	9.2×10^{-5}	3.2×10^{-5}	$[2.8 \times 10^{-5}, 1.6 \times 10^{-4}]$
CMA	2.7×10^{-5}	3.0×10^{-5}	$[-3.2 \times 10^{-5}, 8.7 \times 10^{-5}]$

Table 29: Regime 2-Fama-French Factor Mean Daily Log Returns

Characteristic	**Regime 2 Mean**	**Regime 2 SE**	**Regime 2 CI**
MOM	2.7×10^{-4}	9.2×10^{-5}	$[9.0 \times 10^{-5}, 4.5 \times 10^{-4}]$
SMB	1.4×10^{-4}	5.1×10^{-5}	$[3.9 \times 10^{-5}, 2.4 \times 10^{-4}]$
HML	2.2×10^{-4}	6.6×10^{-5}	$[8.7 \times 10^{-5}, 3.5 \times 10^{-4}]$
RMW	1.6×10^{-4}	4.0×10^{-5}	$[8.4 \times 10^{-5}, 2.4 \times 10^{-4}]$
CMA	2.0×10^{-4}	4.3×10^{-5}	$[1.1 \times 10^{-4}, 2.4 \times 10^{-4}]$

premium only exists in regime 2. As noted previously, this finding is congruent with Ahn et al. (2019)'s result. Similarly, consistent with Sheth and Lim (2017)'s analyses, we find that the investment factor nearly doubles its performance within the turbulent state; it earns annualized returns of 4.09% in regime 2, whereas it earns annualized returns of 2.32% in regime 1.

5.4 Cross-Asset Regime-Conditional Characteristics

Table 30: Regime 1-Mean VIX and Mean 10-year Government Yields

Characteristic	**Regime 1 Mean**	**Regime 1 SE**	**Regime 1 CI**
VIX	**15.36**	0.56	[14.2,16.45]
10-Year US Government Yield	**4.49%**	0.13%	[4.23%,4.75%]

First, we investigate the VIX options market index and a generic US government bond 10-year (annualized) yield index ('GT10'), both of which were obtained from Bloomberg (Bloomberg L.P., 2023). Regime-conditional means for these two variables are contained in tables 30 and 31.

Examining these two tables, as expected, the VIX index is elevated in regime 2. This further reinforces the alignment between VIX and MS-EGARCH. 10-year US Government yields are also larger in regime 2. Based on the confidence intervals and the point-estimates, both of these differences appear to be statistically significant (though again, we can not rigorously prove this).

Next, we shift our attention from government bonds to two corporate bond return indices: 'IBOXHY' and 'IBOXIG', representing US high-yield and US investment grade bonds

Table 31: Regime 2-Mean VIX and Mean 10-year Government Yields

Characteristic	**Regime 2 Mean**	**Regime 2 SE**	**Regime 2 CI**
VIX	**23.76**	0.60	[22.60,24.93]
10-Year US Government Yield	**6.63%**	0.23%	[6.17%,7.09%]

respectively. We convert these to log returns series and investigate their regime-conditional mean daily log returns. We also examine their daily betas with respect to US equity markets (i.e., the NYSE/Nasdaq index). The beta formula we employ is

$$\beta_I = \frac{\widehat{\rho}_{I,M} \cdot \widehat{\sigma}_I}{\widehat{\sigma}_M} \tag{30}$$

where $\widehat{\rho}_{I,M}$ is the sample daily returns correlation between the bond index I and the NYSE/-Nasdaq index, $\widehat{\sigma}_I$ is the sample standard deviation of the bond index, and $\widehat{\sigma}_M$ is the sample standard deviation of the NYSE/Nasdaq index. Tables 32 and 33 contain our IBOXHY and IBOXIG characteristic estimates, along with standard errors and confidence intervals within both regimes.

Table 32: Regime 1-High-Yield and Investment Grade Bond Mean Returns and Betas

Characteristic	**Regime 1 Value**	**Regime 1 SE**	**Regime 1 CI**
IBOXHY Mean	5.9×10^{-5}	4.1×10^{-5}	$[-2.2 \times 10^{-5}, 1.4 \times 10^{-4}]$
IBOXIG Mean	7.0×10^{-5}	4.3×10^{-5}	$[-1.4 \times 10^{-5}, 1.5 \times 10^{-4}]$
IBOXHY to US Beta	0.194	0.022	[0.150,0.238]
IBOXIG to US Beta	0.006	0.004	[-0.003,0.014]

Table 33: Regime 2-High-Yield and Investment Grade Bond Mean Returns and Betas

Characteristic	**Regime 2 Value**	**Regime 2 SE**	**Regime 2 CI**
IBOXHY Mean	2.6×10^{-4}	5.2×10^{-5}	$[1.6 \times 10^{-4}, 3.6 \times 10^{-4}]$
IBOXIG Mean	2.5×10^{-4}	4.5×10^{-5}	$[1.6 \times 10^{-4}, 3.4 \times 10^{-4}]$
IBOXHY to US Beta	0.088	0.024	[0.041,0.136]
IBOXIG to US Beta	-0.017	0.006	[-0.030,-0.005]

In the tables, we observe that both US high-yield and US investment grade bonds earn higher returns in regime 2. In fact, when annualized, high-yield bonds earn 6.58% in regime 2, compared to 1.47% in regime 1, while investment grade bonds earn 6.28% in regime 2 compared to 1.77% in regime 1. This marked improvement in returns is coupled with a decline in high-yield and investment grade betas to the US equity market (NYSE/Nasdaq index), implying that both types of corporate bonds become less correlated with stocks. Our findings here echo those of Connolly et al. (2005) and Chiang et al. (2014).

Overall, in this section, we have elucidated that MS-EGARCH's forecasted shifts in volatility regimes map to differences in cross-asset relationships. These findings illustrate the inter-connected nature of US financial markets. Moreover, they further validate that MS-EGARCH's detected regimes are economically meaningful, out-of-sample.

Chapter 6

Conclusion and Next Steps

In this study, we obtain rich and novel insights by implementing a two-regime Markov-Switching EGARCH model (MS-EGARCH). In an expansive equity volatility forecasting test comprising 19,441 trading days, we demonstrate that MS-EGARCH outperforms single-regime EGARCH and GARCH. Hence, we contribute to the existing literature showing the benefits of incorporating regime-switching into the GARCH framework. Future research is required to investigate whether the forecasting improvements of MS-EGARCH yield meaningful improvements to models that employ volatility forecasts as inputs. For example, further work could test whether MS-EGARCH-based volatility forecasts ameliorate option-pricing models.

We also further the claims of researchers such as Blair et al. (2001) who have shown that option-implied volatilities contain as much information about future volatility as GARCH models. We show that this fact extends to Markov-Switching GARCH models too. An additional area for exploration would be the use of MS-EGARCH in conjunction with the VIX to measure the variance premium and, thereby, predict returns. Researchers like Bekaert and Hoerova (2014) have tried similar experiments using single regime models, so it would be interesting to see whether Markov-switching enhances performance.

We can attribute much of MS-EGARCH's superior predictive abilities to the economic significance of its 1-step ahead regime probabilities. In fact, defining the predicted regime based on whether these probabilities are greater than 50%, we observe clear regime-conditional differences in volatility, kurtosis, and the frequency of large drops. Moreover, we document changes to equity risk-factor performance as well as bond behaviour following a predicted regime shift. Further research should explore the potential of MS-EGARCH regime probabilities in shaping effective trading strategies. Moreira and Muir (2017) showed that scaling risk exposure to trailing volatility yields alpha. Future researchers could examine scaling exposure based on MS-EGARCH regime probabilities.

Finally, another extension of our work would be to apply MS-EGARCH to other geographies and asset classes. Our results suggest that such extensions should reveal fundamental facts about the evolution and market implications of volatility.

References

Ahn, D.-H., Min, B.-K., & Yoon, B. (2019). Why has the size effect disappeared? *Journal of Banking & Finance*, *102*, 256–276. https://doi.org/https://doi.org/10.1016/j.jbankfin.2019.02.005

Andersen, T. G., & Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, *39*(4), 885–905. Retrieved November 28, 2023, from http://www.jstor.org/stable/2527343

Ang, A., & Bekaert, G. (2002). Regime Switches in Interest Rates. *Journal of Business & Economic Statistics*, *20*(2), 163–182. Retrieved December 8, 2023, from http://www.jstor.org/stable/1392056

Ang, A., & Bekaert, G. (2004). How Regimes Affect Asset Allocation. *Financial Analysts Journal*, *60*(2), 86–99. Retrieved December 8, 2023, from http://www.jstor.org/stable/4480559

Ardia, D., Bluteau, K., Boudt, K., Catania, L., & Trottier, D.-A. (2019). Markov-Switching GARCH Models in R: The MSGARCH Package. *Journal of Statistical Software*, *91*(4), 1–38. https://doi.org/10.18637/jss.v091.i04

Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility [Analysis of Financial Data]. *Journal of Econometrics*, *183*(2), 181–192. https://doi.org/https://doi.org/10.1016/j.jeconom.2014.05.008

Blair, B. J., Poon, S.-H., & Taylor, S. J. (2001). Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics*, *105*(1), 5–26. https://ideas.repec.org/a/eee/econom/v105y2001i1p5-26.html

Bloomberg L.P. (2023). GT10, VIX, IBOXHY, and IBOXIG PX_LAST Field [Accessed on 2023-03-01].

Center for Research in Security Prices. (2023). CRSP U.S. Stock Database. Retrieved March 1, 2023, from https://wrds-www.wharton.upenn.edu/pages/get-data/center-research-security-prices-crsp/

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chiang, T., Li, J., & Yang, S.-Y. (2014). Dynamic stock–bond return correlations and financial market uncertainty. *Review of Quantitative Finance and Accounting*, *45*. https://doi.org/10.1007/s11156-013-0430-4

Christensen, B., & Prabhala, N. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, *50*(2), 125–150. https://doi.org/https://doi.org/10.1016/S0304-405X(98)00034-8

Connolly, R., Stivers, C., & Sun, L. (2005). Stock Market Uncertainty and the Stock-Bond Return Relation. *The Journal of Financial and Quantitative Analysis*, *40*(1), 161–194. Retrieved December 8, 2023, from http://www.jstor.org/stable/27647190

Diebold, F., & Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–63. https://EconPapers.repec.org/RePEc:bes:jnlbes:v:13:y:1995:i:3:p:253-63

Douc, R., Moulines, É., & Rydén, T. (2004). Asymptotic Properties of the Maximum Likelihood Estimator in Autoregressive Models with Markov Regime. *The Annals of Statistics*, *32*(5), 2254–2304. Retrieved January 1, 2024, from http://www.jstor.org/stable/3448572

Dueker, M. J. (1997). Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility. *Journal of Business & Economic Statistics*, *15*(1), 26–34. https://doi.org/10.1080/07350015.1997.10524683

French, K. R. (2023). *U.S. Research Returns Data* [Data Library, Tuck School of Business at Dartmouth]. Retrieved March 1, 2023, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research

Gray, S. F. (1996). REGIME-SWITCHING IN AUSTRALIAN SHORT-TERM INTEREST RATES. *Accounting & Finance*, *36*(1), 65–88. https://doi.org/https://doi.org/10.1111/j.1467-629X.1996.tb00299.x

Guidolin, M., & Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, *31*(11), 3503–3544. https://doi.org/https://doi.org/10.1016/j.jedc.2006.12.004

Haas, M., Mittnik, S., & Paolella, M. S. (2004). A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, *2*(4), 493–530. https://doi.org/10.1093/jjfinec/nbh020

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, *57*(2), 357–384. Retrieved December 7, 2023, from http://www.jstor.org/stable/1912559

Hamilton, J. D. (1994). Modeling Time Series with Changes in Regime. In *Time series analysis* (pp. 677–703). Princeton University Press. Retrieved December 9, 2023, from http://www.jstor.org/stable/j.ctv14jx6sm.25

Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, *64*(1), 307–333. https://doi.org/https://doi.org/10.1016/0304-4076(94)90067-1

Henry, Ó. (2009). Regime switching in the relationship between equity returns and short-term interest rates in the UK. *Journal of Banking & Finance*, *33*, 405–414. https://doi.org/10.1016/j.jbankfin.2008.08.001

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, *27*, 363–394. https://doi.org/10.1007/s001810100100

Kritzman, M., Page, S., & Turkington, D. (2012). Regime Shifts: Implications for Dynamic Strategies (corrected). *Financial Analysts Journal*, *68*(3), 22–39. https://doi.org/10.2469/faj.v68.n3.3

LeSage, J. P. (1998, December). ECONOMETRICS: MATLAB toolbox of econometrics functions. https://ideas.repec.org/c/boc/bocode/t961401.html

Maheu, J. M., & McCurdy, T. H. (2000). Identifying Bull and Bear Markets in Stock Returns. *Journal of Business & Economic Statistics*, *18*(1), 100–112. https://doi.org/10.2139/ssrn.146531

Marcucci, J. (2005). Forecasting Stock Market Volatility with Regime-Switching GARCH Models. *Studies in Nonlinear Dynamics & Econometrics*, *9*(4), 1–55. https://EconPapers.repec.org/RePEc:bpj:sndecm:v:9:y:2005:i:4:n:6

Moreira, A., & Muir, T. (2017). Volatility-Managed Portfolios. *The Journal of Finance*, *72*(4), 1611–1643. Retrieved December 12, 2023, from http://www.jstor.org/stable/26652549

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, *59*(2), 347–370. Retrieved November 28, 2023, from http://www.jstor.org/stable/2938260

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708. Retrieved December 11, 2023, from http://www.jstor.org/stable/1913610

Sheth, A., & Lim, T. (2017, December). Fama-French Factors and Business Cycles. https://doi.org/10.2139/ssrn.3082577

Stock, J., & Watson, M. W. (2003). *Introduction to Econometrics*. Prentice Hall.

Tsay, R. S. (2010). Conditional Heteroscedastic Models. In *Analysis of Financial Time Series* (pp. 109–173). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9780470644560.ch3

Vrieze, S. (2012). Model Selection and Psychological Theory: A Discussion of the Differences Between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological methods*, *17*, 228–43. https://doi.org/10.1037/a0027127